

Yawn analysis with mouth occlusion detection

Masrullizam Mat Ibrahim, John J Soraghan, Lykourgos Petropoulakis and Gaetano Di Caterina

Department of Electronic and Electrical Engineering, University of Strathclyde, 204 George Street, G1 1XW, Glasgow, UK

Abstract. One of the most common signs of tiredness or fatigue is yawning. Naturally, identification of fatigued individuals would be helped if yawning is detected. Existing techniques for yawn detection are centred on measuring the mouth opening. This approach, however, may fail if the mouth is occluded by the hand, as it is frequently the case. The work presented in this paper focuses on a technique to detect yawning *whilst also allowing for cases of occlusion*. For measuring the mouth opening, a new technique which applies adaptive colour region is introduced. For detecting yawning whilst the mouth is occluded, Local Binary Pattern (LBP) features are used to also identify facial distortions during yawning. In this research, the Strathclyde Facial Fatigue (SFF) database which contains genuine video footage of fatigued individuals is used for training, testing and evaluation of the system.

Keywords: Yawn, fatigue, mouth covered, mouth opening, database.

1. Introduction

Fatigue is usually perceived as the feeling of tiredness or drowsiness. In a work related context, fatigue is a state of mental and/or physical exhaustion that reduces the ability to perform work safely and effectively. Fatigue can result due to a number of factors including inadequate rest, excessive physical or mental activity, sleep pattern disturbances, or excessive stress. There are many conditions which can affect individuals and which are considered directly related to fatigue such as visual impairment, reduced hand-eye coordination, low motivation, poor concentration, slow reflexes, sluggish response or inability to concentrate. Fatigue is considered as the largest contributor to road accidents, leading to loss of lives. Data for 2010 from the Department of Transport, UK [1], points to 1850 people killed, and 22,660 seriously injured, with fatigue contributing to 20% to the total number of accidents [2]. US National Highway Traffic Safety estimates suggest that approximately 100,000 accidents each year are caused by fatigue [3].

Fatigue is identifiable from human physiology such as eye and mouth observations, brain activity, and by using electrocardiograms (ECG) measuring, for example, heart rate variability. The physical activities and human behaviour may also be used to identify fatigue [4-6]. In this paper yawning, indicating fatigue, is detected and analysed from video sequences. Yawning is an involuntary action where the mouth opens wide and, for this reason, yawning detection research focuses on measuring and classifying this mouth opening. Frequently, however, this approach is thwarted by the common human reaction to hand-cover the mouth during yawning. In this paper, we introduce a new approach to detect yawning by combining mouth opening measurements with a facial distortion (wrinkles) detection method. For mouth opening measurements a new adaptive threshold for segmenting the mouth region is introduced. For yawning detection with the mouth covered, Local Binary Patterns (LBP) features and a learning machine classifier are employed. In order to detect the wrinkles, the edge detector Sobel operator [30] is used. Differently from [32], where the mouth covered detection technique is applied to static images only, the yawn analysis approach proposed in this paper describes a complete system for yawn detection in video sequences. In this research, genuine

yawning fatigue video data is used for training, testing and evaluation. The data is from the Strathclyde Facial Fatigue (SFF) video database, which was explicitly created to aid this research, and which contains series of facial videos of sleep deprived fatigued individuals. The, ethically approved, sleep deprivation experiments were conducted at the University of Strathclyde, with the aid of twenty volunteers under controlled conditions. Each of the 20 volunteers was sleep deprived for periods of 0, 3, 5 and 8 hours on separate occasions. During each session the participants' faces were recorded while they were carrying out a series of cognitive tasks.

The remainder of this paper is organised as follows. Section 2 discusses the related work of this research, while Section 3 explains the SFF database. Section 4 discusses the overall system, while analysis and results are reported in Sections 5 and 6 respectively. Conclusions to the paper are provided in Section 7.

2. Related work

Yawning is a symptom indicated visually by a mouth open state. Yawning detection research focuses mostly on the measurement of the mouth opening [7-9]. The algorithms used must therefore be able to differentiate between normal mouth opening and yawning. Yawn detection can be generally categorised into three approaches; feature based, appearance based, and model based. In the feature based method, the mouth opening can be measured by applying several features such as colour, edges and texture, which are able to describe the activities of the mouth. Commonly, there are two approaches to measure the mouth opening, mainly by tracking the lips movement and by quantifying the width of the mouth. Since the mouth is the wider facial component, colour is one of the prominent features used to distinguish the mouth region. Yao *et al.* [10] transfer RGB colour space to LaB colour space in order to use the 'a' component, which is found as acceptable for segmenting the lips. The lips represent the boundary of the mouth opening that is to be measured. YCbCr colour space was chosen by Omidyeganeh [11] for their yawning detection. In that work, the claim is that colour space is able to indicate the mouth opening area, which is the darkest colour region, by setting a certain threshold. Yawning is detected by computing the ratio of the horizontal and vertical intensity projection of the mouth region. The mouth

opening is detected as yawn when this ratio is above a set threshold. Implementation using the colour properties is easy but, to make the algorithm robust, several challenges must be addressed. The lips colour is different for everyone, and due to the different lighting conditions it is necessary to ensure that the threshold value is adapted to the changes.

Edges are another mouth feature able to represent the shape of the mouth opening. Alioua *et al.* [12] extracted the edges from the differences between the lips region and the darkest region of the mouth. The width of the mouth is measured in consecutive frames, and yawn is detected when the mouth is continuously opening widely more than a set number of times. When using the edges to determine yawn, the challenge is that the changes of the edges are proportional to the illumination changes, making it difficult to set the parameter of an edge detector. In order to locate accurately the mouth boundary, the active contour algorithm can be used as implemented in [8]. However, the computational cost of this approach is very high. The mouth corners have also been employed by Xiao *et al.* [9] to detect yawn. The mouth corners can be reference points in order to track the mouth region. Yawning is classified based on the change of the texture in the corner of the mouth while the mouth is opened widely. Moreover, as in the case implemented for measuring eye activities, Jiménez-Pinto *et al.* [13] have detected the yawn based upon the salient points in the mouth region. These salient points introduced by Shi and Tomasi [14] are able to describe the motion of the lips movement, and yawn is determined by examining the motion in the mouth region. In [7, 11] yawn is detected based on a horizontal profile projection. The height of the mouth opening in profile projection represents the mouth yawn. However, this approach is inefficient for beard and moustache bearing individuals.

In appearance based method, statistical learning machine algorithms are applied and distinctive features need to be extracted in order to train the algorithms. Lirong *et al.* [15] combined Haar-like features and variance value in order to train their system to classify the lips using Support Vector Machine (SVM). With the additional integration of a Kalman filter their algorithm is also able to track the lip region. Lingling *et al.* [16] also applied Haar-like features to detect the mouth region. The yawn is classified using colour properties to measure the breadth of the mouth opening. Elsewhere, Gabor wavelet features that represent texture have been used by Xiao *et al.* [9] for evaluating the degree of mouth opening. The texture features are extracted from the corner regions of the mouth. The texture of the corner is different for mouth opening and closing. The yawn is classified using Linear Discriminant Analysis (LDA). Yang *et al.* [17] employed a back propagation neural network for yawn classification in three states: mouth closed, normal open, and wide open. The mouth region features are extracted based on the RGB colour properties. For this approach a huge number of datasets for images is required in order to obtain adequate results.

The model based approach requires that the mouth or lips are modelled first. An Active Shape Model (ASM) for instance, requires a set of training images and, in every image, the important points that represent the structure of the mouth need to be marked. Then, the marked set of images is used for training using a statistical algorithm to establish the shape model.



Fig. 1. Example images from video footage SFF.

Garcia *et al.* [18] and Anumas *et al.* [19] train the structure using ASM, and the yawn is measured based on the lips opening. Hachisuka *et al.* [20] employ Active Appearance Model (AAM), which annotates each essential facial point. For yawning detection three points are defined to represent the opening of the mouth. In this approach a lot of images are required since every person has different facial structure.

3. Strathclyde Facial Fatigue (SFF) Database

The Strathclyde Facial Fatigue database was developed in order to obtain genuine facial signs of fatigue. Most of the fatigue related researchers have used their own datasets for assessing their developed system. For example, Vural *et al.* [21] developed their fatigue database by using video footage from a simulated driving environment. The participants had to drive for over three hours and their faces were recorded during that time. Fan *et al.* [22] recruited forty participants and recorded their faces for several hours in order to obtain facial signs of fatigue. Some researchers only test their algorithms based on video footage of persons who pretend to experience fatigue symptoms. For example, researchers in [8,11,12] detected yawn based on the width of the mouth opening and the algorithms were not tested on subjects experiencing natural yawning. The SFF database provides a genuine medium of facial fatigue video footage from sleep deprivation experiments which involved twenty participants.

The sleep deprived volunteers were ten male and ten female with ages ranging between twenty to forty years old. Each volunteer had to perform a series of cognitive tasks during four separate experimental sessions. Prior to each session, the participants were sleep deprived for 0 hours (no sleep deprivation), 3, 5 and 8 hours under controlled conditions. The cognitive tasks were designed to test a) simple attention, b) sustained attention and c) the working memory of the participants. These cognitive tasks are associated with sensitivity to sleep loss and are designed to accelerate fatigue signs. Fig. 1 shows some examples of video footage in the SFF database.

4. System overview

A block diagram of the new yawn analysis system is shown in Fig. 2. The system begins with an initialization operation carried out in the first frame of the video. This consists of face acquisition and region initialisation algorithms. During face acquisition, the face, eye and mouth detection algorithms are performed sequentially before a region of interest initialisation algorithm is executed.

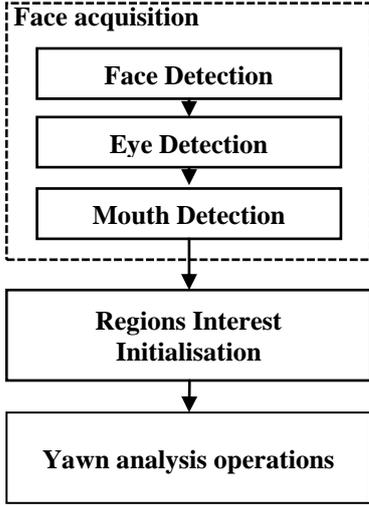


Fig. 2. Yawn Analysis System Overview.

Two regions of interest are initialised: Focused Mouth Region (FMR) and Focused Distortion Region (FDR). These two regions are to be tracked in all the following frames. The FMR region is used in measuring the mouth opening and to detect when the mouth is covered, and FDR is used in measuring the distortions during yawning.

4.1. Region of Interest Evaluation

As is indicated in [32] a facial region of interest needs to be established first. Fig. 3 depicts a face acquisition operation where the eyes and mouth are sequentially detected using a Viola Jones technique [23, 24]. Training is performed through a cascade classifier applied to every sub-window in the input image in order to increase detection performance and reduce computation time. The SFF video footage provided the large number of face, eyes and mouth data needed to train the Haar-like features [25, 26]. In the present method, the face region is firstly detected and then, the eyes and mouth areas are identified within it. The acquisition of the distinctive distances between these facial components is effected through anthropometric measurements, thus accounting for human facial variability. This information is then used in the formation of the two following regions of interest.

4.1.1. Focused Mouth Region (FMR)

This is formed based upon the detected locations of the eyes and mouth. The distance between the centre of the eyes (ED), and the distance between the centre of mouth and the mid-point between the eyes (EMD) are obtained (Fig. 3(a)). The ratio of these distances R_{EM} is then computed as:

$$R_{EM} = \frac{ED}{EMD} \quad (1)$$

From this information the coordinates of the FMR (Fig. 3(b)) are empirically defined as:

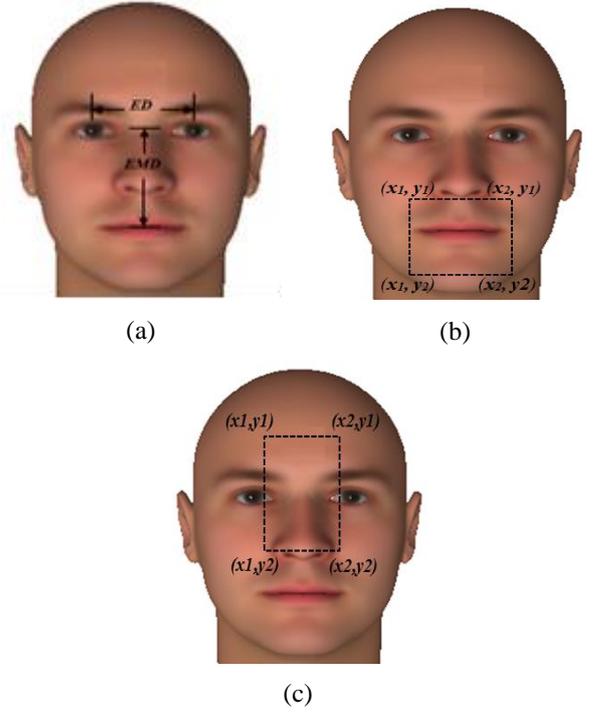


Fig. 3. (a) Anthropometric measurement; (b) Focused Mouth Region (FMR); (c) Focused Distortion Region (FDR).

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_R \\ y_R + 0.75EMD \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_L \\ y_L + 0.75EMD + 0.8ED \end{bmatrix} \quad (3)$$

where (x_R, y_R) and (x_L, y_L) are the centre points of the right eye iris and left eye iris respectively. The FMR depends on the location and distance of the eyes. When the face moves forward the eye distance increases and so does the FMR. Conversely, the FMR reduces following a backward move.

4.1.2. Focused Distortion Region (FDR)

The focused distortion region (FDR) is a region in the face where changes of facial distortion occur. The region, which is most likely to undergo distortions during yawning, was identified based on conducted experiments using the SFF database. The coordinates of the FDR (Fig. 3(c)), are empirically defined as:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_R \\ y_R - 0.75EMD \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_L \\ y_L + 0.75EMD \end{bmatrix} \quad (5)$$

The threshold values 0.75 and 0.8 used in (2)–(5) are based on observation tests, undertaken on the SFF database.

4.2. Yawn Analysis Operations

As indicated in Fig. 2, when all regions of interest have been identified, three operations are carried out using these regions: mouth opening measurement, covered mouth detection, and distortion detection.

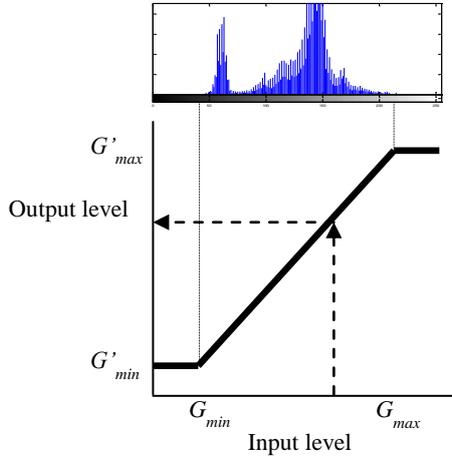


Fig. 4. A linear transform that remaps intensity level input image in between G'_{min} and G'_{max} .

4.2.1. Mouth Opening Measurement

Yawning is an involuntary action that causes a person to open the mouth widely and breathe in. The technique in this paper detects yawning based upon the widest area of the darkest region between the lips. As previously discussed when applied techniques are based on intensities and colour values, the challenge is to obtain an accurate threshold value which will adapt to multi-skin and lips colouration, as well as to different lighting conditions. The proposed algorithm applies a new adaptive threshold in order to meet this challenge.

The adaptive threshold value is calculated during the initialisation operation after the FMR is formed. Due to the dependence of the method on the intensity value of the region, it is necessary to improve the face image in order to contrast enhance the darkest and the brightest facial regions. The enhancement is implemented by transforming the histogram of the image to spread the level of colour values evenly but without changing the shape of the input histogram. This is computed using the following transformation [27]:

$$g'(x, y) = \frac{G'_{max} - G'_{min}}{G_{max} - G_{min}} [g(x, y) - G_{min}] + G'_{min} \quad (6)$$

The linear transformation illustrated in Fig. 4 stretches the histogram of the input image into the desired output intensity level of the histogram between G'_{min} and G'_{max} , where G_{min} and G_{max} are the minimum and maximum input image intensities respectively. From this enhanced histogram illustrated in Fig 5(a), the cumulative histogram H_i is calculated as in (7), thus generating the total number of index pixels in the histogram bins between 0 and 255 as illustrated in Fig. 5(b).

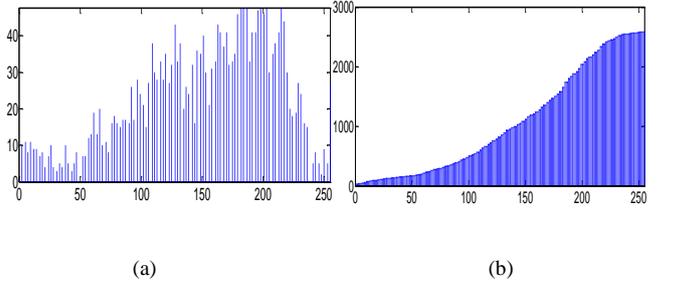


Fig. 5. (a) Enhanced image histogram; (b) Cumulative histogram.

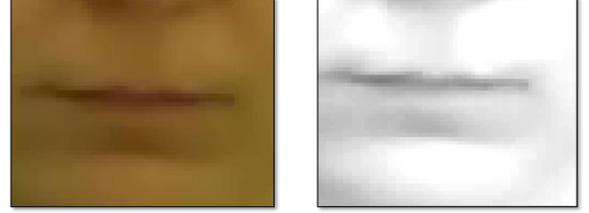


Fig. 6. Focused Mouth Region (FMR).

$$H_i = \sum_{j=1}^i h_j \quad (7)$$

Consequently, the initial adaptive threshold is calculated by reference to the darkest area as shown in Fig. 6. On average, the darkest area constitutes approximately 5%, or less, of the entire obtained region of interest. Based on this percentage, the darkest area in the histogram is situated between 0 and X intensity levels. The intensity value X can be obtained by selecting the first 5% of the total number of pixels contained in the enhanced histogram (Fig. 5(a)) and reading the corresponding right-most intensity value from the x-axis.

Then, the adaptive threshold T_a value is calculated as follows:

$$T_a = \frac{X - P_{min}}{P_{max} - P_{min}} \quad (8)$$

where P_{max} and P_{min} are the maximum and minimum pixel values respectively.

After the adaptive threshold T_a is computed, the FMR is segmented and the mouth length L_M is obtained as illustrated in Fig. 7. The length of L_M must be within the range indicated in (9).

$$0.7ED < L_M < 0.9ED \quad (9)$$

If L_M is outside this range, the value T_a is increased or decreased accordingly, in steps of 0.001. Then the segmentation process is repeated until the value of L_M is within range as in (9).

The adaptive threshold is applied to the next sequence of frames for segmenting the darkest region between the lips. In order to determine the yawning, the height of the mouth H_M , based on the darkest region in FMR, is measured as shown in Fig. 8. Then, the Yawn Ratio (R_Y) of the height H_M over the height of FMR is computed as follows:

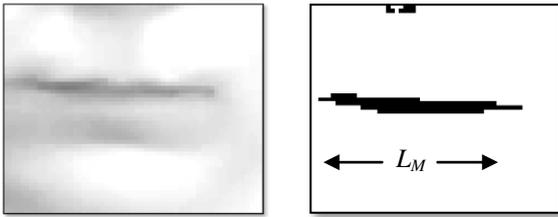


Fig. 7. FMR segmentation using adaptive threshold.

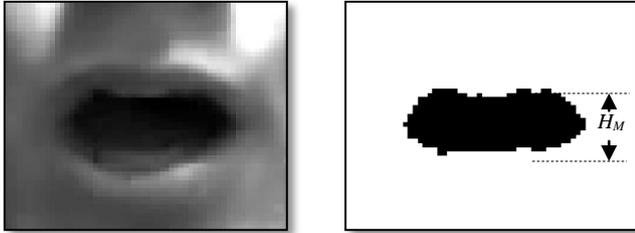


Fig. 8. The measurement of high of mouth opening.

$$R_Y = \frac{H_M}{FMR_{height}} \quad (10)$$

A value of R_Y higher than a certain threshold denotes a mouth open. Such a threshold has been computed empirically as reported in section 4.1.

4.2.2. Covered Mouth Detection

It is not unusual to occlude the mouth with the hand during yawning. This means that information from section 4.2.1 can no longer be used to detect yawning. The approach first introduced in [32] on still images can also be used in video sequence footage. The approach involves that the FMR is examined to determine whether the region is covered or not. Fig. 9 displays images of FMR (a) not covered and (b) covered. From these images the different texture of the FMR regions is clearly obvious in the two cases. In order to differentiate between these two cases, Local Binary Patterns (LBP) are used to extract the FMR texture pattern. LBP features have been proven highly accurate descriptors of texture, are robust to the monotonic gray scale changes, as well as simple to implement computationally [28].

Local Binary Patterns (LBP) was originally applied on texture analysis to indicate the discriminative features of texture. With reference to Fig. 10, the basic LBP code of the central pixel is obtained by subtracting its value from each of its neighbours and, if a negative number is obtained, it is substituted with a 0, else it is a 1.

The LBP operator is limited since it represents a small scale feature structure which may be unable to capture large scale dominant features. In order to deal with the different scale of texture, Ojala *et al.* [29] presented an extended LBP operator where the radius and sampling points are increased. Fig. 11 shows the extended operator where the notation (P, R) represents a neighborhood of P sampling points on a circle of radius R.

The LBP result can be formulated in decimal form as:

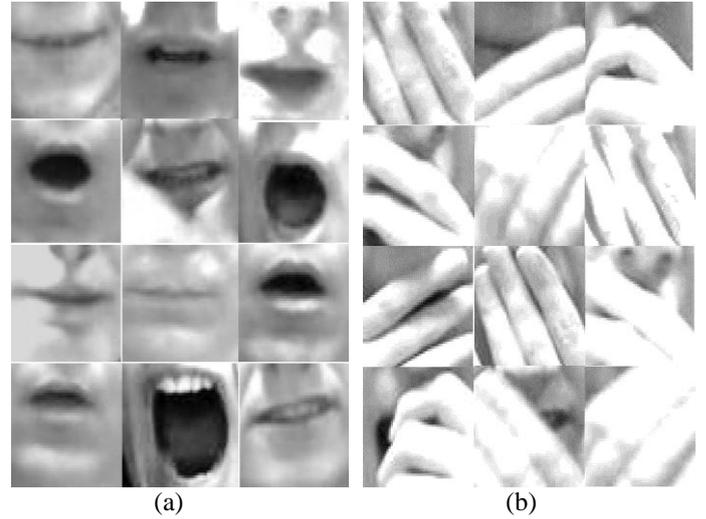


Fig. 9. (a) Uncovered mouth; (b) Covered mouth.

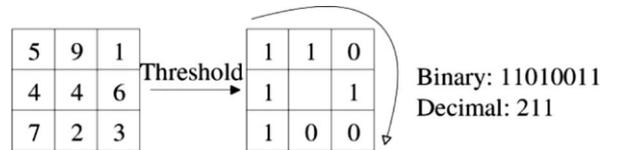


Fig. 10. Example of the basic LBP operator.

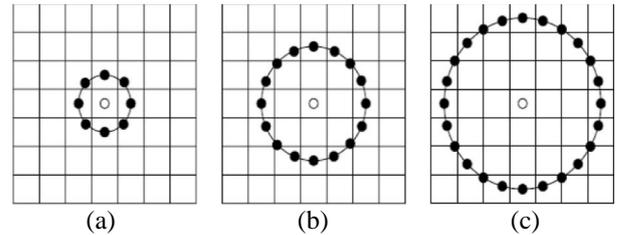


Fig. 11. Example of extended LBP operator. (a) (8,1); (b) (16,2) and (c) (24,3) neighborhoods.

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(i_P - i_c) 2^P \quad (11)$$

where i_c and i_p denote gray level values of the central pixel and P represent surrounding pixels in the circle neighborhood with radius R . Function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (12)$$

The LBP operator, as formulated in (11), has a rotational effect. Hence, if the image is rotated, then, the surrounding pixels in each neighborhood are moving accordingly along the perimeter of the circle. In order to remove this effect Ojala *et al.* [29] proposed a rotational invariant (ri) LBP as follows:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, i) \mid i = 0, 1, \dots, P-1\} \quad (13)$$

where $ROR(x, i)$ performs a circular bitwise right shift on the P -bit number x with i time. A LBP uniform pattern ($u2$), $LBP_{(P,R)}^{u2}$ was also proposed. The LBP is called uniform when it contains a maximum of two bitwise transitions from 0 to 1 or vice versa. For example, 1111111 (0 transition) and 00111100 (2 transitions) are both uniform, but 100110001 (4 transitions) and 01010011 (6 transitions) are not uniform. Another LBP operator, which is a combination of a rotational invariant pattern with a uniform pattern, is $LBP_{(P,R)}^{riu2}$. This operator is obtained by simply counting the 1's in uniform pattern codes. All non-uniform patterns are placed in a single bin. An example of an LBP histogram of the three types of LBP operators for covered and uncovered mouth region is shown in Fig. 12.

A classifier then uses the LBP features to detect mouth covering, as explained in section 6.2.

4.2.3. Distortions Detection

Yawning cannot be concluded solely on the detection of a covered mouth region. It is possible that the mouth is covered accidentally or even intentionally for reasons unrelated to yawning. To minimise the possibility of false yawn detection, the metric of distortions in a specific region of the face, most likely to distort while yawning, is also used [32]. This region is identified based on observations and the experiments conducted using the video footage from the SFF database.

The region most affected by distortion during yawning is the Focused Distortion Region (FDR) (section 4.1.2). From Fig. 13(b, c) it is obvious that there are significant visual differences in the FDR area between normal conditions and during yawning.

To evaluate the changes occurring in FDR, edge detection is used to identify the distortions. Here any edge detector can be used. However, based on experiments carried out, the Sobel operator [30] was chosen since it detected most of the required edges, at a reasonable computational cost. The Sobel operator first calculates the intensity gradient at each point in the region of interest. Then, it provides the direction of the largest possible increase from light to dark and the rate of change in the horizontal and vertical directions. The Sobel operator represents a partial derivative of $f(x, y)$ and of the central point of a 3×3 area of pixels. The gradients for the horizontal G_x and vertical G_y directions for the region of interest are then computed as [31]:

$$G_x = \left\{ \begin{aligned} &f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1) \\ &- \{f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)\} \end{aligned} \right\} \quad (14)$$

$$G_y = \left\{ \begin{aligned} &f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1) \\ &- \{f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)\} \end{aligned} \right\} \quad (15)$$

The FDR distortions are detected based on the gradients in both directions and the results are shown in Fig. 14. The number of edges detected under normal conditions, as illustrated in Fig. 14(a), is significantly different compared to yawning (Fig. 14(b)). In order to identify the distortion in the FDR during yawning, the sum FDR_{SAD} of the absolute values obtained in (16) is used to compute the number of wrinkles in the region. The normalized FDR_{SAD} is calculated

in (17) where W and H denote the width and height of the FDR respectively.

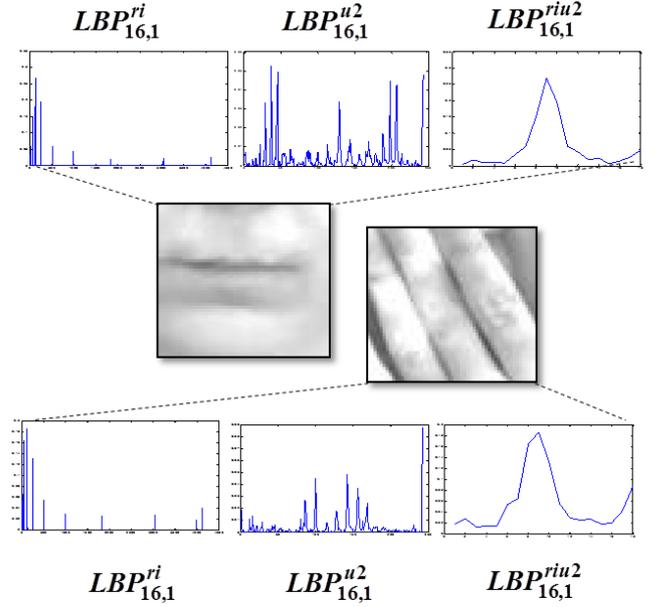


Fig. 12. LBP histogram for not covered mouth and covered mouth regions.

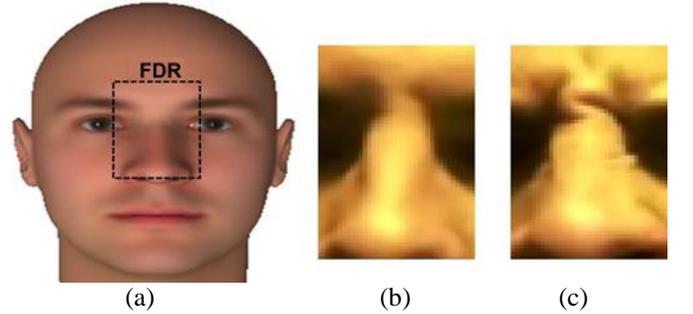


Fig. 13. (a) Focused Distortion Region (FDR); (b) normal condition in FDR; (c) yawning in FDR.

During yawning, the numbers of the detected edges increase as shown in Fig. 15.

$$FDR_{SAD} = \sum_{i,j} |I_1(i, j) - I_2(x+i, y+j)| \quad (16)$$

$$Normalised FDR_{SAD} = \frac{FDR_{SAD}}{255 \times W \times H} \quad (17)$$

5. Yawn analysis

In general, the action of yawning can be categorised into three situations: a) mouth not-covered and wide open, b) mouth wide open for a short time before it gets covered and, c) mouth totally covered with invisible mouth opening. These situations are based on the observations made from the video footage of the SFF database. A yawn analysis algorithm, as depicted in Fig. 16, was therefore developed to deal with these cases by introducing the aforementioned procedures, i.e. mouth opening measurement, mouth covered detection and distortions detection.

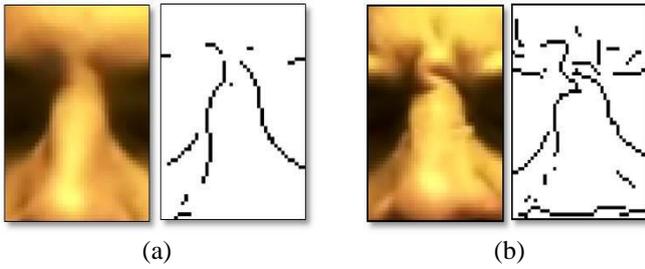


Fig. 14. FDR with input image and edges detected image. (a) normal; (b) yawn.

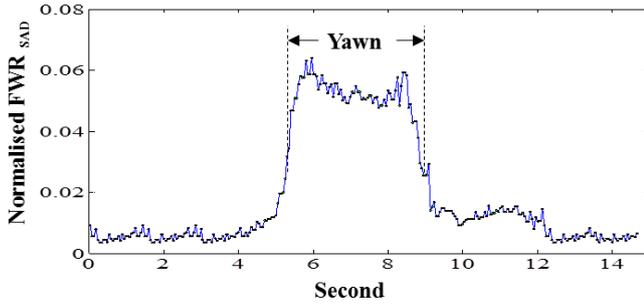


Fig. 15. Plot of Normalised FDR_{SAD} values during yawning.

In this algorithm, the status of mouth opening, mouth covered, and distortions are examined periodically. The chosen time frame for measurements, based on observations from the SFF database, is 30 seconds – a time interval reached by examining the SFF video data. This assumes that yawning does not occur twice within this time period.

The algorithm starts by first examining the mouth opening. Yawning is identified through the mouth opening for longer than a fixed interval τ , usually between 3 and 5 seconds. If this shows no yawning, then the mouth covered detection is triggered. The status of distortions in FDR is only examined when mouth covering is detected.

In the situation where yawning occurs with the mouth wide open for a short time and then hand-covered quickly, the algorithm checks the sequence of events, i.e. mouth open wide followed by mouth covered. As a result of this analysis, a yawn status is detected and the period of yawning is also obtained.

6. Experimental results

The performance of the developed algorithms was tested by using the full SFF video footage. The results presented here were the outcome of substantial real-time experimentation using the video sequences from the whole of the database.

6.1. Mouth Opening Measurement

The mouth opening measurement algorithms are tested using the SFF database. In order to evaluate the accuracy of the mouth opening segmentation using the proposed adaptive threshold, the actual input image is compared with the segmented region of mouth opening. Fig. 17 shows examples of mouth opening images paired with the output image of mouth opening segmentation. A total of 145 mouth opening images under varying lighting conditions were used for testing. From these tests, it was shown that the algorithm

was able to segment the darkest region between the lips as required for detecting yawning.

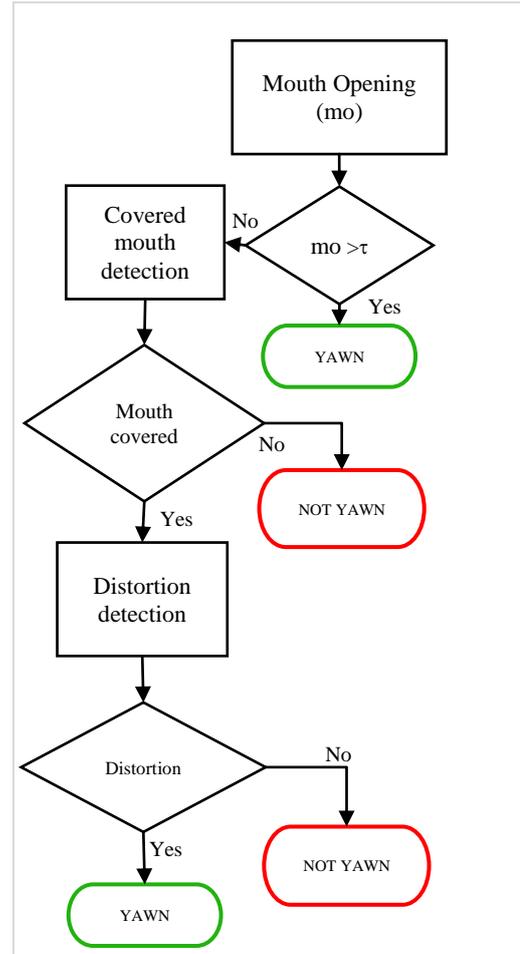


Fig. 16. Yawn analysis algorithm.

As discussed in section 4, yawning is determined by measuring the height of the segmented region of the mouth opening.

In order to obtain the applicable threshold value that represents yawning, 25 of the real yawning scenes from the SFF database were used. From the yawning scenes it was found that the minimum value of R_Y which denotes yawning is 0.5. This ratio value is applied in mouth opening measurements during yawn analysis.

6.2. Mouth Covered Detection

In order to detect the covered mouth, the FMR is examined in every video frame. The LBP features are extracted from the region, and the learning machine algorithm decides the status. The classifiers tested, to determine their suitability for use in the analysis algorithm, comprised Support Vector Machines (SVM) and Neural Networks (NN).



Fig. 17. Mouth opening segmentation images.

A total of 500 covered mouth images and 100 non covered mouth images of size 50x55 pixels have been used to train the NN and SVM classifiers. The LBP operators used for feature extraction were: rotational invariant (ri), uniform ($u2$), and uniform rotational invariant ($riu2$). The radii of the regions examined were either 8, 16, or 24 pixels.

Classification was performed on 300 images. Four different kernel functions were tested for the SVM classifier and two different network layers for the NN. Analytically these were: SVM with a) radial basis function (SVM-rbf), b) linear (SVM-lin), c) polynomial (SVM-pol), and d) quadratic (SVM-quad). The neural networks tested had 5 (NN-5) and 10 neurons (NN-10) respectively in their hidden layer.

It is usual to present such results using graphs of sensitivity (ordinate) against 1-specificity (abscissa), where sensitivity and specificity are given respectively by:

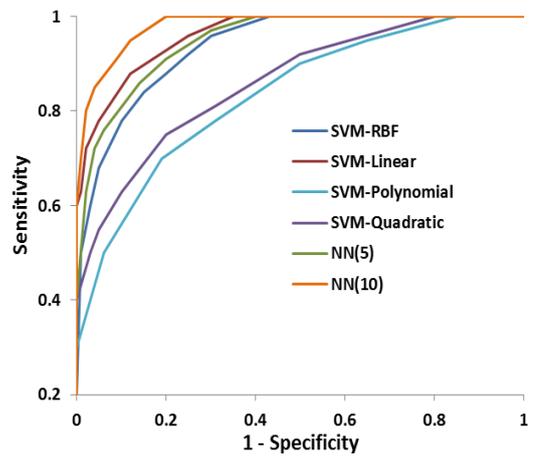
$$Sensitivity = \frac{T_P}{T_P + F_P} \quad (18)$$

$$Specificity = \frac{T_N}{T_N + F_N} \quad (19)$$

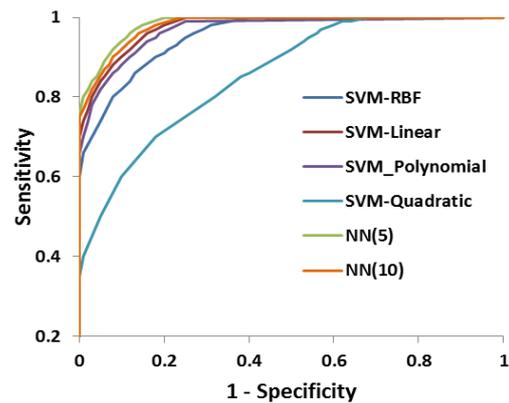
where, T_p represents the number of true positives, F_p the number of false positives, T_n the number of true negatives, and F_n the number of false positives. Sensitivity represents the ability of the developed algorithm to identify the covered mouth. The specificity relates to the ability of the algorithm to detect the non-covered mouth. For maximum accuracy and minimum number of false positives, maximum values of sensitivity and specificity are required. The ROC curves from the results of the classifications are shown in Fig.18.

The legends shown in Fig. 18(a-c) indicate the classifiers used in each case and their respective outputs.

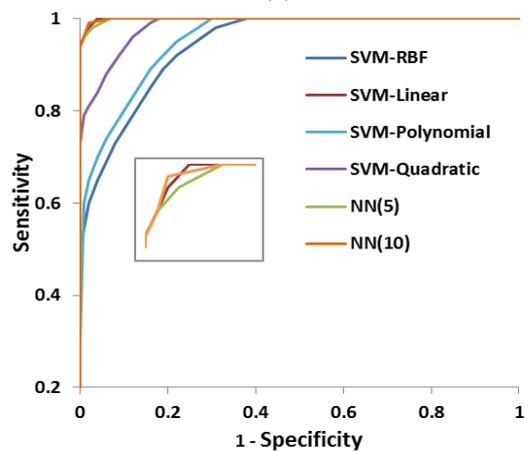
In Fig. 18 the best performance is provided by the LBP-classifier combination which is closer to the upper left part of the graphs where the sensitivity approaches the value of 1 and the value of 1-specificity approaches 0. On this basis, the best performance is illustrated in Fig 18(c), which corresponds to the combinations of LBP with NN10, NN5 and SVM-linear as classifiers, with results close to perfection.



(a)



(b)



(c)

Fig. 18. ROC curves for LBP operator: (a) rotational invariant (ri); (b) rotational invariant with uniform pattern ($riu2$); (c) uniform pattern ($u2$) with zoom-in insert of upper left corner.

6.3. Yawn Analysis

Yawn analysis is a combination of mouth opening measurement, mouth covered detection, and distortions detection.

These operations are examined every 30 seconds. For mouth opening measurements, the best threshold value that represents yawning is 0.5. Based on the results obtained and shown in Fig.19, for mouth covered classification the LBP uniform pattern ($u2$) with radius 16 and a Neural Network classifier with 5 neurons in the hidden layer are chosen to determine the status of FMR.

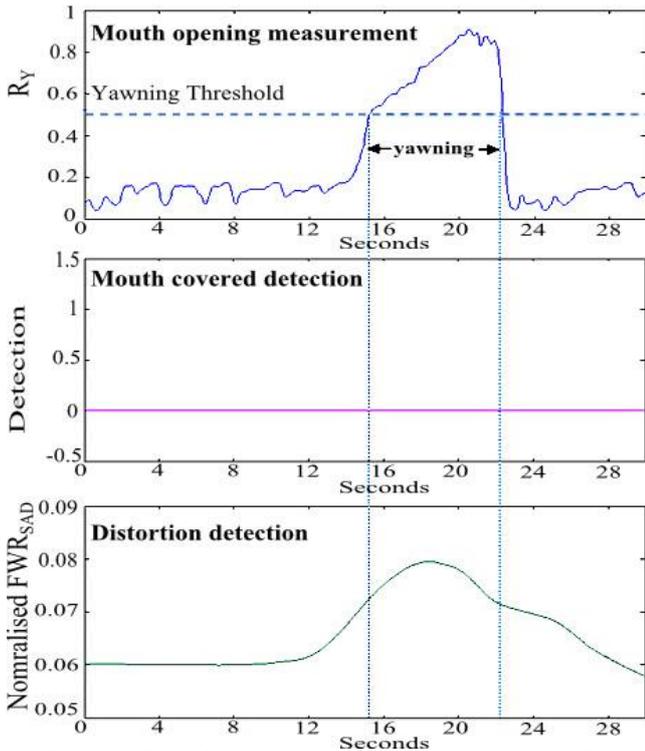


Fig. 19. Plots of yawning detection with mouth not occluded.

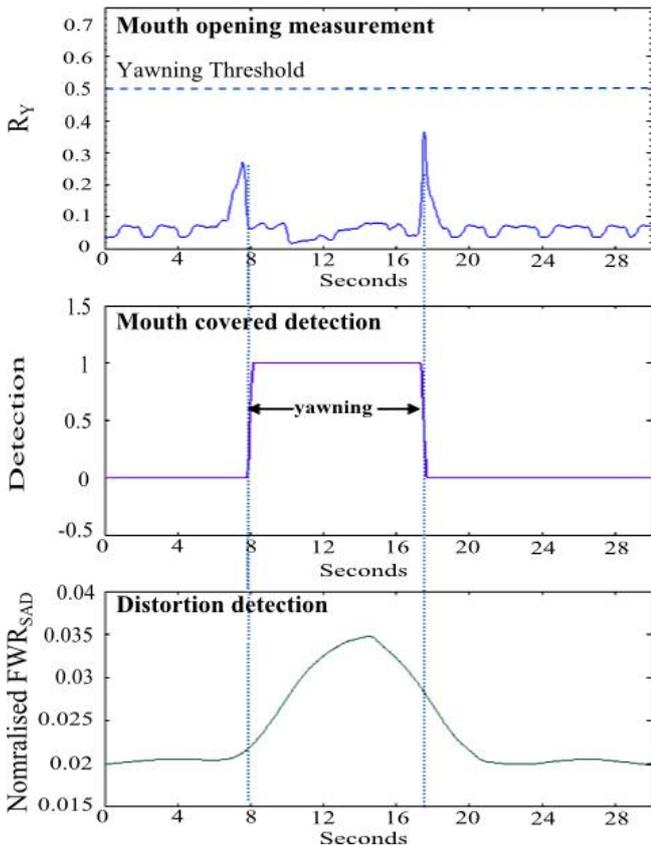


Fig. 20. Plots of yawning detection with mouth occluded.

Genuine yawning images from sequences of 30 video clips from the SFF database are used for evaluation of the algorithm's performance. These videos contain 10 yawning sequences with non-covered mouth scenes, 14 with covered mouth scenes and 6 scenes with both situations present. Examples of the detection results in each situation of yawning are shown in Fig. 19 and 20.

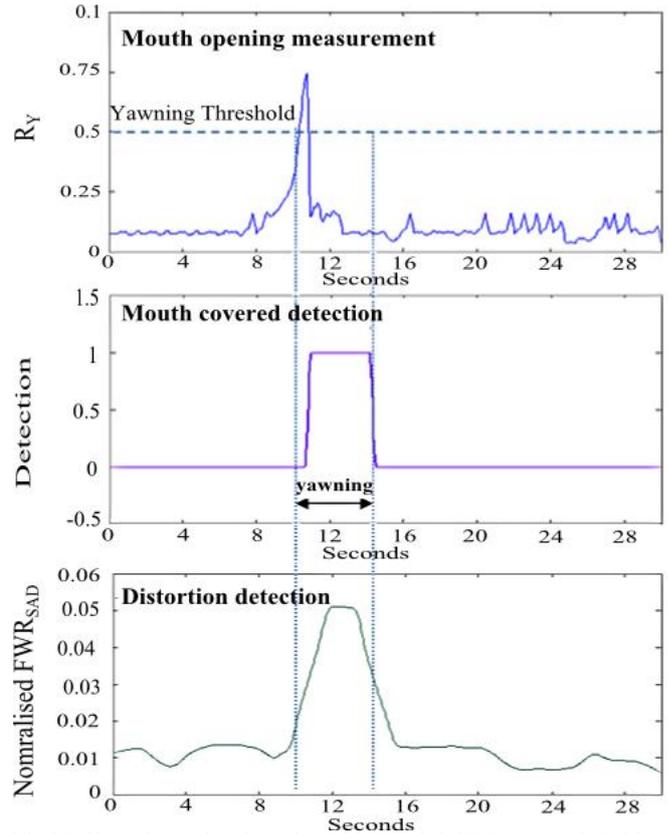


Fig. 21. Plots of yawning detection with mouth initially not occluded but occluded subsequently.

When the mouth is not occluded (Fig. 19), yawning is detected when the height of the mouth opening is equal to or greater than the threshold value. The yawning period is measured from the beginning of the intersection point between the threshold value and the ratio of the mouth opening height to FMR height R_Y . When the R_Y value is below the threshold (Fig. 20), this could be due to mouth being covered during yawning. In this case, the mouth covered detection part of the system is triggered. If mouth occlusion is detected, facial distortion during the period of occlusion is checked. Yawning is assumed to occur when distortion increases substantially within that period.

When the mouth is opened wide over a short period of time before it is occluded (Fig. 21), the value of R_Y exceeds the threshold value for a short period before it drops quickly. In this case, the mouth occlusion detection will be triggered when the R_Y value drops below the threshold. If occlusion is detected, then the period of yawning is measured from the beginning of the first intersection of R_Y with the threshold value, until the end of the mouth occlusion detection.

From the 30 video scenes of yawning, 28 of them were detected successfully with the period measured accurately. However the algorithm failed to detect two yawning scenes where the mouth is completely covered. This is because the FDR did not indicate any distortion.

7. Conclusion

A yawn analysis algorithm designed to deal with both occluded and uncovered mouth yawning situations was presented. This algorithm is a combination of techniques for mouth opening measurement, mouth covered detection and distortions detection. The mouth opening is measured based on the darkest region between the lips. The darkest region is segmented using an adaptive threshold which is introduced

in this paper. For detecting mouth occlusion, an LBP uniform operator is applied to extract features of the mouth region, and the occlusion is evaluated using a Neural Network classifier. To ensure that the mouth occlusion is indeed due to yawning and no other reason, the wrinkles of specific regions on the face are measured. Yawning is assumed to have occurred when distortions in these regions are detected.

In the presented yawn analysis algorithm, the two regions FMR and FDR are monitored every 30 seconds. The analysis results produce the status of yawning, as well as the period of yawning. Genuine sequences of yawning from the SFF database video footage are used for training, testing and evaluating the performance of the developed algorithms. Based on the conducted experiments and results obtained, the overall developed system developed indicated a very encouraging performance.

Acknowledgements

This work was supported by the Scottish Sensor Systems Centre (SSSC), the Ministry for Higher Education of Malaysia and Universiti Teknikal Malaysia, Melaka (UTeM). The SFF database development was funded by a Bridging The Gap (BTG) grant from the University of Strathclyde.

The work which led to the development of the SFF database was in collaboration with the school of Psychological and Health Sciences, University of Strathclyde, the Glasgow Sleep Centre, University of Glasgow, Brooks Bell Ltd.

References

- [1] D. f. T. UK, "Reported Road Casualties in Great Britain: 2010 Annual Report " Department for Transport UK, United Kingdom, Statistical Release - 29 September 2011
- [2] P. Jackson, C. Hilditch, A. Holmes, N. Reed, N. Merat, and L. Smith, "Fatigue and Road Safety: A Critical Analysis of Recent Evidence," ed. London: Department for Transport: Department for Transport, 2011, p. 88.
- [3] R. Resendes and K. H. martin, "Saving Lives Through Advanced Safety Technology," Federal Highway Administration US, Washington DC 2003.
- [4] Daimler. (2012). *Drowsiness-detection system warns drivers to prevent them falling asleep momentarily*. Available: <http://www.daimler.com/dccom/0-5-1210218-1-1210332-1-0-0-1210228-0-0-135-0-0-0-0-0-0-0-0.html>
- [5] M. Michael Littner, C. A. Kushida, D. Dennis Bailey, R. B. Berry, D. G. Davila, and M. Hirshkowitz, "Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002," *Sleep*, vol. 26, p. 337, 2003.
- [6] N. Mabbott, "ARRB Pro-active fatigue management system," *Road and Transport Research*, vol. 12, 2003.
- [7] Y. Lu and Z. Wang, "Detecting Driver Yawning in Successive Images," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, 2007, pp. 581-583.
- [8] M. Mohanty, A. Mishra, and A. Routray, "A non-rigid motion estimation algorithm for yawn detection in human drivers," *International Journal of Computational Vision and Robotics*, vol. 1, pp. 89-109, 2009.
- [9] F. Xiao, Y. Bao-Cai, and S. Yan-Feng, "Yawning Detection for Monitoring Driver Fatigue," in *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, pp. 664-668.
- [10] W. Yao, Y. Liang, and M. Du, "A real-time lip localization and tracking for lip reading," in *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*, 2010, pp. V6-363-V6-366.
- [11] M. Omidyeganeh, A. Javadtalab, and S. Shirmohammadi, "Intelligent driver drowsiness detection through fusion of yawning and eye closure," in *Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2011 IEEE International Conference on*, 2011, pp. 1-6.
- [12] N. Alioua, A. Amine, M. Rziza, and D. Aboutajdine, "Driver's Fatigue and Drowsiness Detection to Reduce Traffic Accidents on Road Computer Analysis of Images and Patterns." vol. 6855, P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, and W. Kropatsch, Eds., ed: Springer Berlin / Heidelberg, 2011, pp. 397-404.
- [13] J. Jiménez-Pinto and M. Torres-Torriti, "Driver alert state and fatigue detection by salient points analysis," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, 2009, pp. 455-461.
- [14] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, 1994, pp. 593-600.
- [15] W. Lirong, W. Xiaoli, and X. Jing, "Lip Detection and Tracking Using Variance Based Haar-Like Features and Kalman filter," in *Frontier of Computer Science and Technology (FCST), 2010 Fifth International Conference on*, 2010, pp. 608-612.
- [16] L. Lingling, C. Yangzhou, and L. Zhenlong, "Yawning detection for monitoring driver fatigue based on two cameras," in *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, 2009, pp. 1-6.
- [17] Y. Yang, J. Sheng, and W. Zhou, "The Monitoring Method of Driver's Fatigue Based on Neural Network," in *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*, 2007, pp. 3555-3559.
- [18] H. García, A. Salazar, D. Alvarez, and Á. Orozco, "Driving fatigue detection using active shape models," *Advances in Visual Computing*, pp. 171-180, 2010.
- [19] S. Anumas and S.-c. Kim, "Driver fatigue monitoring system using video face images & physiological information," in *Biomedical Engineering International Conference (BMEiCON), 2011, 2012*, pp. 125-130.
- [20] S. Hachisuka, K. Ishida, T. Enya, and M. Kamijo, "Facial expression measurement for detecting driver drowsiness," *Engineering Psychology and Cognitive Ergonomics*, pp. 135-144, 2011.
- [21] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," *Human-Computer Interaction*, pp. 6-18, 2007.
- [22] X. Fan, B.-C. Yin, and Y.-F. Sun, "Dynamic Human Fatigue Detection Using Feature-Level Fusion," in *Image and Signal Processing*, ed: Springer, 2008, pp. 94-102.
- [23] P. Viola and M. Jones, "Robust real-time face detection," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, pp. 747-747.
- [24] P. Viola and M. J. JONES, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57(2), pp. 137-154, 2004 2004.
- [25] P. I. Wilson and J. Fernandez, "Facial feature detection using Haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, pp. 127-133, 2006.
- [26] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, vol. 18, pp. 130-140, 2007.
- [27] E. R. Davies, *Machine vision: theory, algorithms, practicalities*: Morgan Kaufmann, 2004.
- [28] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 915-928, 2007.
- [29] T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 971-987, 2002.
- [30] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis* vol. 3: Wiley New York, 1973.
- [31] J.-Y. Zhang, Y. Chen, and X.-x. Huang, "Edge detection of images based on improved Sobel operator and genetic algorithms," in *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*, 2009, pp. 31-35.
- [32] Mat Ibrahim, M.; Soraghan, J; Petropoulakis, L, "Mouth covered detection for yawn", IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2013. pp. 89-94.