DEPARTMENT OF
APPLIED PHYSICS AND ELECTRONICS
UMEÅ UNIVERISTY, SWEDEN

DIGITAL MEDIA LAB

A video based real time fatigue detection system

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-2004:3
ISSN Number: 1652-8441

Report Date: September 20, 2004

[1]

[2]Supported by ICA banken.

Abstract

Fatigue could be detected or predicted by the dynamic facial expression event: yawn. Facial expression parameters are extracted from real time video containing faces and used as observations for the underlying Hidden Markov Model for recognizing the yawn event. Promising results from tests by many users and their evaluations are reported.

Keywords

Fatigue detection, yawn, Hidden Markov Model, real time, facial expression event.

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 [7], [1]. 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 [10]. Obviously, for traffic and driving safety, it is very important to detect and recognize the driver in fatigue and to give him/her a warning, "you should stop for coffee".

The importance of detecting fatigue has been realized for years. Many approaches have been proposed to handle the fatigue detection problem. A recent review [14] both gives a overview of current fatigue detection technologies and raises good thoughts about how to implement such system in right way. According to [17] 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 techniques, and the task is treated as a pattern recognition problem. Fatigue detection will be done through observing the driver's face and examining facial expressions, so our technology could be classified into group 4) mentioned before. 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. 1.

Figure 1: 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 [10]:

- 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 E then fatigue can be represented as Fatigue = $E_1, E_2, , E_n$, 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 use 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 [16], head nodding [14]. We choose to detect yawn instead as an evidence of fatigue, there are several advantages of doing so:

1. Yawn happens in earlier stage compare to eyelid dropping or head nodding, so an earlier warning is possible and will be safer.
2. 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.
3. Yawn event is relative a longer facial movement and therefore easier to be recognized if modelled carefully.
4. Methods based on eyelid dropping usually need to solve the driver’s eyeglasses problem, which is usually not the case for yawn detection.
5. 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.

2 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 [**?**] 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 [4], [11], [12], [13]. 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 [12], [13], muscle-based representation of facial motion by using a detailed physical model of the skin and muscles was employed in [4], local parameterized models of facial motion was tried in [11], and FACS (Facial Action Coding System) based representation was used in [2],[5]. 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 [3], 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 model-based image coding. 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 where real facial motion is almost never completely localized and detecting a unique set of action units for a specific facial expression is not guaranteed [4]. A similar strategy to employ HMMs has been successfully used in speech recognition [8], head gesture recognition [6], and sign language recognition [9].

In summary, we suggest a hierarchical way to handle human mental state estimation. This is shown in Fig. 2.

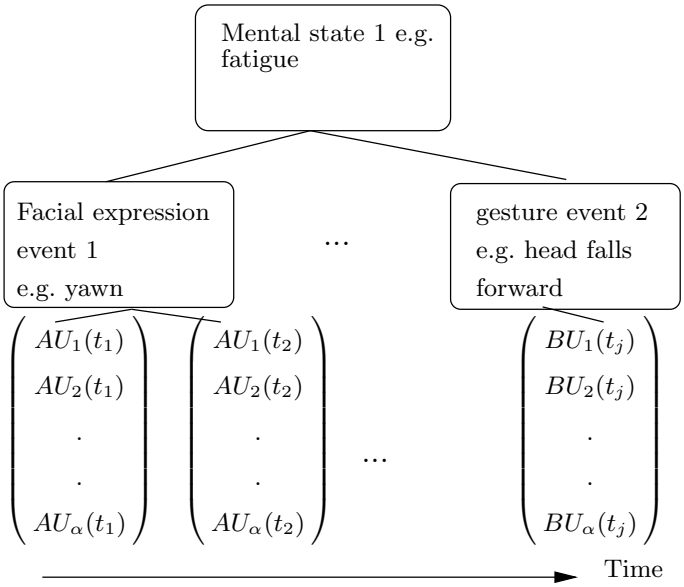

Figure 2: The hierarchical framework for human mental state estimation. The action unit vector $[AU1(tj)AU2(tj)AUa(tj)]^T$ specifies the facial expression at time t_j The position of the body is specified in a similar manner by the body action unit vector $[BU1(tj)BU2(tj)BUb(tj)]T$. Based on the time 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 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. In this paper, Hidden Markov Models (HMM) are used for this purpose.

It could be nice if we use action units as the observations, but currently the calculation of action units is expensive and sensitive to noise. So in our demo implementation, instead of using action units, we want to find other observation that could describe the yawn motion, and also it is easier to track and stable. After many tests we found that simply using the ”dark region” (Fig. 3) within mouth region as observation is robust and fast enough for real time purpose. The reason lies in that for most people when they stifle a yawn, typical motion includes head raising, mouth opening and tone moving back, this often make a ”dark hole” inside mouth which is often a quite stable feature, even with a front light source, this dark area is often observable and easy to be tracked. We further use the rectangular area outside the ”dark hole” to approximate the real dark area. Our labs have shown better tracking quality with this observation compared to using the mouth color block, mouth edge, hight and width ratio, etc. instead.

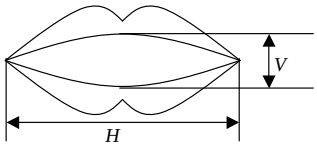

Figure 3: Observations $O = V \times H$ is a one dimensional signal

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 HMMs, see [8] for more about HMM.

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.

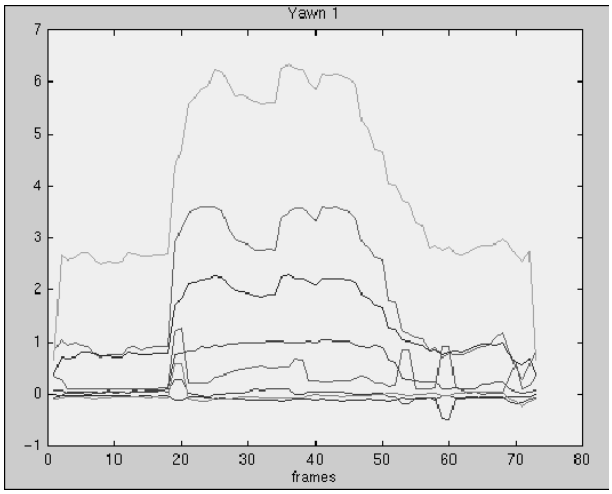

Figure 4: Action units measured from a yawn sequence.

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.

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). In the following subsections, we show how to train a HMM and how to employ it for yawn detection.

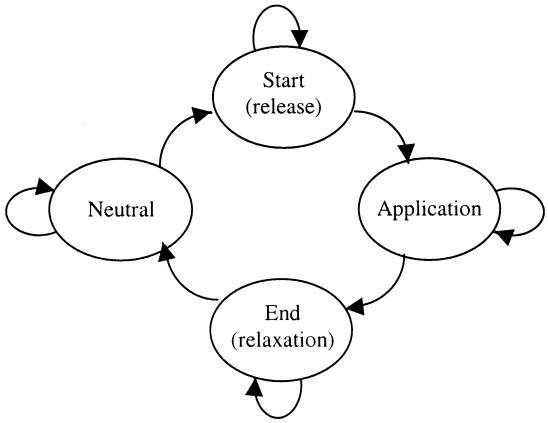

Figure 5: Hidden Markov Model of a facial expression event.

4 Real Time Implementation of The System

Our real time system consists of three main function blocks: the head detection, dark area within mouth region locating, HMM estimation.

4.1 Head detection

Head detection itself is a big research topic and is the key in our work. All subsequent estimation is based on the collect finding of the ”open mouth” region, which is defined by us as the dark area with mouth region, the haar like feature based face detection system [15] 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.

4.2 Mouth Motion Extraction

After tracking the head, the estimated mouth area is also found, we then look for ”dark area” to locate the open mouth area. A predefined darkness threshold is used to threshold the mouth region to isolate the open mouth region. The area should be in the shape of an ellipse, but due to tooth, tone, light condition, the located block area could be quite random in shape, one typical example is shown in Fig. 6. The white area within the white square

is the located open mouth area. We use the white square area instead of white pixels inside it to represent the open mouth region, this tend to smooth the observation signal. A simple color blob tracking algorithm is used to track this dark area, so that when the face detection module loses track of the face, the color blob tracking module could still track the open mouth region. To cope with the scale problem, we further use the ratio of the white square region with respect to the estimated mouth region (the dark gray square outside the white square region in Fig. 6) as the input observation for HMM.

Figure 6: A typical found open mouth area (the white area within the white square).

The output from the mouth motion extraction is a 1D real value data. For the purpose of recognition, a quantization process of the computed real value was made. After experiments, we quantized the signal with only five levels for simplicity. The quantized value form the observations which are used as input to the HMM.

In the program, we used a buffered list to save the observations, a mouth open and close event trigged the estimation procedure, it then retrieve buffered observations, and use it as input for a HMM estimator procedure.

4.3 HMM Estimation

Yawn Model Training

The task of the training is to find the HMM parameters state transition matrix , observation matrix , and initial states [8]. 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. In the system implementation, the HMM structure is a 1D HMM look like in Fig. 7. The states s1, s2 s5 are correspond to the neutral, start, application, end, neutral states mentioned before.

Yawn Detection

Using the solution for the HMM problem 1 [8], we got the result probability, as it's larger than one threshold value, we signaled a yawn event. It has been found that due to different person's yawn habit, the result probability is within certain range when use a trained HMM model, so it's wise to signal the yawn event for a range of probability. Since the yawn detection in our system is depended on the "dark area" detection, the threshold of darkness should be adjust according to the light condition. As mentioned before, 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

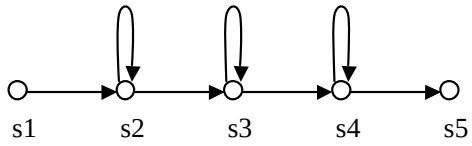

Figure 7: The 1D HMM structure used in our system.

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

module. On a 700 MHz PC the program runs at 15.2 frames per second. (The demo program is attached with the paper)

5 Testing and Subject 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 beeps as the calculated probability of the mouth motion is within a predefined range, the range was settled by our experience. Inspired by [14], we also designed questions and ask the attendant do a subjective evaluation. We thought this could be helpful and interesting to people working with 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 on the dashboard?
3. Do you think you will follow such kind of system's suggestion or not (you would like to trust yourself more)?

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 many testing users said it's not easy 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 [16], which found for every 3 drivers who were strongly in favor of onboard safety monitoring including alertness monitoring, four were completely opposed to it. This difference is maybe due to [16] 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. 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.

List of Figures

List of Tables

References

[1] L. Cheboub,“Computer Recognition of Fatigue by using Hidden Markov Models”. Master’s Thesis, Linköping University, 1998.

[2] C. Choi, H. Harashima and T. Takebe, “Analysis and Synthesis of Facial Expressions in Knowledge-Based Coding of Facial Image Sequences”, in Proc. ICASSP’91, May 1991.

[3] P. Ekman and W. Friesen, Facial Action Coding System, Palo Alto, Calif.: Consulting Psychologists Press, Inc., 1978.

[4] I. Essa and A. Pentland, “Facial Expression Recognition using a Dynamic Model and Motion Energy”, In Proc. International Conference on Computer Vision, Boston, MA, June 1995.

[5] Haibo Li, P. Roivainen and R. Forchheimer, “3D Motion Estimation in Model-Based Facial Image Coding”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No.6, June, 1993.

[6] C. Morimoto, et al, “Recognition of Head Gestures using Hidden Markov Model”, Proceeding of the Second International Conference on Face and Gesture Recognition, 1996.

[7] A. Pentland and A. Liu, “Modeling and Prediction of Human Behavior”, MIT Media Lab Perceptual Computing Technical Report No.433, 1997.

[8] L. Rabiner and B. Juang, “An Introduction to Hidden Markov Models”. IEEE ASSP Magazine, p.4-16, Jan. 1996

[9] T. Starner, Visual recognition of American Sign Language using Hidden Markov Models, Master’s thesis, MIT Media Laboratory, Feb. 1995.

[10] STR, Preparing for Your Driving License, 1995.

[11] M. Black and Y. Yacoob, “Recognizing facial expressions in image sequences using local parameterized models of image motion”, Inter. Journal of Computer Vision 25(1), 23-48, 1997.

[12] K. Mase, “Recognition of facial expressions for optical ”, IEICE Transactions, Special Issue on Computer Vision and Its Applications, E 74(10), 1991.

[13] Y. Yacoob and L. Davis, “Computing Spatio-Temporal Representations of Human Faces”. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1994.

[14] T. Horerry, L.Hartley, etl. ”Fatigue Detection Technologies for Drivers: A Review of Existing Operator-Centred System”, IEEE Conference on Human Interfaces in Control Rooms, 2001.

[15] Rainer Lienhart and Jochen Maydt, ”An Extended Set of Haar-like Features for Rapid Object Detection”. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.

[16] 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.

[17] Dinges, D.F. and Mallis, 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.