Driver Monitoring System based on Distracted Driving Decision Algorithm

Sehwan Jeon[1], Seeyoung Lee[2], Eunjeong Lee[3] and Juseok Shin[4]*
[1234]Dept. of AI Engineering
Korea Polytechnics
Daegu, Korea
[1]fxlimsin@gmail.com, [2]ai.engr.sy@gmail.com, [3]viej0102@gmail.com, [4]jsshin7@kopo.ac.kr
*Corresponding Author

2022 13th International Conference on Information and Communication Technology Convergence (ICTC) | 978-1-6654-9939-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICTC55196.2022.9952837

***Abstract*— Distracted driving and drowsy driving are the main causes of car accidents. We proposed the Distracted Driving Decision Algorithm (DDDA) proposed whereby binary results from blink detection and gaze estimation are computed with the sliding window algorithm. The proposed algorithm is expected to be efficient in a vehicle environment system. The DDDA runs based on computational algorithms instead of using a machine learning model; ultimately, the process of determining the status of a driver can be performed in an efficient and accurate way. The average accuracy of the algorithm and time for it to process one frame was 83.5% and 42ms respectively. By improving the accuracy, the DDDA can contribute to provide easier access to the DMS (Driver Monitoring System).**

Keywords— Driver Monitoring Systems, Distracted Driving Detection, Drowsy Driving Detection, Sliding Window Algorithm, Lightweight System, Advanced Driving Assistance System

I. Introduction (*Heading 1*)

Such as falling asleep or looking at a cellphone, a small mistake a driver makes while driving can lead to a significant consequence. For example, the road traffic deaths in South Korea from 2015 to 2019 were mainly caused by drowsy driving and distracted driving, accounting for 67.6% of the 1,079 reported cases according to Korea Expressway Corporation [1].

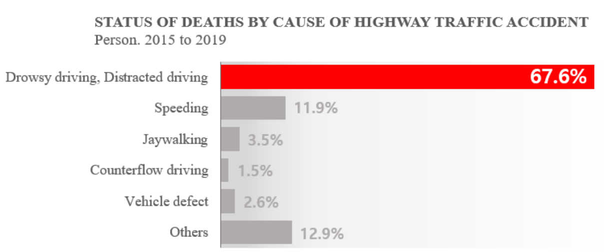

Fig 1. Causes of the road traffic deaths from 2015 to 2019 in South Korea

As the interest in safe driving increases worldwide, research related to the DMS (Driver Monitoring Systems) has been actively conducted and the systems are already on market. However, as more ADAS (Advanced Driving Assist Systems are used in cars, resources are limited over and over in a vehicle embedded environment.

To optimize the DMS system in the environment, the DDDA is proposed in this paper which is based on the window sliding algorithm. The proposed algorithm processes data from blink detection, gaze estimation, and head pose estimation and determine whether a driver is in a proper status while driving. According to the result of the experiment, the DDDA showed an 83.5% accuracy and a 42ms one frame process speed.

II. Proposal Method

We proposed the DDDA that monitors a driver's status in real time and that gives an alarm when driver is not in a proper status. In this paper, the behaviors of distracted driving and drowsy driving is defined as the sideward gaze, downward gaze, lowered head, closing eyes for 3 seconds. The structure of the proposed decision algorithm is shown in Figure 2.

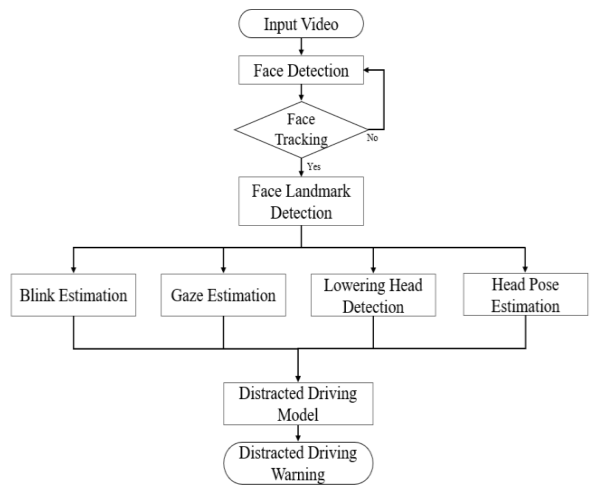

Fig 2. Distracted driving estimation algorithm

Once a driver's video is sent from the camera, the algorithm starts localizing the driver's face and returns the coordinates of the entire face, eyes, nose, and mouth. The data are used for calculations needed to detect a driver's blinks, gaze direction, and head direction. Once the calculations are completed, the DDDA makes a prediction to determine whether the driver is in an appropriate status based on the calculations.

A. Face and Landmark Detection

For the face and landmark detection, models of the dlib were used [2]. For more efficient landmark detection, the algorithm sets the face as the ROI (Region of Interest) from the first frame and subsequently tracks it instead of exploring the entire image in the following frames. In case the tracking process fails, face-redetection is performed at an interval of 30 frames. Finally, if more than one person are detected, only the person with the largest face image is identified as the driver as shown in Figure 3.

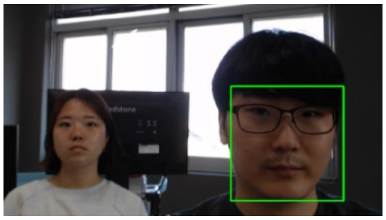

Fig 3. Face detection

B. Blink Detection

The area, width, and height of each eye are used as the main data for blink detection. When the eyes are closed, the area and height of the eyes decrease while the area of the face and width of the eye tend to remain at a similar level. Since these values can change rapidly depending how close to or far from the camera, the ratios in Equation 1 and 2 are used instead of using only the area and height for blink detection.

$$areal\ ratio\ = \frac{area\ of\ each\ eye}{area\ of\ face} \tag{1}$$

$$aspect\ ratio\ = \frac{height\ of\ each\ eye}{width\ of\ each\ eye} \tag{2}$$

For more accurate results, two ratios are calculated from each eye and a total of four ratios are obtained from one frame. If these ratios vary depending on drivers, it is difficult to make an accurate prediction based on a fixed value of the ratios. To overcome the potential error, the minimum and maximum ratios over the entire frames are constantly updated and every ratio is normalized with Equation 3.

$$normalized\ ratio\ = \frac{calculated\ \ ratio - minimum\ ratio}{maximum\ ratio - minimum\ ratio} \tag{3}$$

As these normalized ratios were below 0.4 when the eyes were closed, the algorithm is designed to determine whether the eyes are closed in case once more than two of ratios are below the point.

C. Gaze Estimation

Given the location of pupils is one of the reliable indicators to represent the direction of gaze, the algorithm finds the center of the driver's pupils. To obtain accurate coordinates of both pupils, eye images are masked and only the center area of each pupil appears white. Subsequently, the algorithm returns the coordinates of the center of the pupils.

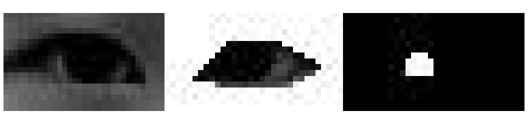

Fig 4. Image preprocessing

Subsequently, two virtual vertical lines are created at both ends and another line is created in the middle of them as shown in Figure 5. Once three lines are created, the distance to each line is calculated and the algorithm determines the gaze direction, considering the closest line represents the direction as shown in Equation 4.

$$direction\ = min(distancce\ to\ left, middle, right) \tag{4}$$

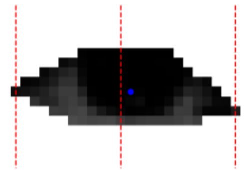

Fig 5. Pupil location and gaze estimation result

D. Head Pose Estimation

To predict the direction of a driver's face, the algorithm converts 2D coordinates of the landmarks to the real world coordinates consisting of the yaw, pitch, and roll axis [3]. The direction of the face is determined using the x-coordinate values and of the roll and the center, the criteria for the determination are shown in Equation 5:

$$\begin{aligned} right &: center\ axis - roll\ axis\ >=\ 6 \\ center &: -6\ <\ Center\ axis - roll\ axis\ <\ 6 \\ left &: Center\ axis - roll\ axis\ <=\ -6 \end{aligned} \tag{5}$$

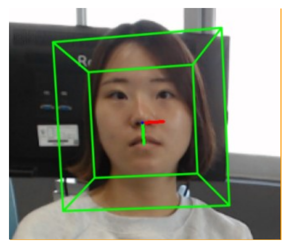

Fig 6. Head pose estimation result

E. Lowering Head Detection

When a driver uses a cellphone or falls asleep, the head tends to be lowered. In the case of lowering the head, parts on the bottom side such as the mouth appear smaller in the video while parts on the top side remain at a similar level. Hence the distance between the nose to chin is used to determine whether the head is lowered or not. For reliable results regardless of varying distances between the camera and driver, the algorithm returns a ratio as shown in Equation 6.

$$chin\ distance\ ratio\ = \frac{middle\ of\ the\ fore\ head\ to\ nose}{nose\ to\ chin} \tag{6}$$

This ratio, however, can vary depending on drivers and it is not advisable to determine the status with a fixed threshold value. To handle this, the algorithm saves the ratio from each frame and finds the mode of the ratio. The mode is the ratio indicating the normal status and the ratio with the least counts after the mode is considered the optimal threshold because the point is where the peak ends. The algorithm is designed to consider that the head is lowered when the ratio is below the optimal threshold

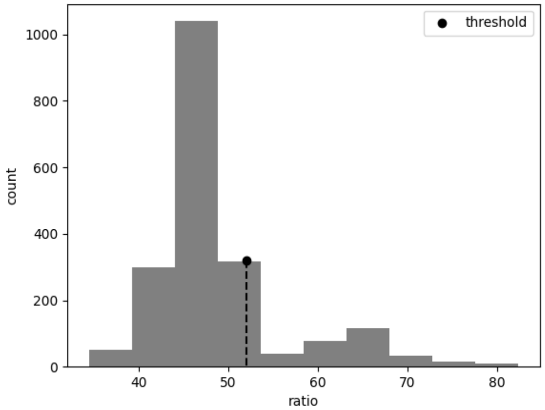

Fig 7. Ratio distribution and mode ratio peak range

F. Post Processing Algorithm

The DDDA makes a prediction based on the four binary results from the blink detection, gaze and face direction estimation, and lowered head detection. Prioritizing closed eyes, lowered head, sideward direction of gaze and face, if any of the results returns true, the system sends an alarm to the driver.

However, small differences in the calculations can cause errors in the aforementioned binary classification results, leading to an incorrect prediction. To address this problem, if the average over all frames is calculated, unnecessarily excessive use of the RAM is inevitable. Therefore, the proposed algorithm uses the sliding window technique to improve the calculation speed and reduce the error in the final prediction.

```
if len(short_queue) == shortLen:
    [short | long]_sum += (now_state - [short | long]_queue.leftpop())
    [short | long]_queue.append(now_state)
else:
    [short | long]_queue += now_state
    [short | long]_queue.append(now_state)

if short_sum < short_Threshold and long_sum < long_Threshold:
    return False
else:
    return True
```

Fig 8. Sliding Window algorithm

For example, incorrect binary results occur sporadically in individual frames while eyes are closed. Considering that a blink takes 208 to 607ms [4], the incorrect results can be ignored if the majority of the results indicate they are closed in the time period. In this paper, it is assumed that eyes remaining closed for 90 frames means that the driver is not in a proper status and the status is confirmed if 80% of the last 90 frames indicate that the eyes are closed.

In summary, the four binary results for the last 90 frames are saved in a queue and the algorithm makes a prediction based on the values in the queue. In addition, for more accurate results, the four of the binary results for the last 30 frames are saved in another queue and determine the status in the same way. By performing this method, the DDDA has more reliable data and finally can make a more accurate prediction about a driver's status. Finally, the whole process is performed more efficiently without excessive use of the RAM and the comparison results are shown in Figure 9.

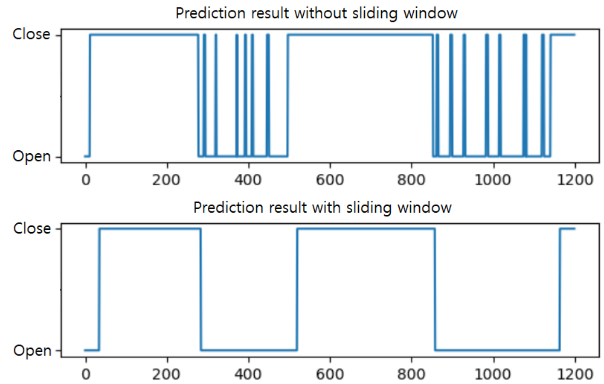

Fig 9. Closed eye prediction result without and with window sliding

III. RESULT

A total of four participants in their 20s and 30s participated in the algorithm evaluation, and the distance between the camera and the participants was maintained at 0.8m in consideration of the distance between the driver and the dashboard. Participants were asked to turn their heads, gaze to left and right, close their eyes, and lower their heads for a certain period of time. Figure 10 to 13 show the prediction results of the DDDA.

Fig 10. Correct prediction results when turning head to left and right

Fig 11. Correct prediction results when eyes move and closed

Fig 12. Correct prediction results when nodding

Fig 13. Detection failure and incorrect prediction results

As shown in Table 1, the average face detection rate of the DDDA was approximately 94% at a total of 50,450 frames.

TABLE 1. FACE DETECTION PERFORMANCE

	Total Frames	**Face Detected Frames**	**Face Detection Rate**
Participant 1	10,281 frames	9,051 frames	88 %
Participant 2	12,778 frames	12,353 frames	96 %
Participant 3	13,257 frames	12,584 frames	94 %
Participant 4	14,134 frames	13,946 frames	98 %

Moreover, the algorithm detected the driver's negligence when the participants turned their head to left and right, closed their eyes, and looked at their cell phone. As Table 2 shows, it took 42.75ms on average for the DDDA to process one frame and the average accuracy in the prediction about driver's negligence was approximately 84%.

TABLE 2. CALCULATION AND PREDICTION PERFORMANCE

	Total Frames	**Process Time**	**Accuracy**
Participant 1	10,281 frames	41ms	73 %
Participant 2	12,778 frames	44ms	90 %
Participant 3	13,257 frames	42ms	79 %
Participant 4	14,134 frames	44ms	94 %

IV. CONCLUSION

In this paper, a driver's negligence is defined as the behavior of closing eyes, lowering head, turning head to left or right, and gazing sideward for 3 seconds. In addition, the DDA detects these behaviors and warns the driver in the case of detection.

The algorithm determines the status of a driver based on facial movement, blinking, and pupil movement with landmark coordinates. This algorithm is expected to have strengths in the vehicle embedded environment in which resources are limited because the DDDA imposes much less burden with computational algorithms and sliding window technique in lieu of using a machine learning model for predictions.

The algorithm needs to be improved because it fails to operate properly when a face is not detected being covered by hands or smartphone, and when inaccurate landmarks are extracted due to a distorted face shape caused by lowering the head too low. If the DDDA is improved by addressing the aforementioned problems, its performance will be significantly improved.

REFERENCES

[1] KoROAD, 2021, 『2021 edition (2019 statistics) OECD member countries Comparison of Traffic Accidents』, pp.8-11.

[2] Dlib Python API Tutorials [Electronic resource], [online] Available: http://dlib.net/python/index.html

[3] Cıvır, Cevdet, and Cihan Topal. "Robust Landmark Selection for 3D Face Pose Estimation." 2019 27th Signal Processing and Communications Applications Conference (SIU). IEEE, 2019.

[4] Hömke, Paul, Judith Holler, and Stephen C. Levinson. "Eye blinks are perceived as communicative signals in human face-to-face interaction." PloS one 13.12 (2018): e0208030.