

Visual Based Drowsiness Detection Using Facial Features

Quang N. Nguyen¹, Le T. Anh Tho¹, Vo Van Toi¹, Hui Yu², Nguyen Duc Thang¹

¹Biomedical Engineering Department, International University, VNU-HCM, Vietnam

²School of Creative Technologies, University of Portsmouth, United Kingdom

Abstract— In this paper, a camera-based system is proposed to detect and monitor drowsiness of a car driver in real time. The system utilizes an RGB image to track the drivers' face and their eyes to detect sleepy sign. For the face detection and segmentation, a robust method based on Haar features is applied. Within the segmented areas of faces, random forest is utilized to locate eye regions. Once the eyes are located, the local region of eyes is extracted to yield binary images of the eye silhouettes in which the open and close stages of the eyes are revealed. The portion of the close states of the eyes during a certain number of frames is calculated to track the drowsiness signs. If this portion exceeds a predefined threshold, the system concludes that the driver tends to falling asleep and generate alert to the users.

Keywords— Human-machine interface; Drowsiness monitoring; Face detection; Decision tree.

I. INTRODUCTION

Drowsy and fatigue driving have been accounted to a major portion of accidents and other fatalities. According to the National Sleep Foundation's 2005 Sleep in America poll, 60% of adult drivers – about 168 million people – say that they have experienced driving with feeling drowsy, and more than one-third have actually fallen asleep at the wheel. The National Highway Traffic Safety Administration estimates that 100,000 police-reported crashes are directly resulted by driver fatigue each year. This results in an estimated 1,550 deaths, 71,000 injuries and \$12.5 billion in monetary losses. Therefore, the need of a drowsiness detection system is enormous. This research proposes a camera-based solution to address this issue.

There are several approaches to tackle this problem. Renner and Mehring suggested in [1] that we could detect drowsiness through brain or eye activities as the signals from our ability of information processing or ability of perception, respectively. This study follows the latter that is to monitor drowsiness through eye blinks.

One effective way to track eye states is to use electrooculogram (EOG), the measurement of electric activities of eyes' muscles. EOG is reliable due to its high sampling rate (up to 250Hz) and has been widely used in the study of drowsiness [1]. However, using EOG may not seem practical since it requires at least three electrodes for

proper operation, which makes user rather unpleasant. Moreover, EOG's signal tends to deteriorate when the contact of the electrodes with skin decreases over time. Thus, the research community is switching to video-based methods for a more practical approach.

In this paper, we present a drowsiness detection system using visual camera. Another contribution of this work is that we propose a novel method of eye region extracting using a state-of-the-art machine learning technique called decision tree (DT). Unlike traditional approaches which are based solely on image processing to localize eye silhouettes [2], decision tree allows us to determine the relative location of the eyes in the face with a wider range of skin color, light intensity, and camera angle.

II. DROWSINESS MONITORING SYSTEM

A. Face Detection

For rapid development and effectiveness we adopted the Viola and Jones work [3] on a robust object detection framework. In this study, we applied the method to detect face's location on each frame independently.

B. Eye Region Extraction

a) Dataset:

Over 100 facial images were collected for training stage. Data for eye localization included over facial images with various angles, skin color and light intensity.

b) Feature Selection:

Textons: This feature provided a compact representation for the rage of different appearances of an object. Textons have been proven effective in categorizing materials as well as generic object classes [4]. Textons of a particular pixel in the image is taken by calculating gray intensity difference of two neighboring pixels. Specifically, let denote a certain pixel in the image by m , its relative coordinate in the image (x, y) and corresponding gray value I . Then the Textons feature of m will be determined as:

$$M = f(x, y) = I(x + a, y + b) - I(x - a, y - b) \quad (1)$$

where a and b are arbitrary integer numbers. $I(x + a, y + b)$ and $I(x - a, y - b)$ are the gray intensities of the local pixels.

Haar-like feature: This type of feature share some similarities with Haar basic functions. In general, each Haar-like feature consists of two or three jointed “black” and “white” rectangles. The value of the feature is the difference between the black and white rectangular regions’ summation of all pixel gray intensity. Although there are a great number of possible Haar-like features, we only use five of them in this study.

c) Decision Tree and Random Forest:

In [5] A.Crimisi and his colleagues presented a comprehensive, unified model of a decision tree. The idea of the algorithm is to arrange chosen features into a binary tree structure so that when data passes from one node to another, the randomness of the data is reduced as much as possible.

In this study, a decision tree was used to classify pixels in the image into five classes: right eye, left eye, nose and mouth. The initial intention was to localize all primary parts of the face, not just eyes so that more information of the facial expressions was derived. However, this paper focused on eye region only.

Entropy and Information Gain:

These two concepts are fundamental for the working principle of a decision tree. Entropy represents the impurity of a node, the higher the entropy is, the less certain outcome of the node is. On the other hand, information gain (IG) measures the expected reduction in the entropy when data flows from one node to its children ones. Mathematically, two terms are defined as:

$$H(X) = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log_b p(x_i) \quad (2)$$

$$I = H(S) - \sum_{i \in \{1,2\}} \frac{|S^i|}{|S|} H(S^i).$$

In DT, we can use information gain to select the split node which yields the highest IG in the final distribution.

Split function and weak learner models:

Each split node in a tree is a binary function whose output is either 0 or 1:

$$h(v, \theta_i) \in \{0,1\} \quad (3)$$

where v is data vector input and θ is the associated weak learner. Each weak learner model is characterized by a set of parameters (ϕ, ψ, τ) . ϕ (filter function) selects some features of choice out of the entire data vector v . ψ defines the geometric primitive used to separate the input data. τ is the

parameter vector that holds the thresholds of the inequalities used in binary test.

Training objective function

To train a certain j^{th} node, in all possible weak learner θ , we choose the θ^* such that it maximizes the information gain of that node:

$$\theta_j^* = \arg \max_{\theta_j} I_j \quad (4)$$

The leaf prediction model

Each leaf node contains the distribution over classes associated to the subset of training data that has reached that leaf. The probabilistic leaf predictor model for a single tree is $p(c|v)$ where $c \in \{c_k\}$ indexing possible classes. In a classification task, the leaf output is not a probability, but rather the class that has highest probability.

$$c^* = \arg \max_c p(c|v) \quad (5)$$

Random forest model

Basically, a random forest model is constructed by several decision trees. The training and the testing phases operate in the same way as presented above. The significant feature which makes this model become powerful is the addition of the randomness model and the ensemble model.

The key aspect of decision random forest is based on the principle that its component trees are all randomly different from one another. This leads to de-correlation between the individual tree predictions and, in turn, to improve generalization. Forest randomness also helps achieve high robustness with respect to noisy data.

Randomness is injected into the trees during the training phase. Two of the most popular ways to do so are to randomize training data and randomize node optimization. These two techniques are not mutually exclusive and could be used together.

The ensemble model tells that in a forest with T and all trees are trained independently, tree testing can be done in parallel to achieve high computational efficiency. Combining all tree prediction into a single forest prediction may be done by a simple averaging operation:

$$p(c|v) = \frac{1}{T} \sum_{t=1}^T p_t(c|v) \quad (6)$$

C. Drowsiness Monitoring

Eye blink:

After eyes region was extracted, the next step was to quantify their activities. In this work, we applied two image processing techniques, global thresholding and dilation to

monitor eye states, particularly, whether they're open or not. The result was depicted in Fig. 1.

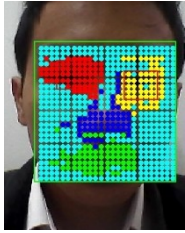


Fig.1 Facial component recognition.

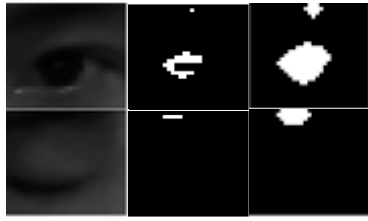


Fig.2 Eye detection.

By taking the normalized average of all pixels' intensities, we can easily create a real-time signal whose values fluctuate between two states clearly as shown in Fig.2. The formula of each value of the signal is shown as follows:

$$m = \frac{1}{WL} \sum_{y=1}^{W-1} \sum_{x=1}^{L-1} f(x,y) \quad (7)$$

where $f(x,y)$ is the binary eye image, W and L are short for the width and the length of the image, respectively.

Drowsiness detection

There are various ways to judge fatigue condition, the authors in [2] suggest that dozing or drowsiness is detected when eyes close over a specific number of consecutive. This approach is straightforward however it's likely to miss other evidences of fatigue, which reduces the accuracy of the system. Sometimes, frequent of short blinks during a certain amount of time is also a strong indicator of drowsiness besides a single long eye-closed duration.

A compromising solution is posed by calculating the portion of eye closed state within a certain number of frames. As shown in Fig. 3, the red dots represent the drowsiness level of the driver. In both discussed cases, long closed duration and frequent short blinks, the method is able to evaluate the fatigue level. When the measurement exceeds a predefined threshold the driver is considered dozing or drowsy.

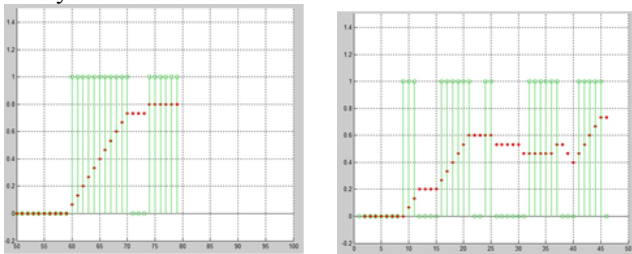


Fig. 3. The left image indicates a long closed eye duration which is clearly a sign of fatigue while The right one indicates frequent short blinks.

III. RESULT S

a) Experimental Setup:

The proposed driver drowsiness system uses an A4Tech PK-835G color camera to capture driver's images. The system is tested under the environment of Intel Atom 1.61 GHz CPU and 1.99 GB RAM. The format of input video is 320x240 RGB color. Eye tracking can achieve the speed of 15 frames per second. A Graphical User Interface was built to for a faster and easier set-up procedure. The hardware component and GUI were showed in Fig. 4

4 subjects were involved in this study. Each was asked to sit still comfortably in normal light condition while the camera was recording their activities.

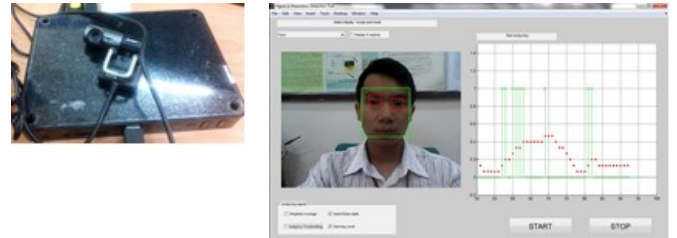


Fig. 4 Hardware setup and software interface.

b) Evaluation:

The effectiveness of the system is judged by 2 parameters: the correct rate of eye tracking and that of the eye's states detection.

The first parameter is defined in the following equation,

$$\text{Correct Rate}_{\text{eye tracking}} = \frac{\text{Total Frames} - \text{Tracking Failure}}{\text{Total Frames}} \quad (8)$$

The correct rate of eye's states detection is the portion of correct eye states prediction over the total successful eye tracked frames. Wrong prediction is the sum of false positives (eye closed but predict opened) and false negatives (eye opened but predict closed). Table 1 and Table 2 show the experimental results tested on 4 videos.

TABLE 1: RESULT OF EYE TRACKING

Video	1	2	3	4	5	6	7
Total Frames	500	500	500	500	500	500	500
Tracking Failure	2	3	0	5	0	0	73
Correct Rate (%)	99.6	99.4	100	99	100	100	85.4
Average (%)	97.6						

TABLE 2: RESULT OF EYE-STATE PREDICTION.

Video	1	2	3	4	5	6	7
Total Frames (eye tracked)	498	497	500	495	500	500	427
False Positives	11	5	3	6	3	43	2
False Negatives	20	0	29	5	56	5	0
Wrong Prediction	31	5	32	11	59	48	2
Correct Rate (%)	93.8	99.0	93.6	97.8	88.2	90.4	99.5

IV. DISCUSSION

In this work, a non-invasive visual based system to localize the eyes and monitor drowsiness was developed. Information about the head and eyes position is obtained through face detection with Haar features and random decision forest. During the monitoring, the system is able to decide if the eyes are opened or closed. When the eyes have been closed for too long and the portion of closed eyes during a certain amount of time exceeds a warning level, drowsiness sign of a driver is detected.

The indirect drowsiness testing yields good results with high accuracy and at acceptable speed. Direct testing with drowsy driver will be deployed in the further research.

Despite of the optimistic results shown above, the system still remains some significant drawbacks needed to address in the future study. First, the signal from the eyes usually corrupts when dealing with driver wearing glasses. Second, although the system is able to work under various light conditions, the responses in too dark or too bright ones are usually weak and unstable. Finally, if the distance between face and the camera becomes too far, the signal also turns out distorted which results in a number of wrong prediction.

Since the decision tree has successfully detected other primary parts on the face, the future work is to use information from other regions to produce more information on human behaviors.

ACKNOWLEDGMENT

This research is funded by Newton Research Collaboration Programme, reference number NRCP1516/1/74.

REFERENCES

1. S.Mehring, G.R.a.: 'Lane departure and drowsiness - two major accident causes - one safety system', in Editor (Ed.)^(Eds.): 'Book Lane departure and drowsiness - two major accident causes - one safety system' (Transport Research Laboratory 1997, edn.), pp.
2. Królak, A., and Strumiłło, P.: 'Eye-blink detection system for human-computer interaction', Universal Access in the Information Society, 2011, 11, (4), pp. 409-419.
3. Viola, P., and Jones, M.: 'Rapid object detection using a boosted cascade of simple features', in Editor (Ed.)^(Eds.): 'Book Rapid object detection using a boosted cascade of simple features' (2001, edn.), pp. I-511-I-518 vol.511.
4. Malik, J., Belongie, S., Leung, T., and Shi, J.: 'Contour and Texture Analysis for Image Segmentation', Int. J. Comput. Vision, 2001, 43, (1), pp. 7-27.
5. Criminisi, A.a.S., J. and Konukoglu, E.: 'Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning.', in Editor (Ed.)^(Eds.): 'Book Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning.' (Microsoft Research, 2011, edn.).

Corresponding Author: Nguyen Duc Thang
 Institute: International University-VNUHCM
 City: Ho Chi Minh
 Country: Vietnam
 Email: ndthang@hcmiu.edu.vn