2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Driver Drowsiness Detection through HMM based Dynamic Modeling

Eyosiyas Tadesse, Weihua Sheng, Meiqin Liu

Abstract—**Drowsiness is one of the main causes of severe traffic accidents occurring in our daily life. In order to reduce the number of drowsiness-induced accidents, various researches have been conducted with the aim of finding practical and non-invasive drowsiness detection systems by using behavioral measuring techniques. Many of the previous works on behavioral measuring techniques have mainly focused on the analysis of eye closure and blinking of the driver. It is recently that more attention started to shift to inclusion of other facial expressions and only few, among those researches, have been done on the analysis of temporal dynamics of facial expressions for drowsiness detection. In this paper we propose a new method of analyzing the facial expression of the driver through Hidden Markov Model (HMM) based dynamic modeling to detect drowsiness. We have implemented the algorithm using a simulated driving setup. Experimental results verified the effectiveness of the proposed method.**

Keywords

Drowsiness detection, facial expression, SVM, HMM

I. Introduction

According to the US National Highway Traffic Safety Administration, approximately 100,000 crashes occur in US each year due to drivers' drowsiness [1]. In an effort to prevent such crashes, the U.S. Department of Transportation has taken a notable initiative in the making of intelligent vehicles. In this context, the development of robust and practical drowsiness detection system is a crucial step. Many researches are being undertaken to develop better ways of detecting drowsiness, such as the behavioral, physiological changes of the driver, the steering wheel movement or vehicle responses, etc. It is critical that a drowsiness detection system should be accurate and reliable when they are deployed for commercial use. Even if vehicle based drowsiness detection systems are noninvasive, they have been found to be very unreliable as they depend on the nature of the road, the vehicle, the traffic, the way the driver drives and other external factors. Behavioral measuring methods are more reliable than vehicle based systems and are also noninvasive and easier to be implemented. However, many of the commercially available behavioral measuring methods mainly focus on eye closure and not on other facial expressions.

In this paper, we propose a drowsiness detection method that includes other facial motions and behavioral changes in addition to eye closure. We also adopted a dynamic model for analyzing the facial expressions to determine drowsiness which will significantly improve the reliability of drowsiness detection. We first developed a frame based drowsiness detection algorithm. Then we introduced drowsiness detection based on temporal analysis of facial expression and demonstrated its advantage over frame based drowsiness detection through experiments. We have optimized the system parameters to maximize the accuracy and speed of detection. We conducted the experiments in a simulated driving environment.

II. Related Work

Real time drowsiness detection has been implemented using different detection techniques analyzing various types of input data. The first approach is analyzing the measurement of physiological activities of the human body, such as brain wave (EEG), heart rate or pulse rate [2]. Even though the measurements and their correlation with the alertness of the driver is quite accurate, they are not practical as it would require the driver to always wear the sensing devices and the hardware cost is too high for commercial use.

The second approach makes use of vehicle based measuring techniques to detect the drowsiness of the driver. In this approach, the driver's drowsiness is measured by analyzing the different controller signals of the vehicle, such as steering wheel movement, pressure from the gas and brake pedal, speed of the vehicle, change in shift lever and deviation from lane position [3]. The measurements of these signals are obtained from sensors equipped in the vehicle. Among the vehicle based metrics that have been used to determine drowsiness, steering wheel movement has been shown to give better detection capability [4]. The steering angle is constantly measured by a sensor and the change in angle movement is checked if it is within or exceeds a specified threshold. Even though vehicle based approaches are noninvasive, they are not reliable in detecting drowsiness as their performance is highly affected by the nature of the road, the way the driver drives, the traffic or a driving impediment other than being drowsy.

The third approach is behavioral measuring that makes use of computer vision techniques to detect the changes in driver's facial expressions [5]. Existing works in this area have mainly relied on analyzing the percentage of closure (PERCLOS) of the driver's eyes. The first step in such systems is face and eye detection. Li *et al*. [6] performed successive image filtering techniques such as image subtraction, morphologically closed operations and binarization, and finally counted the number of pixels around the eyes region to detect eye closure. Liu *et al*. [7] extracted simple features from the temporal difference image and used

This work is partially supported by the NSF grant CISE/IIS 1231671 and National Natural Science Foundation of China under Grants 61328302 and 61222310. Eyosiyas Tadesse and Weihua Sheng (contacting author) are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater,OK74078, USA (e-mail: weihua.sheng@okstate.edu). Meiqin Liu is with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China.

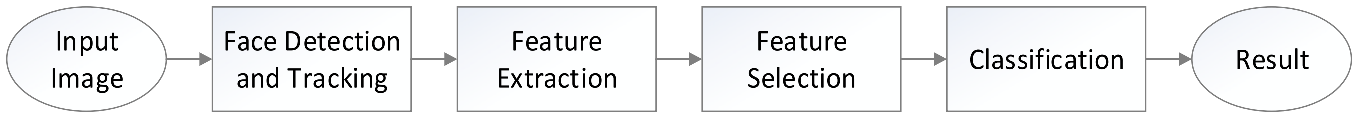

Figure 1. The system diagram of drowsiness detection using facial expression.

them to analyze the rules of eyelid movement during drowsiness. Flores *et al.* [8] computed the binary, gradient and logarithm image of eyes region, obtained random samples around the region and used an elliptic shape to represent the eyes. They then used an SVM classifier [9] to decide whether the eyes are closed or not. Garcia *et al.* [10] have also presented a non-intrusive approach to drowsiness detection. They used an IR illumination system and a high resolution camera to accept a stream of images and perform face and eye detection.

In recent years, attention has started to shift from analysis of eye blinks to facial expressions. One of the first studies was conducted by Gu & Ji [11] where they included certain facial expressions other than blinks. These facial expressions were used as action units and a dynamic Bayesian network (DBN) is adopted to detect fatigue. Vural *et al.* [12] employed machine learning methods to analyze facial motion from video. They developed fully automated facial expression analysis based on the Facial Action Coding System (FACS) [13]. These facial motions include blinking, yawn motions, eye gaze movements and other movements.

Many of the researches on behavior based drowsiness detection system used frame based classification techniques that give decision using the spatial features extracted from one input image frame. While this is simple to implement, it lacks efficiency in situations where there is non-uniform change in transition between drowsy and non-drowsy episodes which actually is the case in most real life scenarios. Moreover, analysis of image sequences gives more accurate description of facial expressions and frame based classification approaches do not utilize all the information available in image sequences. The dynamic Bayesian network in Gu and Ji's [11] work consists of a first order HMM along with the Bayesian network to capture the temporal dynamics of the facial movements during drowsiness across consequent frames in a specific period of time. Yin *et al.* [14] have also implemented dynamic drowsiness detection system using Local Binary Pattern (LBP) operators on multi-scale Gabor features of image sequences. Generally, there is still a challenge in extracting dynamic features of facial expressions for drowsiness detection and there have only been few researches done in this area thus far.

In this paper, we propose a new and efficient method of extracting and processing the dynamic features of facial expressions through HMM modeling. The remainder of this paper is organized as follows. In Section III, we present the algorithm of frame based drowsiness detection using facial expression recognition. In Section IV, we present the algorithm of drowsiness detection using HMM based dynamic modeling. We then discuss the experimental setup in Section V and give the experimental results in Section VI. Finally, the conclusion and future work are given in Section VII.

III. Drowsiness Detection Using Single Frame Based Analysis

Facial expressions based drowsiness detection makes use of computer vision to detect and recognize the facial motion and appearance changes during drowsiness. The system diagram is shown in Figure 1. It accepts a stream of input images from a camera in front of the driver at a rate of 20 frames per second. The stream of images then passes through the four main image processing stages: face detection and tracking, feature extraction, feature selection and classification.

1. *Face Detection and tracking*

Each frame is first converted to grayscale. Then the system performs histogram equalization to increase the contrast of the image for better face detection. Then it uses the Viola-Jones robust real time face detection algorithm [15] implemented in OpenCV [16] to detect the face of the driver. However, the face detector is not reliable to effectively localize the face when the driver's head rotates to certain angles, suddenly moves or turns to certain directions which frequently happens when the driver is drowsy. Hence we implemented Camshift tracking algorithm [17] to track the face of the driver under different circumstances where the face detector fails to detect. The final face region from Camshift is then passed to the next processing stage to extract features.

2. *Feature extraction*

The grayscale image is passed to the Matlab engine along with the locations of the detected face. We have two options for feature extraction:

- Crop the detected face of the driver
- Crop the region where the eyes are most likely located

The input image is reshaped to a fixed size and its Gabor features are extracted through Gabor Wavelet Decomposition [18]. We preferred Gabor wavelet features for detection because they can represent changes in surface textures such as wrinkles, bulges and changes in feature shapes and they are relatively more robust to illumination changes and random head movement.

3. *Feature Selection – Adaboost weak learning Algorithm*

The facial features from the Gabor decomposition are too many to be used for classification in their entirety and hold a significant amount of redundant information. Hence we used the Adaboost weak learning algorithm [19] to select the most important features for classification.

Adaptive boosting is an algorithm for constructing a "strong" classifier as linear combination of "weak" classifiers $h_t(x)$.

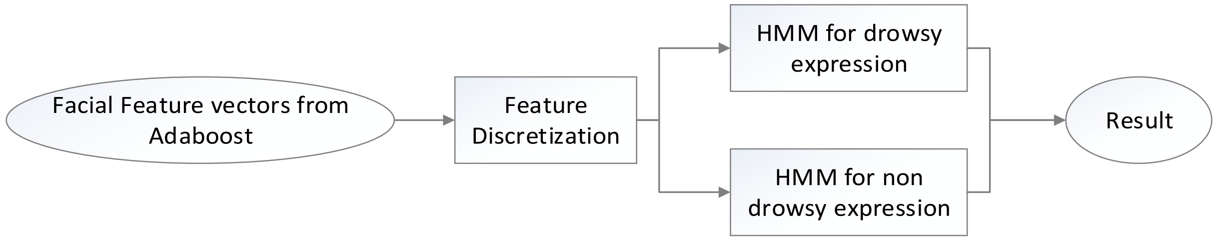

Figure 2. The system diagram of drowsiness detection using HMM based dynamic modeling.

$$f(x)=\sum_{t=1}^{T}\alpha h_t(x) \tag{1}$$

The weak classifier used here is a simple threshold function $h_j(x)$ consisting of only one feature $f_j(x)$ [20].

$$h_j(x)=\begin{cases}1 & if \quad p_j f_j(x)<p_j\lambda_j\\ -1 & otherwise\end{cases} \tag{2}$$

where λ_j is a threshold and p_j is a parity to indicate the direction of the inequality.

We compute the threshold value in two different ways:

- *Averaging:* It can be computed as the average of the mean value of the positive samples and the mean value of the negative samples on the j^{th} feature response

$$\lambda_j=\frac{1}{2}\left(\frac{1}{m}\sum_{p=1}^{m} f_j(x_p \mid y_p=1)+\frac{1}{l}\sum_{n=1}^{l} f_j(x_n \mid y_n=-1)\right) \tag{3}$$

- *Searching-maximum:* We can also choose a threshold among the j^{th} feature of all the samples that maximizes separation between the classes:

$$\lambda_j=\max\left(\arg\min\left\{S^+ + \left(T^- - S^-\right), S^- + \left(T^+ - S^+\right)\right\}\right) \tag{4}$$

Where S^+ is the number of positive samples below threshold, S^- is the number of negative samples below threshold, T^+ is the total number of positive samples and T^- is the total number of negative samples.

4. Classification

We are essentially dealing with a two-class problem (drowsy or non-drowsy). We chose SVM as it is generically used for binary classification problems and has attributes that make it a perfect fit to our problem. SVM does not depend on the dimensionality of the input space, is less prone to over-fitting and always gives an optimum global solution during training. In addition to the SVM, we have also used Adaboost cascaded classifier for comparison.

i. Adaboost cascaded classifier

We linearly combine the weak classifiers working on each selected feature to get a strong classifier and obtain the classification output $H(x)$ as follows:

$$H(x)=sign\left(\sum_{t=1}^{T}\alpha h_t(x)\right) \tag{5}$$

ii. Support Vector Machines (SVM)

We feed the selected features to the support vector machine for nonlinear classification by using the kernel method which proved to have a gain in performance over the linear combination of the Adaboost weak classifiers.

IV. Drowsiness Detection Using HMM Based Dynamic Modeling

It has been shown that facial expressions are better recognized through sequences of frames [21]. This is because they have a unique temporal pattern of behavioral changes that can be easily recognized. Similarly, drivers have certain temporal changes of facial expressions when they are feeling drowsy. By using Hidden Markov Models (HMMs), we captured the temporal information of the facial expressions of the driver which leads to more accurate classification results as compared to single frame based drowsiness detection.

HMM is modeled to have a set of unobservable stochastic processes (hidden states) that produce a sequence of observations. While the sequences of observations in HMM are essentially discrete symbols, the input signal of our system is a multidimensional feature vector extracted from the detected face of the driver. Hence, we quantized the feature vectors to discrete symbols through Gaussian mixture models which uses Expectation Maximization (EM) algorithm to cluster the feature vectors to different classes corresponding to the observation symbols. After having the observation symbols, we adopted two HMMs for drowsy and non drowsy facial expression detections. The detail explanation for HMM is presented in [22].

In our implementation, we quantized the Gabor features selected through Adaboost to definite discrete observation symbols. We used the same centroids to cluster the drowsy and non-drowsy image sequences and trained both models with their respective observation sequences. We used Viterbi algorithm to estimate the state transition and state-to-observation. The block diagram of the system is shown in Figure 2.

V. Experimental Setup

To experiment and evaluate our proposed approach, we have set up a G27 Logitech racing wheel system and a Logitech Communicate QuickCam STX webcam to implement our system as shown in Figure 3. We also used a driving simulation software Simuride [23] which gives a visual display of the actual traffic scenes along with the car dashboard. By using Simuride, we made two users drive in different scenarios in both drowsy and non-drowsy conditions and collected training and testing images for the analysis and evaluation of our approach.

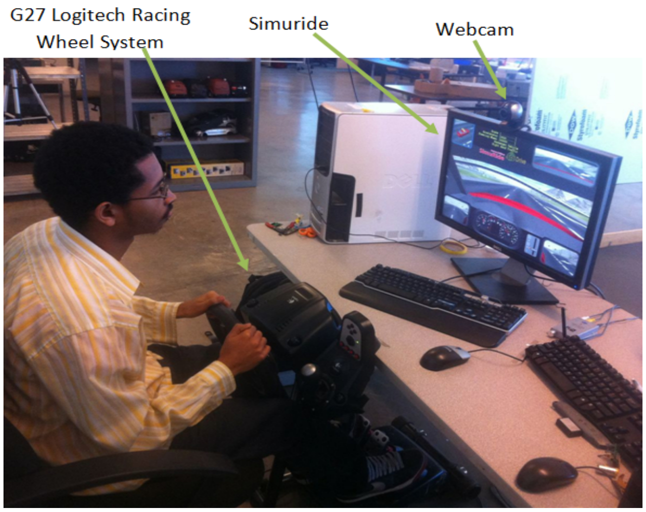

Figure 3. Experimental Setup.

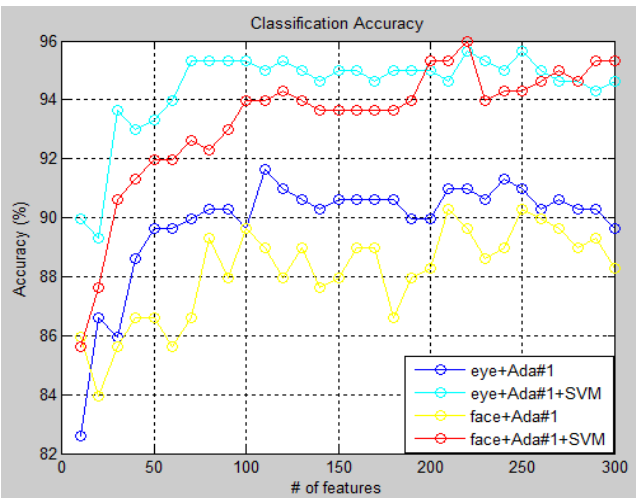

Figure 4. Classification accuracy for approach# 1 threshold computation

VI. Experimental Results

A. *Drowsiness Detection Based on Single Frame Analysis*

In the training stage, we have selected two thirds of the image frames of the labeled video data (177 non-drowsy and 179 drowsy images) from two videos of two different drivers. For testing, on third of the image frames of the video data recorded (95 non-drowsy images and 84 drowsy images) have been used. All input images are normalized to a matrix of 100x100 if the input is the detected face and a matrix of 200x80 if the input is the cropped eyes region. The Gabor wavelet is of 2 scales and 4 orientations filter bank.

During classification, we computed the average accuracy of drowsiness detection using facial expression recognition with different number of features, choice of region of interest, threshold computation, and classification techniques (i.e. Adaboost or SVM). We increased the number of features selected by the Adaboost from 10 to 300 with an interval of 10 and observe the variation in performance.

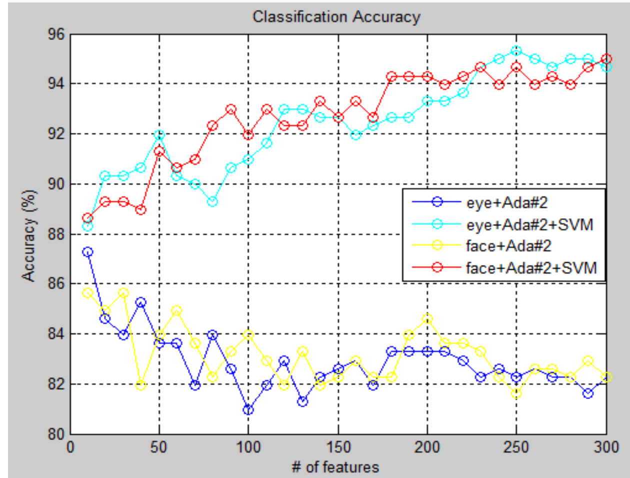

Figure 5. Classification accuracy for approach# 2 threshold computation

- *Averaging Threshold Calculation – Approach #1*

As the number of facial features selected for classification increases, the performance saturates to the accuracy values shown in Figure 4 and 5 for the different system parameter settings. From Figure 4, we can see that the accuracy roughly increases to near-maximum at a rapid rate until the number of features reaches 100. For averaging threshold computation, the accuracy of classification using Adaboost is generally much lower than that of the SVM classifier with Gaussian Radial Basis Function (RBF) kernel for the same region of interest chosen (either eye region or detected face). Moreover, for the same classification technique chosen (either Adaboost or SVM), the classification accuracy of the system using eye region as ROI is much better than using the detected face. Hence choosing eye region as the ROI and SVM as the classifier gives much better result for a broad range of selected facial features as can be seen in Figure 0. We have obtained a maximum accuracy of 95.99% for 220 facial features selected, detected face selected as ROI and SVM classification.

- *Searching-maximum Threshold Computation – Approach #2*

As shown in Figure 5, by using the Adaboost classification method, the classification accuracy gets lower as the number of features increases. It shows that the more features are selected using Adaboost of searching-maximum threshold computation, the features being added to the cascaded combination will have less significance to the detection and the overall performance deteriorates. The SVM classifier with RBF kernel, however, provides a nonlinear decision hyper-plane that better classifies the two sets of inputs. Hence, as the number of facial features increases, the classification accuracy general increases. We have obtained a maximum accuracy of 95.32% for 250 facial features, eye region selected as ROI and using SVM classification.

B. *Drowsiness Detection through HMM dynamic modeling*

Next, we implemented the dynamic approach of recognizing drowsy and non-drowsy facial expressions from

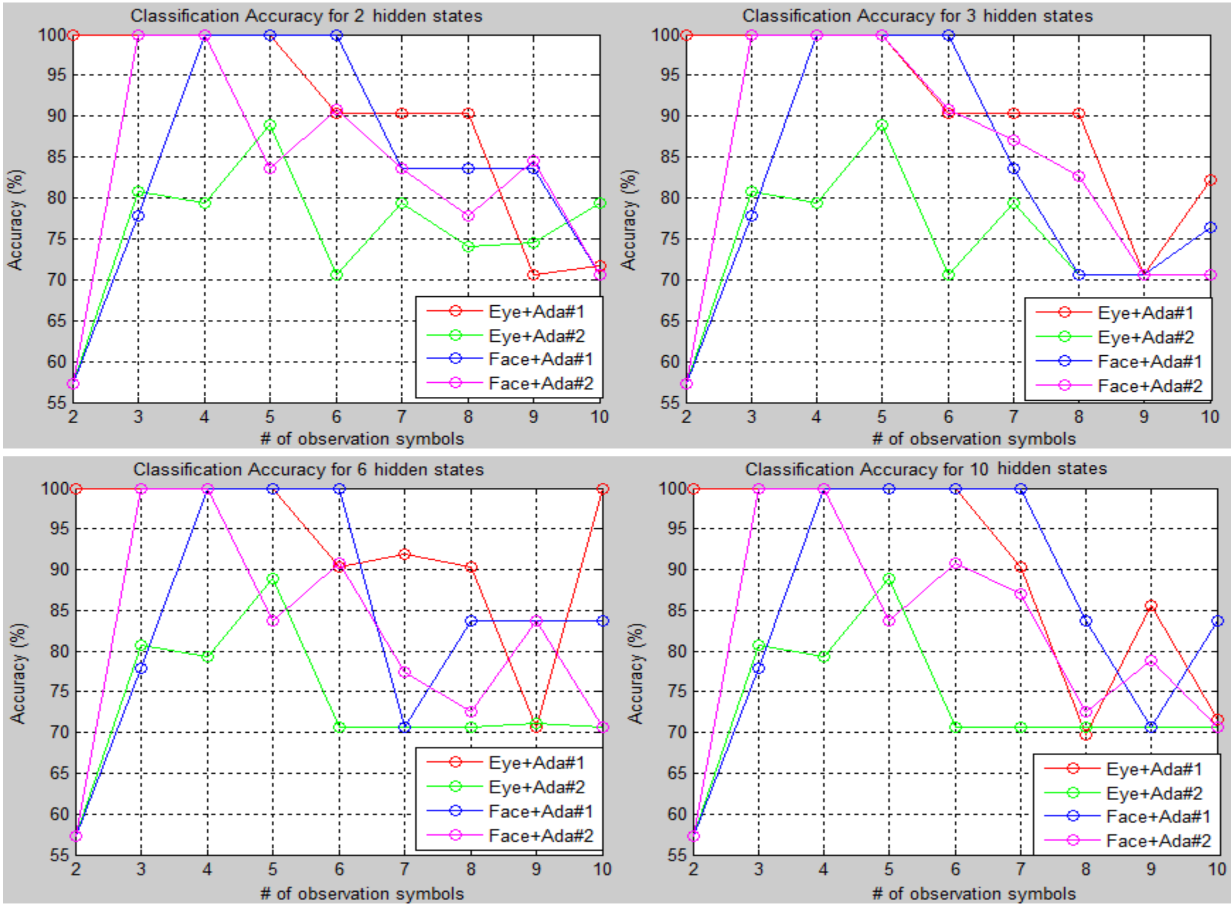

Figure 6. Classification accuracy for different number of hidden states and observation symbols.

sequences of images. We trained two HMMs for drowsy and non-drowsy scenarios with similar sets of training image sequences. We first optimized the number of observation symbols to the hidden states while keeping the same number of facial features selected and the window size constant.

We kept the number of features at 100 which is minimum yet optimum as shown in Figure 4 and 5. We also kept the window size to 20 which means that decision is given for a sequence of images captured in a second. As shown in Figure 6, we have plotted variations of classification accuracies for different number of observation symbols while keeping the number of hidden states constant for each case. It can be noticed that it optimizes in the range of 3 to 6 observation symbols for all the cases. The number of observation symbols plays a vital role in HMM modeling as it reflects the transition in the facial expression to which it is trained for.

We also optimized the window size of the sequences of images while keeping the number of features to 100 and the number of hidden states and observation symbols to the optimum combination that gave us the maximum classification accuracy. Optimizing video segmentation is a main challenge in temporal classification. The window size may be too small to properly classify the facial expression. It may also be too large that definite transitions between facial expressions will be overwhelmed in one window size and the facial expression may not be correctly classified. We picked 10 hidden states to 5 observation symbols combination. In Figure 8, the cases for Adaboost approach #2 have lower accuracies and those for Adaboost approach #1 have high accuracies. In Figure 7, it attests that an increase in window size increases the accuracy and reaches saturation at 10 and 15 for cases using Adaboost approach #2. On the other hand, for Adaboost approach #2, an increase in window size deteriorates the accuracy as shown in Figure 7.

In Table 1, we compare the average accuracies achieved by the two drowsiness detection approachess. It can be seen that

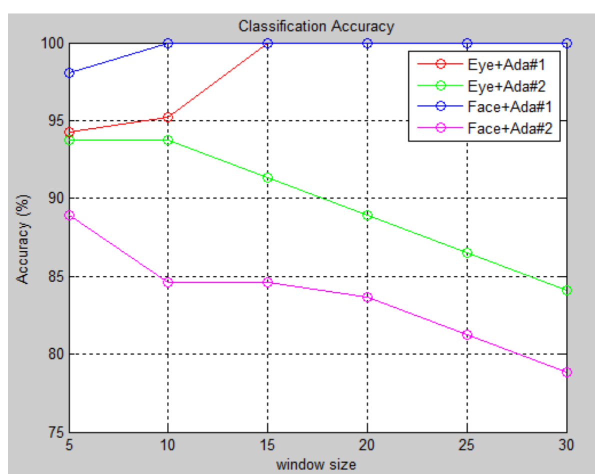

Figure 7. Classification accuracy for approach# 1 threshold computation

in average the dynamic approach gives better classification accuracy even for a smaller number of facial features selected than the single frame based classification approach.

Types of approaches	Accuracy
Facial expression recognition	90%
HMM based dynamic modeling	97%

Table 1. Classification accuracy of the two approaches

VII. Conclusion

This paper investigates the driver drowsiness detection using facial expression recognition for single frame based analysis and HMM based dynamic modeling. We have independently implemented the two methods and evaluated their performances for different system parameters. The performance advantage of the dynamic approach over the single frame based analysis entails that facial expressions are better recognized through the analysis of sequence of frames. We have also optimized the number of observation symbols and hidden states. The variation in the number of hidden states has little effect on the classification accuracy of the system. The number of observation symbols was optimized to a range of values which roughly shows better accuracy results for the different cases of parameter settings. Currently, the system has been trained and tested using inputs from two users. In order to increase the robustness of the system, collecting more training data from more users will be conducted. In our implementation, we have preset the window size to a fixed value and optimizing it for different parameter settings was difficult. In the future, automatic segmentation of the sequences of images to adapt to the different transitions in facial expressions will be considered.

References

1. Hartman, K. and J. Strasser, *Saving Lives Through Advanced Vehicle Safety Technology: Intelligent Vehicle Initiative Final Report*. 2005, Department of Transportation: Washington, DC. p. 12.
2. Akin, M., et al., *Estimating vigilance level by using EEG and EMG signals.* Neural Computing and Applications, 2008. **17**(3): p. 227-236.
3. Sahayadhas, A., K. Sundaraj, and M. Murugappan, *Detecting driver drowsiness based on sensors: a review.* Sensors (Basel), 2012. **12**(12): p. 16937-53.
4. Eskandarian, A. and R. Sayed, *Unobtrusive drowsiness detection by neural network learning of driver steering.* Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2001. **215**(9): p. 969-975.
5. Vural, E., et al., *Machine Learning Systems for Detecting Driver Drowsiness*, in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, et al., Editors. 2009, Springer US. p. 97-110.
6. Xing, L., et al. *A new method for detecting fatigue driving with camera based on OpenCV*. in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*. 2011.
7. Danghui, L., et al. *Drowsiness Detection Based on Eyelid Movement*. in *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*. 2010.
8. Flores, M.J., J.M. Armingol, and A. Escalera. *Real-time drowsiness detection system for an intelligent vehicle*. in *Intelligent Vehicles Symposium, 2008 IEEE*. 2008.
9. Cristianini, N. and J. Shawe-Taylor, *An introduction to support Vector Machines: and other kernel-based learning methods*. 2000: Cambridge University Press. 189.
10. Garcia, I., et al. *Vision-based drowsiness detector for real driving conditions*. in *Intelligent Vehicles Symposium (IV), 2012 IEEE*. 2012.
11. Haisong, G. and J. Qiang. *An automated face reader for fatigue detection*. in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*. 2004.
12. Vural, E., et al. *Automated Drowsiness Detection for Improved Driving Safety*. in *The International Conference on Automotive Technologies*. 2008.
13. Bartlett, M.S., et al., *Automatic recognition of facial actions in spontaneous expressions.* Journal of Multimedia, 2006. **1**(6): p. 22-35.
14. Yin, B.-C., X. Fan, and Y.-F. Sun, *Multiscale dynamic features based driver fatigue detection.* International Journal of Pattern Recognition and Artificial Intelligence, 2009. **23**(03): p. 575-589.
15. Viola, P. and M. Jones. *Rapid object detection using a boosted cascade of simple features*. in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001.
16. Bradski, G., *The OpenCV Library.* Dr. Bobb's Journal of Software Tools, 2000. **25**(11): p. 120-+.
17. Donghe, Y. and X. Jinsong. *Face Tracking Based on Camshift Algorithm and Motion Prediction*. in *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*. 2009.
18. Movellan, J.R., *Tutorial on Gabor Filters.* Tutorial paper http://mplab.ucsd.edu/tutorials/pdfs/gabor.pdf, 2008.
19. Freund, Y. and R. Schapire, *A short introduction to boosting.* Japonese Society for Artificial Intelligence, 1999. **14**(5): p. 771-780.
20. Shen, L. and L. Bai, *AdaBoost Gabor Feature Selection for Classification*. 2004.
21. Cohen, I., et al., *Facial expression recognition from video sequences: temporal and static modeling.* Comput. Vis. Image Underst., 2003. **91**(1-2): p. 160-187.
22. Rabiner, L., *A tutorial on hidden Markov models and selected applications in speech recognition.* Proceedings of the IEEE, 1989. **77**(2): p. 257-286.
23. *SimuRide Home Edition (HE) Driving Simulation Software manual*. Available from: http://www.aplusbsoftware.com/simuride-he.html.