

Driver Fatigue Detection Based on Real-Time Eye Gaze Pattern Analysis

Wilbert G. Aguilar^{1,3,4(✉)}, Jorge I. Estrella², William López³,
and Vanessa Abad⁵

¹ Dep. DECEM, Universidad de las Fuerzas Armadas ESPE, Sangolquí,
Ecuador

wgaguilar@espe.edu.ec

² Dep. Eléctrica y Electrónica, Universidad de las Fuerzas Armadas ESPE,
Sangolquí, Ecuador

³ CICTE Research Center, Universidad de las Fuerzas Armadas ESPE,
Sangolquí, Ecuador

⁴ GREC Research Group, Universitat Politècnica de Catalunya,
Barcelona, Spain

⁵ Universitat de Barcelona, Barcelona, Spain

Abstract. This paper introduces a real time non-intrusive method to determine driver fatigue by analyzing eye gaze patterns. Using a standard webcam and a personal computer, the proposed method combines different techniques in order to keep a low computational cost without a loss of performance. Facial features are identified in a reference image to extract a region of interest, around the eyes of the user, and tracked by an optical flow algorithm in subsequent frames. Color segmentation on the resulting images allow the system to extract data needed to determine ocular following, blink detection, frequency, and percentage of eye lid closure over time (PERCLOS). This approach, while simple, proves to be very efficient and accurate for the hardware restricted setup, allowing faster information processing on modest specifications systems. For safety reasons, our experiments are limited to different subjects, simulating fatigue in laboratory conditions as well as a real time test on a moving vehicle to analyze the blinking patterns.

Keywords: Fatigue detection · Real-time · Driver assistance · Eye-gaze patterns · PERCLOS

1 Introduction

Over the last decade, there is a growing concern in the increased rate of road accidents. The world health organization statistics shows road injury as one of the top ten causes of death worldwide. Last year, the registered number of deaths on the road was estimated at 1.25 million, reaching the same number of deaths caused by diseases like malaria or AIDS. The amount of road accidents has led to the creation of the Decade of Action for Road Safety in 2011 and is a case of study for the Global Status Reports on Road Safety (GSRRS). The Sustainable Development Goals include a target of 50% reduction in road traffic deaths by 2020 through campaigns on law enforcement,

education and technology. The technological field of road safety has seen a continuous development on vehicles by the research of reactive methods, like the ABS breaking systems, and proactive methods, like road, driver monitoring, and autonomous navigation systems [1–5]. It is of particular interest to further develop the latter, considering almost 50% of the accidents involve car occupants and nearly 20% are related to driver fatigue [6, 7]. Driver fatigue is characterized by reduced alertness associated with diminished cognitive and motor performance of an individual [8] and can be monitored to an extent by various methods. Driver physical and physiological measures have been the subject of several studies, and vary according to their implementation [9]. The most relevant are, intrusive methods, which rely on biometrics like electroencephalogram (EEG) and electrocardiograph (ECG) signals [10, 11]; semi intrusive methods like eye tracking systems embedded on glasses, as seen in commercial applications such as Optalert or SmoothEyes; and non-intrusive methods, based on remote sensing and analysis of data. While intrusive and semi-intrusive methods present a highly reliable solution to the fatigue detection issue, their invasive nature and high price range turns away the population most affected by road accidents (low- and middle-income countries [12]). The rapid increase in computational power on compact electronic devices is starting to give remote sensing systems an edge over intrusive ones by reducing implementation costs and the ability to retrofit in any vehicle. In this scenario, vision based driver fatigue systems should continue to be researched to make them affordable and reliable worldwide.

At present, computer vision based systems [13, 14] can track several facial cues to determine the state of alertness and drowsiness of the driver. Some approaches focus on head position and face orientation by tracking feature points on subsequent frames [15] or use color segmentation techniques to identify facial expressions [16]. Other cues like nodding and yawning have also been analyzed in driver fatigue [17]. Finally, most recent studies focus on eye gaze features like eyelid movement, blinking frequency and percentage of eye closure over time (PERCLOS) [18], addressing more accurately the micro-sleep issue [19]. An updated survey of current and ongoing investigations on this subject can be found at [20]. While most of these methods use near-IR cameras to track the eye gaze features and patterns, few specify the limitations it has on daylight use [9, 18]. Considering various researches show that there are two daily peaks when fatigue accidents occur due to cognitive and motor performance, one early in the morning (2:00 AM to 6:00 AM), and the other between 2:00 PM and 4:00 PM [21, 22], the use of conventional cameras and methods should not be discarded.

In this paper, we propose a non-invasive fatigue monitoring system based on eye gaze activity and estimation patterns using a standard single webcam connected to a mobile computer. Our approach focuses on mobility and performance of common hardware to determine the accuracy and response of the methods in a real environment. The system starts by acquiring a single reference image of the driver where face and eyes are identified by the well-known Viola-Jones method [23]. This image is used to find key points in the eye region of the driver using the ORB algorithm [24]. After classifying the reference key points, they are introduced in a Kanade Lucas Tomassi (KLT) algorithm to track the eyes of the driver on subsequent frames. The optical flow of the points in the eye region allows the system to determine ocular following while a color map segmentation creates a binary mask to extract the pixel area of the eyes as

well as the blinking patterns of the driver. Experiments involving a group of 5 persons aged 25–65, in laboratory conditions and on the road, show a high accuracy of blinking patterns in both scenarios, which allow us to calculate PERCLOS. It is to be noted that this method requires a steady frame rate and a low execution time to be reliable, which is the main reason we decided to focus on computational cost and performance, metrics rarely reported in current literature [20].

2 Our Proposal

The general architecture of our system can be divided in three stages. (1) Feature extraction, (2) Eye region tracking and (3) Eye gaze pattern analysis.

2.1 System Initialization (Feature Extraction)

The system starts by acquiring a reference image of the test subject inside the vehicle while his level of attention is high under stable lighting conditions, to quickly identify features by conventional methods.

The captured image goes through a facial feature detection algorithm and will be labeled as reference image once it contains all the necessary parameters to start the tracking algorithm. If it does not meet this criteria, another image will have to be captured and processed.

Over the last decade the Viola-Jones method has been continuously researched and well documented on the facial features detection field. By using a series of boosted cascades of classifiers to identify simple features this method proved to be very efficient on static images [23]. However, with a video input, the rate at which images are analyzed drops considerably, more so when trying to identify specific features like eyes, lips or nose. In a real time facial feature detection system, where lighting conditions and background content varies frame to frame, this results in a considerable decrease in performance over time. For this reason, we do not use the Viola-Jones method to track the facial features. Instead we only use it to identify the face and the eyes of the driver, creating a mask image to find key points which can be later tracked by a more efficient method. This image is obtained either on system start up or after a conditioned reset to re-acquire the reference parameters. Figure 1 shows the different stages to obtain the reference image using the Viola-Jones algorithm.

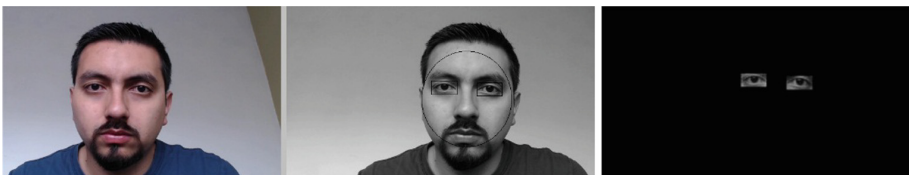


Fig. 1. Facial feature extraction using the Viola Jones algorithm. From left to right: input image is converted to greyscale to identify the face and the eyes of the subject. Image on the right shows the “eyes mask image” obtained.

After extracting the eye mask, points of interest are extracted to finally determine if the image acquired can be classified as a reference image. Feature extraction methods focus on specific regions of an image to determine whether it can be classified as a feature or not. They differ in aspects such as attribute invariance, compute efficiency, robustness among others. A survey on the various methods to extract features and descriptors can be found in [25]. For our application, the most convenient method was ORB due to its open source nature, low processing cost and fast execution time. As stated in [25], “ORB is a highly optimized and very well engineered descriptor, since the ORB authors were keenly interested in compute speed, memory footprint and accuracy.” ORB can be used in real-time practical applications [26–29]. However, like most feature extractors and descriptors, ORB by itself is not designed to match or track specific facial features frame to frame. In order to do so, some pre-processing of the images is needed to help the detector find the interest points, increasing the computational cost and reducing the frame rate. Considering the latter, we used ORB only to extract the feature points in a reference image candidate.

We set up the ORB algorithm to extract 300 key points within the eye region. The Features From Accelerated Segment Test (FAST) extractor inside ORB, filters them and we select 50 key points with the highest response rating. The response rating specifies how strong a key point is, according to:

$$R = \text{Det}(\mathbf{H}) - k(\text{Trace}(\mathbf{H}))^2$$

Where,

R is the response rating

H is the matrix of the sums of the products of derivatives at each pixel

$$\mathbf{H}(x, y) = \begin{pmatrix} S_{x2}(x, y) & S_{xy}(x, y) \\ S_{xy}(x, y) & S_{y2}(x, y) \end{pmatrix}$$

The higher the response value, the more likely the feature will be recognized among several instances of an object. In our system, the key points stack around the eye contour, easing the identification of outliers to set up reset conditions later on (Fig. 2).

The feature and key point extraction flowchart can be seen in Fig. 3.

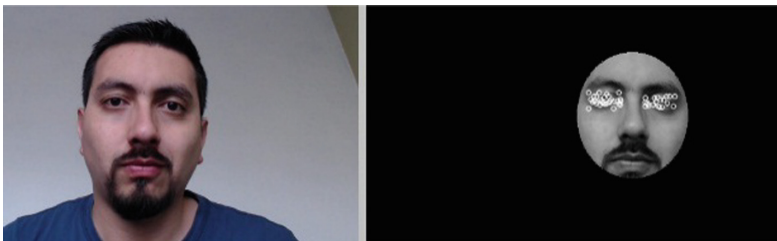


Fig. 2. Key point extraction in the eye mask image using the ORB algorithm. 50 key points are selected from a set of 300 points.

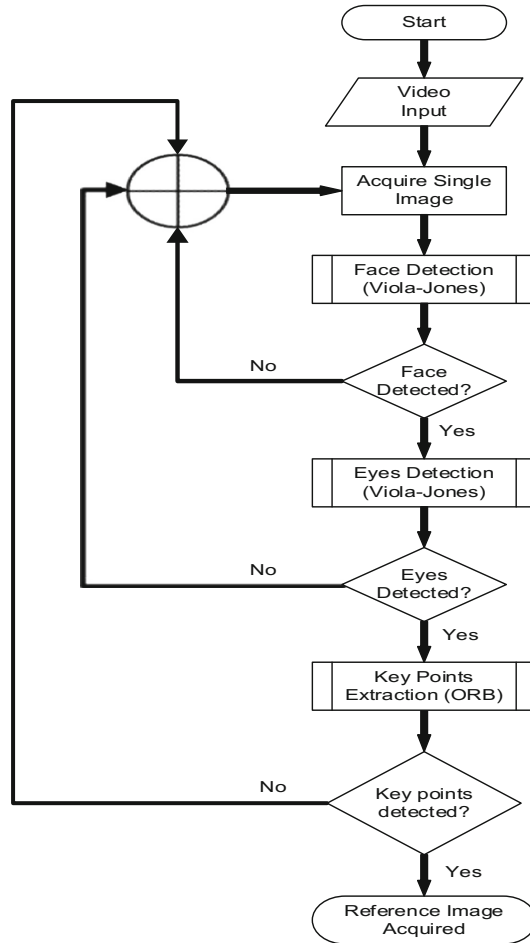


Fig. 3. Feature extraction algorithm flow chart. Several conditions have to be met before declaring an image as Reference Image.

2.2 Eye Region Tracking

Once the key point extraction is successful, the resulting points are inserted into an optical flow loop to track them over time in consecutive frames. We chose the KLT method for our purpose due to the nature of the experiment. The tracking of the facial features of a driver consists in slow head displacements and small eye movements, a temporal persistence. The brightness of the analyzed frames is rather constant in time, and the points we extracted before in the k instant ($S_k[j]$, $j = 50$) are classified according to the spatial coherence of the reference image.

By definition, the KLT algorithm relies on the local information of the surroundings of a key point to determine the relative position of the same point in subsequent frames. The surroundings of every key point in the k instant will be analyzed in a 30×30

pixels window, any size larger will improve the tracking at the expense of increased computational cost as stated in [30].

To extract both eye regions in every frame, we classify the coordinates of each tracked key point in the instant $k + 1$ ($S_{k+1}[j], j = 50$) in ascending order. A rectangular region of interest around the contour of each eye is drawn by using the maximum and the minimum of the (x, y) value of the coordinates, as shown in Fig. 3. We need to separate the eye regions for two reasons; first, to control the parameters of each eye, such as brightness level and color segmentation; second, to determine positive tracking by the amount of outliers in the region of interest (Fig. 4).

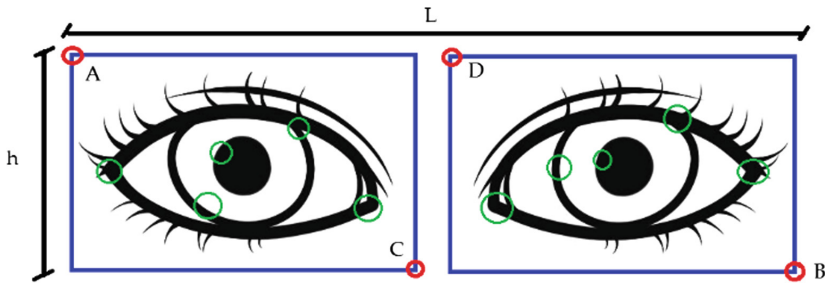


Fig. 4. Diagram of the eye regions of interest. In green, the tracked points, in red, the minima and maxima of the key points coordinates. (Color figure online)

The Left Eye Region (LER) and Right Eye Region (RER) are defined by two corners:

$$LER \left\{ \begin{matrix} A(X_{min}, Y_{min}) \\ C(X_{min} + \frac{X_{max}-X_{min}}{2}, Y_{max}) \end{matrix} \right\}; RER \left\{ \begin{matrix} D(X_{max} - \frac{X_{max}-X_{min}}{2}, Y_{min}) \\ B(X_{max}, Y_{max}) \end{matrix} \right\}$$

After extracting the eye region in real time, we create a mask containing LER and RER, the Tracked Eye Region (TER). The height and length of the eye region in the reference image are stored as reference dimension values, and allow us to estimate ocular following of the driver by comparing the dimensions of the TER with the reference eye region. If the values in the x or y axis vary over a threshold value, it means the distance between the reference key points has changed over time too, according to the optical flow equation:

$$\varphi_{j,k+1} = \frac{(P_{j,k+1} - P_{j,k})}{T_s}$$

Since the tracked key points remain stacked around the eye region, then the variation of the distance, over the threshold, between two points $P_{j,k+1} - P_{j,k}$ can be interpreted as a movement of ocular following, this is, if a person moves the head towards a point in the borders of his view range, the eyes will follow the same direction in order to maintain awareness. Using the same reasoning, we defined the limits of the distance variation to maintain the eye pattern analysis relevant only when the TER is within them.

2.3 Eye Gaze Pattern Analysis

The state of alert of the driver can be estimated by several factors such as distraction, drowsiness and fatigue. Our method can identify distraction through ocular following, but the most accurate way to determine state of alert, is through monitoring drowsiness and fatigue levels. We propose to use two parameters to determine driver fatigue: blinking patterns and PERCLOS.

To determine the blinking frequency and duration, the eye Region of Interest (ROI) of the reference image is transformed from the Blue Green Red (BGR) to Hue Luminosity Saturation (HLS) color space. By modifying the value of the L channel, we successfully threshold the eye ROI to obtain a binary image where the eye contour is represented by white pixels, as seen in Fig. 5.



Fig. 5. Top: TER with open eyes converted to the HLS color space. Bottom: TER with closed eyes converted to the HLS color space

A pixel count in the area formed by the eye mask sets the reference value of the non-zero pixels when the driver eyes are open (NZ_{Ref}). The same procedure takes place in successfully tracked eyes ROI ($NZ_{Tracked}$). Lower white pixel values indicate driver eyelid closure, while values well above the reference set point indicate the tracking conditions have been altered and a reset is required. The NZ values over time form a curve as (Fig. 6):

Studies on measuring the eye closure have determined the average time of normal blinks to take less than 300 ms, while slow blinks take over 300–500 ms [31] Since we are working at 20FPS, each frame is analyzed every 50 ms, enough time to distinguish both types of blinks.

“PERCLOS is the percentage of eyelid closure over the pupil over time” and several studies demonstrate the high correlation between this measure and the level of fatigue of an individual [32], according to:

$$PERCLOS = \frac{t_c}{t_c + t_o}$$

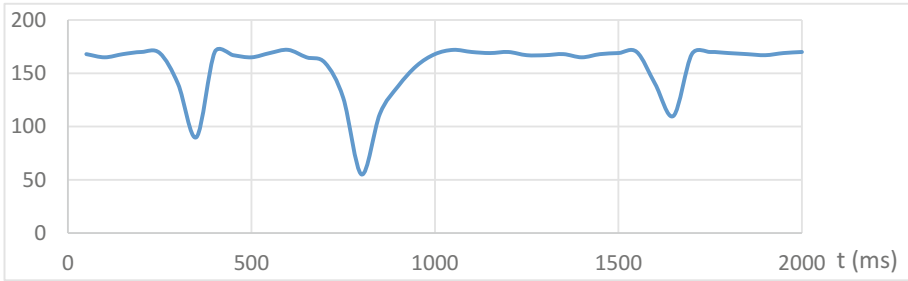


Fig. 6. White pixels over time in the TER of the driver. Valleys represent blinks, plateaus represent open eyes.

Where t_c is the time the eyes are closed and t_o is the time the eyes are open. The common time frame to compute this measure is between 30 s and 1 min.

By calculating the area covered by each eye in every incoming binary image, and the variation of it in relation to the reference image, the system determines blinking frequency and duration accurately enough to calculate blinking duration, frequency and PERCLOS.

2.4 Reset Conditions

To keep the information relevant throughout time, the algorithm resets on several conditions. A soft reset replaces the key points in the tracking algorithm with the reference key points. A hard reset attempts to acquire a new reference image, calculates new key points and new HLS values for the new tracking parameters.

- Maximum number of outliers allowed reached inside the KLT tracking algorithm: occurs on partial occlusion, on driver distraction and on large lighting variations. A hard reset is required.
- Number of frames limit reached (refreshing): soft reset.
- The area of the eye region greatly exceeds the reference value: the Luminosity parameter on the HLS color map is recalculated.

3 Results

3.1 System Performance

The main focus of our research is to validate an algorithm able to perform on low cost commercial devices without losing accuracy. Although it can still be optimized, our algorithm worked using 6% of a single thread in the CPU and 125 MB of RAM memory. Without needing to resize the input images, our approach works at 16.1 FPS on a live video feed using the webcam and at 20.9 FPS working on a pre recorded video. Table 1 compares this results with the Viola Jones implementation, extracting facial features on every frame instead of tracking it.

Table 1. Average Frames Per Second using the Viola Jones method and using the KLT tracker + color segmentation method.

	Viola Jones	Our approach
Webcam IN	9.8	16.1
Recorded video IN	13.1	20.9

The execution time of the different parts of our algorithm is detailed in Table 2.

Table 2. Execution time of the various methods used in our system in comparison to the state of the art.

Execution time [ms]				
Feature extraction		Eye detection/tracking		Eye gaze analysis
Shi - Tomassi	ORB	Template matching	KLT tracker	
27	6	18	8	5

By adapting modern methods like ORB to the tracking optical flow algorithm, we managed to reduce the total execution time of our method to an average 42.5 ms/frame, guaranteeing real-time performance.

3.2 Detection and Tracking

Table 3 compares detection rate and tracking of facial features over time using our approach and using the Viola–Jones method on every frame. The recordings were segmented in 5 min videos.

Table 3. Identification and tracking of facial features over time.

Identification/tracking average rate [%]		
Subject	Viola Jones	ORB + Optical flow
1	35.7	92.1
2	38.5	90
3	47.8	96.4
4	42.1	93.8
5	44.9	96.1

Using the KLT tracking approach, the detection rate of facial features in subsequent frames is highly reliable due to the partial tracking of features that takes place during periods of occlusion. Soft reset conditions keep the eye region relevant throughout time during long periods of occlusion and allow rotation of the face to a certain degree before losing the tracking points. This allows our method to constantly compare the area of the eye region in subsequent frames for the PERCLOS calculation. If the eye region is not tracked on a high rate, the blinking pattern algorithm is rendered inaccurate.

3.3 PERCLOS and Blinking Parameters Calculations

The average accuracy rating of blinking and PERCLOS of our system is shown in Table 4 for every subject. It is to be noted that, while blinks can be counted to evaluate the detection rating, the PERCLOS measure cannot be manually measured. Instead we decided to calculate the number of false positives due to false tracking.

Table 4. Average accuracy rating (%).

Subject	Blinking	PERCLOS
1	97.2	98.1
2	93.4	92.7
3	96.7	95.3
4	96.1	94.2
5	95.4	95.8

4 Conclusions

Our proposal for driver fatigue detection proved to be very cost effective, reducing the execution times compared to state of the art methods on standard commercial hardware. Real-time performance at over 15 FPS was achieved without reducing the quality or the size of the input image. The experiment on vehicle showed that false positives encountered during the calculation of PERCLOS are mainly due to false tracking. This can occur during fast and strong lighting changes, when crossing below a bridge for instance, or during partial occlusion moments. To prevent this, we only calculate eye gaze parameters while the dimensions of tracked eye region are similar to the reference ones. A better tracking of the features would improve the PERCLOS calculation by having a more constant influx of data.

References

1. Aguilar, W.G., Verónica, C., José, P.: Obstacle avoidance based-visual navigation for micro aerial vehicles. *Electronics* **6**(1), 10 (2017)
2. Aguilar, W.G., Verónica, C., José, P.: Obstacle avoidance for low-cost UAVs. In: IEEE 11th International Conference On Semantic Computing (ICSC), San Diego (2017)
3. Aguilar, W.G., Morales, S., Ruiz, H., Abad, V.: RRT* GL based optimal path planning for real-time navigation of UAVs. In: Rojas, I., Joya, G., Catala, A. (eds.) IWANN 2017. LNCS, vol. 10306, pp. 585–595. Springer, Cham (2017). doi:[10.1007/978-3-319-59147-6_50](https://doi.org/10.1007/978-3-319-59147-6_50)
4. Aguilar, W.G., Morales, S.: 3D environment mapping using the Kinect V2 and path planning based on RRT Algorithms. *Electronics* **5**(4), 70 (2016)
5. Cabras, P., Rosell, J., Pérez, A., Aguilar, W.G., Rosell, A.: Haptic-based navigation for the virtual bronchoscopy. In: 18th IFAC World Congress, Milano, Italy (2011)
6. The Royal Society for the Prevention of Accidents: The Royal Society for the Prevention of Accidents. <http://www.rospa.com/road-safety/advice/drivers/fatigue/road-accidents/>. Accessed 17 Aug 2016

7. Road Safety Observatory: Road Safety Observatory. <http://www.roadsafetyobservatory.com/Summary/drivers/fatigue>. Accessed 17 Aug 2016
8. Williamson, A., Friswell, R.: Exploratory study of fatigue in light and short haul transport drivers in NSW Australia. *Accid. Anal. Prev.* **40**(1), 410–417 (2008)
9. Dawson, D., Searle, A.K., Paterson, J.L.: Evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Med. Rev.* **18**, 1–12 (2013)
10. Zhao, C., Zhao, M., Liu, J., Zheng, C.: Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accid. Anal. Prev.* **45**, 83–90 (2012)
11. Saroj, A.C., Lal, K.L.: A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* **55**, 173–194 (2001)
12. World Health Organization: Global Status Report on Road Safety (2015). http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/. Accessed 17 Aug 2016
13. Aguilar, W.G., Luna, M.A., Moya, J.F., Abad, V., Ruiz, H., Parra, H., Angulo, C.: Pedestrian detection for UAVs using cascade classifiers and saliency maps. In: Rojas, I., Joya, G., Catala, A. (eds.) IWANN 2017. LNCS, vol. 10306, pp. 563–574. Springer, Cham (2017). doi:10.1007/978-3-319-59147-6_48
14. Aguilar, W.G., Luna, M., Moya, J., Abad, V., Parra, H., Ruiz, H.: Pedestrian detection for UAVs using cascade classifiers with meanshift. In: IEEE 11th International Conference on Semantic Computing (ICSC), San Diego (2017)
15. Liu, K., Luo, Y., Gyomei, T.E.I., Yang, S.: Attention recognition of drivers based on head pose estimation. In: Vehicle Power and Propulsion Conference (2008)
16. Sanchez-Cuevas, M.C., Aguilar-Ponce, R.M., TecpanecatI-Xihuitl, J.L.: A comparison of color models for color face segmentation. *Procedia Technol.* **7**, 131–141 (2013)
17. Li, L., Chen, Y., Li, Z.: Yawning detection for monitoring driver fatigue based on two cameras. In: 12th International IEEE Conference on Intelligent Transportation Systems (2009)
18. Bergasa, L.M., et al.: Visual monitoring of driver inattention. *IEEE Trans. Intell. Transp. Syst.* **1**(1), 63–77 (2006)
19. SleepDex: “Sleepdex.” <http://www.sleepdex.org/microsleep.htm>. Accessed 18 Aug 2016
20. Al-Rahayfeh, A.M.E.R., Faezipour, M.I.A.D.: Eye tracking and head movement detection: a state-of-art survey. *IEEE J. Transl. Eng. Health Med.* **1**, 2100212 (2013)
21. Eskandarian, A., et al.: Advanced driver fatigue reasearch. US Department of Transportation (2007)
22. Horne, J., Reyner, L.: Vehicle accidents related to sleep: a review. *Occup. Environ. Med.* **56**, 289–294 (1999)
23. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Conference on Computer Vision and Pattern Recognition (2001)
24. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: ICCV (2011)
25. Krig, S.: Interest point detector and feature descriptor survey. In: Computer Vision Metrics: Survey, Taxonomy, and Analysis. Apress (2014)
26. Aguilar, W.G., Angulo, C.: Real-time model-based video stabilization for microaerial vehicles. *Neural Process. Lett.* **43**(2), 459–477 (2016)
27. Aguilar, W.G., Angulo, C.: Real-time video stabilization without phantom movements for micro aerial vehicles. *EURASIP J. Image Video Process.* **1**, 1–13 (2014)
28. Aguilar, W.G., Angulo, C.: Robust video stabilization based on motion intention for low-cost micro aerial vehicles. In: 11th International Multi-conference on Systems, Signals & Devices (SSD), Barcelona, Spain (2014)

29. Aguilar, W.G., Casalglla, V.P., Pólit, J.L., Abad, V., Ruiz, H.: Obstacle avoidance for flight safety on unmanned aerial vehicles. In: Rojas, I., Joya, G., Catala, A. (eds.) IWANN 2017. LNCS, vol. 10306, pp. 575–584. Springer, Cham (2017). doi:[10.1007/978-3-319-59147-6_49](https://doi.org/10.1007/978-3-319-59147-6_49)
30. Tomasi, C., Kanade, T.: Detection and Tracking of Point Features. Carnegie Mellon University, Pittsburgh (1991)
31. Hammoud, R.I.: Passive Eye Monitoring, Algorithms. Applications and Experiments. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-75412-1](https://doi.org/10.1007/978-3-540-75412-1)
32. Dinges, D.F., Grace, R.: PERCLOS: a valid psychophysiological measure of alertness as assessed by psychomotor vigilance. US Department of Transportation (1998)