Drowsiness Detection Based on Facial Landmark and Uniform Local Binary Pattern

Dini Adni Navastara, Widhera Yoza Mahana Putra, and Chastine Faticah
1Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
dini_navastara@if.its.ac.id,

Abstract. When driving a vehicle, it is often challenging for someone to force his condition to keep driving even though in sleepy condition, thus causing a traffic accident. One of the characteristics of drowsy drivers is the eyes are closed for a certain period. This research proposes a system to detect drowsiness, thus can alert the drowsy driver. The first step is to detect the face using a Funnel-structured cascade algorithm. And then extract the facial landmark features on the face to get the eyes location. The features of eyes are extracted by using a Uniform Local Binary Pattern (ULBP) and the Eyes Aspect Ratio (EAR). EAR is the distance between points at eye landmarks. After the features have been extracted, the system classifies the eyes, whether closed or open by using Support Vector Machine (SVM) method. The system calculates the percentage of eye closure (PERCLOS) to detect drowsiness. Based on the experimental results, the proposed method yields the best accuracy of 95.5% and the optimal value of PERCLOS in drowsiness detection is greater than or equal to 60% with a period of 20 frames.

Keywords: Drowsiness Detection, Facial Landmark, Funnel-structured cascade, PERCLOS, real-time, Support Vector Machine, Uniform Local Binary Pattern.

1. Introduction
When driving a vehicle, it is often the case that someone forces his condition to stay driving even though he is in a drowsy condition which results in a traffic accident. Fatigue in the driver can endanger other drivers, so we need a way to handle it. One of the ways to detect drowsiness is through observation of the eye condition of drivers. If the divers have a closed eye condition in a certain period, the driver can be classified as being sleepy.

In a previous study, Omar Rigane, et al [1] conducted a drowsiness detection study using neural networks to detect the eyes and fuzzy methods to classify whether the user slept or not. Another study was conducted by Anjali K U, et al [2]. They did drowsiness detection by starting with face and eye detection using an algorithm from Viola-Jones and doing a sleepy classification by calculating the time when the eye is not detected because it closes. Basically, drowsiness is closely related to eye activity. Activities of the eyes themselves include closed eyes and open eyes. Krystyna Malik and Bogdan Smolka [3] conducted a study to classify open and closed eyes by comparing histograms of LBP features of closed eyes and open eyes.

In this research, we implement the detection of sleepy drivers. The first step is to do face detection using the Funnel-structured cascade (FuSt) method. After the facial area is obtained, facial feature
extraction is performed to calculate the Eye Aspect Ratio (EAR) and using the Uniform Local Binary Pattern (ULBP). Then a closed and open eye is classified by the Support Vector Machine (SVM) method. Users are said to be sleepy if their eyes are closed until a set time. The results of this research are expected to be able to classify drivers who are drowsy or not in real-time by observing the driver’s eye activity.

2. Literature Review

This section explains the basic theory used in conducting research to provide a general description of the research being undertaken.

2.1. Face Detection

Face detection is an activity to classify faces in an image. One algorithm that is very popular and is still used today is the Viola-Jones algorithm. This algorithm was first introduced in 2001 by Paul Viola and Michael Jones [4].

Viola-Jones is the basis for the creation of a new face detection algorithm at this time. One of the face detection algorithms is Funnel-structured cascade (FuSt) which is introduced by Shuzhe Wu, et al [5]. The FuSt algorithm uses a pyramid structure to detect faces. Each level has a certain stage, namely Fast Locally Assembled Binary (LAB) Cascade classification, Coarse Multilayer Perceptron (MLP) Cascade classification, and Fine MLP Cascade classification.

The first stage is the Fast LAB cascade. At this stage, the sliding window paradigm is carried out. The location of the window at the same position in different images has different results. If a window is needed to detect a face of 20x20 than for a 640x480 size image, it takes around one million windows to be inspected. It certainly will take time and a high cost. Optimization had been done by cascade classification which was initiated by Viola and Jones [4]. Besides, the LAB feature works by only considering the relationship with the Haar feature which makes it have a constant complexity [5].

The next stage is a coarse MLP cascade classification. A more sophisticated classification is done by using MLP with Speeded-Up Robust Feature (SURF) to get the face area. The SURF feature is more expressive than the LAB feature. MLP is a neural network consisting of input layer, output layer, and one or more hidden layers. An n-layer in MLP $F(\bullet)$ is formulated in Equation (1).

$$F(x) = f_{n-1}(f_{n-2}(\ldots f_1(x)$$

where value $x$ is the input value of the SURF feature of the window and $b_i$ are the weight and bias of the relationship between layers $i$ and $i+1$. The activation function $σ(\bullet)$ is usually a nonlinear function like Equation (3).

$$σ(x) = \frac{1}{1+e^{-x}}$$

In the implementation of face detection in real-time, the face position cannot be consistent because the head sometimes faces forward or occasionally faces the other side. Moving the face position makes a face have a unique characteristic. If the face faces the front, the secant cheek is not visible. If the face is facing sideways, the cheeks will be seen clearly. To overcome this problem, FuSt adopted the concept of fine MLP cascade with shape-indexed features extracted in a symmetrical position. There are four semantic positions used as markers, namely the right eye, left eye, nose, and mouth. The MLP is able to make a classification between faces and non-faces more accurately and able to detect face from various
2.2. Uniform Local Binary Pattern
Uniform Local Binary Pattern (ULBP) is a modification of the Local Binary Pattern (LBP) algorithm. LBP was first introduced by Ojala et al [7]. LBP works by doing a 3x3 neighbourhood thresholding with each pixel as the average value. The results of this process will be changed to a binary value called Local Binary Pattern (LBP). This process will also produce 256-bin LBP histograms. LBP histogram is used as a texture descriptor [8][10].

In LBP, some patterns can be used for LBP optimization and one of these patterns is the uniform pattern. Uniform patterns can reduce the length of LBP feature vectors. Uniform patterns emerge because texture often appears in several binary patterns. LBP is said to be uniform if there are at most two transitions 1-0 or 0-1. Examples of cases for uniform patterns in binary patterns such as 00010000 which has two transitions so that the pattern is expressed as a uniform pattern. Examples of patterns that are open uniform are 01010100 because the pattern consists of 6 transitions. The length of the uniform pattern vector features is 59 features. Fifty-eight binary uniform patterns consist of integers 0, 1, 2, 3, 4, 6, 7, 8, 12, 14, 15, 16, 24, 28, 30, 31, 32, 48, 56, 60, 62, 63, 64, 96, 112, 120, 124, 126, 127, 128, 129, 131, 135, 135, 143, 159, 191, 192, 193, 195, 199, 207, 223, 224, 225, 225, 227, 239, 240, 241, 243, 247, 248, 249, 251, 252, 253, 254 and 255. Between 0 to 255 binary patterns, there are binary patterns that are not used. Examples of binary patterns that are not used are 9. Binary from number 9 is 00001001. The binary pattern has three transitions so that binary 9 is not included in the uniform pattern [9].

2.3. Facial Landmark
Facial landmark functions as a sign for part of the face. Commonly used key points are the angle of the eye, the tip of the nose, the angle of the nostril, the corner of the mouth, the endpoint of the eyebrow arc, the outline of the ear, and the chin as in Figure 1 [6].

Facial landmarks on the eye area can show eye condition when closed and open. The process in which the eyes are closed and immediately reopen is called eye blink. Every individual has characteristics in eye blinking. Blinking time only lasts 100-400 millisecond [7]. Figure 2 shows the eye aspect ratio (EAR) based on the eyes landmarks. The EAR value of open eyes is always higher than closed eyes.

Each eye has landmarks $P1$ through $P6$. The eyes landmarks are used to calculate the EAR formulated in Equation (4).

$$ EAR = \frac{|P2-P6|+|P3-P5|}{2|P1-P4|} \tag{4} $$

The value of EAR tends to be constant when the eyes are open and will approach 0 if the eyes are closed. EAR values for open eyes are around 0.25 and closed eyes are around 0.05 [7].
2.4. Estimated Head Pose

Pose estimation is the process to determine the position of an object at real coordinates based on the coordinates of the object on the camera. In estimating poses, there is a problem called the Perspective-n-Point problem or often called PNP when calibrating the camera. The goal of solving this problem is to get a pose from an object when we have camera calibration and we know the location of the $n$ points of 3D on objects with 2D projections on the images [11].

Three Dimension (3D) objects only have two movements based on the perspective of the camera. The first movement is a translation. The translation is the movement of a camera from a 3D location $(X, Y, Z)$ to a new 3D location $(X', Y', Z')$. The second movement is a rotation of the camera with the center $X$, $Y$, or $Z$. To get the estimated pose, it requires 6 points such as three points for rotation and three points for translation [12].

The estimation of the head pose requires 2D points on face images, 3D locations at 2D points, and camera calibration. The 2D points on face images are on the left edge of the left eye, the right edge of the right eye, the tip of the nose, the left edge of the mouth, the right edge of the mouth, and the tip of the chin [13]. Figure 3 shows the example of a frame that applied the head pose estimation.
2.5. **Drowsiness Detection**

Drowsiness detection using image processing is one of the best methods. The driver's sleepiness can be determined because the driver tries to close his eyes [15]. With this method, the duration when the eyes are closed can be calculated to determine whether the driver is asleep or awake. One of the most frequently used indices for calculating drowsiness is PERCLOS (percentage of eye closure). According to a study by Walter Wierwille and colleagues [15], PERCLOS is one of the most important real-time warning steps for vehicle drowsiness detection systems. The PERCLOS formula is shown in Equation (5).

\[
PERCLOS = \frac{\text{closed eyes time}}{\text{closed eyes time} + \text{open eyes time}} \times 100
\]

(5)

3. **DESIGN**

There are several stages in the drowsiness detection system as shown in Figure 4.

3.1. **Read Frame**

The first stage is reading video as input frame by frame. Each frame is resized to 480x270 pixels using scaling with area interpolation. Decreasing image resolution needs to be done so that the computing process becomes faster.

3.2. **Face Detection**

The resized frame is converted to grayscale. Then the image is processed to detect the face in the image using FuSt algorithm described in Section 2.1.

3.3. **Localize the Eye Area**

In this study, we determined that the smallest face size was 50x50 pixels after conducting an experiment by considering the distance of the camera to the driver. If there are faces that are detected less than 50x50 pixels, the object will be ignored and considered not a face. The next process is the system will extract facial landmark features. Not all landmarks will be used. The landmarks used are landmarks on the eye only. The left eye is at index 36 to 41, while the right eye is at index 42 to 47. The length of the eye area is the length between index 36 and 39 for the left eye while the right eye length is the length between index 42 to index 45. The height of the eye area is equal to the length of the eye area.

3.4. **Feature Extraction**

The stage before classification is feature extraction. The features that are used in this research are Uniform Local Binary Pattern (ULBP) and Eyes Aspect Ratio (EAR).

3.4.1. **Uniform Local Binary Pattern (ULBP)**
The texture of the left and right eye are detected using the Uniform Local Binary Pattern (LBP). The images of the right and left eye are first resized to 29x29. After that, both eyes are converted to a grayscale image and be processed by using ULBP.

3.4.2. Eyes Aspect Ratio (EAR)
The EAR is used to measure the extracted eye landmarks. EAR is the distance between points at eye landmarks. If the eyes are closed, the EAR will get closer to zero. The EAR will move away from zero when the eyes are open.

3.5. Eyes Blink Detection
The system must be able to distinguish the condition of the eye being closed and the condition of the eye that is being open from the video data that has never been trained at all using the model obtained after conducting the training. The model was obtained from learning 1192 face data with eyes closed and 1231 face data with eyes open. Blink detection based on combining the features of ULBP and EAR is done by using Support Vector Machine (SVM) method.

3.6. Drowsiness Detection
Humans are said to be asleep if the eyes do not open immediately after closing their eyes. Every closed and open eye will be counted in the period by using PERCLOS formula in Equation (5). The system will issue a warning if within a certain period there are more or equal to 80% of the eyes closed.

4. Experimental Result and Evaluation

4.1. Dataset
The training data used in this research is data collection from F. Song et al, Nanjing University of Aeronautics and Astronautics [16]. The training dataset contains the face images with a size of 100x100. Dataset is divided into two conditions: closed eyes and open eyes.

The testing data is the data collection that the authors take themselves. The testing data are videos with a ratio of 16:9 in the RGB channel. There are three types of videos. Videos that have all frames with open eyes and all frames with closed eyes are used to measure the performance of models and video simulations of people falling asleep are used to measure the performance of drowsiness detection. Each subject in the video will be involved in two conditions, namely glasses and without glasses. The positions of the subject's head are various, such as: normal, looking down, looking up, tilting left, and tilting right.

4.2. Experiment on ULBP Features
This experiment used ULBP as a feature. There are \( P \) and \( R \) parameters that will be tested on ULBP. \( P \) is the neighbourhood of uLBP and \( R \) is the radius of ULBP. The highest accuracy of 95.50% is obtained ULBP \((8,5)\), therefore ULBP with radius 5 is able to reach all the information in the image.

4.3. Experiment on ULBP and EAR Features
This experiment contained a comparison of the model using 10-fold cross validation. Making the model used SVM with a linear kernel. This experiment used ULBP and EAR as features. Similar to adding EAR to LBP, adding EAR to ULBP also did not change the accuracy of the model.

4.4. Experiment on Video without Glasses and with Glasses
This experiment is comparing the value of precision, recall, and accuracy in the video of people with glasses and without glasses who have never entered the training stage. The features that used in this
experiment are ULBP with a \( P \) value of 8 and an \( R \) value of 5. The results of the experiment are in Table 1. In the bespectacled video, there is a slight decrease in accuracy because the eyeglass frame can be considered as the eye.

### Table 1. Test Results Using ULBP (8, 5) on Video Without Head Movement

| Condition   | Recall (%) | Precision (%) | Accuracy (%) |
|-------------|------------|---------------|--------------|
| with glasses| 97.90      | 100           | 98.88        |
| without glasses | 100         | 100           | 100          |

### Table 2. Test Results Using ULBP (8, 5) on Videos with Head Movements

| Condition   | Pose        | Recall (%) | Precision (%) | Accuracy (%) |
|-------------|-------------|------------|---------------|--------------|
| without glasses | Look down   | 74.75      | 99.56         | 86.78        |
| without glasses | Tilt left   | 25.95      | 52.01         | 34.82        |
| without glasses | Tilt right  | 29.57      | 79.88         | 49.63        |
| without glasses | Look up     | 94.65      | 88.89         | 94.65        |
| with glasses   | Look down   | 66.20      | 99.30         | 83.87        |
| with glasses   | Tilt left   | 51.84      | 29.68         | 34.41        |
| with glasses   | Tilt right  | 31.08      | 77.49         | 38.40        |
| with glasses   | Look up     | 81.17      | 98.37         | 90.22        |

### 4.5. Experiment on Videos with Head Movement Without Estimated Pose

This experiment is comparing the value of precision, recall, and accuracy in the video with head movements that have never entered the training stage. The movement of the head to be tested is to look down, tilt left, tilt right and look up. The features that will be used in this experiment are ULBP with \( P \) value of 8 and \( R \) value of 5. The results of the experiment are shown in Table 2.

The decrease in accuracy of the head pose looks down and up because the eyes seem to constrict when in the pose of the head. So that, it produces a texture like eyes closed. In the left and right sloping poses, the low accuracy is caused by the head being too tilted so that the landmarks are unable to determine eye coordinates correctly.

### 4.6. Experiment on the video with head movements without pose estimation

This experiment is comparing the value of precision, recall, and accuracy in the video with head movements that have never entered the training stage and the system used the head pose estimation. The movement of the head to be tested is to look down, tilt left, tilt right and look up. The feature that will be used in this experiment is ULBP with \( P \) value of 8 and \( R \) value of 5. The results are shown in Table 3.

### Table 3. Test Results Using ULBP (8, 5) Size 1 of Video with Head Movement and Pose Estimation

| Condition   | Pose        | Recall (%) | Precision (%) | Accuracy (%) |
|-------------|-------------|------------|---------------|--------------|
| without glasses | Look down   | 77.10      | 99.13         | 87.80        |
| without glasses | Tilt left   | 80.98      | 99.62         | 90.39        |
| without glasses | Tilt right  | 98.80      | 99.04         | 98.65        |
| without glasses | Look up     | 88.44      | 100           | 94.44        |
| with glasses   | Look down   | 64.90      | 99.27         | 83.29        |
| with glasses   | Tilt left   | 91.35      | 94.07         | 92.78        |
The use of pose estimation is able to improve accuracy on the left and right sloping poses because with the estimation of the pose the system is able to obtain a tilt angle from the head and is able to rotate the frame at the angle of the head tilt and make the head seem to be not tilted.

4.7. **Comparative Experiment of Accuracy in Drowsiness Detection**

This experiment used ULBP (8, 5) with pose estimation. Testing data in the form of video simulations of people who will fall asleep. The system detects the condition of the person in the video by calculating PERCLOS in Equation (5). There are PERCLOS limit parameters and PERCLOS calculation periods to be tested. The PERCLOS limit with closed eye time intervals and alarms with the shortest delay of 2.67 seconds is found on PERCLOS greater than or equal to 60% in a 20 frame period.

In determining the limit, PERCLOS used the minimum threshold to tolerate errors in classification and in the shortest possible period to avoid too long a delay. In real-time sleepiness detection, room lighting affects the classification. A room that is too dark can make the eyes open to have a texture like closed eyes. Besides lighting, the distance between the face and the camera can also affect the classification process. Face distance that is too far away from the camera can make the eyes become narrower and make the eyes open has a texture like the eyes closed. To overcome this, resizing of the frame can be done to make it bigger but enlarging the image resolution can be at risk with the longer the system computing time.

5. **Conclusion**

Face detection can be done using Funnel-structured cascade (FuSt) with a minimum face size of 30x30 and input in the form of frames that have been converted to grayscale images. Based on the experimental results, model accuracy is performed using 10-fold cross-validation with linear kernel SVM with the highest accuracy using ULBP features (8, 5) at size 1 of 95.50%. And the optimal value of PERCLOS to detect drowsiness is greater than or equal to 60% in 20 frames. The use of EAR in features does not affect the accuracy of the model. The use of glasses can affect the performance of blink detection because the lens of the glasses gives a disturbance to the texture of the eye and under certain conditions the eyeglass frame can be considered as the eye. The use of head pose estimation in the data pre-processing stage can maximize the extraction of eye landmark features when the head pose is tilted left and right. Room lighting and face distance from the camera affect the results of sleepiness detection in real-time.

References

[1] R. Omar, A. Karim, A. Chokri and M. Mohamed, "A Fuzzy Based Method for Driver Drowsiness Detection," IEEE/ACS 14th International Conference on Computer Systems and Applications, 2017.

[2] K. U. Anjali, K. T. Athoramol, V. Athira and K. R. Bindhu, "Real-Time Nonintrusive Monitoring and Detection of Eye Blinking in view of Accident Prevention due to Drowsiness," International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016.

[3] M. Krystyna and S. Bogdan, "Eye Blink Detection Using Local Binary Patterns," IEEE, 2014.

[4] M. J. Paul Viola, "Rapid Object Detection using a Boosted Cascade of Simple Features," Accepted Conference On Computer Vision And Pattern Recognition, 2001.

[5] W. Shuzhe, K. Meina, H. Zhenliang, S. Shiguang, and C. Xilin, "Funnel-structured cascade for multi-view face detection with alignment awareness," Neurocomputing 221, p. 138–145,
2017.

[6] S. Tereza and C. Jan, "Real-Time Eye Blink Detection using Facial Landmarks," 21st Computer Vision Winter Workshop Luka Čehovin, Rok Mandeljc, Vitomir Struc (eds.) Rimske Toplice, 2016.

[7] T. Ojala, M. Pietikäinen and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," Proceedings of the 12th IAPR International Conference on Pattern Recognition, vol. I, pp. 582-285, 1994.

[8] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast High Dimensional Vector Multiplication Face Recognition," IEEE International Conference on Computer Vision, p. 1960, 2019.

[9] J. S. R. Feldman, "The Text Mining Handbook," New York: Cambridge University Press, 2007.

[10] R. Abdur, H. Najmul, W. Tanzillah, and A. Shaiful, "Face Recognition using Local Binary Patterns (LBP)," Global Journal of Computer Science and Technology Graphics & Vision, 2013.

[11] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of the ACM, pp. 281-395.

[12] G. P. Meyer, S. Gupta, I. Frosio, D. Reddy and J. Kautz, "Robust Model-based 3D Head Pose Estimation," IEEE International Conference on Computer Vision (ICCV), pp. 3649-3657, 2015.

[13] K. Fornalczycy and A. Wojciechowski, "Robust face model based approach to head pose estimation," Federated Conference on Computer Science and Information Systems (FedCSIS), p. 1291, 2017.

[14] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 22, pp. 1330-0334, 2000.

[15] E. E. Galarza, F. D. Egas, F. M. Silva, P. M. Velasco, and E. D. Galarza, "Real Time Driver Drowsiness Detection Based on Driver’s Face Image Behavior Using a System of Human Computer Interaction Implemented in a Smartphone," Advances in Intelligent Systems and Computing, 2018.

[16] F. Song, X. Tan, X. Liu and S. Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, Pattern Recognition, 2014.