DRIVER FATIGUE MONITORING SYSTEM USING SUPPORT VECTOR MACHINES

Matthew Sacco, Reuben A. Farrugia

Department of Communications and Computer Engineering
University of Malta
Msida, MSD 2080, Malta
msac0005@um.edu.mt, reuben.farrugia@um.edu.mt

ABSTRACT

Driver fatigue is one of the leading causes of traffic accidents. This paper presents a real-time non-intrusive fatigue monitoring system which exploits the driver's facial expression to detect and alert fatigued drivers. The presented approach adopts the Viola-Jones classifier to detect the driver's facial features. The correlation coefficient template matching method is then applied to derive the state of each feature on a frame by frame basis. A Support Vector Machine (SVM) is finally integrated within the system to classify the facial appearance as either fatigued or otherwise. Using this simple and cheap implementation, the overall system achieved an accuracy of 95.2%, outperforming other developed systems employing expensive hardware to reach the same objective.

Index Terms— Driver Fatigue Monitoring, Fatigue Detection, Support Vector Machines, Computer Vision.

1. INTRODUCTION

The U.S. National Highway Traffic Safety Administration (NHTSA) estimates that between 2 and 23% of all vehicle crashes can be attributed to driver fatigue [1]. Every year, around 100,000 traffic accidents and 71,000 injuries related to driver drowsiness are reported in the United States, out of which more than 1,300 are fatal [2]. The NHTSA also reports that in the year 2005 alone, there were approximately 5,000 road fatalities which were caused by fatigue [3]. Moreover, around 16,000 other fatalities were caused by lane-keeping failure, one of the indirect effects of fatigue on drivers. These alarming statistics pointed to the need to design and implement systems capable of tracking and analysing a driver's facial characteristics or body states, issuing a warning signal at the earliest possible indications of fatigue to try and prevent the likely occurrence of an accident.

Many different approaches for systems tackling the problem of driver tiredness have been studied and implemented over the past few years. Earlier devices, such as those presented in [4], [5] tended to be rather intrusive,

requiring physical contact to measure fatigue features while driving. These features included heart rate variability, analysis of brain signals as well as the driver's physiological state. Other systems studied the relation of driver drowsiness to steering grip and vehicle movements [6], with some also employing lane tracking facilities [7]. However, the focus nowadays is more towards autonomous non-intrusive systems that work in the background without distracting the driver in any way, having the ability to detect and track head and eye movements by means of one or more cameras mounted on the vehicle's dashboard. Several fatigue detection approaches have been attempted in recent years, including the use of neural networks [8] and probabilistic models [9]. In the system presented in this paper, a real-time non-intrusive approach for driver fatigue monitoring is adopted, exploiting the driver's facial expression using computer vision techniques and a Support Vector Machine for classification purposes.

The rest of this paper is organized as follows: in Sections 2 and 3, the concepts behind the face and facial feature detection and fatigue classification techniques are outlined. Section 4 discusses the design and implementation details at every stage of the application, whereas Section 5 presents the results achieved using different evaluation methods. A number of conclusions are finally drawn in Section 6.

2. VIOLA-JONES RAPID OBJECT DETECTION

The rapid object detection method presented in [10] forms the basis of the face and facial feature detection approach considered in this work. Instead of analysing an image pixel by pixel to determine corresponding RGB values, this technique is based on Haar-like features, which resemble the Haar wavelets. This leads to a major reduction in the required computational time. The Viola-Jones object detection framework is based on three important principles: i) working out the integral image representation of a grayscale image to enable quick evaluation of Haar-like features; ii) modifying slightly the AdaBoost learning algorithm to produce a boosted classifier, making use of only the most critical features; and iii) organizing classifiers in ascending order of complexity (forming a cascade) to

quickly eliminate the useless regions in the image, shifting the focus instead on those regions likely to contain the required object. This cascade of classifiers is required to be able to achieve more accurate detection due to the fact that Haar-like features on their own are considered as weak classifiers, that is, classifiers whose performance is only marginally better than that of random classifiers.

The process of training these classifiers starts with the collection of both positive and negative images of the object to be detected. For the positive training set, a list of the images in that set needs to be placed in a text file containing the image directory, the number of objects to be detected in that image and the x and y coordinates of the location as well as the dimensions of each object. A similar list is also created for the negative training set, containing only the image directory.

The Open Source Computer Vision Library (OpenCV) developed by Intel Corporation [11] provides a command prompt training utility called haartraining which generates a classifier in XML format when given positive and negative examples of the object to be detected. The duration of the training process depends heavily on the choice of parameters for this utility, and it normally lasts between 5 to 7 days. However, the recognition phase can be applied in real-time.

3. SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are very often used for the classification of data in a wide variety of machine learning applications. Given a set of training examples, the task of the SVM is to generate a model that enables it to classify unseen examples into the appropriate targets. In general, for a training set D consisting of m examples,

$$D = \{(\vec{x}_1, y_1), (\vec{x}_2, y_2),...,(\vec{x}_m, y_m)\}\, \vec{x} \in \Re^N, y \in \{-1, 1\} \quad (1)$$

where $\vec{x}_i$ are the N-dimensional feature vectors, y_i are the corresponding target values for those vectors and m is the total number of feature vectors [12].

By applying what is known as the kernel trick, a non-linear mapping function maps the original data points onto a higher-dimensional feature space in which linear separability could be attained. The SVM then linearly classifies the transformed data points in the new feature space, even though the input space may not be linearly separable. The mapping function is known as the kernel function, and it is defined in terms of a transform φ, where the dot product between two vectors in the input space is being considered. Both Linear and Radial Basis Function (RBF) kernel models were considered and used during the training and testing phases. The Linear kernel function is given by:

$$K(x_i, x_j) = x_i * x_j \quad (2)$$

whereas the RBF kernel is defined as follows:

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2)\, \gamma > 0 \quad (3)$$

When compared to other types of classifiers, apart from being capable of classifying data that is not linearly separable, SVMs also perform particularly well when the available training set is limited. Moreover, it is known that SVMs provide better generalization and are not affected by over-fitting.

4. DESIGN AND IMPLEMENTATION

The developed fatigue monitoring system consists of five main stages: pre-processing of the original image, face and facial feature detection, template matching, feature extraction and SVM fatigue state classification. The flowchart shown in Figure 1 illustrates the major processes taking place. The application was implemented in C++, making extensive use OpenCV, which provides many tools specifically designed for tasks involving computer vision.
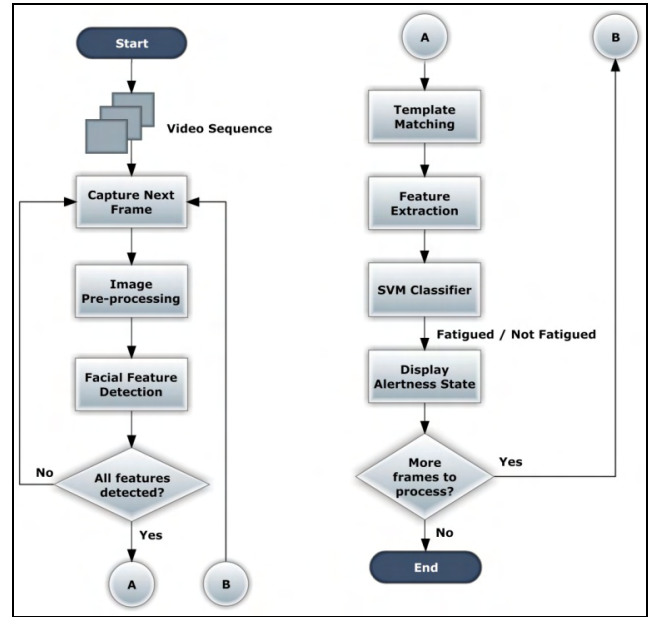

Figure 1: High-level flowchart of the developed fatigue monitoring system.

4.1. Image Pre-processing

Before face and facial feature detection, the captured frame is pre-processed to normalize the illumination and reduce image noise. The first step is to convert the RGB image into grayscale to be able to analyze single pixel intensity values. Histogram equalization is then performed to increase image contrast and reduce the effects of different illumination conditions as much as possible. The 3×3 median filter is then used to reduce the effect of noise.

4.2. Facial Feature Detection

Three separate cascades of classifiers based on the Viola-Jones object detection technique are used for face, eye and mouth detection. To reduce the number of false positives and the computational complexity, the concept of regions of interest was adopted. After identifying the location of the face region, subsequent processing takes place only within that region of interest. In particular, the eye and mouth regions are estimated heuristically with respect to the width and height of the detected face, and the search for the corresponding features is restricted only to these regions.

To further improve the facial feature detection rate, it is assumed that when at least one of the features is not found, the driver's facial position has not moved in a significant manner since the last correctly detected frame. Hence, the feature locations and dimensions of that frame are retained and also used for the current frame. This is particularly useful to keep detecting eyes even when they are in a closed state.

4.3. Template Matching

After successful face and facial feature detection, eye and mouth states are determined in every frame using the correlation coefficient template matching method. Due to the required computational complexity of this method, the reference images are restricted to the same search regions as those used for the facial feature detections, with each region being compared to two corresponding open and closed instances. The six grayscale equalized templates adopted in this work are illustrated in Figure 2.

Figure 2: The six templates considered.

All templates are initially resized according to the size of the actual detected features. Each template is then slid across the appropriate search regions, with the resulting correlation scores being raised to the power of a constant to get more pronounced values, further discriminating between open and closed features. Finally, the maximum correlation score in that image is retrieved, allowing comparison between pairs of correlation scores to establish the state of each feature (open/closed) in every frame.

4.4. Feature Extraction

The features adopted in this work were designed by observing a set of videos containing fatigued and non-fatigued subjects. These features will then be used by the Support Vector Machine classifier for classification purposes. The feature values therefore needed to be discriminative between these two situations. The average eye closure interval, the PERcentage eye CLOSure over time (PERCLOS) and the degree of mouth opening were the chosen features to be extracted for the developed system.

The eye closure interval feature calculates the duration between successive closed eye states. This feature was found to be significantly correlated with fatigue. Every time both eyes are detected to be closed, the elapsed time since the last eye closure is measured. PERCLOS is another widely used measure to discriminate between changing levels of attention. Its calculation simply involves dividing the number of frames in which the eye is closed by the total amount of processed frames in a given period of time. Drowsiness often leads to an increase in the number of closed frames, mainly because the individual tends to take a longer time to re-open his eyes. Similarly, the degree of mouth opening could suggest a decreased alertness level, indicating an increase in the yawning frequency.

4.5. SVM Fatigue State Classification

The extracted features are finally combined and presented to a Support Vector Machine classifier, resulting in either one of two possible driver states. Two class labels were identified: +1 to indicate a normal state and -1 for a drowsy state, and the three features were all scaled to values between 0 and 1. The video sequences considered in this paper include sequences generated within a controlled environment (constant illumination) and others captured within a real vehicle (variable illumination). The SVM classifier was trained using 10-fold cross-validation on a database containing a total of 3,000 labeled vectors. The trained parameters are then used by the SVM classifier for recognition purposes. This classifier was integrated within the OpenCV library to recognize fatigued drivers in real-time.

The concept of overlapping fatigue windows was adopted in order to quickly determine a fatigued state. 20-second windows were considered, whose timer is activated successively every second for the first 20 seconds of the running application. The timer for each window is immediately restarted once the corresponding 20-second period elapses, and the SVM classification for the feature values in that window is subsequently issued. In this manner, the driver's state is constantly updated.

5. RESULTS AND EVALUATION

In order to evaluate the performance of the Viola-Jones classifiers for face, eye and mouth detection, a testing set consisting of 3,000 positive and 3,000 negative images was used. The positive images (frontal faces) were collected from the FERET face database [13], whereas the negative set consisted of non-facial images of size 100×100 captured from a video clip recorded inside a vehicle. The accuracy,

precision and recall results for detection are illustrated in Table 1. It must be noted that the accuracy of the eye and mouth detection classifiers was significantly improved by limiting the search within a heuristically defined region.

TABLE 1
DETECTION RESULTS

	Accuracy	Precision	Recall
Face Detection	99.9%	99.9%	100%
Eye Detection	91%	83.8%	99.5%
Mouth Detection	91.7%	86.4%	99.1%

Four different test subjects evaluated the developed system. The average eye template matching accuracy rate was 96% while the corresponding template matching value for the mouth region reached 91.9%. When single fatigue features were analysed, accuracy rates of 93% (PERCLOS), 72.7% (mouth opening) and 76.9% (average eye closure interval) were attained. As can be seen in Tables 2 and 3, the final results for system accuracy using all three features were 93.2% and 95.2% for Linear and RBF SVM kernel functions respectively. These results confirm that applying multiple features together with a robust classifier such as SVM provides better results than when considering single features with heuristic thresholds.

TABLE 2
SVM CLASSIFICATION – LINEAR KERNEL

	Constant Illumination		Changing Illumination		Overall Accuracy
#	Attentive	Fatigued	Attentive	Fatigued	
1	100%	100%	62.66%	100%	90.67%
2	100%	100%	100%	72.22%	93.06%
3	94.59%	84.29%	91.22%	98.45%	92.14%
4	100%	100%	100%	88.40%	97.10%
Overall System Accuracy					**93.24%**

TABLE 3
SVM CLASSIFICATION – RBF KERNEL

	Constant Illumination		Changing Illumination		Overall Accuracy
#	Attentive	Fatigued	Attentive	Fatigued	
1	100%	100%	94.15%	97.13%	97.82%
2	100%	100%	100%	69.54%	92.39%
3	100%	90%	88.02%	97.53%	93.89%
4	100%	95.80%	100%	90.94%	96.69%
Overall System Accuracy					**95.20%**

An illustrative example of the developed real-time fatigue detection system is shown in Figure 3. In both cases, the face, eyes and mouth are surrounded by boxes and a message is written describing whether the mouth or eyes are closed. An alarm message is launched if the test subject is detected to be fatigued.
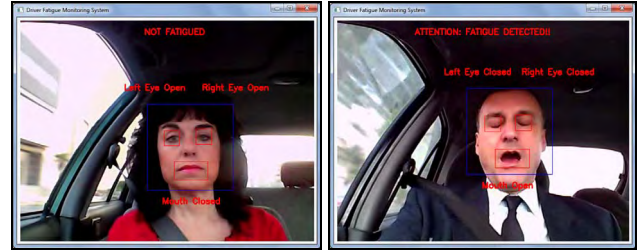

Figure 3: Fatigued and non-fatigued classification in real-time.

6. CONCLUSION

A real-time implementation of a driver fatigue monitoring system was presented in this paper. A number of techniques were used in the development process, including histogram equalization and median filtering for image pre-processing. The Viola-Jones object detection framework was then used to detect the face, eyes and mouth in successive frames, together with correlation coefficient template matching to determine feature states. Support Vector Machine classification based on a combination of three fatigue features was then used to detect the overall fatigue level of the driver being captured by the camera in real-time at 15 frames per second with a 640×480 resolution. In general, as demonstrated in the previous section, the computer vision approach adopted was very successful in classifying the visual appearance of the driver, achieving an average recognition rate of 95.2%. The results further demonstrate the feasibility of deploying non-intrusive fatigue monitoring systems for commercial applications.

REFERENCES

[1] D. F. Dinges, G. Maislin, J. W. Powell, and M. M. Mallis, "Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management", 1998. Available: http://ntl.bts.gov /lib/jpodocs/edlbrow/7d01!.pdf.

[2] National Highway Traffic Safety Administration, "NHTSA Vehicle Safety Rulemaking and Supporting Research Priorities: Calendar Years 2005-2009", 2005. Available: http://www.nhtsa.gov/cars/ rules/rulings/priority plan-2005.html.

[3] National Highway Traffic Safety Administration, "Traffic Safety Facts 2005: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System". Available: http://www-nrd.nhtsa.dot.gov/pubs/tsf2005.pdf.

[4] J. Cavuoto, "Alertness Monitoring Devices Emerge from San Diego". Available: http://www.neurotechreports.com/pages/alert ness.html.

[5] J-S Co. Neurocom, "Engine Driver Vigilance Telemetric Control System EDVTCS". Available: http://www.neurocom.ru/en2/pdf/edv tcs_adv_eng.pdf.

[6] Fatigue Management International, "ASTiD: Advisory System for Tired Drivers". Available: http://www.fmig.org/ASTID%20Informa tion%20Document.pdf.

[7] AssistWare Technology, "Tired of Facing Another Night Alone? SafeTRAC can help". Available: http://www.assistware.com/Down loads/SafeTRAC-Fleet%20Datasheet.pdf.

[8]   S. Ribarić, J. Lovrenčić, and N. Pavešić, "A Neural-Network-Based System for Monitoring Driver Fatigue", i*n Proc. of the 15th IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 1356-1361, 2010.

[9]   Q. Ji, Z. Zhu, and P. Lan, "Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue", *IEEE Trans. Veh. Technol. (VT)*, 53(4):1052-1068, 2004.

[10] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 511-518, 2001.

[11] G. Bradski, "The OpenCV Library", *Dr. Dobb's Journal of Software Tools*, 2000. Available: http://sourceforge.net/projects/opencvlibrary.

[12] H. Yu and S. Kim, "SVM Tutorial: Classification, Regression and Ranking", 2009. Available: http://hwanjoyu.org/publication/svm tutorial.pdf.

[13] National Institute of Standards and Technology, "The Color FERET database". Available: http://face.nist.gov/colorferet/.