DEPARTMENT OF APPLIED PHYSICS AND ELECTRONICS UMEÅ UNIVERISTY, SWEDEN

DIGITAL MEDIA LAB

A Networked Yawn Detector

Zhengrong Yao [1]
Dept. Applied Physics and Electronics
Umeå University
SE-90187, Umeå Sweden
e-mail: Zhengrong.Yao@tfe.umu.se

Haibo Li [2]
Dept. Applied Physics and Electronics
Umeå University
SE-90187, Umeå Sweden
e-mail: haibo.li@tfe.umu.se

DML Technical Report: DML-TR-2003:3

ISSN Number: 1652-8441

Report Date: November 20, 2003

1
2

Abstract

The driver's mental state could be estimated from visual clues. The typical driver's fatigue event could be detected or predicted by the dynamic facial expression events such like yawn. This paper demonstrate a networked surveillance system, where the driver's facial expression parameters are extracted from real time video of face in car and sent via wireless network to a surveillance center, where the parameters could be evaluated to find if the driver is under fatigue situation. The parameter extraction using the Model-based coding (MBC) technique. A Hidden Markov Model (HMM) is used for recognizing the yawn event which characterize a typical fatigue event. A prototype of such a networked system was set up and subjected to user tests. Promising results from user tests and their subjective evaluations are reported.

Keywords

Fatigue detection, Hidden Markov Model, Model-based coding (MBC).

1 Introduction

In the near future, cars will be equipped with intelligent control systems. The main idea behind intelligent control systems is to enhance the driver's situation awareness [20], [4]. To do so, these systems must take into consideration not only the physical traffic and road situation, but also the driver's behavior. A dangerous situation is when the driver is under fatigue. Studies show that fatigue is the cause of about 40 percent of all highway deaths [23]. Obviously, for traffic and driving safety, it is very important to detect and recognize the driver in fatigue and to give the driver a warning, in the extreme case, stop the engine in safe manner.

The importance of detecting fatigue has been realized for years. Many approaches have been proposed to handle the fatigue detection problem [9], [19], [6], [14]. A review of fatigue detection technologies was given in [8]. A recent work [10] describe a real time prototype of drive-fatigue monitor using multiple cues. According to [6] there are four classes of fatigue detection and prediction technologies:

- Readiness-to-perform and fitness-for-duty technologies;
- Mathematical models of alertness;
- Vehicle-based performance technologies;
- In-vehicle, on-line, operator status monitoring technologies.

In this paper, we are approaching fatigue detection based on computer vision and image coding techniques. Instead of rely on in-vehicle technologies, we explore the feasibility of implementing a networked driver's mental state surveillance system, which consists of both the driver's vehicle and a remote surveillance system connected via wireless communication networks. The paper is organized as following, section 2 reasons the motivation of using a networked system for fatigue detection, section 3 describes using the action unit from the yawn event as key clue for fatigue detection, section 4 describes using a probabilistic HMM model as underlying estimating engine, section 5 describes a prototype of the network system and user tests result with subjective evaluation, section 6 concludes the paper.

2 A networked mental state surveillance system

Our idea of using a networked driver's mental state surveillance system for fatigue detection is motivated by the following reasons:

- Driver's fatigue is a complex event which involves many factors such as vigilance level, physiological conditions, visual behaviors, etc. A reasonable detector of such event should employ multiple cues and probably use machine learning technology [10] to study the mathematical model of fatigue event. Due to the diversity of drivers, the training process should be personal-dependent and online. It would be extreme expensive to train such a system for each driver in reality, what's more, the human's aid is almost inevitable for training such an intelligent system and this would be impossible to do in vehicle. A more reasonable solution is to employ a remote surveillance server whose job is to train the driver's model and serve for the driver when it is necessary.
- As the development of the modern communication technology, nowadays it becomes increasingly popular to access mobile device, GPS system, cameras and many other devices within vehicle. The emerging 3G wireless network provide the possible availability of communication channel between the moving vehicle and the remote surveillance server, although the channel capacity is still limited.
- A networked system is a wanted feature for future traffic management, think about such scenarios: a driver's family want to watch on the situation of the driver, the truck company, or the police want to check driver's

situation. Although legal and privacy issues must be taken into consideration in reality, obviously there is need for such functionality.

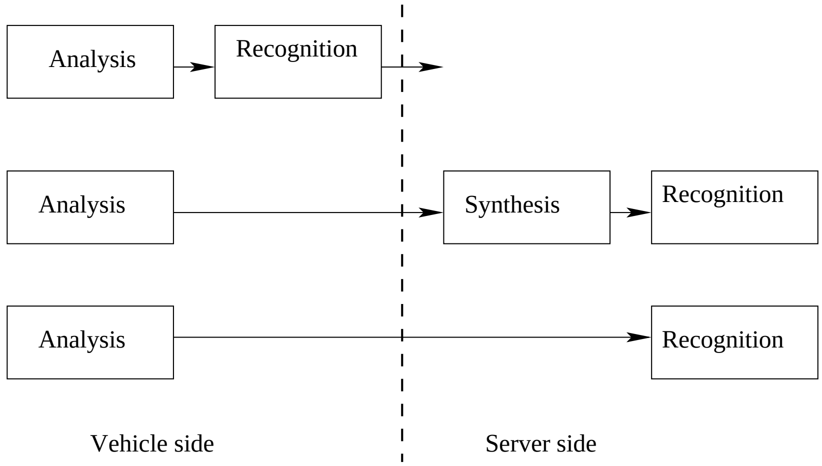

Figure 1: Three different schemes for a networked surveillance system. Upper row: the analysis and recognition of fatigue is performed at vehicle side. Middle row: the video is analyzed at vehicle side and synthesized at server side and the recognition is performed at server side. Lower row: the analysis is done at vehicle side and the recognition is done in server side without synthesis module. All the three schemes only show the information streaming direction from the vehicle side to the server side.

Generally speaking, three kinds of networked structure could be possible as shown in Fig. 1. The upper row scheme is similar to the traditional in-vehicle scheme with a network connection to the surveillance server, it consists of the one analysis module that extract visual parameters from driver's face, and a recognition module whose task is to recognize the fatigue event.

The middle row scheme is to send the video to the surveillance server that take care of the recognition of the fatigue event. Due to the limited communication channel, streaming service often valid at rate 50-1500 kbps [25]. The common steaming service schemes could not handle the video at acceptable quality at such bitrate. A very low bit rate video coding scheme, such as the Model-based coding (MBC) is a must [12] for such case. The overall scheme is very similar to a MBC scheme with an adding on recognition module. The analysis module at the vehicle side extract the parameters and the synthesis module at the server side could reconstruct the video using the received parameters followed by the recognition module.

Due to the legal issue and privacy consideration, at the server side the synthesis module could often be omitted if the primary aim is only to detect fatigue event. This is the third case shown in the lower row in Fig. 1. The analysis module at the vehicle side is unchanged compared to the middle row case, the recognition module at the server side is performed without the synthesis module. The analysis technique from MBC still could be utilized for the task. In this paper, we adopt the third scheme for our system design.

3 Yawn event

As most other works that based on visual sensor, the fatigue detection task in our work is treated as a pattern recognition problem. Fatigue detection will be done through observing the driver's face and examining facial expressions. To do so, a camera system by which the face of the driver can be captured and analyzed has to be mounted inside the car. The motivation for doing this is based on the observation that human face usually carries a lot of implicit behavior information. A somewhat exaggerated example of a facial expression associated with fatigue is illustrated in Fig. 2.

Figure 2: An exaggerated example of a facial expression associated with the mental stage fatigue.

Similarly, serious traffic situations can also be read from other types of facial expressions, e.g. fright. If the frightened facial expression can be detected, an instant decision can be made for the driver. This can greatly reduce the reaction time. The behavior can be recognized by carefully reading facial expressions.

To recognize fatigue, we have to have a good knowledge about it. Studies show that fatigue is a complicated temporal behavior, which can be divided into stages and consists of a series of events. A typical fatigue, as an example, consists of the following basic stages [23]:

- have a feeling of disinterest and a certain slowness of thought;
- stifle a yawn;
- become cold and drowsy;
- start to yawn more frequently;
- the eyelids begin to droop;
- hallucinations may appear;
- neck muscles slacken;
- head falls forward.

If we denote the event with $\mathbf{E}$ then fatigue can be represented as $\mathbf{E} = (E_1, E_2, \cdots, E_n)$, where E_i is the ith stage mentioned above.

We should note that fatigue is not a simple temporal ensemble of events. The order of the events is logical and meaningful. It is heavily constrained by psychological causality. Therefore, if the psychological causality can be learnt from the training set, then checking if the detected events are in a causal order can easily recognize fatigue. In reality, we have problems with the detection of the individual events. This is because these events are under defined and hard to measure quantitatively.

In our study, we used computer vision techniques to handle the fatigue recognition problem. Previous vision based studies have chosen to attack this problem through checking the fatigue events list before, such as eyelid dropping [19], head nodding [8]. We chose to detect yawn instead as an evidence of fatigue, there are several advantages of doing so:

- Yawn happens in earlier stage compare to eyelid dropping or head nodding, so an earlier warning is possible and will be safer.
- Although there is no known ”critical” point for fatigue, driver stifle more yawns compare to eyelid dropping and head nodding before a ”critical” point, this gives more space for error tolerance design of the detection system, since small number of yawn detection error could be not critical.
- Yawn event is relative a longer facial movement and therefore easier to be recognized if modelled carefully.
- Methods based on eyelid dropping usually need to solve the driver’s eyeglasses problem, which is usually not the case for yawn detection.
- The yawn is a dominant facial expression event associated with fatigue

Therefore, we chose to attack the fatigue detection problem through recognizing a typical facial expression event, the yawn. Here we need to distinguish facial expression from facial expression event. Facial expression is a static appearance of a face while facial expression event is a dynamic process consisting of sequential facial expressions. This naturally suggests a way to infer facial expression events from facial expressions.

3.1 Facial Expressions, Facial Expression Events, and Mental States

Facial expressions represent changes in neuromuscular activity that lead to visually detectable changes in facial appearance. Facial expressions convey rich information about mental states. Quantitatively characterizing facial expressions is the key step to achieve facial expression event identification and recognition. Bassili [11] observed and verified that facial motion dominates facial expressions. He showed that facial expressions can be identified by facial motion cues even without any facial texture information. This observation has been accepted and tested by most researchers of facial expression recognition [3], [2], [15], [26]. Therefore, facial motion is vital to characterize facial expressions. This has been the motivation for the approach used in this work: to detect facial expression events instead of facial expressions.

The next problem is how to represent facial motion. The low-level representation is the employment of a 2D motion field, for example an optical flow field or a displacement field of facial feature points. In reality, there are some problems with the employment of such a 2D motion field, whose lack of semantic makes it very difficult to manipulate. A commonly used way to improve this problem is to introduce an intermediate description to govern the 2D motion field. Several such intermediate descriptions have been suggested. Typically, the motion of facial muscles from optical flow was used in [15], [26], muscle-based representation of facial motion by using a detailed physical model of the skin and muscles was employed in [3], local parameterized models of facial motion was tried in [2], and FACS (Facial Action Coding System) based representation was used in [5],[12]. In this paper, unlike in the mentioned facial expression recognition approaches where several typical facial expressions, such as happiness and sadness were targeted, we focus on the recognition of one facial expression event, the yawn, through which we will show that action units defined in the FACS could be suitable for this purpose.

Ekman and Friesen introduced the Facial Action Coding System [7], FACS, to describe all visually distinguishable facial movements. In FACS, action units are defined to account for changes in facial expressions relative to a neutral face. The combination of these action units results in a large set of possible facial expressions. For example, a happiness expression can be synthesized by a linear combination of pulling lip corners (AU12+13) and mouth opening (AU25+27) with upper lip raiser (AU10). FACS has been successfully used to characterize facial expressions, especially in the area of MBC. It can further be observed from current standardization activities that MPEG-4 has developed a facial animation parameter (FAP) system, a low-level FACS, to describe human faces. Now an interesting problem is if it is possible to use FACS to characterize dynamic facial expression events.

According to our common experience, most facial expressions can be classified properly by human beings from static pictures of faces. This observation has been successfully utilized by Ekman and Friesen to invent their FACS. A reasonable interpretation of how human emotion can be guessed from static images is that a neutral face is always implicitly defaulted in your mind when you watch a picture containing an expressed face. The difference between the expressed face and the neutral face in fact tells the dynamic information which is used implicitly by humans for emotion recognition. Therefore, the real problem of how to handle facial expression events is that the events are dynamic and time-varying. Since a facial expression event consists of sequential facial expressions and individual facial expressions can be specified by action units, the key to characterizing facial expression events is to exploit a temporal combination of action units specifying individual facial expressions. The analysis of facial expression events becomes a problem of how to identify such temporal rules which govern facial expression variation behind expression events. The temporal behavior of expression events can be extracted based on the observation that the measured action units at each frame look apparently random. However, they are fully controlled by invisible, internal states. Therefore, it is natural to employ Hidden Markov Models (HMM) to model and specify facial expression events.

Action units could be chosen as observations for the underlying Hidden Markov Models. Through the HMM framework, action units are probabilistically coupled to facial expression events, which is very suitable for the real applications. The real facial motion is almost never completely localized and detecting a unique set of action units for a specific facial expression is not guaranteed [3]. A similar strategy to employ HMMs has been successfully used in speech recognition [21], head gesture recognition [17], and sign language recognition [22].

In summary, we suggest a hierarchical way to handle human mental state estimation. This is shown in Fig. 3.

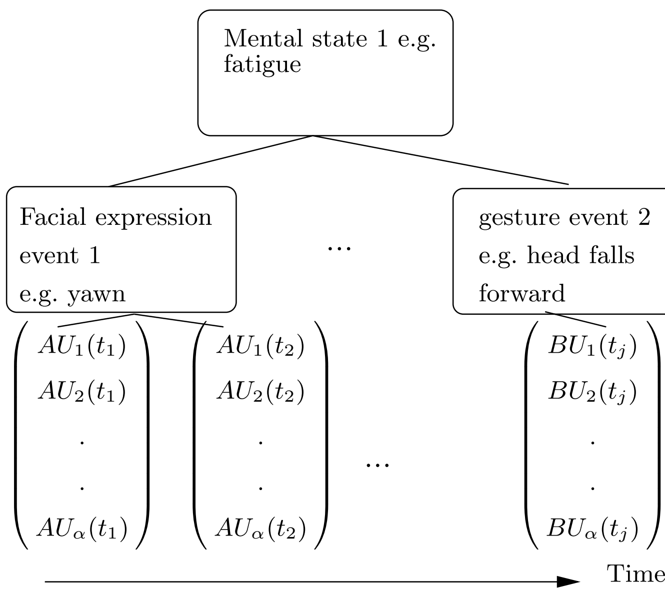

Figure 3: The hierarchical framework for human mental state estimation. The action unit vector $[AU1(tj)AU2(tj)AUa(tj)]^T$ specifies the facial expression at time t_j The position of the body is specified in a similar manner by the body action unit vector $[BU1(tj)BU2(tj)BUb(tj)]T$. Based on the time sequence of action unit vectors, facial expression events, e.g. yawns and smiles, can be detected. Similarly, gesture events can be detected from body action units. In the final step, the mental state can be estimated from the facial expression events and the gesture events.

The hierarchy consists of the layers

1. mental state, e.g. fatigue, happiness
2. facial expression event, e.g. yawn, laugh and gesture event, e.g. head falls forward.
3. facial expression and pose, a facial expression is a time sample of a facial expression event and a pose is a time sample of a gesture event. Facial expression and pose are completely described by the AU and *BU* vector, respectively.

3.2 Action unit extraction

To extract the action units is related to the extraction of 3D motion of head object. Estimating 3D motion through two consecutive frames has a long history in computer vision community. Two major representative solution are the ”exact methods” and the ”approximation method” [24]. Our discussion follows the ”approximation method” as used in MBC [12].

Let $\mathbf{s} = (x,y,z)^T$ denote a 3D vertex point of a generic wireframe model, and $\mathbf{s}' = (x',y',z')^T$ denote the corresponding point on the target face. For the nonrigid facial motion, the change from $\mathbf{s}$ to $\mathbf{s}'$ could be written as [12]:

$$\mathbf{s}' = \mathbf{Rs} + \mathbf{T} + \mathbf{Ds} \tag{1}$$

where $\mathbf{R}$ is called rotation matrix, $\mathbf{T}$ is the translation matrix, $\mathbf{D}$ is a deformation matrix that describes non-rigid motions caused by facial expressions. It is known that any facial expression can be analyzed into a weighted linear combination of a set of typical Action Units [7], thus the $\mathbf{Ds}$ term in (1) can be written as

$$\mathbf{Ds} = \mathbf{E}\blacksquare \tag{2}$$

where $\blacksquare$ is the vector of m facial expression parameters. The $3 \times m$ matrix $\mathbf{E}$, built from Action Units, determines how a certain point $\mathbf{s}$ is affected by $\blacksquare$.

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ e_{31} & e_{32} & \cdots & e_{3m} \end{bmatrix} \tag{3}$$

Let $\mathbf{p} = [\phi_1, \phi_2, \ldots, \phi_m, \Omega_x, \Omega_y, \Omega_z, T_x, T_y, T_z]^T$ denote the $n \times 1$motion parameter vector ($n = m+6$), where ϕ_i are expression vector, Ω_x, Ω_y and Ω_z are angular velocities about the x,y and z axis respectively, T_x, T_y, T_z are the three translation velocities along x,y and z axes. Suppose the geometry projection from the 3D space onto the 2D image plane is a perspective projection and that the focal length is 1. The optical flow field (u,v) induced by the 3D motion of the face is [12]:

$$\begin{aligned} & u = x' - x \\ & = \tfrac{1}{z}\textstyle\sum_{i=1}^{m}(e_{1i} - xe_{3i})\phi_i + xy\Omega_x - (1+x^2)\Omega_y + y\Omega_z - x\frac{T_z}{z} + \frac{T_x}{z} \\ & = \textstyle\sum_{i=1}^{i=n}(c_{ui}p_i) \end{aligned} \tag{4}$$

$$\begin{aligned} & v = y' - y \\ & = \tfrac{1}{z}\textstyle\sum_{i=1}^{m}(e_{2i} - xe_{3i})\phi_i + (1+y^2)\Omega_x - xy\Omega_y + x\Omega_z - y\frac{T_z}{z} + \frac{T_y}{z} \\ & = \textstyle\sum_{i=1}^{i=n}(c_{vi}p_i) \end{aligned} \tag{5}$$

Note here that the depth z is known if we have a known 3D model such as the Candide model [1]. The recovery of 3D motion from (u,v) is thus reduced into a linear problem. In [12], different solutions for solving the 3D motion parameters including action units are suggested for the small motion and large motion respectively. The recent AAM techniques [16],[1] also suggest themselves good candidate for recovering motion parameters.

4 Yawn Modelling

Yawn is a spontaneous behavior, which is a universal and culture-independent facial expression event. Typical yawns contain an opening of the mouth and a closing of the mouth. However, this does not mean that yawns can be identified if we can detect these two mouth actions. In fact, mouth opening and closing are also typical actions in a smile. The important difference between a smile and a yawn lies in the dynamic properties of each mouth action. Therefore, the key in yawn modelling is to choose a suitable framework that can recover the implicit and temporal rules governing a yawn facial expression event.

Hidden Markov Models (HMM) is used as the basic stochastic framework yawn modelling. An observation $\mathbf{O}$ of the facial expression is represented by a sequence of action units, which consists of the observation feature vectors $\mathbf{O} = (o_1, o_2 ... o_T)$, where o_t is the feature vector extracted at time instant t. It is assumed that the feature vectors follow continuous probability distributions which is modelled by a mixture of Gaussian distributions and the temporal changes during facial expression changing are stationary and follow a first order Markov process [21].

To specify a HMM for yawns, we have to study the physical mechanisms of yawns. As is known, facial expressions are generated by contraction of muscles. Contraction of muscles can be taken as the physical basis on which to employ HMM.

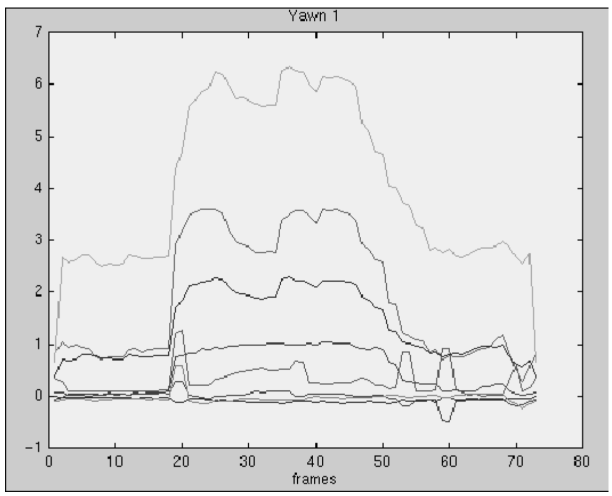

Figure 4: Action units measured from a yawn sequence.

We now show how to build a HMM for yawn. A complete muscle contraction can be divided into different stages: release, application, and relaxation. Examining a facial expression event, we see that they follow the same development principle: starting from a neutral expression, there is the beginning of the expression, the release period; then the expression gets complete and lasts approximately constant for some time, this state is called the application period; after that, the expression decreases gradually, the relaxation period; finally, the face comes back to the neutral position. This evolution process can be observed in Fig. 4.

To characterize a yawn event with Hidden Markov Models, we can specify an expression action into four physical states:

1. N: the Neutral state;
2. S: the Start (release) state;
3. A: the Application state;

4. E: the End (relaxation) state.

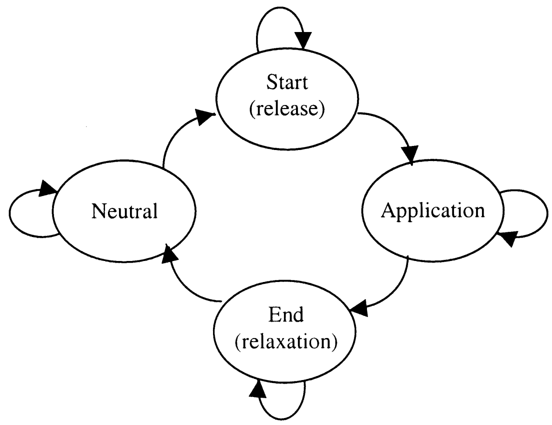

Figure 5: Hidden Markov Model of a facial expression event.

A Hidden Markov Model to describe yawns based on these four states is shown in Fig. 5. Facial expression events can be expected to have a sequence of states that will look like (NNNNSSSAAAAEEENNN). The HMM only allowed self-loops and sequential transitions from the current to the next state as represented in Fig. 5. The initial state probabilities are set to zero for all other states but to one for the first state. The remaining parameters are estimated from the extracted model parameters of the training set. The prior probabilities of the models P_i are assumed to be uniform for all models. The fatigue event recognition is performed based on the maximum posterior probability (MAP) criterion given by

$$\arg\max_i P(\Lambda|O) \tag{6}$$

where Λ represents yawn event and O the observation sequence. The posterior probability can be estimated from the likelihood and the prior probabilities using Bayes rule

$$P(\Lambda|O) = \frac{P(O|\Lambda)P(\Lambda)}{P(O)} \tag{7}$$

The prior probabilities $P(O)$ are equal for all action unit sequence during recognition. The Viterbi algorithm is used to calculate the most likely state sequence [21].

5 Real Time Implementation of The System

Our real time system at the vehicle side consists of three main function blocks: the head detection, action unit extraction, HMM estimation.

Head detection itself is a big research topic and is the key in our work. The haar like feature based face detection system [13] is adapted into our system, it work fine with a near front face, which is the case in the driver's face locating if the camera is correctly mounted in front of the driver in the car. It has been estimated that the face detection module count for more than 95 percent of the computation time of the whole system.

The task of the training is to find the HMM parameters state transition matrix , observation matrix , and initial states [21]. Usually, an initial estimation has to be done and then improved by the use of the Baum-Welch iterative algorithm. Since it's a real time system, the training process became very simple and fast.

Using the solution for the HMM problem 1 [21], we got the result probability, as it's larger than one threshold value, we signaled a yawn event. The speed of the system is mainly restricted by the face detection module, on a 400MHz PC the haar like feature based face detector could run as 5.2 frames per second on a 160x120 pixels size video, and runs at the same speed when added the HMM recognition module. On a 700 MHz PC the program runs at 15.2 frames per second. In reality, this training job could be done at the surveillance server side.

5.1 Network simulation

For the networked system, we built a demonstrating system which consist of a client side and server side using a CORBA solution [18]. CORBA is a proven technique for doing function calls over the network. Besides being a tool for distributing functions over a network, CORBA also offers a standardized way of declaring the input and output variables. It consists of a web client (a web-browser), a web server and a CORBA server as illustrated in the Fig. 6. The fatigue detection module is on a web page. The web page includes a uploading function for sending testing action unit sequence. The client specifies the input parameters and executes the method. A web page containing the results is then loaded. The web server communicates through a script (CGI, PHP etc) with the CORBA client (e.g. passing the parameters, converting the output results etc.) which then requests execution from the CORBA server (e.g. the actual module). The results are passed from the CORBA server to the CORBA client. The script prepares the web page, which is communicated, to the web client.

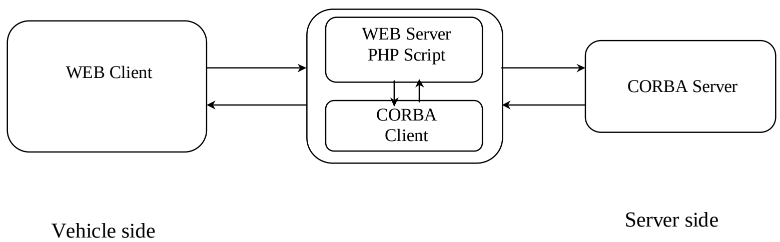

Figure 6: Simulating networked driver mental state surveillance system. The left side is the web client, which act like the vehicle side, the right side is the server side.

The web page for the demo looks like Fig. 7. The testing users could upload their testing action units sequence to the server side to see if it is a yawn sequence.

5.2 User tests and subjective evaluation

We test our system publicly, 26 students from the Applied Physics and Electronic Engineering Department of Ume University tried the system. They all have driving experience. Each attendant was requested to perform at least 10 yawns, the correctly recognized number was counted. The system was trained using the training data collect from out lab group member's yawn sequences. The system beeps as the calculated probability of the mouth motion is within a predefined range, the range was settled by our experience. We also designed questions and ask the attendant do a subjective evaluation. We thought this could be helpful and interesting for those who implement such kind of system. The major questions include:

1. Do you think yawn detection really could be used to indicate driver's fatigue?
2. If a alertometer based on yawn detection could be successfully designed instead of a audio feedback as in demo, do you accept to add such device into your car?
3. Do you think you will follow such kind of system's suggestion or not (you would like to trust yourself more)?

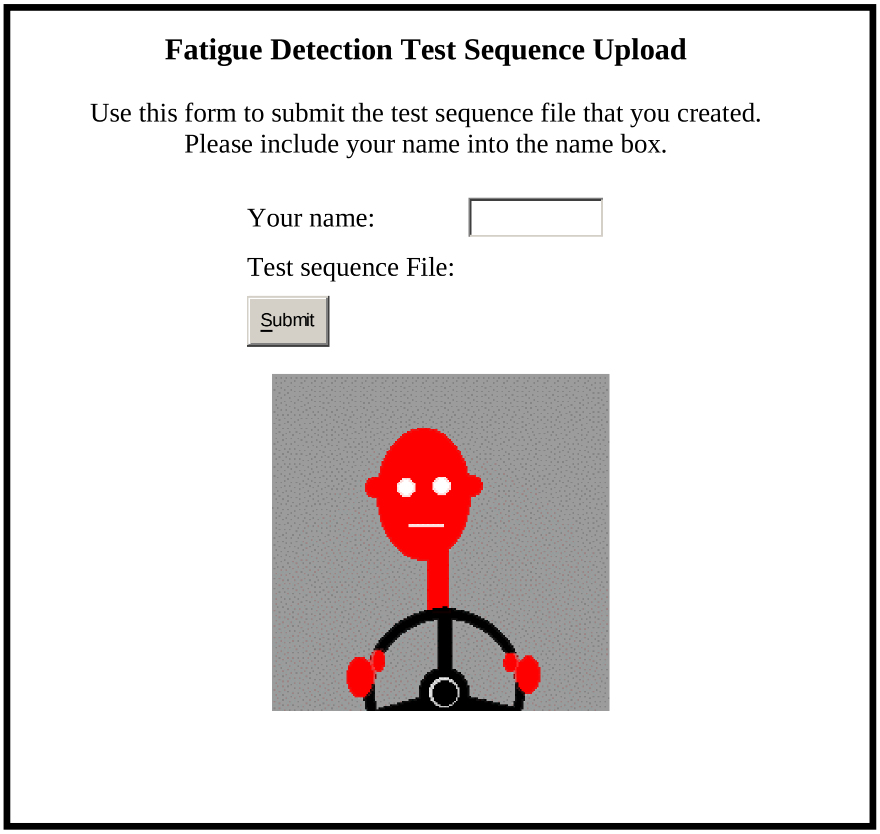

Figure 7: A demo CORBA system web page, the user could upload testing sequence to server for testing.

Among them, question 1 is a survey of whether the technology should be put into use of fatigue detection, question 2 is about feedback and user acceptance, question 3 is about user's belief about the functionality of such automatic system. For question 2, we explain to the attendants that although our demo is a audio feedback, but we intend to design the feedback as a alterometer which based on the yawns number and frequency, and we explain that we want the system to be just a private security device and they only need to check it when they want to do so, in this way, the system is highly non-intrusive. Table I shows the results of user test and subject evaluation.

The tests show promising results of the system performance with detecting a yawn like mouth motion instead a true yawn, since the users are asked to fake a yawn when they are not tired.

From the answers to the designed question, we could conclude that most people believe the yawn detection could be utilized to detect or predict driver's fatigue (question 1). The answers to question 2 show the wide acceptance of such device in car, this is interesting according to [19], which found for every 3 drivers who were

Table I: RESULTS OF 26 USER TEST AND SUBJECT EVALUATION

Total yawns performed	270
Correctly recognized	200
Question 1	Yes: 23, No: 1, Not sure: 2
Question 2	Yes: 25, Not sure: 1
Question 3	Yes: 3, No: 6, Both: 17

strongly in favor of onboard safety monitoring including alertness monitoring, four were completely opposed to it. This difference is maybe due to [19] investigate mainly bus and track drivers, who are more subject to privacy invading from companies. Although we found that pure private usage of such device is appealing to car drivers, it's really hard to say how to prevent it from ”wrong” usage from authority like police or insurance companies. The answers to question 3 is expected, most people would like to trust the system only when they really feel tired.

6 Conclusion

This paper addresses the problem of driver's fatigue detection using a networked driver's mental state surveillance system. The focus is on how to recognize the yawn facial expression event, which is highly associated with fatigue. In this paper the Facial Action Coding System is suggested to be used to represent facial expression variations. We have demonstrated that probabilistic coupling of mouth actions with Hidden Markov Models is a promising way to handle dynamic facial expressions. The hierarchical framework developed in this paper can also be extended to handle the general human mental state estimation problem. User tests and subject evaluation show the technology is appealing for car users, and also give hints for further development of such kind of system.

Acknowledgment

This research described in this paper is supporte by the European project ”Interface”.

References

[1] J. Ahlberg. Model-based coding - extraction, coding, and evaluation of face model parameters, dissertation no. 761, 2002.

[2] M. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion, 1995.

[3] Michael J. Black and Yaser Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *ICCV*, pages 374–381, 1995.

[4] L. Cheboub. Computer recognition of fatigue by using hidden markov models, master's thesis, linkoping university, 1998.

[5] Chang Seok Choi, Hiroshi Harashima, and Tsuyosi Takebe. Analysis and synthesis of facial expressions in knowledge-based coding of facial image sequences. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 2737–2740, 1991.

[6] D.F. Dinges, Mallis, and M.M. Managing fatigue by drowsiness detection: Can technological promises be realised? In *Proc. of the 3rd International Conference on Fatigue and Transportation, Fremantle, Western Australia*, 1998.

[7] P. Ekman and W. Friesen. *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1978.

[8] T. Horerry and etl. L.Hartley. Fatigue detection technologies for drivers: A review of existing operator-centred system. In *IEEE Conference on Human Interfaces in Control Rooms*, 2001.

[9] T. Ishii, M. Hirose, and H. Iwata. Automatic recognition of driver's facial expression by image analysis. In *Journal of JSAE*, volume 41, pages 1398–1403, 1987.

[10] Q. Ji, Z. Zhu, and P. Lan. Real-time nonintrusive monitoring and predection of driver fatigue. 53(4):1052–1068, 2004.

[11] J.N.Bassili. Facial motion in the perception of faces and emotional expression. *Inter. Journal of Experimental Psychology*, 4:373–379, 1978.

[12] Haibo Li, Pertti Roivainen, and Robert Forchheimer. 3-d motion estimation in model-based facial image coding. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 15(6):545–555, June 1993.

[13] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *Proceedings of International Conference on Image Processing*, 2002.

[14] N. Mabbott, M. Lydon, L. Hartley, and P. Arnold. Procedures and devices to monitor operator alertness whilst operating machinery in open-cut coal mines. stage 1: State-of-art review, arrb transport res. rep. rc 7433, 1999.

[15] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions, Special Issue on Computer Vision and Its Applications*, 74(10):3474–3483, 1991.

[16] Iain Matthews and Simon Baker. Active appearance models revisited. Technical Report CMU-RI-TR-03-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, April 2003.

[17] C. Morimoto, Y. Yacoob, and L. Davis. Recognition of head gestures using hidden markov models international conference on pattern recognition, 1996.

[18] Robert Orfali, Dan Harkey, and Jeri Edwards. *Instant CORBA*. John Wiley and Sons, 1997.

[19] Penn and Schoen. User acceptance of commercial vehicle operations services, task b, critical issues relating to acceptance by interstate truck and bus drivers, final report, contact no.dtfh61-94-r-00182.

[20] Alex Pentland and Andrew Liu. Modeling and prediction of human behavior, mit media lab perceptual computing technical report no.433, 1997.

[21] L. Rabiner and B. Juang. An introduction to hidden markov models, 1996.

[22] T. Starner. Visual recognition of american sign language using hidden markov models, master's thesis, mit media laboratory, 1995.

[23] STR. Preparing for your driving license, 1995.

[24] R. Tsai and T. Huang. Estimating three-dimensional motion parameters of a rigid planar patch i. *IEEE Transactions of Acoustics, Speech, Signal Processing,*, ASSP-29:1147–1152, 1981.

[25] Thomas Wiegand, Gary J. Sullivan, and Ajay Luthra. Overview of the h.264/avc video coding standard. *IEEE Trans. circuits and system for video technology*, 13(7):560–576, 2003.

[26] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1994.