10th International Workshop on Innovative Simulation for Healthcare
18th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0371 ISBN 978-88-85741-65-2 © 2021 The Authors.
doi: 10.46354/i3m.2021.iwish.004

EpiMon: Vision-Based Early Warning System for Monitoring Uprising Epileptic Seizures During Night

Gerald A. Zwettler[1,2,*], Anna Reichhardt[2,3], Michael Stradner[1,2], Christoph Praschl[2] and Emmanuel Helm[2,3]

[1]Department of Software Engineering, Faculty for Information, Communication and Media, University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg, 4232, Austria
[2]Research Group Advanced Information Systems and Technologies (AIST), University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg, 4232, Austria
[3]Department of Bio and Medical Informatics, Faculty for Information, Communication and Media, University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg, 4232, Austria

[*]Corresponding author. Email address: gerald.zwettler@fh-hagenberg.at

Abstract

At a prevalence of almost 1%, potential epileptic seizures manifest a significant health risk for many juvenile patients. Thus, monitoring is essential to set early counteractive measurements to prevent from damage. The sensor-based monitoring systems mainly address epileptic seizures indicated by a change in the muscle tonus but cannot be utilized for patients that show Prévost's-sign only. To monitor initiating Prévost's-sign with opened-eyes as critical visual feature, the applicability of deep-learning eye detection systems on night vision images is evaluated in this paper as basis for modelling and classifying the eye state (closed, opened, not visible). A holistic research prototype is presented as proof of concept, showing the applicability of state-of-the-art face detection on night vision images as well as multi-variate feature analysis on Graph segmentation pre-fragmentation, applicable to detect the state of the eye in a robust way. Results show a single frame accuracy in face/eye detection of 73.91% and 94.44% for classification of the opened eyes as indication of a potentially initiating epileptic seizure. The monitoring system is based on a Raspberry computation unit with two ELP night vision cameras attached and a smart phone app for user-interaction and configuration besides on-demand visual monitoring. Future work will show that the single frame detection rate is sufficient for building up a rule-based monitoring state machine at user predefined sensitivity and specificity by analysing the visual content as time-series rather than single images.

Keywords: night vision monitoring; feature-based classification; face detection; epileptic seizures; Prévost's-sign

1. Introduction

Epileptic strokes are a very severe medical seizure that can lastly affect the health condition of the patient if no immediate counteractive measures are taken. The alert systems available on the market only address specific kinds of epileptic diseases as epilepsy in general is a multilayered and complex illness showing a high level of patient-specific variability. Thus, only the most prominent groups of epileptic diseases are currently covered by medical alert systems available on the markets.

For many patients suffering from this disease, an initiating epileptic seizure is indicated by changes in

the muscle tonus. Tools such as *Epi-Care* from German company Epitech GmbH (2021) or *NightWatch* by LivAssured B.V. (2021) claim to be able to detect changes in the patient's heart-beat rhythm, acceleration of motion or muscle contraction in general by utilizing AI paradigms for detection. Unfortunately, these systems can be utilized for patients only, that show a change in the muscle tonus. If this sensory implication is not significant for a patient, common visual baby monitoring systems can be utilized for manual observation by the parental overseer. The permanent monitoring of the diseased children marks a significant burden and responsibility for the parents and can only be marginally automated by utilizing monitoring systems. While the caretakers try to avoid harm to the juvenile patient suffering from epilepsy at all costs, this permanent monitoring of potentially overcautious custody represents a huge burden and constriction of daily life for both, the care-taking parents or siblings as well as the teen patients according to findings of Tsuchie et al. (2006). Thus, the demand for an imperceptible and automated monitoring system is huge to allow for a higher level of freedom and quality of life while not increasing the patient's stress level by disturbing sound or immersive lighting.

With the audio analytic system *Lollipop Care* by Masterwork Aoitek Tech Corp Ltd. (2021) or the visual analysis of movement or children leaving their beds as offered by tool *HelloBaby HB25* of HELLOBABY Ltd. (2021) only the monitoring of healthy babies or infants is possible without hesitation. But for highly critical epilepsy monitoring with only marginal and fine-grained visual implications such as opening of the eyes or Prévost's-sign these systems are not applicable.

With video EEG, a gold standard in the diagnosis of epileptic seizures is available as discussed by Asano et al. (2005) and Schulc et al. (2009). Nevertheless, the hard-wired EEG hardware that can hardly be built up at home for daily routine monitoring must be characterized as significant limitation of freedom for the patient.

For many epileptic seizures, no change in the muscle tonus can be identified at all. Thus, the monitoring systems available on the market cannot be utilized. These patients often show the Prévost's-sign (*fr. Déviation conjuguée, de. "Herdblick"*) named according to Prévost (1868), i.e. a non-influenceable aberrancy of the viewing direction with both eyes wide open and permanently gazing during a pathological seizure of the frontal brain area. In case of strokes or specific epileptic seizures, both eyes are turned to the direction of the lesion in a parallel way, i.e. without parallax-conditioned strabismus.

Consequently, the demand for a non-invasive visual monitoring system is high to allow for detection of early signs of initiating epileptic seizures for patients with eyes opened and gazing during night as early indication.

1.1. Medical Background

Epilepsy can be named the most frequent chronic neurological disease at a prevalence of [0.5; 1.0]% and is extremely dangerous due to sever seizures recurring in an unpredictable way as analyzed by Sander (2003) and Forsgren et al. (2005). At an incidence of around [40; 70] per 100.000 residents and year, the chance of an individual to contract epilepsy can be assessed by 3-4% according to Baumgartner et al. (2019). Regarding the incidence per age group, the bi-modal histogram indicates that there are two evident age groups when people get affected by epilepsy, namely between infancy and adolescence with the second peak starting at seniority 60+ years according to MacDonald et al. (2000).

Eileptic seizures with observable motoric symptoms that are indicated by changing muscle tonus can be detected based on sensor hardware, namely from micro-electro-mechanical systems (MEMS) as in depth analyzed by Schulc et al. (2009).

1.2. State of the Art

One critical, visual feature to detect uprising epileptic seizures is related with the state of the eye, defining if they are closed or opened. Algorithms for this kind of scenario are commonly used in the area of driver drowsiness systems based on visual models as described by Tian and Qin (2005) or Hong et al. (2007) or using Haar Cascades and circular Hough transformations as presented by Fitriyani et al. (2016). Alternative methods use cascade regressions as shown by Gou et al. (2017) or aspect ratios of facial landmarks as described by Pandey and Muppalaneni (2021). Next to such classic image processing approaches also machine learning based methods are used in the area of driver drowsiness detection as described by Dua et al. (2021) and Zhang et al. (2017), that use convolutional neural networks to detect the states of a person's eyes. The area of application of the latter approach is comparable to ours in terms of using a infrared system instead of a RGB system to detect the state of the eyes and results in a high recognition accuracy. In difference to ours, all approaches focuse on the eye state detection to evaluate a person's blinking characteristics to determine the state of drowsiness. These methods are mostly based on daylight grayscale as well as RGB images in the context of driver monitoring systems, while our system is placed within night vision images of sleeping persons created with infrared cameras.

Besides these eye-specific detection and classification approaches, the vitreous body glaring in night vision images can be detected by generic image processing strategies, too. With a pre-fragmentation of the input image according to sub-image regions of similar intensities or enclosed by high gradients, the task of instance segmentation can be formulated as

generic multi-variate feature-classification problem that is applicable to various 2D and 3D image analysis domains in medicine as in depth analyzed by Zwettler (2014), too.

Based on a pre-fragmentation, for all of the sub-image regions a bag of features is calculated per class with the class similarity to be calculated based on CDF (cumulated density function), PDF (probability density function) or the Mahalanobis distance as introduced by Mahalanobis (1936) to incorporate the correlation of the features within the bag-of-feature vector. In case of an insufficient amount of input samples thus leading to a sparse multi-dimensional feature density space, Monte Carlo simulation can be applied as delineated by Atanassov and Dimov (2008).

1.3. Night-Vision Based Monitoring of Uprising Epileptic Seizures

In this research work we address several research questions that are fundamental for the development of a vision-based epilepsy monitoring system, namely:

- *Is the performance of state of the art deep-learning (DL) face detection models on night vision images sufficient for practical application without need for re-training or transfer learning?*
- *Are conventional image processing concepts for image pre-segmentation suitable to classify the eye state in a reliable manner by applying a rule-based bag-of-feature classification?*

While these research questions are addressed and answered by implementing and evaluating a research prototype, the applicability of this approach in the medical domain, namely "*Is the robust analysis of the eye state at night a reliable and robust indication for uprising epileptic seizures related to the Prévost's-sign?*" can only be answered by carrying out a clinical study together with our medical partners what needs to be achieved in future.

2. Material

The developed algorithms for monitoring uprising epileptic seizures are tested based on multiple night vision recordings of human faces. These images are acquired using different setups with the following camera systems: (I) camera *eufy Security SpaceView T8300-M Babyphone* and (II) camera *ELP 1080P Nightvision IP CAMERA*. Additionally, three persons, one child and two adults, are used as test subjects for the monitoring system. The images are inhomogeneous in different aspects. There are differences according to the dimensions and bit depth and for this have a range from 330×277 pixels up to 1920×1080 pixels and a bit depth of 24, respectively 32 bit. Additionally, the orientation of the pictures varies between landscape and portrait

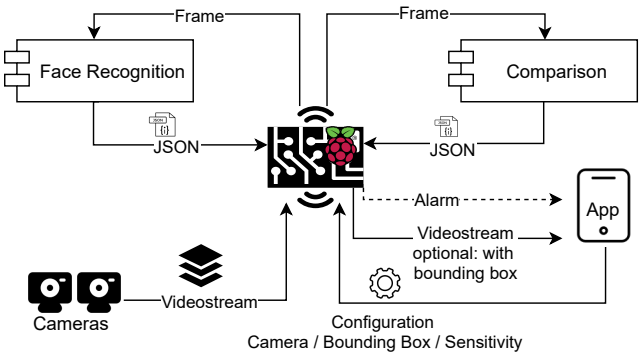


Figure 1. Concept of the monitoring system prototype. Two camera systems monitor the patient from different views and send the videostream to the Raspberry computing unit that carries out the deep-learning based face recognition. The frames are then fragmented utilizing Graph segmentation before they get classified according to a pre-calibrated eye model (open, closed) at 1 fps. In case of an emergency, an alarm is reported onto the monitoring app.

captures, too. Next to the technical differences also the images' content differs according to the brightness and the included sections of the test persons. The test images contain situations, where the person's face is completely visible, to pictures, where parts of the face are occluded e.g. by a pillow or a blanket. Since the desired ideal situation of the presented approach targets images containing the uncovered face of the monitored person, we have separated the images into two test data sets. The first data set contains 37 images from camera system (I) and the second data set contains 50 images from camera system (II). The images from the second data set have a higher quality than the ones from the first data set, but they are rather too bright for the eye state classification due to multiple areas of specular reflection besides the eyes.

3. Methodology

An overview of the hardware and infrastructure setup of the epilepsy monitoring research prototype is shown in figure 1. The three main parts, namely the face recognition, the model comparison as well as the model calibration are all addressed in this paper. A high level of modularization on the software architecture allows to replace the components in a generic way. The central system receives the video stream from the cameras and gets information from the modules *Face Recognition* and *Comparison* to get the necessary information about the position of the face and eyes and the state of the person. Thereby, the state means if there is a person visible and if the eyes are opened or closed. Dependent on the state the system triggers an alarm on the smartphone, so that the user can handle a probably upcoming seizure.

3.1. Hardware Setup of the Monitoring System

For the monitoring system hardware setup, a Raspberry Pi 4 is utilized as the central processing unit of the system. It is connected with two cameras for gathering the video stream from varying views. The cameras from ELP feature night-vision at high image resolution of 1080 pixels. The connection to the Raspberry Pi is established by utilizing USB connections. The cameras have to be placed around the person to be monitored so that sleeping on the back and both sides is at least covered within the cumulative camera viewing frustum.

3.2. Pre-Fragmentation with Graph Segmentation

For input images $\mathcal{I}$ where the face detection succeeds, local regions $\mathcal{R} \subset \mathcal{I}$ are the basis for subsequent feature analysis. The regions thereby get pre-fragmented according to local adjacent intensity homogeneity. Thereby, potential eye region $\mathcal{R}_i$ are derived from Graph segmentation applied to all pixels within the bounding box from face detection. The Graph segmentation as presented by Zahn (1971) and recently discussed by Felzenszwalb and Huttenlocher (2004) allows for intensity/edge based pre-fragmentation of grayscale input images similar to conventional watershed transformation e.g. known from Beucher and Lantuéjoul (1979). For the Graph segmentation utilized in this research work, the parameters *sigma* = 0.4 for smoothing the image via gaussian filter and k = 80 as constant for the threshold function are utilized. Besides, *min_size* = 15 as minimum component size, where components smaller than *min_size* are joined together, is chosen for the segmentation to achieve a well-balanced fragmentation granularity.

3.3. Modelling the Eye State

A feature $\mathcal{F}$ is generally defined as a function evaluating a real-valued result for an input region $\mathcal{R}_i$ as $\mathcal{F} : \mathcal{I}[\mathcal{R}] \rightarrow \mathbb{R}$ for regions $\mathfrak{R}_1 = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n\}$ belonging to the same class $\mathcal{X}$. Values of feature $\mathcal{F}$ typically are statistically distributed around mean μ with standard deviation σ and thus can be statistically approximated by a normal distribution as $\mathcal{F} \sim \mathcal{N}(\mu, \sigma^2)$. In the calibration phase the statistical deviation of the feature values per class $\mathfrak{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m\}$ are generally derived from a training set $\mathcal{D} = \{(\mathcal{R}_k, x_k)\}_{k=1}^{N}$ with N regions assigned a certain class label of $\mathfrak{X}$.

Thus for feature $\mathcal{F}_l$ and class $\mathcal{X}_o$, the mean is calculated as $\mu_{l_o} = mean\, D(x_o)$ and standard deviation as $\sigma_{l_o}^2 = std\, D(x_o)$, approximating the feature value distribution as normal distribution with $\mathcal{F}_{l_o} \sim \mathcal{N}(\mu_{l_o}, \sigma_{l_o}^2)$. The features are then normalized to range $[0; 1]$.

Generally, not a single feature $\mathcal{F}_l$ is feasible for classification of test regions, but several different features are required to be incorporated to cover different aspects and characteristics as bag-of-features vector.

For each of the classes $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$, the model is calculated as delineated above from several camera-specific training images. For each image, the position of the face via *Face Recognition* and the pre-fragmented regions from Graph segmentation are denoted as regions $\mathcal{R}_i$. For each region $\mathcal{R}_i$ of the training data sets the following features $\mathcal{F}_l$ are calculated and utilized for modelling:

- $\mathcal{F}_1$: *Region size*, the average pixel count per region relative to the size of the face calculated by the bounding box.
- $\mathcal{F}_2$: *Region Intensity*, the average scalar pixel value per region.
- $\mathcal{F}_3$: *Relative Region Intensity*, the average scalar pixel value relative to the average intensity value within the entire face bounding box.
- $\mathcal{F}_6$: *Region Intensity Variability*, the scalar value variance within the region.
- $\mathcal{F}_7$: *Circularity*, the region perimeter relative to the region area compared to a perfect sphere.

With these intra-region features being independently calculated per region $\mathcal{R}_i$, the discrimination between eye states $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$ is featured. As the pre-fragmentation based on Graph segmentation leads to several candidate regions around the eye-position, the subsequent features are calculated on an inter-region basis to evaluate and validate the most prominent eye region candidates in the sense of *meta-features*, namely:

- $\mathcal{F}_4$: *Vertical Relative Position*, the vertical relative position calculated w.r.t. the detected bounding box per face.
- $\mathcal{F}_5$: *Horizontal Relative Position*, the horizontal relative position calculated w.r.t. the detected bounding box per face. To support detection of both, left and right eye regions, the horizontal position always gets mirrored to left side, i.e. $[0; 0.5]$.

As explained before, the feature values are calculated relative to the size of the face determined by the bounding box to ensure scale invariance. Furthermore, the feature values are all normalized to $[0; 1]$ per feature-model as stated before.

With this modelling, for the classes $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$ the expected feature statistics are modelled as normal deviation. These models can be used for the classification provided by module *Comparison* to determine the state of the eyes by calculating the probability of an unknown region $\mathcal{R}_i$ w.r.t. the features of the modelled classes. The concept of *Calibration* is shown in the figure 2.

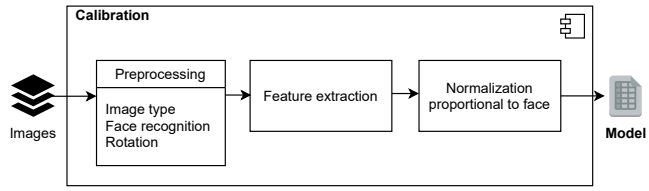


Figure 2. Concept of the module *Calibration* for training custom classification models for the patient's image state. Features are calculated for regions from Graph segmentation applied within the face detection bounding box. Normalized feature values for the states open/closed are then statistically evaluated to form a classification model.

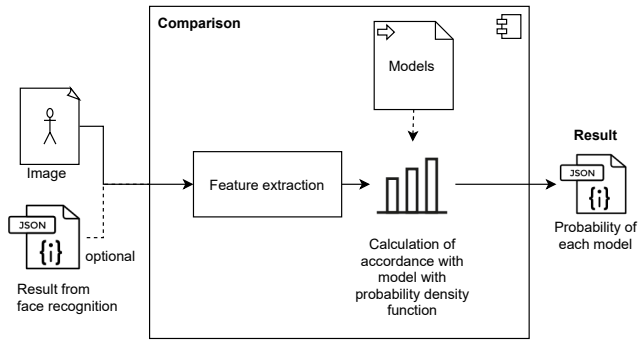


Figure 3. Concept of *comparison* to classify the state of the eyes according to the calibrated patient-specific models.

3.4. Face and Eye Detection

The eye detection service uses a python module called *mtcnn* to detect face, eyes and nose. Based on the result from the module the important key points are extracted and returned. Multi-task Cascaded Convolutional Networks, short MTCNN, is a process which is built on three stages of convolutional networks, that are able to recognize faces and landmarks such as eyes, mouth and nose as stated by Zhang et al. (2016).

3.5. Classification of the Eye State

For multivariate feature classification, a vector $\mathfrak{F} = \{\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n\}$ with n features is utilized for classification as delineated by Backhaus (2008) and Izenman (2008). In our model, $n = 5$ inter-region features are utilized with their similarity per class $sim(\mathcal{X}_i, \mathcal{R}_k) = P(\mathcal{R}_k|\mathcal{X}_i)$ calculated from the probability density function (PDF). For each feature $\mathcal{F}_l$ it is expected that the Bayesian inference probability calculated from the PDF of $\mathcal{N}_o$ for all classes $\mathcal{X}_o \in \mathfrak{X}$, defined as

$$pdf(x, \mathcal{X}_o, \mathcal{F}_{l_o}) = \frac{1}{\sqrt{2\pi\sigma_{l_o}^2}} \exp\left(-\frac{(x-\mu_{l_o})^2}{2\sigma_{l_o}^2}\right) \quad (1)$$

with $x \in \mathbb{R}$ for $\mathcal{F}_{l_o} \sim \mathcal{N}(\mu_{l_o}, \sigma_{l_o}^2)$, at least partially overlap. Thus, a *Bayes error* is introduced and no classification at 100% accuracy can be performed.

Features $\mathfrak{F}_o$ of a class $\mathcal{X}_o$ are considered to be independ and uncorrelated in all dimensions. Thus, multivariate probability density function pdf_{ind} is calculated with

$$pdf_{ind}(X, \mathcal{X}_o, \mathfrak{F}_o) = \frac{1}{|\mathfrak{F}_o|} \cdot \sum_{\mathcal{F}_{i_o} \in \mathfrak{F}_o} pdf(x_i, \mathcal{X}_o, \mathcal{F}_{i_o}) \quad (2)$$

as average of the probability density functions for each feature with function evaluation at positions in vector $X = \{x_1, \ldots, x_d\} \in \mathbb{R}^d$.

The module *Comparison* can be used with a default model which defines the possible values of a certain state. For modelling the state of a new image, the process starts with *Face Recognition*, to get the position of the face and the eyes. In the area of the face, indicated by ROI derived from the face detection, Graph segmentation is used to get the most relevant eye candidate regions. This approach is especially useful for night vision images, because open eyes can be detected easily by their brightness. The detected regions are compared with the models and an accordance to each feature of each model is calculated with the probability density function. The result of the comparison are two probabilities, which shows the accordance to each model (open/closed) and to the state "no eyes", like in table 6. The concept of the module *Comparison* is shown in Figure 3.

Table 1. Overview of the python modules which are utilized for the implementation.

Module	Usage	Version
OpenCV	Image processing	4.5.1
MTCNN	Face detection	0.1.0
Tensorflow	Face detection	2.5.0-rc1
FLASK	REST interfaces	1.1.2

4. Implementation

The implementation of the modules *Face Recognition*, *Comparison* and *Calibration* is based on Python in version 3.8+. Table 1 shows the utilized modules and their usage. Most of the operations are supported by modules which are common in use: *Pandas*, *SciPy*, *Matplotlib* and *NumPy*. These modules provide multiple methods for calculations on image data and useful data structures like `DataFrame` and `ndarray`. For an independent and encapsulated implementation, the project provides REST endpoints for each module. The results are provided in JSON format. The implementation of the REST endpoints is written with FLASK, a web application framework.

Table 2. Accuracy of *Face Recognition* for two data sets compared with Microsoft Face Detection.

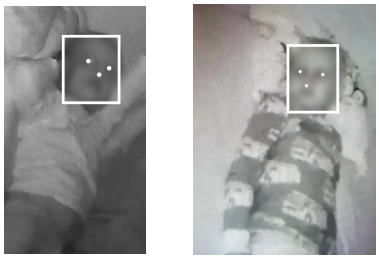
	Accuracy *Face Recognition*	Accuracy Microsoft Tool
Data set *a*	56.76%	81.08%
Data set *b*	73.91%	80.00%

5. Results

In the following sections, results from the particular processing steps are presented and the overall classification is evaluated.

5.1. Face and Eye Detection

In figure 4 the results from *Face Recognition* are annotated as a bounding box and dots for the key points, respectively. For anonymization, the images plotted in this paper are blurred. The module *Face Recognition* provides the coordinates of the detected face and the key points in JSON format for independent subsequent use.



(a) (b)

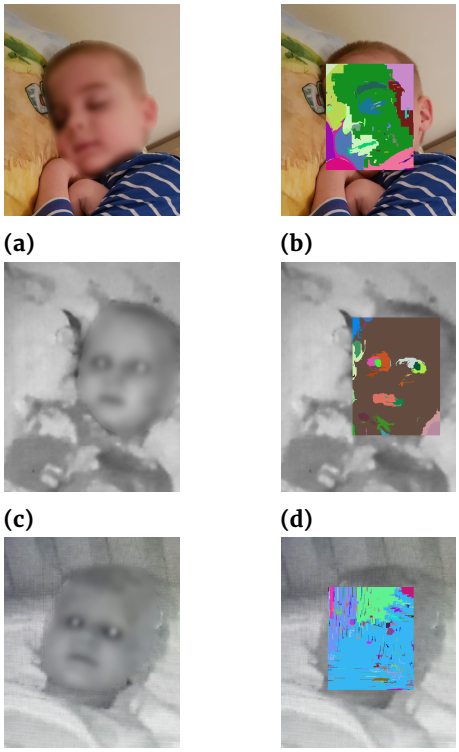
Figure 4. Image (a) and image (b) are blurred for anonymization and show the face detected from night vision input images as bounding box, as well as eye and nose position marked by points.

The usage of the module *mtcnn* is compared with an online tool for face recognition from the Microsoft Corp. (2021). For testing, there are two different data sets in use. Data set "a" is provided by the project initiator, where the images are a picture of the display of the baby phone and contains for face recognition tests 37 images. The second data set "b" is provided by students from the bio- and medical informatics study program and contains 50 face recognition test images. The results of *Face Recognition* in table 2 are compared between the two data sets and with the online tool (Microsoft Face Detection).

In table 6 you can see the results from the module *Comparison* of the two data sets. The comparison makes only sense if the Face Recognition leads to sufficient quality of results. Thus, these calculations are only performed on images where the Face Recognition works. Data set *a* contains 18 images and data set *b* 26 images for the classification of the eye state. The modules are implemented in a generic and encapsulated way, so the exchange and improvement of the module *Face Recognition* can be done in an independent way.

Table 3. Accuracy of the classification of the eye state for two data sets.

	Accuracy *Eye State classification*
Data set *a*	94.44%
Data set *b*	57.69%



(a) (b) (c) (d) (e) (f)

Figure 5. Illustration of the Graph Segmentation pre-fragmentation. Images (a), (c), (e) are blurred for anonymization. The Graph Segmentation is applied to the detected face bounding boxes only to reduce the computational effort. For open eyes as seen in (c) and (e), the resulting segmentation in (d) and (f) shows almost circular-shaped regions. For closed eyes as seen in (a), the segmentation results in a larger region for the left eye of the person and a narrower region for the right eye.

5.2. Pre-Fragmentation with Graph Segmentation

To allow for multivariate feature analysis as described in the methodology section of this paper, a robust pre-fragmentation utilizing Graph segmentation is inevitable. The parameters for the Graph segmentation are tuned to feature a reasonable number of resulting regions, i.e. especially having the opened eyes discriminated as congruent regions, see figure 5 (d) and (f). With closed eyes, the eyes can tend to merge with other face areas, cf. figure 5 (b).

5.3. Classification of the Eye State

The classification of the eye state uses the probability density function for each feature and each model. The values of the new image are compared with a model containing the feature values of the state "Open Eyes" and with a model for the state "Closed Eyes". With these calculations and comparison of each feature, the mean accordance to each model can be calculated.

(a) (b)

Figure 6. Image (a) showing a child with opened eyes and image (b) with a child having the eyes closed. Night vision images as basis for eye state classification. Both images are blurred for anonymization.

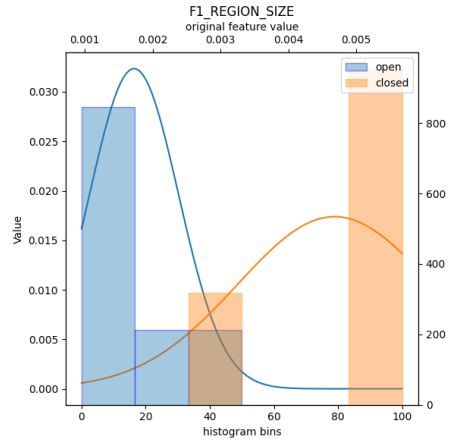
The features to be evaluated per region $\mathcal{R}_j$ are shown for the state models $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$ in figure 7 and figure 8. All of these features show a high level of selectivity, thus leading to a small Bayes error. Especially the mean relative intensity/brightness level (F_3, 8.b) as well as the mean relative region size (F_1, 7.a) besides the sphericity (F_7, 7.b) prove to be adequate for classification. Combining these features in the sense of multivariate feature analysis, a clear distinction between the eye states open/closed becomes possible.

For the final decision between the open/closed eye-state, several candidate regions $\mathcal{R}_j$ around the potential eye position can be taken into consideration. To allow for robust selection of the most relevant regions for the left and right eye, the two meta features from figure 9, namely F_4 (a) and F_5 (b) are utilized.

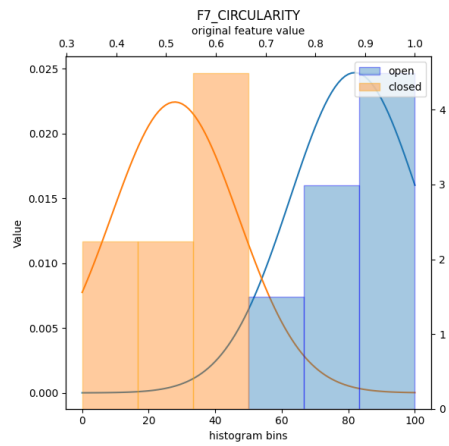
The local features allow for a high single feature accuracy, cf. table 4, while the meta features F_4 and F_5 are utilized for determination of eye regions only, cf. table 5.

Table 4. Single local feature classification accuracy for discrimination of $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$.

	F_1	F_2	F_3	F_6	F_7
acc.	0.9132	0.8527	0.9652	0.8507	0.9091



(a)



(b)

Figure 7. Comparison of the local features F_1 (a) and F_7 (b) calculated from reference images with a person showing eyes in state open/closed. The feature statistics of the training data set are thereby illustrated as histogram. Thereby, the model distribution as $\mathcal{X}_{eyeOpen}$ as well as $\mathcal{X}_{eyeClosed}$ show a low level of overlapping only. These features are calculated per region $\mathcal{R}_j$ and combined to further reduce the Bayes error.

Table 5. Single meta-feature classification accuracy for discrimination of $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$.

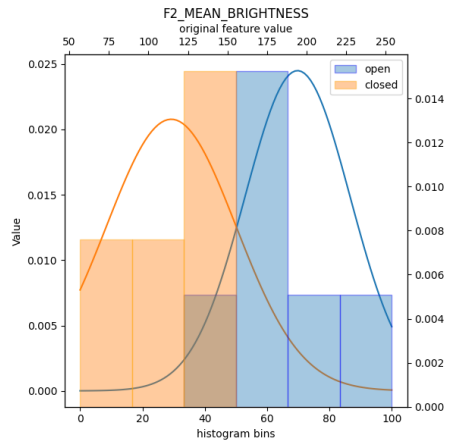
	F_4	F_5
acc.	0.6758	0.6930

Table 6. Classification of the eye state of figure (a) with open eyes and figure (b) with closed eyes.

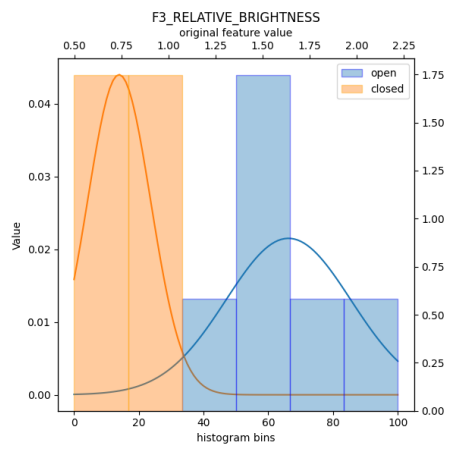
	Open Eyes	Closed Eyes
Figure a - open eyes	0.8545	0.2408
Figure b - closed eyes	0.3587	0.5004

6. Discussion

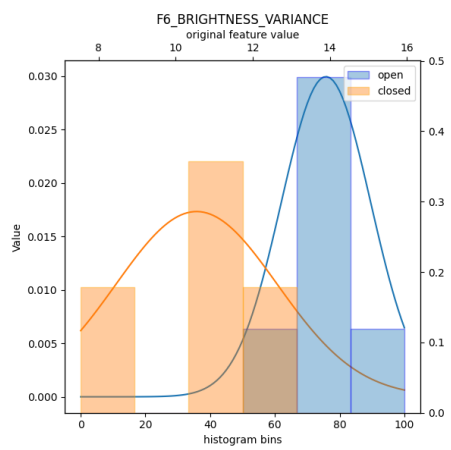
As shown in the table 2, the results are in need of improvement regarding stability of the face detection. The position of the cameras is crucial for the accuracy of



(a)



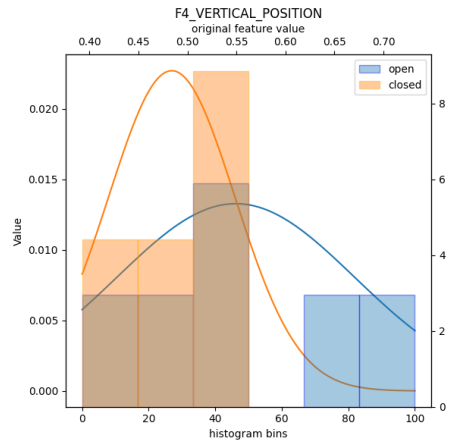
(b)



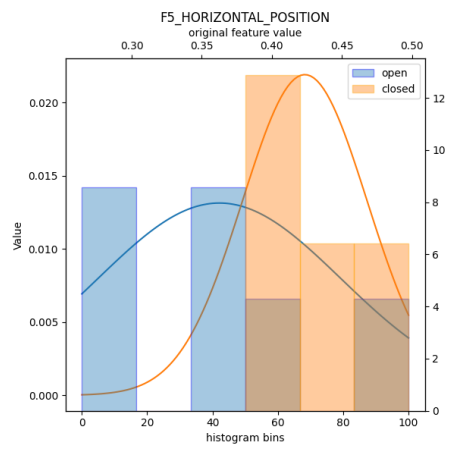
(c)

Figure 8. Comparison of the local brightness/intensity features F_2 (a), F_3 (b) and F_6 (c) calculated from image with a person showing eyes in open and closed state, respectively.

the module. The quality of the images of data set *a* is not that good, because the images are taken by the smartphone from the display of the baby phone. The images of the student group in data set *b* are better in



(a)



(b)

Figure 9. The meta-features F_4 (a) and F_5 (b) for relative horizontal and vertical position show a very high level of overlap. They are not intended to discriminate the eye state $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$ respectively per region $\mathcal{R}_j$. Instead, candidate regions for left and right eye can be selected in a more robust way.

quality but the position of the camera can be enhanced for further improvement. As you can see the results are not as good as the online tool from Microsoft Face Detection. The reason why we used the module *mtcnn* instead of the Microsoft tool was the crucial data privacy aspect of the private images in sleeping situations. Another reason was the small impact of the module which should run on the Raspberry Pi 4 in future.

Regarding classification, the results for data set *a* are quite promising as shown for features F_1 and F_3. The default models $\mathcal{X}_{eyeOpen}$ and $\mathcal{X}_{eyeClosed}$ allow for a high level of predictability, i.e. showing almost no overlap. Consequently, there is no absolute need to interpret the sparse bag of features vector as a multivariate distribution.

In future usage, the user can provide a certain specifically-trained model of the concerned person, which may provide better results. Consequently, uti-

lizing specific custom models, variability of the image acquisition hardware can be compensated for. For the comparison process, the brightness of the images are relevant as you can see from the results of the data set *b*. Besides, region size and sphericity are good features to identify closed eyes. With eyes closed, the Graph segmentation cannot precisely discriminate the eyes often leading to massive over-segmentation. This significant deviation in region size and circularity can be precisely detected by utilizing features $\mathcal{F}_1$ and $\mathcal{F}_7$, respectively.

The results are not perfectly satisfying because the images with the modern camera setup are too bright and thus sometimes falsify the result of the state of the eyes. For this problem a further approach is to normalize the histogram of the image before calculating the state or to a priori dampen the lighting emission by applying a sun-glass-like foil onto the camera objective.

7. Outlook

While this paper focuses on a research prototype for analysing single night vision images for both, model training and eye state classification, future work will focus on the evaluation of entire video sequences in real-world scenarios. Thereby, in a first phase, different sleeping locations with heterogeneous camera distances and lighting conditions will be evaluated. The healthy testing persons thereby will reconstruct different emergency situations according to a script prepared from medical expert knowledge to simulate relevant emergency situations. From a usability point of view the key focus will thereby be laid onto the balance between sensitivity and specificity of the detection system besides human-computer-interaction for training of custom models.

If these testing series can be evaluated in a positive way, subsequent clinical studies for actual epileptic patients can be taken into consideration for future.

8. Funding

Our thanks to the Austrian Research Promotion Agency FFG for facilitating project *EpiMon* with innovation check funding program, fed by research budget of the Federal Republic of Austria.

9. Acknowledgements

Many thanks to the project initiator and business partner, Mr. *Christian Seidl*, for visionary spirit of enterprise, providing test data and his practice-oriented expertise.

Many thanks to Priv. Doz. Dr. *Gudrun Gröppel* from the Kepler University Clinics in Linz for accompanying this project as medical specialist in neurology and neuropaediatrics.

Special thanks to the highly motivated and enthusiastic students from the bio- and medical informatics study program, represented by *Clara Diesenreiter*, for great vision and effort regarding development of the research prototype.

Further thanks to *Andreas Pointner* for valuable input on the code quality and implementation performance by providing precise code reviews.

References

Asano, E., Pawlak, C., Shah, A., Shah, J., Luat, A. F., Ahn-Ewing, J., and Chugani, H. T. (2005). The diagnostic value of initial video-eeg monitoring in children—review of 1000 cases. *Epilepsy Research*, 66(1):129–135.

Atanassov, E. and Dimov, I. T. (2008). What monte carlo models can do and cannot do efficiently? *Applied Mathematical Modelling*, 32(8):1477–1500. Special issue on numerical and computational issues related to applied mathematical modelling.

Backhaus, K. (2008). *Multivariate Analysemethoden*. Springer, Berlin, Germany, 12 edition.

Baumgartner, c., Feichtinger, M., and Pataraia, M. (2019). Dfp-literaturstudium: Epilepsie. *Österreichische Ärztezeitung*, (19).

Beucher, S. and Lantuéjoul, C. (1979). Use of watersheds in contour detection. *In International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation*, 132.

Dua, M., Singla, R., Raj, S., Jangra, A., et al. (2021). Deep cnn models-based ensemble approach to driver drowsiness detection. *Neural Computing and Applications*, 33(8):3155–3168.

Epitech GmbH (2021). Epilepsieüberwachung epicare.

Felzenszwalb, P. F. and Huttenlocher, D. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59:167–181.

Fitriyani, N. L., Yang, C.-K., and Syafrudin, M. (2016). Real-time eye state detection system using haar cascade classifier and circular hough transform. In *2016 IEEE 5th Global Conference on Consumer Electronics*, pages 1–3. IEEE.

Forsgren, L., Beghi, E., Õun, A., and Sillanpää, M. (2005). The epidemiology of epilepsy in europe – a systematic review. *European Journal of Neurology*, 12(4):245–253.

Gou, C., Wu, Y., Wang, K., Wang, K., Wang, F.-Y., and Ji, Q. (2017). A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 67:23–31.

HELLOBABY Ltd. (2021). ellobaby hb25 wireless digital video baby monitor video recording.

Hong, T., Qin, H., and Sun, Q. (2007). An improved real time eye state identification system in driver drowsiness detection. In *2007 IEEE International Conference on Control and Automation*, pages 1449–1453.

Izenman, A. J. (2008). *Modern multivariate statistical techniques : regression, classification, and manifold learning*. Springer, New York, USA.

LivAssured B.V. (2021). Epilepsie und sicherer schlaf.

MacDonald, B., Cockerell, O., Sander, L., and Shorvon, S. (2000). The incidence and lifetime prevalence of neurologic disorder in a prospective community-based study in the uk. *Brain : a journal of neurology*, 123 (Pt 4):665–76.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India*, 2:49–55.

Masterwork Aoitek Tech Corp Ltd. (2021). Lollipop care - sleep tracking beta.

Microsoft Corp. (2021). Facial recognition.

Pandey, N. N. and Muppalaneni, N. B. (2021). Real-time drowsiness identification based on eye state analysis. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 1182–1187.

Prévost, J. L. (1868). *De la déviation conjugée des yeux et de la rotation de la tête dans certains cas d'hémiplégie*. PhD thesis.

Sander, L. (2003). The epidemiology of the epilepsies revisited. *Current opinion in neurology*, 16:165–70.

Schulc, E., Hilbe, J., Saboor, S., Ammenwerth, E., Unterberger, I., and Them, C. (2009). Automatische detektion epileptischer anfälle basierend auf beschleunigungsmessungen - literaturübersicht. *PRInternet*, 10:517–525.

Tian, Z. and Qin, H. (2005). Real-time driver's eye state detection. In *IEEE International Conference on Vehicular Electronics and Safety, 2005.*, pages 285–289.

Tsuchie, S. Y., Guerreiro, M. M., Chuang, E., Baccin, C. E., and Montenegro, M. A. (2006). What about us?: Siblings of children with epilepsy. *Seizure*, 15(8):610–614.

Zahn, C. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers*, C-20(1):68–86.

Zhang, F., Su, J., Geng, L., and Xiao, Z. (2017). Driver fatigue detection based on eye state recognition. In *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, pages 105–110.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.

Zwettler, G. A. (2014). *General Model-Based Segmentation Strategy for Holistic Analysis of Tomographic Medical Image Data in 3D Radiology*. PhD thesis, University of Vienna.