Pattern Recognition Letters 31 (2010) 234–243

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Gabor-based dynamic representation for human fatigue monitoring in facial image sequences

Xiao Fan [a,b,*], Yanfeng Sun [a], Baocai Yin [a], Xiuming Guo [c]

[a] Beijing Key Laboratory of Multimedia and Intelligent Software, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China
[b] Cogent Beijing R&D Center, Beijing 100089, China
[c] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China

ARTICLE INFO

Article history:
Available online 3 September 2009

Communicated by N. Pears

Keywords:
Human fatigue
Multi-scale
Dynamic feature
Feature fusion
AdaBoost algorithm

ABSTRACT

Human fatigue is an important reason for many traffic accidents. To improve traffic safety, this paper proposes a novel Gabor-based dynamic representation for dynamics in facial image sequences to monitor human fatigue. Considering the multi-scale character of different facial behaviors, Gabor wavelets are employed to extract multi-scale and multi-orientation features for each image. Then features of the same scale are fused into a single feature according to two fusion rules to extract the local orientation information. To account for the temporal aspect of human fatigue, the fused image sequence is divided into dynamic units, and a histogram of each dynamic unit is computed and combined as dynamic features. Finally, AdaBoost algorithm is exploited to select the most discriminative features and construct a strong classifier to monitor fatigue. The proposed method was tested on a wide range of human subjects of different genders, poses and illuminations under real-life fatigue conditions. Experimental results show the validity of the proposed method, and an encouraging average correct rate is achieved.

1. Introduction

According to the estimated data from the National Highway Traffic Safety Administration (NHTSA, 2005), 100,000 police-reported crashes are directly caused by driver fatigue each year, which result in an estimated 1550 deaths, 71,000 injuries, and $12.5 billion losses. In China, driver fatigue resulted in 3056 deaths in vehicular accidents in 2004, and caused 925 deaths in highway accidents that amounted to about 14.8%. Human fatigue has become an important factor for many traffic accidents. Therefore, it is essential to develop novel methods for monitoring human fatigue in order to improve transportation safety.

1.1. Previous work

In fact, there have been many attempts to achieve reliable fatigue monitoring for reducing the number of automobile accidents due to human fatigue in the last decade. These methods can be divided into three major categories as follows.

1.1.1. Driver physiological parameters

This method focuses on measuring physiological changes of drivers. It can accurately, validly, and objectively to determine fatigue and sleep of the drivers. A significant effort has been made to measure them in laboratory. The popular physiological parameters include electroencephalogram (EEG) (Abdul-Latif et al., 2004; Parikh and Micheli-Tzanakou, 2004; Wu et al., 2004a,b; Lin et al., 2005a,b), electrocardiogram (ECG) (Hayashi et al., 2005), EOG (Galley and Schleicher, 2004), and electromyography (EMG) Bonato et al., 2001. EEG is found to be useful in determining the presence of ongoing brain activity, and its measures have been used as the reference point for calibrating other measures of sleep and fatigue. Abdul-latif et al. (2004) found the mean RMS of EEG bands were increased during fatigue compared to the RMS value in the case of relaxation before fatigue, and the RMS value was seen to be greatest in the beta band and lowest in the gamma band. In literature (Parikh and Micheli-Tzanakou, 2004), Alpha waves (8–13 Hz) are observed with increasing amplitude when fatigue. Wu et al. (2004a,b) describes a system that combines EEG power-spectrum estimation, principal component analysis (PCA), and fuzzy neural network model to estimate/predict drivers' drowsiness level in a driving simulator. Lin et al. (2005a) proposed a system that combines EEG power spectra estimation, independent component analysis (ICA) and fuzzy neural network models to estimate

* Corresponding author. Address: Beijing Key Laboratory of Multimedia and Intelligent Software, College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China. Fax: +86 10 67396568.
E-mail address: wolonghongni@yahoo.cn (X. Fan).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.08.014

drivers' cognitive state in a dynamic virtual-reality-based driving environment. Lin et al. (2005b) developed a drowsiness-estimation system based on EEG by combining ICA, power-spectrum analysis, correlation evaluations, and a linear regression model to estimate a driver's cognitive state when he/she drives a car in a virtual-reality-based dynamic simulator. Unfortunately, most of these physiological parameters are obtained intrusively, making them unacceptable in practical applications.

1.1.2. Vehicle based performance

Fatigue can also be characterized by the behaviors of the vehicle that a driver operates. The vehicle based performance methods detect the behaviors of the drivers by monitoring the transportation hardware systems under the control of the drivers, such as steering wheel movements (Takei and Furukawa, 2005), driver's grip force on the steering wheel (Thum et al., 2003), speed, acceleration, lateral position, turning angle, changing course, braking and gear changing, etc. Thum et al. (2003) described an automobile driver fatigue detection method by monitoring the driver's grip force on the steering wheel, based on the variation in steering grip force due to fatigue or loosing alertness. In Takei and Furukawa (2005), the chaos theory was applied to explain the changes of steering wheel motion. If there is chaos in the motion, a strange trajectory called attractor can be found by applying the Takens' theory of embedding. The chaos characteristics are used to estimate a driver's fatigue. While these methods may be implemented non-intrusively, they are subject to several limitations, including the vehicle type, driver experiences, and driving conditions.

1.1.3. Driver physical conditions

These methods focus on detecting driver's physical changes during drowsiness by image-processing techniques. People in fatigue exhibit certain visual behaviors that are easily observable from changes in facial features. Visual behaviors that typically reflect a person's fatigue level include slow eyelid movement, smaller degree of eye openness (or even closed), frequent nodding, yawning, gaze (narrowness in the line of sight), sluggish in facial expression, and sagging posture. These image-processing based methods use optical sensors or video cameras to get visual fatigue cues.

Many efforts have been reported in the literatures on developing image-processing fatigue monitoring systems. When fatigue, the frequency and time of eye closed would increase. Much attention is paid to eye's features for fatigue detection. In 1998, based on the data of the Federal Highway Administration (Dinges et al., 1998), percentage of eyelid closure (PERCLOS) (Dinges and Grace, 1998) was taken as the most reliable and valid measure of a person's alertness level among several drowsiness detection measures. Liu et al. (2002) incorporated Kalman filtering and mean shift to track eyes, extracted eye's motion information as driver features. Hamada et al. (2003) extracted the driver's stage of drowsiness by means of the blink measurement with motion picture processing. Wang et al. (2003) used Gabor wavelets to extract texture features of drivers' eyes, and used neural network classifier to identify drivers' fatigue behavior. The doze stage was judged when the area of the iris becomes below a threshold (Miyakawa et al., 2004). Dong and Wu (2005) decided whether the driver was fatigue by detecting the distance of eyelids. Wang and Qin (2005) combined gray scale projection, edge detection with Prewitt operator and complexity function to judge whether the driver had his eyes closed. Fan et al. (2008) extracted LBP features of eye areas and used AdaBoost algorithm to determine whether a driver was fatigue.

When fatigue, people often yawn. Mouth features are extracted to detect fatigue (Wang et al., 2004; Wang and Shi, 2005). Wang et al. (2004) took the mouth region's geometric features to make up an eigenvector as the input of a BP ANN, and they acquired the BP ANN output of three different mouth states that represent normal, yawning or talking state, respectively. Wang and Shi (2005) represented the openness of the mouth by the ratio of mouth height to width, and detected yawning if the ratio was above 0.5 in more than 20 frames. Lu and Wang (2007) used directional integral projection to locate the midpoint of nostrils, recognized yawn by calculating the vertical distance between the midpoint of nostrils and the chin. To acquire accurately or reliably fatigue monitoring with the change in time, environment, or different persons, systems that can extract multiple visual cues which typically characterize the alertness level of a person and systematically combine them have been introduced (Bergasa et al., 2006; Zhu and Ji, 2004). Study shows that the performance of methods based on driver physical conditions is comparable with those methods using physiological signals. The major benefits of the visual measures using computer vision technologies are that they can be acquired non-intrusively.

Among those different methods, the best detection accuracy is achieved with methods that measure physiological parameters. Requiring physical contact with drivers (e.g., attaching electrodes), the methods based on driver physiological parameters are intrusive, causing annoyance to drivers. Good results have also been reported with methods that monitor driver physical conditions. These methods are non-intrusive and become more and more practical and popular with the rapid development of camera and computer vision technology. Most of these methods are spatial approaches. The visual features obtained from a single face image are used for classification. Although spatial approaches can achieve good recognition in some cases, they do not model the dynamics of fatigue and therefore do not utilize all information available in facial image sequences.

In facial expression recognition, according to psychologists (Bassili, 1979), an analysis for an image sequence produces more accurate and robust facial expression recognition. The facial motion is fundamental to the facial expression recognition. Therefore, more attention (Zhao and Pietikainen, 2007; Yang et al., 2007; Tong et al., 2007) has been shifted particularly towards modeling of dynamic facial expressions.

Human fatigue is a cognitive status that is developed over time. Dynamic features which capture the temporal pattern should be the optimal features to describe fatigue. To account for the temporal aspect of human fatigue, Ji et al. (2006) introduced a probabilistic framework based on dynamic bayesian networks (DBN) for modeling and inferring human fatigue by integrating information from various sensory data and certain relevant contextual information. States of nodes in a DBN satisfy the Markovian condition that is, the state at time t depends only on its immediate past. The dynamic fatigue model integrates the fatigue evidences spatially and temporally, therefore, leading to a more robust and accurate fatigue modeling and inference. But, in nature, no dynamic features are extracted in the system. In summary, there is limited research in extracting dynamic features from image sequences for fatigue monitoring. High accuracy in fatigue monitoring is still a challenge due to the complexity and variety of facial dynamics.

1.2. Overview of the proposed method

In this paper, the attention is focused to the dynamics of fatigue and using the spatial and temporal information from continuous image sequences to monitor the fatigue. To account for the temporal character of human fatigue and the multi-scale character of different facial behaviors, a novel Gabor-based dynamic representation with feature level fusion is proposed to monitor human fatigue from image sequences.

Fig. 1 gives an overview of the architecture of the proposed method. The system can be divided into training process and test

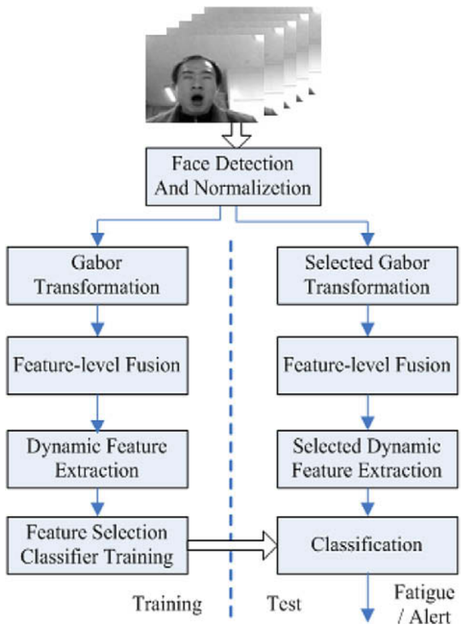

Fig. 1. Architecture of the proposed method.

process. In the training process, after each image in a facial image sequence is preprocessed by face detection, geometric normalization and cropping, Gabor wavelets are first used to extract multi-scale and multi-orientation features from each image in the sequence because of the multi-scale character of different facial behaviors. Then, to extract the local orientation information and reduce the dimension of the features, multi-orientation features of the same scale are fused according to the proposed fusion rules to produce a single feature. To get the dynamic features of human fatigue, the fused image sequence is divided into rectangle sub-image sequences as dynamic units, and a histogram of each dynamic unit is computed and combined as dynamic features. Considering that these features are redundant for classification, weak classifiers are constructed on the dynamic features and AdaBoost algorithm is applied to select a subset of the most discriminative dynamic features and build a strong classifier for fatigue monitoring. In the test process, only the selected Gabor wavelets (corresponding to the selected dynamic features) are used for the Gabor feature extraction. Accordingly, only the selected dynamic features are extracted. Finally, the trained AdaBoost classifier is used to determine whether the subject in a test facial image sequence is fatigue.

The rest of this paper is organized as follows. Section 2 introduces Gabor transformation and the Gabor-based representation of image sequences. Feature-level fusion and multi-scale dynamic feature extraction are showed in Section 3. In Section 4, feature selection and classifier learning are described. Experimental and analytic results are presented in Section 5, and finally, conclusions are drawn in Section 6.

2. Gabor-based representation for a facial image sequence

When a subject is fatigue, features of different facial behaviors have different scales. For example, behaviors of yawning are motion of a big area, so they can be analyzed in a big scale. Glassy-eyed facial expression is tiny change, thus they must be analyzed in a small scale. Therefore, multi-scale methods should be used to analyze human fatigue. Gabor wavelet is a powerful tool of multi-scale analysis and two dimensional Gabor wavelets can decompose an image in several directions at different scale. Hence, Gabor wavelets are exploited to decompose the input facial image sequence and get the multi-scale representation of the sequence for human fatigue monitoring.

2.1. Gabor wavelets

Gabor wavelets (Wiskott et al., 1997), whose kernels are similar to the 2-D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains (Daugman, 1980). Due to their biological relevance and computational properties, they have been successfully used in a variety of image-processing applications such as face recognition (Zhang et al., 2005; Liu and Wechsler, 2002). The two dimensional Gabor wavelets can be defined as:

$$\varphi_j(x) = \frac{\|\boldsymbol{k}_j\|^2}{\sigma^2}\exp\left(-\frac{\|\boldsymbol{k}_j\|^2\|\boldsymbol{x}^2\|}{2\sigma^2}\right)\left[\exp(i\boldsymbol{k}_j\cdot\boldsymbol{x}) - \exp\left(-\frac{\sigma^2}{2}\right)\right]$$

where, $\boldsymbol{k}_j == \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v\cos\phi_u \\ k_v\sin\phi_u \end{pmatrix}$, $k_v = 2^{-\frac{v+2}{2}}\pi$, $\phi_u = u\frac{\pi}{K}$, v determines the frequencies (scale) of the Gabor wavelets, u determines the orientations of the Gabor wavelets. $\|\cdot\|$ denotes the norm operator. Gabor wavelets with a certain orientation respond to the edges and bars in this orientation, and Gabor wavelets with a certain frequency extract the corresponding frequency (scale) information.

The Gabor transformation of a face image can be obtained by convolving the face image with a family of Gabor wavelets. For an image with the intensity distribution of $L(\boldsymbol{x})$, at a given point $\boldsymbol{x} = (x, y)$, the Gabor representation of the image can be defined as:

$$O_j(\boldsymbol{x}) = L(\boldsymbol{x}) * \varphi_j(\boldsymbol{x})$$

where $*$ denotes the convolution operator. In the experiments, to obtain the multi-scale Gabor features, Gabor wavelets of five different scales $v \in \{0, 1, \ldots, 4\}$ and eight orientations $u \in \{0, 1, \ldots, 7\}$ are used, as in (Wiskott et al., 1997; Zhang et al., 2005; Liu and Wechsler, 2002).

2.2. Representation for an image sequence

Before Gabor wavelets are applied to each image in the sequence, original face images are preprocessed so that they are aligned in a predefined way. The alignment and cropping are done according to the eye centers from face detection. In the present experiments, face images are gray scale and normalized with a size of 64∗64.

The Gabor-based representation for a face image sequence can be derived by convoluting them with the multi-scale and multi-orientation Gabor wavelets. Each image in the sequence is convolved with the 40 Gabor wavelets (five scales and eight orientations) to generate the Gabor features. For each pixel position in the face image, 40 complex values can be calculated as Gabor features, which will totally result in 40 multi-scale and multi-orientation images. Evidently, this will result in a multi-scale presentation with a dimension of 40 times of the original face image size. Fig. 2 illustrates the procedure of the multi-scale and multi-orientation representation extraction of a face image.

For a given image sequence I with n images, each image is labeled with I_i, where i is the index of the image. This paper labels the multi-scale and multi-orientation features of image I_i with $G_{i,u,v}(x, y)$, where i is the index of the image and v is the index of the Gabor wavelet scale, u is corresponding to the Gabor wavelet orientation. Based on each multi-scale and multi-orientation feature $G_{i,u,v}(x, y)$, a multi-scale and multi-orientation representation G for the image sequence I is obtained as

$$\{G_{i,u,v}(x, y) : \; i \in (0, \ldots, n-1), u \in (0, \ldots, 7), v \in (0, \ldots, 4)\}.$$

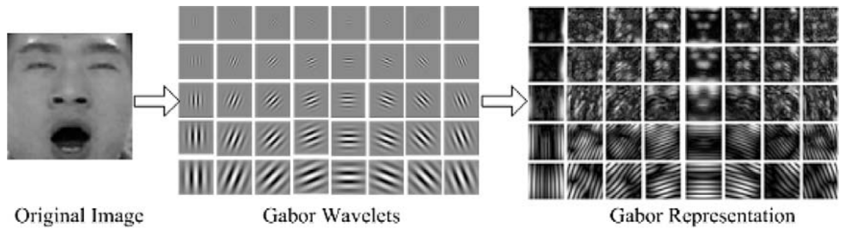

Fig. 2. Multi-scale and multi-orientation representation (magnitudes of the Gabor features) extraction of a face image.

3. Dynamic features extraction with feature-level fusion

Multi-orientation Gabor features in the same scale are fused into a single feature according to the fusion rules to extract the local orientation information and reduce the dimension of the features. To obtain the Gabor-based dynamic features of human fatigue, the fused image sequence is divided into sub-image sequences as dynamic units. Then, a histogram of each dynamic unit is computed and combined as dynamic features.

3.1. Feature fusion

In order to extract the local orientation information and reduce the feature dimension, the multi-orientation features are merged into a single feature (Fig. 3). The fusion is based on the local orientation. The Gabor wavelets adopted here include eight distinct orientations, and the corresponding eight Gabor features are exploited to estimate the local orientation. This paper proposes two fusion rules to merge the eight multi-orientation Gabor features.

3.1.1. Rule I

The local orientation is estimated by thresholding the 8-orientation features of each pixel into binary codes and considering the result as a binary number. Gabor features are thresholded to {0, 1} codes by:

$$T_{i,u,v}(x,y) = \begin{cases} 1 & \text{if } G_{i,u,v}(x,y) > 0 \\ 0 & \text{if } G_{i,u,v}(x,y) \leqslant 0 \end{cases}$$

where $G_{i,u,v}(x,y)u \in (0,\ldots,7)$ corresponds to the values of the 8-orientation Gabor features at the pixel (x, y). The resulting eight codes are considered as a binary number. The decimal form of the fused code can be expressed by:

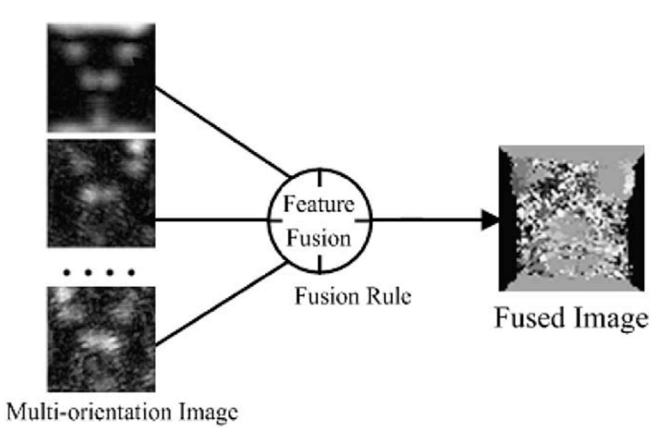

Fig. 3. Multi-orientation Gabor features fusion.

$$C_{i,v}(x,y) = \sum_{u=0}^{7} T_{i,u,v}(x,y) * 2^u$$

There are a total of 256 possible values and each code value represents a type of local orientations. Decimal forms of the fused codes are computed at each scale. Finally, the paper gets 5-scale fused images for one image (Fig. 4). In fact, the result of rule *I* is similar to that of the method proposed by Zhang et al. (2007).

3.1.2. Rule II

In Rule II, the orientation of a local region is estimated by the maximal magnitude of the 8-orientation Gabor features of each pixel:

$$m = \arg\max_u \{\|G_{i,u,v}(x,y)\|\}, \quad u \in (0,\ldots,7)$$

where $G_{i,u,v}(x,y)u \in (0,\ldots,7)$ also corresponds to the 8-orientation Gabor features at the pixel (x, y), and m is called the winning index of orientations. The paper uses the winning index as the fused code.

$$C_{i,v}(x,y) = m(i \in (0,\ldots,n-1),\ v \in (0,\ldots,4))$$

There are eight possible values and each code value also represents a type of local orientations. Finally, the paper can get 5-scale fused images for an image similarly (Fig. 4). From left-to-right, the scale is increased. Each scale image contains the information of the original image in the corresponding scale, and it is more notable for the real parts fused by the Rule II. The dimension of the fused features is 1/8 of the original features.

3.2. Dynamic features extraction

This paper defines the dynamic feature as a combination of histograms from each dynamic unit. To obtain the dynamic features of human fatigue and enhance their shape information, the fused image sequence is further divided into small non-overlapping rectangle sub-image sequences with a specific size as dynamic units. Local histograms of each dynamic unit are computed and combined into a single extended histogram as dynamic features to capture the temporal information. The procedure of the dynamic features extraction is illustrated in Fig. 5.

In the experiments, this paper divides each fused image into 8∗8 sub-images. For a fused image $C_{i,v}(x,y)(i \in (0,\ldots,n-1), v \in (0,\ldots,4))$, each rectangle sub-image can be denoted by

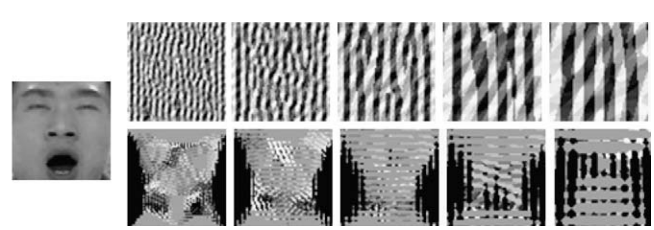

Fig. 4. Original image and multi-scale fused images (real part) (first row fused by Rule I, second row fused by Rule II).

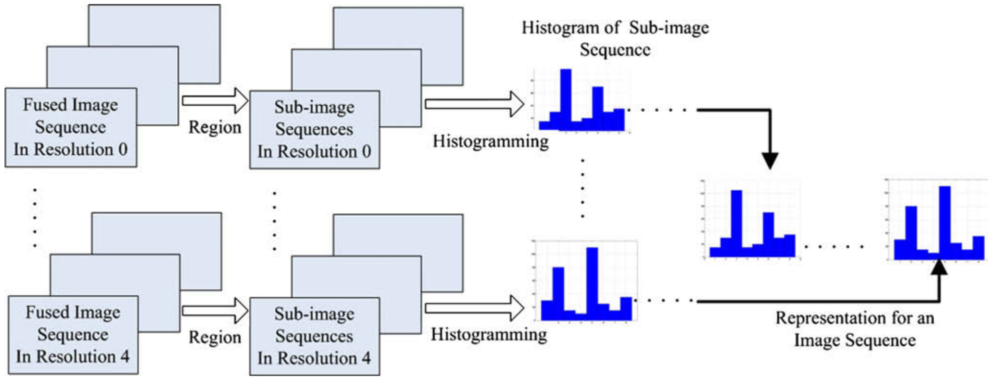

Fig. 5. Dynamic features extraction.

$R_{i,v,r}(x,y)(i \in (0,\ldots,n-1), v \in (0,\ldots,4), r \in (0,\ldots,63))$. The same sub-images in the fused image sequence are concatenated into a dynamic unit (sub-image sequence) $D_{v,r}(x,y)$ as $\{R_{0,v,r}(x,y), R_{1,v,r}(x,y),\ldots,R_{n-1,v,r}(x,y)\}$. The histogram of a dynamic unit can be defined as:

$$h_{i,v,r} = \sum_{x,y} I\{D_{v,r}(x,y) = i)\}, \quad i = 0,\ldots,k-1$$

where $I\{A\} = \begin{cases} 1, A & \text{is true} \\ 0, A & \text{is false} \end{cases}$, and k is 256 (Rule I) or 8 (Rule II).

Each histogram bin is the number of occurrences of the corresponding code in a dynamic unit. There are k bins for each dynamic unit. The resulting histograms of each dynamic unit are combined yielding the extended histogram for an image sequence. The histogram describing an image sequence can be defined as:

$$H = \{h_{i,v,r} : \ i \in (0,\ldots,k-1), v \in (0,\ldots,4), r \in (0,\ldots,63)\}$$

where k is 256 (Rule I) or 8 (Rule II). With this histogram, the paper obtains the Gabor-based dynamic features that can effectively describe the dynamic nature of a face image sequence at spatial and temporal levels.

4. Statistical learning of best features and classifiers

The dimension of the Gabor-based dynamic features for an image sequence is 81,920 (256∗64∗5) or 2560 (8∗64∗5), which is too high dimensional for fast extraction and accurate classification. At the same time, most of these features are redundant for classification. Therefore, it is necessary to reduce the dimension of the dynamic features. In the proposed method, AdaBoost algorithm is used to select a small set of dynamic features and train the classifier at the same time.

4.1. Weak classifiers

AdaBoost algorithm performs classification based on a set of weak classifiers and features. In this case, the paper uses threshold weak classifiers and the multi-scale dynamic features (histogram features), and one histogram feature corresponds to one weak classifier. The weak classifier is of the form:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j\theta_j \\ 0 & \text{otherwise} \end{cases}$$

where θ_j is the classification threshold of the weaker classifier, $p_j = \pm 1$ is a parity for inequality sign, and $f_j(x)$ is a dynamic feature.

4.2. AdaBoost algorithm

Due to its good generalization capability, fast performance and low implementation complexity, AdaBoost introduced by (Freund and Schapire, 1997) has achieved great success in face detection (Viola and Jones, 2001; Hou et al., 2006), face recognition (Li et al., 2007), face expression recognition (Yang et al., 2007) and other applications as a classification tool.

AdaBoost algorithm can select features and build a strong classifier at the same time. It provides a simple yet effective stage-wise learning approach for feature selection and nonlinear classification. It learns a sequence of easily learnable weak classifiers, each of which needs only slightly better than random guessing, and boosts them into a single strong classifier by a linear combination. AdaBoost algorithm is an ensemble classifier learning algorithm that works by creating a sequence of weak classifiers in iterative fashion, where each weak classifier is selected based on its performance on the training set. In each iterative step, the weight distribution over the training samples is updated in a way that forces the weak classifiers to focus on the examples that are hard to classify. This leads to a classifier with low training error and good generalization performance.

After all T iterations of the algorithm, the final classifier is obtained. The final classifier of AdaBoost algorithm is in the form of:

$$H(x) = \begin{cases} +1 & \text{if } \sum_{t=1}^{T} \alpha_t h_t(x) - \frac{1}{2}\sum_{t=1}^{T}\alpha_t \geqslant 0 \\ -1 & \text{otherwise} \end{cases}$$

The strong classifier is a linear combination of the T weak classifiers. α_t is the selected weight for classifier $h_t(x)$, as chosen by the AdaBoost algorithm, and $x = (x_1, x_2, \ldots, x_T)$ denotes all the selected dynamic features for the input facial image sequences. The AdaBoost learning procedure is aimed to derive α_t and $h_t(x)$.

5. Experiments

The goal of this section is to experimentally and scientifically demonstrate the validity of the proposed method. This paper tests the proposed method on a fatigue face database to gain knowledge of the proposed method about the performance. First, some details about the test dataset are showed. Then, comparison methods and experimental results are presented. Finally, some analysis about the proposed method is showed.

5.1. Database

There are many public databases for face detection, face recognition, and face expression recognition, however, there is no public database to test methods of human fatigue monitoring. The test sets in a lot of research are often very small with only several subjects. To test the fatigue monitoring methods, this paper built a fatigue face database. The paper used web cameras to capture videos of about forty persons. The videos of each person last several hours which are captured indoors. Behaviors of each subject are spontaneous. There are variations in illuminations (Fig. 7a), facial accessories (Fig. 7b) and poses (Fig. 7c) in the videos.

Totally, about 50 GB videos in AVI format are obtained compressed by MPEG-4. The paper selected thirty subjects' face fatigue videos from the original ones. Then the paper extracted the fatigue image sequences from the videos and made up the fatigue face database. Totally, there are 600 image sequences of 10 female subjects and 20 male subjects. Each subject has 10 alert image sequences and 10 fatigue image sequences. There are 5 images with a resolution of 320$*$240 in each image sequence (Fig. 6).

Randomly, the paper selects 300 sequences of 15 persons for the training set and the other 300 sequences of 15 persons for the probe set. Then the paper exchanges the training set and probe set to make a cross validation.

5.2. Methods for comparison

For comparison, the paper presents two statistical learning methods of fatigue monitoring, one based on LDA (PCA + LDA classifier) and the other on HMM (PCA + HMM classifier) for comparison.

The PCA + HMM classifier employs HMM to classify the sequence of PCA coefficients extracted by PCA projection from the face image sequence to monitor fatigue. This approach exploits temporal facial feature changes. HMM (Rabiner, 1989) is frequently used in the literature to perform spatio-temporal classification. It has good performance in many science and engineering fields such as speech recognition and sign language recognition. The PCA coefficients from an image sequence which includes five images are combined into a coefficient sequence and the sequence is classified by HMM. The paper models alert state and fatigue state with left-to-right HMMs. The two HMMs are first trained by the Baum–Welch–Algorithm (Baum, 1972) using the extracted coefficient sequences as observation vectors. For recognizing, the Viterbi-Algorithm is applied. Given a test coefficient sequence, states of drivers are recognized by comparison the likelihood coming from the fatigue model and that from the alert model.

Fig. 6. Examples of alert (first row) and fatigue (second row) facial image sequences in the database.

(a) Illumination variation

(b) Accessory (eyeglasses)

(c) Pose variation

Fig. 7. Examples from the fatigue face database.

The PCA + LDA classifier employs Linear Discriminant Analysis (LDA) to classify the PCA coefficients of face images to detect fatigue. LDA is one of the most widely used discriminant analysis techniques in classification and dimension reduction. Two experiments were made on the LDA based classifier. In the first one, LDA is used to classify the PCA coefficients from a single image. In the other, the PCA coefficients from an image sequence which includes five images are combined into a feature vector and the vector is classified by LDA.

5.3. Experimental results

In the experiments, the paper implemented Real AdaBoost scheme. Real AdaBoost (Schapire and Singer, 1999) is the generalization of a basic AdaBoost algorithm. The proposed method with Rule I was applied to the real and imaginary parts of the Gabor features. The proposed method with Rule II was applied to the real parts, imaginary parts and the magnitudes of the Gabor features. The 5-scale fused images of the same face image are showed in Fig. 8.

The commonly used performance evaluating criterion, correct rate, is adopted defined as the following formula. In the formula, $n_{\text{alert}-\text{alert}}$ represents the number of alert samples classified as alert ones, and $n_{\text{fatigue}-\text{fatigue}}$ represents the number of fatigue samples classified as fatigue ones. n_{alert} and n_{fatigue} are the numbers of alert and fatigue test samples respectively.

$$\text{Correct rate} = \frac{n_{\text{alert}-\text{alert}} + n_{\text{fatigue}-\text{fatigue}}}{n_{\text{alert}} + n_{\text{fatigue}}} * 100\%$$

The performance of the proposed method with different numbers of weak classifiers is showed in Fig. 9. When the number of weak classifiers is small, the correct rates increase quickly with the number of weak learners. When the number of weak classifiers becomes big, the correct rates increase slowly. Even in some instances, the correct rates decrease. Approximately, the performance of the method with Rule II on the real parts of the Gabor features (line Rule II (real)) is better than the others. Method II is better than method I, and real parts of the Gabor features (line Rule I (real) and line Rule II (real)) include more discriminative information for classification than that of the imaginary parts (line Rule I (imaginary) and line Rule II (imaginary)). Magnitudes of the Gabor features can be regarded as a fusion of the real and imaginary parts of the Gabor features, and the performance of the magnitudes with Rule II is between that of the real and imaginary parts.

The recognition results of all the classifiers are listed in Table 1. The classifying methods include PCA + HMM, PCA + LDA (single), PCA + LDA (sequence) and the proposed method. The average correct rates of the comparison methods are 75.50%, 79.77%, and 82.17%, respectively. The best average correct rates achieved by the proposed method with the dynamic features are 97.50% (Rule I (real)), 84.67% (Rule I (imaginary), 99.33% (Rule II (real)), 91.83% (Rule II (imaginary)), and 98.17% (Rule II (magnitudes)). The numbers of the weak classifiers selected by the final classifiers are 78, 146, 34, 98, and 19, respectively.

The proposed method with Rule II on the Gabor real parts (Rule II (real)) achieves a very encouraging correct rate (99.33%) which is much better than that of the others. At the same time, the number (34) of the selected weak classifiers is smaller than most of the oth-

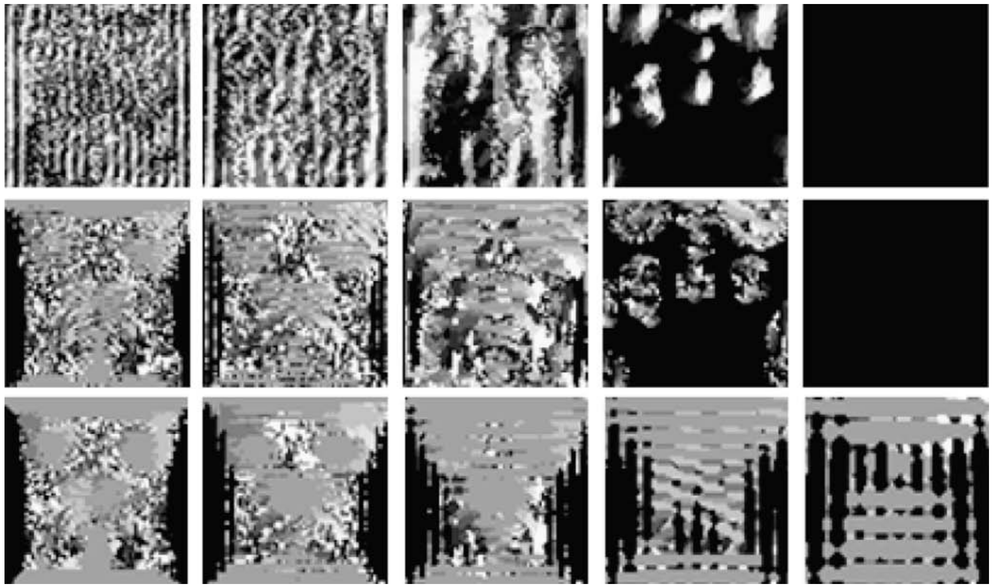

Fig. 8. Multi-scale fused images: first row, imaginary parts fused by Rule I; second row, imaginary parts fused by Rule II; and third row, magnitudes fused by Rule II.

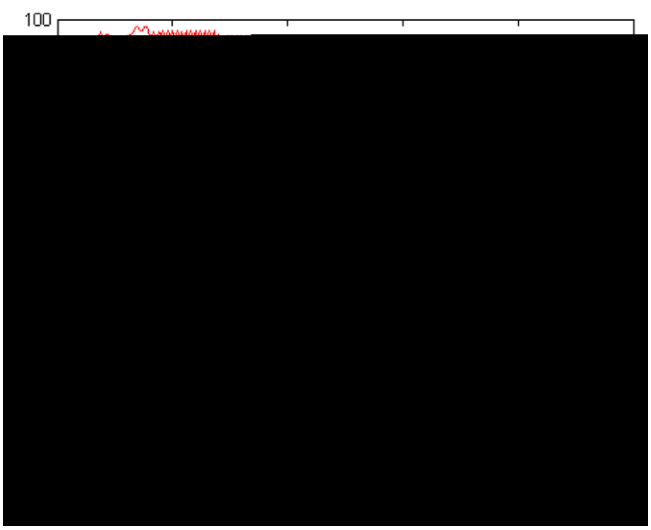

Fig. 9. Performance of the proposed method with different numbers of weak classifiers.

Table 1
Confusion matrix for fatigue detection results.

Method	Data	Alert (%)	Fatigue (%)
PCA + HMM	Alert	74.33	25.67
	Fatigue	23.33	76.67
PCA + LDA (single)	Alert	78.27	21.73
	Fatigue	18.73	81.27
PCA + LDA (sequence)	Alert	82.00	18.00
	Fatigue	17.67	82.33
Rule I (real)	Alert	97.00	3.00
	Fatigue	2.00	98.00
Rule I (imaginary)	Alert	79.00	21.00
	Fatigue	9.67	90.33
Rule II (real)	Alert	100.00	0
	Fatigue	1.33	98.67
Rule II (imaginary)	Alert	94.33	5.67
	Fatigue	10.67	89.33
Rule II (magnitude)	Alert	99.67	0.33
	Fatigue	3.33	96.67

ers. The recognition correct rate of the alert sequences is 100.00% while that of the fatigue sequences is 98.67%. PCA + LDA (sequence) is better than PCA + LDA (single) in performance. This may be account for the dynamic characteristic in the sequences. The performance of PCA + HMM is poor. The reason may be that HMM needs enough training samples.

5.4. Computational cost analysis

In a real fatigue monitoring application, the real-time ability of the system is very important. In this section, the computational cost analysis of the system is presented.

In the training process, all the Gabor wavelets ($5*8$) are used for each pixel of each image. And all the multi-scale dynamic features are extracted and used for the classifier training. Therefore the training is time-consuming. However, the training can be done off-line, so it is not a barrier to the whole system.

In the training process, AdaBoost algorithm is used to select a subset of the dynamic features to construct the final strong classifier to monitor fatigue. The paper calls pixels that are useful for the final selected dynamic features as selected pixels, and the Gabor wavelets that are useful for the final selected dynamic features as selected Gabor wavelets. In the testing process, only the selected pixels and the selected Gabor wavelets are used in the Gabor transformation.

In the proposed approach, the number of the selected features is 34. It is much smaller than that of the whole dynamic features (2560). This is appropriate for fast and accurate classification. At the same time, Gabor transformation is not needed for the non-selected pixels and the non-selected Gabor wavelets that are useless for the final selected dynamic features. This will save much time for the dynamic feature extraction.

5.5. Analysis of the selected features

In the proposed method, AdaBoost algorithm is used to select the most discriminative features and construct a strong classifier to monitor fatigue. The selected features from the training samples should be the most informative features discriminating the different states of the face image sequences and low dimensional. To observe the characteristics of the selected features intuitively, some statistics of the features selected by the proposed method (Rule II on the real parts of the Gabor features) with the best performance are given below.

5.5.1. Scale analysis of the selected features

The scale distribution of the selected dynamic (histogram) features is showed in Fig. 10. The selected features are distributed in most of the scales, and this illustrates that multi-scale analysis is necessary for human fatigue monitoring. The smaller scale features contribute little to the final AdaBoost classifier. The bigger scale features contribute much for human fatigue monitoring and the features from the 3-scale Gabor kernels are more than 55% of all the selected features. The reason may be that yawning as a fatigue behavior is a movement of the mouth area, and need analysis with relatively big scales.

5.5.2. Orientation analysis of the selected features

Different orientations also contribute differently to the classification. Fig. 11 shows the orientation distribution of the selected features. It seems that the orientation distribution is somewhat uniform compared with the scale distribution. However, the vertical Gabor kernels (with $u = 0$) have extracted about 25% of all the selected features, and the Gabor kernels near the horizontal orientation (with $u = 3, 4, 5$) also contribute much to the final classifier.

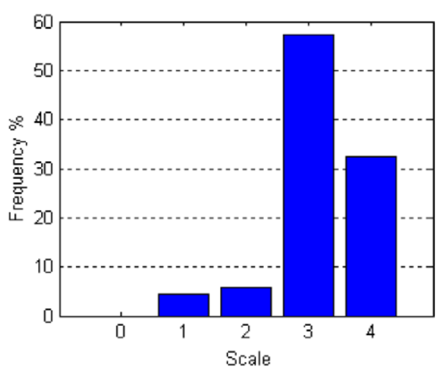

Fig. 10. Scale distribution of the selected features.

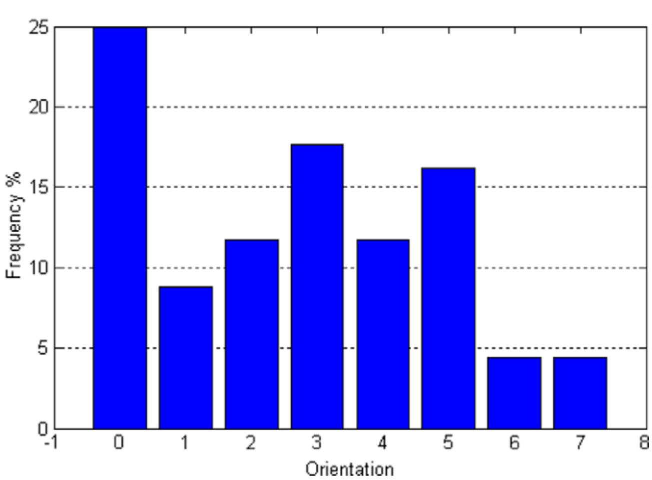

Fig. 11. Orientation distribution of the selected features.

5.5.3. Region analysis of the selected features

Fig. 12 illustrates the region (sub-image) distribution of the selected features for the final AdaBoost classifier. From Fig. 12, it can be safely concluded that features of different regions contribute to the classification quite differently. More than 15% of the selected features are from the No. 60 region, while more than half of the regions have no selected features. Fig. 13 shows the regions of face images and the region distribution of the selected features. Light regions mean that more features for the fatigue monitoring classifier are selected from the regions. The No. 60 region is lighter than the other regions for more selected features, i.e. the Gabor transformation of this region contributes more for the final classifier when compared with the others.

5.6. Parameter measurement for one of the test image sequences

In this section, the paper will show how to apply the proposed method in a real-life condition. Before the Gabor wavelets are applied to each image in the sequence, the original face images are preprocessed so that they are aligned in a predefined way. The paper uses the Gravity-Center template (Miao et al., 1999) to detect human faces. Face detection based on the Gravity-Center template can produce the location information about the faces and the facial organs such as eyes, noses and mouths. The alignment is done according to the eye centers. When the system gets an image sequence of five aligned facial images non-overlapped from a video, the selected Gabor wavelets are used for the Gabor transformation on the aligned facial image sequence and the selected 34 dynamic features are extracted by the proposed algorithm and the trained final classifier is used to get the final parameters. The subject in the sequence is fatigue or alert can be judged by the signs of the parameter values.

Fig. 14 shows some frames of a test image sequence. It was captured in a real-life condition and without directions to the subject. The sequence lasts about 6 s which includes 125 frames. The middle frame of each five frames from the sequence is extracted, and 25 frames are obtained as showed in Fig. 14. This is a representa-

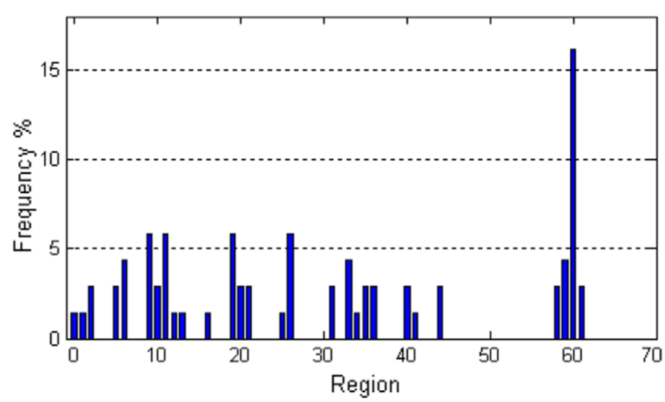

Fig. 12. Region distribution of the selected features.

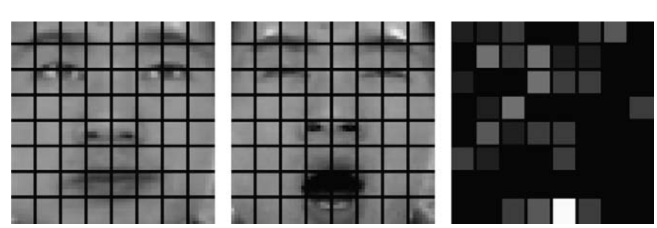

Fig. 13. Regions of the face images and region distribution of the selected features.

Fig. 14. Some frames of the test image sequence.

tive test example where the subject makes fatigue behaviors between two alertness periods. Fig. 15 depicts the parameter values for the image sequence. The states of the subject when the system receives five new frames can be judged by the signs of the values (+1 corresponding to fatigue and −1 corresponding to alert). The faces near each value in Fig. 15 are cropped from the middle frames of each five frames. As can be seen, until the 6th frame (in fact, it is the 28th frame of the image sequence), and from the 20th frame to the 25th frame, the parameters are below 0, indicating an alertness state. From the 7th frame to the 19th frame, the parameters are over 0, indicating a fatigue state. From the face images near the values, it can be found that the signs of the parameter values are consistent to the states of the subject. In some sense, the values can be looked as the fatigue levels of the subject. The higher the value is, the more fatigue the subject is as in the 13th frame the value is very high (0.544214). The lower the value is, the more alert the subject is as in the 24th frame the value is −0.254563. When the value is near 0, it means than it is difficulty to judge whether the subject is fatigue or alert as the 7th frame (0.005651), 19th frame (0.093110) and the 20th frame (−0.076523). Approximately, the values of the parameter are consistent with the fatigue levels of the subject, but there are some deviations in some frames such as the 8th frame. Those deviations may be the result of the face misalignment from the inaccurate positions of the eye centers. A more accurate eye location algorithm can improve the belief of the computed values. Using real-time video input from a USB digital camera, the face detection method can process an image within 45 ms (without tracking) on a PC with a Pentium IV 3.0 GHz CPU, 1G MB memory and Windows XP operating system. The whole system (with tracking) is currently running in real-time (19 frames/s) on the same platform. The resolution of each image from the video is 320∗240 pixels.

6. Conclusions and future work

Human fatigue is one of the most important safety concerns in the modern transportation. Monitoring and preventing human fatigue are crucial to improve the transportation safety. Besides a review of the previous works about human fatigue monitoring, the presented method makes several contributions to this issue. First, a novel multi-scale dynamic feature is presented to account for the multi-scale, spatial, and temporal aspects of human fatigue in image sequences. Second, to extract the local orientation information and reduce the dimension of the features, two fusion rules at the feature level are proposed to fuse the original features of the same scale into a single feature. Finally, AdaBoost algorithm is used to extract the most critical features from the dynamic feature set and construct a strong classifier for fatigue monitoring. The proposed method is validated on 600 image sequences from thirty people in a real-life fatigue environment. Experimental results show the validity of the proposed method, and a promising average correct rate is achieved which is much better than the other methods. Some statistics of the features selected for the final classifier is presented.

Magnitudes of the Gabor features are fusion of the real and imaginary parts of the Gabor features, but the performance is not better than that of the real parts. In future work, efforts will be focused on how to combine the multi-scale dynamic features from the Gabor real parts and imaginary parts to get a better performance. If possible, the paper would take a hybrid classifier to fuse visual features from single images and continuous image sequences in the future work.

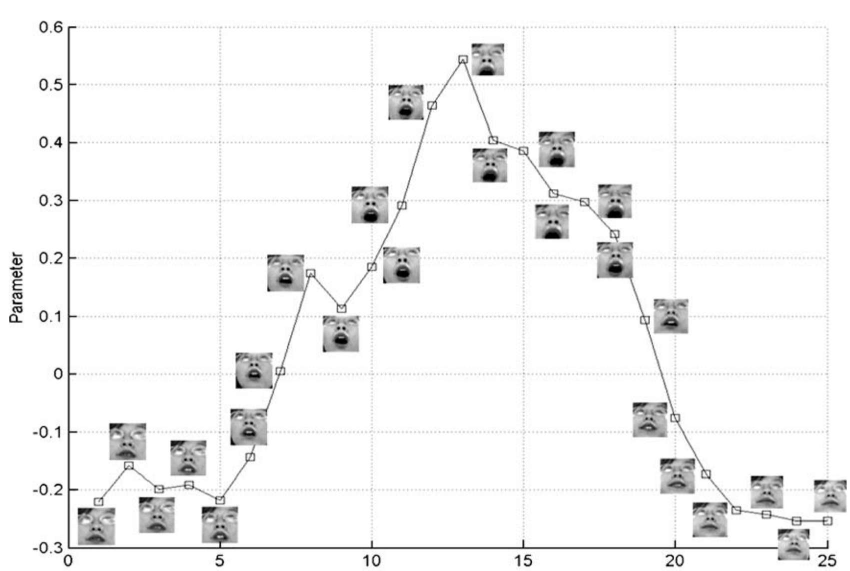

Fig. 15. Parameter values for the test sequence.

Acknowledgements

This work is sponsored partly by National Natural Science Foundation of China (No. 60533030, 60825203, 60973057), and National Key Technology R&D Program (2007BAH13B01).

References

Abdul-Latif, A.A., Cosic, I., Kumar, D.K., Polus, B., Da Costa, C., 2004. Power changes of EEG signals associated with muscle fatigue: The root mean square analysis of EEG bands. In: Proc. 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, pp. 531–534.

Bassili, J., 1979. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. J. Pers. Social Psychol. 37, 2049–2059.

Baum, L.E., 1972. An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes. Inequalities 3, 1–8.

Bergasa, L.M., Nuevo, J., Sotelo, M.A., Barea, R., Lopez, M.E., 2006. Real-time system for monitoring driver vigilance. IEEE Trans. Intell. Transport. Syst. 7 (1), 63–77.

Bonato, P., Heng, M.S.S., Gonzalez-Cueto, J., Leardini, A., O'Connor, J., Roy, S.H., 2001. EMG-based measures of fatigue during a repetitive squat exercise. IEEE Eng. Med. Biol. Mag. 20 (6), 133–143.

Daugman, J., 1980. Two-dimensional spectral analysis of cortical receptive field profiles. Vis. Res. 20, 847–856.

Dinges, D.F., Grace, R., 1998. Perclos: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance. US Department Transportation, Federal Highway Administration, Washington, DC, Technical Report Publication No. FHWA-MCRT-98-006.

Dinges, D.F., Mallis, M., Maislin, G., Powell, J.W., 1998. Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management. Department of Transportation Highway Safety Publication, Washington, DC, Technical Report. 808 762.

Dong Wenhui, Wu Xiaojuan, 2005. Fatigue detection based on the distance of eyelid. In: Proc. 2005 IEEE International Workshop on VLSI Design and Video Technology, pp. 365–368.

Fan Xiao, Yin Baocai, Sun Yanfeng, 2008. Nonintrusive driver fatigue detection. In: 2008 IEEE Internat. Conf. on Networking, Sensing and Control. Sanya, China, pp. 905–910.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Galley, N., Schleicher, R., 2004. Subjective and optomotoric indicators of driver drowsiness. In: The 3rd Internat. Conf. on Traffic and Transport Psychology. Nottingham, UK.

Hamada, T., Ito, T., Adachi, K., Nakano, T., Yamamoto, S., 2003. Detecting method for drivers' drowsiness applicable to individual features. In: Proc. 2003 IEEE Intelligent Transportation Systems, vol. 2, pp. 1405–1410.

Hayashi, K., Ishihara, K., Hashimoto, H., Oguri, K., 2005. Individualized drowsiness detection during driving by pulse wave analysis with neural network. In: Proc. 2005 IEEE Intelligent Transportation Systems, pp. 901–906.

Hou Xinwen, Liu ChengLin, Tan Tieniu, 2006. Learning boosted asymmetric classifiers for object detection. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 330–338.

Ji, Qiang., Lan, P., Looney, C., 2006. A probabilistic framework for modeling and real-time monitoring human fatigue. IEEE Trans. Syst., Man Cybernet., Part A 36 (5), 862–875.

Li, Stan Z., Chu, RuFeng, Liao, ShengCai, Zhang, Lun, 2007. Illumination invariant face recognition using near-infrared images. IEEE Trans. Pattern Anal. Machine Intell. 29 (4), 627–639.

Lin Chin-Teng, Chen Yu-Chieh, Wu Ruei-Cheng, Liang Sheng-Fu, Huang Teng-Yi, 2005a. Assessment of driver's driving performance and alertness using EEG-based fuzzy neural networks. In: IEEE Internat. Symposium on Circuits and Systems, vol. 1, pp. 152–155.

Lin Chin-Teng, Wu Ruei-Cheng, Liang Sheng-Fu, Chao Wen-Hung, Chen Yu-Jie, Jung Tzyy-Ping, 2005b. EEG-based drowsiness estimation for safety driving using independent component analysis. In: IEEE Transactions on Circuits and Systems I: Regular Papers, 52(12), 2726–2738.

Liu, Chengjun, Wechsler, Harry, 2002. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. IEEE Trans. Image Process. 11 (4), 467–476.

Liu, X., Fengliang, X., Fujimura, K., 2002. Real time eye detection and tracking for driver observation under various light conditions. In: Proc. 2002 IEEE Intelligent Vehicle Symposium, Versailles, France.

Lu Yufeng, Wang Zengcai, 2007. Detecting driver yawning in successive images. In: Proc. 1st Internat. Conf. on Bioinformatics and Biomedical Engineering, pp. 581–583.

Miao, Jun., Yin, Baocai., Wang, Kongqiao., Shen, Lansun., Chen, Xuecun., 1999. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. Pattern Recognit. 32 (7), 1237–1248.

Miyakawa, T., Takano, H., Nakamura, K., 2004. Development of non-contact real-time blink detection system for doze alarm. In: SICE 2004 Annual Conference, vol. 2, pp. 1626–1631.

NHTSA, 2005. Drowsy Driver Detection and Warning System for Commercial Vehicle Drivers: Field Proportional Test Design, Analysis, and Progress. National Highway Traffic Safety Administration, Washington, DC. Available from: <http://www.nhtsa.dot.gov/>.

Parikh, P., Micheli-Tzanakou, E., 2004. Detecting drowsiness while driving using wavelet transform. In: Proc. IEEE 30th Annual Northeast Bioengineering Conference, pp. 79–80.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77 (2), 257–286.

Schapire, R.E., Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. Mach. Learn. 37 (3), 297–336.

Takei, Y., Furukawa, Y., 2005. Estimate of driver's fatigue through steering motion. In: IEEE Internat. Conf. on Systems, Man and Cybernetics, vol. 2, pp. 1765–1770.

Thum Chia Chieh, Mustafa, M.M., Hussain, A., Zahedi, E., Majlis, B.Y., 2003. Driver fatigue detection using steering grip force. In: Proc. 2003 Student Conf. on Research and Development, pp. 45–48.

Tong Yan, Liao Wenhui, Ji Qiang, 2007. Facial action unit recognition by exploiting their dynamic and semantic relationships. IEEE Trans. Pattern Anal. Machine Intell. 29 (10), 1683–1699.

Viola, P., Jones, M., 2001. Robust real time object detection. In: IEEE ICCV Workshop on Statistical and Computational Theories of Vision. Vancouver, Canada.

Wang Fei, Qin Huabiao, 2005. A FPGA based driver drowsiness detecting system. In: IEEE Internat. Conf. on Vehicular Electronics and Safety, pp. 358–363.

Wang Tiesheng, Shi Pengfei, 2005. Yawning detection for determining driver drowsiness. In: Proc. 2005 IEEE Internat. Workshop on VLSI Design and Video Technology, pp. 373–376.

Wang Rong-Ben, Guo Ke-You, Shi Shu-Ming, Chu Jiang-Wei, 2003. A monitoring method of driver fatigue behavior based on machine vision. In: Proc. 2003 IEEE Intelligent Vehicles Symposium, pp. 110–113.

Wang Rongben, Guo Lie, Tong Bingliang, Jin Lisheng, 2004. Monitoring mouth movement for driver fatigue or distraction with one camera. In: Proc. 7th Internat. IEEE Conf. on Intelligent Transportation Systems, pp. 314–319.

Wiskott, L., Fellous, J.M., Kuiger, N., von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 775–779.

Wu Ruei-Cheng, Lin Chin-Teng, Liang Sheng-Fu, Huang Teng-Yi, Jung Tzyy-Ping, 2004a. EEG-based fuzzy neural network estimator for driving performance. In: 2004 IEEE Internat. Conf. on Systems, Man and Cybernetics, vol. 4, pp. 4034–4040.

Wu Ruei-Cheng, Lin Chin-Teng, Liang Sheng-Fu, Huang Te-Yi, Chen Yu-Chieh, Jung Tzyy-Ping, 2004b. Estimating driving performance based on EEG spectrum and fuzzy neural network. In: Proc. 2004 IEEE Internat. Joint Conference on Neural Networks, vol. 1, pp. 585–590.

Yang Peng, Liu Qingshan, Metaxas, D.N., 2007. Boosting coded dynamic features for facial action units and facial expression recognition. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–6.

Zhang Wenchao, Shan Shiguang, Gao Wen, Zhang Hongming, 2005. Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Internat. Conf. on Computer Vision. Beijing, pp. 786–791.

Zhang, Baochang, Shan, Shiguang, Chen, Xilin, Gao, Wen, 2007. Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Trans. Image Process. 16 (1), 57–68.

Zhao, Guoying, Pietikainen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Machine Intell. 29 (6), 915–928.

Zhu Zhiwei, Ji Qiang, 2004. Real time and non-intrusive driver fatigue monitoring. In: Proc. 7th Internat. IEEE Conf. on Intelligent Transportation Systems, pp. 657–662.