Development of Yawning Detection Algorithm for Normal Lighting Condition and IR Condition

Nur Fatin Izzati Y., M.M. Ibrahim, Nur' Afifah S.B., N.A. Manap
Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
izzatifatin@gmail.com

***Abstract*— Drowsiness monitoring system has been widely used in this current technology to monitor the driver's state while driving. This paper presents a drowsiness detection method through the activity of yawning for both normal lighting condition and Infrared (IR) condition. Development of the algorithm consists of several steps. Initially, the detection of face and mouth implementing the Viola-Jones algorithm takes place. For IR condition, the mouth is detected by applying the geometrical measurements of a face. After the detection process is done, the tracking process for both face and mouth takes place utilizing the Kanade-Lucas-Tomasi (KLT) algorithm which is basically a point tracking algorithm. Based on the tracked mouth, the region of interest (ROI) is selected which is to be used as an input image in the image processing step in order to get a clearer image of the mouth. From the finalized mouth image in the preprocessing step, the properties of the image are further used in the yawning detection step. In the indication of yawning, the height of the mouth opening reading score is observed. The performance of the proposed method is tested on 5 subjects and achieved an overall accuracy of 98.89% for normal lighting condition and 95.29% for IR condition.**

***Index Terms*—Drowsiness Detection, Height Of Mouth Opening, IR Lighting Condition, Yawning Detection.**

I. Introduction

Recently, despite the current safety features and legislation, the statistic has shown that many fatal car accidents still occur around the entire world. One of the many reasons towards the increasing number of car crashes is drowsy drivers which have become a crucial issue to the society. According to the statistic produced by 'Polis Diraja Malaysia, PDRM', 240,703 cases were recorded from January to June 2015 has shown an increment of 6,942 cases from 233,761 cases recorded from the previous year. From this huge amount, 45% of the cases were recorded due to drowsy drivers.

Drowsy can be considered as the weakness lethargy that frequently happens either due to mental or physical health issues. This symptom may occur due to the immoderate physical activity or lack of quantity and quality of sleep. A drowsy driver has a very low-level concentration, thus make it difficult for them to maneuver the car as they have a high tendency to fall asleep without realizing. Yawning is one of the face features that are always used as an indication of drowsy in drowsiness detection beside the eye. It is an involuntary action of mouth opening triggered when a person feel sleepy or bored. Various techniques have been introduced in detecting yawning. Some approaches exploited the feature of the mouth such as the edge of the mouth, the shape of mouth and the intensity of mouth and some approaches employed machine learning algorithm in detecting yawning.

M.M. Ibrahim et. al. [1] utilized the darkest region of the mouth between the lips in order to detect yawning. Yawn ratio is computed between the height of focus mouth region and the height of mouth based on the darkest region. The mouth is classified as open if the yawn ratio value is higher than a certain threshold. N. Alioua et. al. [2] proposed a new edge detector that will be able to detect only a wide open mouth. With the aid of the remarkable intensity variations between the lips and the dark region of the mouth, the proposed edge detector managed to extract the edge of the wide open mouth. Circular Hough Transform (CHT) is implemented in contemplation to acquire the highest radius of the mouth opening to represent the wide open mouth. Yawning is indicated when the count of wide opening mouth greater than 2 seconds. By utilizing the machine learning algorithm, Zhang et al. 2015 [3] developed a yawning detection system based on neural network which possessed six input features. Four of the inputs are deduced from the computed mouth region. By computing the gradient features for both left half and right half of the mouth, yawning indicates when there are significant changes in the values. Based on the analysis, for the normal state, the gradient sum values vacillated round the value of 8 and the value remarkably increased if there yawning is detected. Both Abtahi et. al.[4] and Omidyeganeh et. al. [5] implement the computation of the histogram of the mouth in their yawning detection process. The histogram is obtained by evaluating the frequency number of occurences for each gray level in an image array. The first frame of the input image is appointed as reference histogram. A comparison between the reference histogram and the histogram for the next frames is compared and the back projection theory will be applied. Throughout this theory, the differences between histogram is determined and the value changes whenever the mouth goes from a normal state to yawning state.

Most yawning detection research focus on the detection itself while excluding some of the important factors that might limit the performance of the algorithm. One of the factors are lighting condition. Lighting conditions are one of the many factors that affect the appearance of the facial component in the yawning detection system. Furthermore, illumination problem has always been as one of the challenge for a system that is based on image processing techniques [6]. As the car crashes cause by drowsy drivers keep increasing annually, a robust drowsiness detection system which can work in various lighting condition is highly anticipated.

The aim of this project is to develop a yawning detection algorithm which can robustly work for both normal lighting condition and Infrared (IR) lighting condition by employing image processing techniques. The proposed algorithm has slight differences in the preprocessing technique for both lighting conditions. The yawning detection analysis is computed through the height of opening and closing of the mouth. As an indication for yawning, this project chose the height of the opening and closing of mouth for its obvious signs for drowsiness. Furthermore, the computation is fast, non-complex and can be implemented in real time. Since there are limited access to yawning database for various lighting condition, this project developed its own database and the algorithm is tested on the database.

The rest of the paper is organized as follows : Section 2 will further interpret the procedure of the proposed algorithm. Section 3 displayed the results along with the analysis and the discussion. Section 4 will provide the conclusion and further work for this project.

II. METHODOLOGY

Primarily, yawning detection algorithm procedure consists of few different stages. These stages are classified as follows and each stage will be explained in detail in the following sections:

(i) Face Detection and Tracking
(ii) Mouth Detection and Tracking
(iii) Region of Interest (ROI) Selection
(iv) Preprocessing
(v) Yawning Detection

The comprehensive system illustration is shown in Figure 1. Further explanation of each step will be clarified in the following subsections.

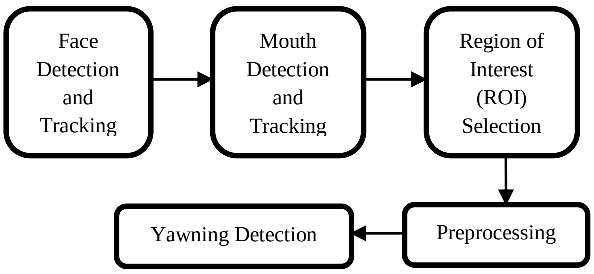


Figure 1: System Flowchart

A. Face Detection and Tracking

From the video frames, the aim of face detection is to determine entire image regions, which consist of the information such as its orientation and position. Since identifying the yawning condition in the whole frame with various light and background condition is a challenging task, thus the experiment scope was reduced by identifying the face region in the video frame. The face detection is done by employing Viola-Jones algorithm [7]. The Viola-Jones algorithm was used because this algorithm yielded fast with an accurate detection, with accuracy rates that are above 90%. Viola-Jones approach used the Cascade object detector to trace the face region in a frame. It is not impractical to use the cascade object detector on each frame, yet it might lead to costly computational. In contrast, there are limitations of using the Viola Jones; which is the result may turn inaccurate, especially when a person tilts or move their head as for example turning up and down. Due to this limitation, an approach of Kanade-Lucas-Tomasi (KLT) algorithm [8] is used for the tracking purpose. Once the face is detected, it can be used as a template to define faces components in an upcoming frame. Face tracking is the computation to allocate the face movement through a sequence of image. Face tracking aims to track and detect faces in all frames of the video. KLT algorithm has the potential to trace face components over each of the video frame with less computational cost. Besides, this approach is able to recognize faces even when the face is turned around.

The approaches used for face detection and tracking are the same for both normal lighting condition and IR lighting condition. Figure 2 displayed the successful face detection and tracking for both lighting conditions.

(a) Normal lighting condition

(b) IR lighting condition

Figure 2: Face Detection and Tracking

B. Mouth Detection and Tracking

The process is continued with mouth detection and tracking process. In contrary to face detection and tracking, this procedure has a different algorithm for normal lighting condition and Infrared(IR) lighting condition since there are intensity differences. However, for both lighting conditions, the detection and tracking of face procedure must be done initially to reduce the false detection and error tracking for mouth.

i. Normal Lighting Condition

The process of mouth detection and tracking for the normal lighting condition is exactly the same as face detection and tracking which utilized the used of Viola-Jones algorithm for detection procedure and KLT algorithm for the tracking procedure. A rectangle bounding box is drawn around the mouth region to mark the detected mouth and the tracked mouth as shown in Figure 3.

Figure 3 : Mouth detection and tracking for normal lighting condition

ii. IR Lighting Condition

The infrared images have the limitation in term of its illumination that capable to trigger out the color intensity in an image. Thus, to overcome this issue, instead of using Viola-Jones algorithm and (KLT) algorithm for detection and tracking process, an approach called the geometrical technique is used. This approach applied geometrical knowledge of face features. Compared to the normal lighting condition, where the mouth component was automatically detected using the constructor function in the KLT algorithm, the mouth bounding box for the IR lighting condition was created using geometrical formula from the detected face region. The coordinates acquired from the detected mouth area will be used to keep track of the mouth by updating the change in coordinates if there are any changes while the video frame running. The coordinates will be used to draw a rectangle bounding box to indicate the detected mouth and the tracked mouth as displayed in Figure 4.

Figure 4 : Mouth detection and tracking for IR lighting condition

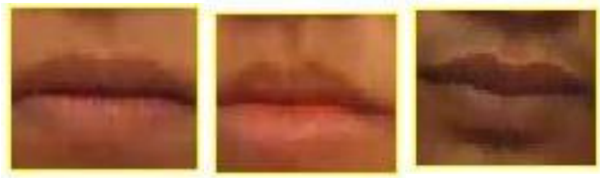
C. Region of Interest (ROI) Selection

Region of interest refers to the desired region or area that will further be used in the preprocessing steps. For this proposed algorithm, the ROI selection is focused on the mouth region which is selected from the mouth tracking process. Based on Figure 5, the image is cropped using the mouth tracking coordinates and is further used for the next step.

Both normal lighting condition and IR lighting condition used the same procedure in contemplation of the ROI selection.

D. Preprocessing

Preprocessing step ensure that the ROI image is processed so that the desired feature which is the mouth will be distinctive so that it can be used as an indication for yawning detection. The project proposed different algorithm for preprocessing procedure for both lighting conditions since there some illumination variation. Figure 6 and Figure 7 simply visualize the preprocessing steps for both lighting conditions.



(a) Normal lighting condition

(b) IR lighting condition

Figure 5 : ROI selection

i. Normal Lighting Condition

For normal lighting condition, the input image is transformed to grayscale image. One of the purposes of converting from RGB to grayscale images is to reduce the computational time where grayscale tends to process image faster compared to the RGB images [9]. Next, the image goes through an image enhancement process where the greyscale intensity range can be adjusted until there is a difference between the mouth opening region intensity with the background. This process is crucial since it will be used to classify the mouth state either it is yawning or not. The mouth image will then go through binarization transformation where the property of blob is utilized in order to mark the difference between an open mouth and a close mouth. Next, a morphological reconstruction of the image is done to remove noises such as small blobs or to fill in the hole inside the mouth blob. For the final step, a bounding box is drawn which is to extract useful information from the mouth blob that is used to evaluate the height of the opening and closing mouth respectively.

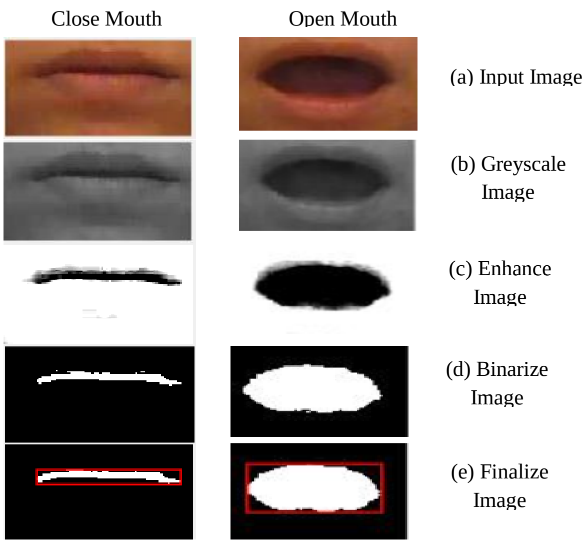


Figure 6 : Preprocessing steps for normal lighting condition.

ii. IR lighting condition

In contrary, for the normal lighting condition process, infrared images act as the input images in this process. The early preprocessing stage for IR lighting condition is the same as the normal lighting condition where the IR input image is transformed to grayscale image and then will go through an image enhancement process. Since the IR lighting images are low in luminance and invariant, Canny edge detector is chosen to filter the edge of the mouth region. Canny edge detector is selected for the image segmentation due to its reliability [9]. Once the canny edge detector is applied, the images go is transformed into a binary image. Morphological technique is applied that involved the operation of dilation and erosion process to remove all the connected components that are fewer than 200 pixels from the mouth while identifying the wide opening mouth of the yawning area respectively. In order to make the mouth region more distinctive, a morphological closing operation is performed. Similar as the normal lighting condition, a bounding box is drawn surrounding the mouth blob to compute the height of mouth opening and closing.

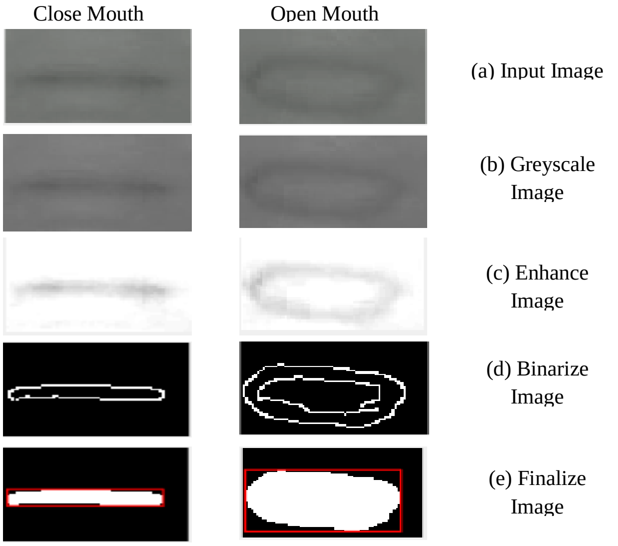


Figure 7 : Preprocessing steps for IR lighting condition

E. Yawning Detection

Basically, the developed algorithm detects yawning based on the darkest region between the lips which is computed through the height of mouth opening and closing evaluate from the height of the bounding box drawn to mark the mouth blob as displayed in Figure 6 (e) and Figure 7 (e). The height of mouth opening and closing is computed across a series of frames and a graph is plotted between the height of mouth and no. of frames to analyze yawning as shown in Figure 8. Based on the graph, the process of yawning can be identified which started with mouth completely close, the start of yawning, yawning, and the end if yawning, ending with mouth completely close. The analysis done on yawning detection has similar procedures for both normal lighting condition and IR lighting condition. However, different lighting condition obtained different value to indicate the beginning of yawning. This will further discuss in the next section.

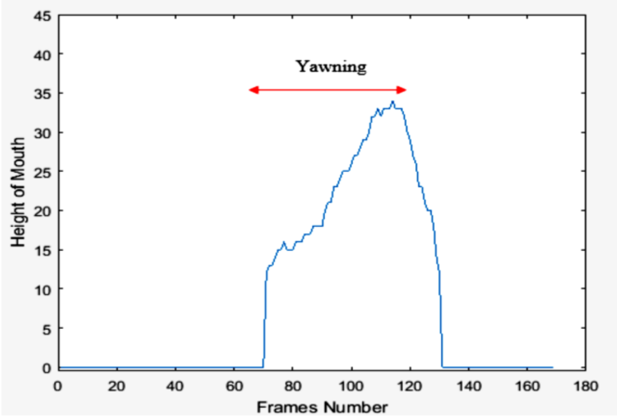


Figure 8 : Graph plotted between the height of mouth and frames number to analyse yawning.

III. Result and Analysis

The developed algorithm is tested on 5 subjects; 4 females and one male respected to the normal and Infrared (IR) lighting condition. The video footage is acquired by placing an Intel Real Sense SR300 camera as shown in Figure 9, together with the monitor in front of subjects to monitor their yawning activities. The Intel Real Sense SR300 camera is developed with a built in IR camera. The subjects are presented with a video showing bunch of people yawning to trigger subjects to yawn as according to facts, yawning is a contagious activities. The video was recorded in normal lighting condition first followed by IR lighting condition.

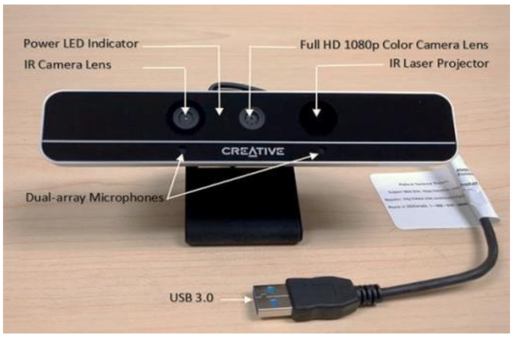


Figure 9 : Intel RealSense SR300 camera mode

As discussed earlier, the analysis for yawning detection is similar for both lighting conditions. However, both lighting condition possessed different value for the indication of the beginning of yawning. It can be seen that from Figure 10, for normal lighting condition, mouth is indicated to be closed if the height of mouth hit 0, and the process of yawning can be observed when the height of mouth showed slight increase and reached the highest peak of mouth openness which is 35 and the height of mouth is slightly decreased, ending with 0 height if mouth which informed that the process of yawning has stopped.

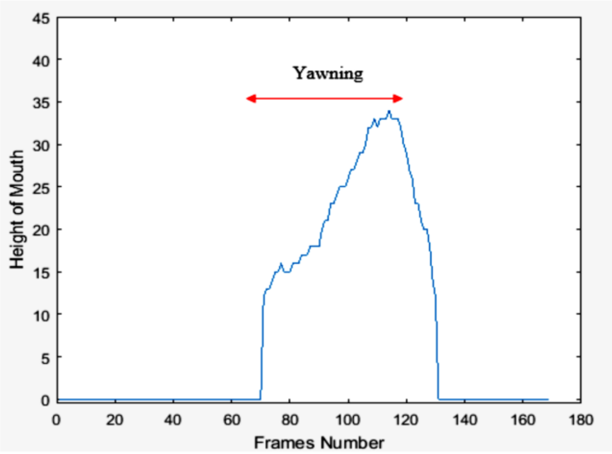


Figure 10: The process of yawning for normal lighting condition.

On the contrary, for IR lighting condition, the indication for closed mouth is varied by referring to Figure 11. It can be observed that the first frame possessed value of 17 for height of close mouth reading. Although the height of a close mouth is varied, it did not surpass the value of height of mouth openness when it is yawning, thus make it easy to analyse the yawning process.

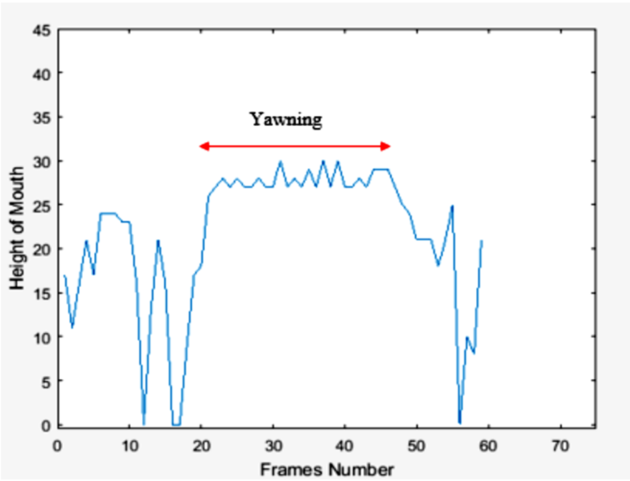


Figure 11: The process of yawning for IR lighting condition.

The issue regarding the different value for the height of the mouth for a close mouth during the yawning process is due to the edge detection technique implemented during the preprocessing step for IR lighting condition. Since there are a great intensity difference between the lips and the darkest region for normal lighting condition, the algorithm managed to acquire only the height of the darkest region. On the other side, for IR lighting condition, there are not much different for the intensity between the darkest region and the lips, thus, when Canny edge detector is utilized, it detected the edge of lips together with the darkest region inside the mouth. The visualization can be seen in Figure 7 (d) and Figure 7 (e).

Towards overcome this issue, the initial value of rising and falling of yawning from each video frames were observed. From the observation, the initial height score for yawning is from the frame number 21 with the height score of 26. Since the initial reading is 26, any value that falls below 26 is indicated as not-yawn. While, any value of height score that achieve 26 and more is considered as yawn. Thus, this is how the analysis for the IR lighting condition (IR) being done. Note that, every subject has different face size along with the features coordination. Thus, the height of the mouth is different from one to another.

IV. Performance Analysis

The analysis process is continued with the performance of the developed algorithm. The algorithm performance being analyzed based on a few parameters such as the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) as shown in Table 1. TP indicates a correctly yawn detected. FP indicates not-yawn but mistakenly detected yawn, while FN indicates yawn but mistakenly detect not-yawn; lastly TN indicates not-yawn and correctly detected yawn.

The purpose of this section is to validate that the developed algorithm administer authentic outcome to identify yawning in various light condition. In order to evaluate performance of this algorithm, observation evaluation has been done.

Table 1
Indication of TP, FP, TN and FN

	Yawn	Not-Yawn
Detect Positive	TP	FP
Detect Negative	FN	TN

In addition, the accuracy, precision and sensitivity of the proposed method is computed based on the equation (1) to (3)

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (2)$$

$$\text{Sesitivity}=\frac{TP}{TP+FN} \quad (3)$$

The performance analysis process is divided into two stages which are the analysis done for normal lighting condition and the analysis done for Infrared (IR) lighting condition. Based on Table 2 and Table 3, there are slight difference between the accuracy, precision and sensitivity for the performance of the proposed algorithm. This situation might occur due to the edge detection technique as being discussed earlier.

Table 2
Performance analysis based on the proposed algorithm (Normal lighting condition)

Subjects	Accuracy (%)	Precision (%)	Sensitivity (%)
Subject 1	100	100	100
Subject 2	100	100	100
Subject 3	100	100	100
Subject 4	95.06	85.7	100
Subject 5	99.4	99	100
Overall	**98.89**	**96.94**	**100**

Table 3
Performance analysis based on the proposed algorithm (IR lighting condition)

Subjects	Accuracy (%)	Precision (%)	Sensitivity (%)
Subject 1	100	100	100
Subject 2	95.67	95.65	98.51
Subject 3	80.8	71	100
Subject 4	100	100	100
Subject 5	100	100	100
Overall	**95.29**	**93.33**	**99.7**

From the computed overall accuracy for both lighting conditions, the overall accuracy for the proposed algorithm is evaluated and has resulted in 99.84% Precision is a calculation measurement amongst the positive values which referred to the total number of correctly detected yawn amongst the sum of correctly detected yawn and the false detected yawn. The total precision of the proposed method is computed as 97.88%. Sensitivity indicates the ability of the proposed method to correctly detect yawn. The proposed system had managed to obtain 99.85% for overall sensitivity.

V. CONCLUSION

Based on the past research, a lot of yawning detection techniques are introduced. However, the developed techniques focused in an environment with a normal lighting condition. Up to this day, there is no technique introduced for yawning detection in Infrared (IR) lighting condition.A development of yawning detection based on various lighting condition is high in demand since illuminance variances are still considered as the major limitation in yawning detection. Furthermore, there is no publicly available database for yawning detection in IR lighting condition. Moreover, the availability of the yawning database under normal illumination is very limited. Hence, in this project, a small scale of the yawning detection database was developed for research purpose.

This paper proposed a development of algorithm to detect yawning for normal lighting condition and Infrared (IR) lighting condition. The developed algorithm focused on the mouth region and the height score reading of the mouth opening was measured based on the darkest region between a person’s lips. The algorithm developed for normal lighting condition utilized the darkest region between the lips to indicate yawning and the algorithm for night lighting condition implement Canny edge detector and utilize the edge of the lips and the darkest region between the lips to indicate yawning which has resulted a slight difference between the accuracy. Nevertheless, based on the experiment conducted, the proposed algorithm managed to identify yawning robustly for both lighting conditions which possessed an accuracy of 99.84% for the overall algorithm accuracy.The developed algorithm has been tested on 5 subjects.

For future work, the number of subjects could be increased as to testify the robustness of the algorithm. Furthermore, the lighting condition can vary in the future such as dim lighting condition or bright lighting condition. In addition, the developed yawning detection algorithm could be implemented in drowsiness detection systems as a step to prevent car crashes.

ACKNOWLEDGMENT

Authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Higher Education for supporting this research under RAGS/1/2015/TK0/FKEKK/02/B00101.

REFERENCES

[1] M. M. Ibrahim, J. J. Soraghan, L. Petropoulakis, and G. Di Caterina, “Yawn Analysis with Mouth Occlusion Detection,” *Biomedical Signal Processing and Control*, vol. 18, pp. 360–369, 2015.
[2] N. Alioua, A. Amine, and M. Rziza, “Driver ’ s Fatigue Detection Based on Yawning Extraction,” *International Journal of Vehicular Technology*, 2014.
[3] W. Zhang, Y. L. Murphey, T. Wang, and Q. Xu, “Driver Yawning Detection Based On Deep Convolutional Neural Learning and Robust Nose Tracking,” *Proceedings of the International Joint Conference on Neural Networks*, 2015.
[4] S. Abtahi, B. Hariri, and S. Shirmohammadi, “Driver drowsiness monitoring based on yawning detection,” *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pp. 1606–1610, 2011.
[5] M. Omidyeganeh *et al.*, “Yawning Detection Using Embedded Smart Cameras,” *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 570–582, 2016.
[6] W. A. Mustafa and H. Yazid, “Illumination and Contrast Correction Strategy using Bilateral Filtering and Binarization Comparison,” *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 8, no. 1, pp. 67–73, 2016.
[7] P. Viola and M. Jones, “Robust real-time face detection,” *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
[8] B. D. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision,” 1981.
[9] N. A. Razak, C. Liong, A. A. Jemain, N. A. Ghani, S. Zakaria, and H. M. Sulaiman, “Firing Pin Impression Segmentation using Canny Edge Detection Operator and Hough Transform,” *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 1, pp. 23–26, 2017.