Dynamic Human Fatigue Detection Using Feature-Level Fusion

Xiao Fan, Bao-Cai Yin, and Yan-Feng Sun

Beijing Key Laboratory of Multimedia and Intelligent Software, College of Computer Science and Technology, Beijing University of Technology, Beijing 100022, China
`wolonghongni@emails.bjut.edu.cn`

Abstract. Driver fatigue is a significant factor in many traffic accidents. We propose a novel dynamic features using feature-level fusion for driver fatigue detection from facial image sequences. First, Gabor filters are employed to extract multi-scale and multi-orientation features from each image, which are then merged according to a fusion rule to produce a single feature. 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 concatenated as dynamic features. Finally a statistical learning algorithm is applied to extract the most discriminative features and construct a strong classifier for fatigue detection. The test data contains 600 image sequences from thirty people. Experimental results show the validity of the proposed approach, and the correct rate is much better than the baselines.

Keywords: Computer vision, human fatigue, Gabor filters, fusion, AdaBoost.

1 Introduction

Driver fatigue is a significant factor in many traffic accidents. In China, driver fatigue resulted in 3056 deaths in vehicular accidents in 2004, and caused 925 deaths in highway accidents which amounted to about 14.8%. When a driver fatigues, many special visual features will appear on his face and body. These features reflect a person's level of fatigue. Many computer vision based approaches have been proposed for fatigue detection in the last decade.

The frequency and time of eye closed all increase when driver fatigue. Much attention is paid to eye features for fatigue detection. Based on the study by the Federal Highway Administration, percentage of eyelid closure (PERCLOS) has been found to be the most reliable and valid measure of a person's alertness level among many drowsiness detection measures. Wang *et al.* [1] used Gabor filters to extract texture features of 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 in [2]. Dong *et al.* [3] decided whether the driver was fatigue by detecting the distance of eyelids.

Yawning is also an important character of fatigue. Wang *et al.* [4] 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 *et al.* [5] 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.

Most of these methods are spatial approaches. Although spatial approaches can achieve good recognition in some cases, they do not model the dynamics of fatigue and therefore do not utilize all of the information available in image sequences. In facial expression recognition, according to psychologists [6], analyzing an image sequence produces more accurate and robust facial expression recognition. Psychological researches suggest that facial motion is fundamental to facial expression recognition. Therefore, more attention [7, 8] has been shifted particularly towards modeling dynamic facial expressions.

Fatigue is a cognitive state that is developed over time. It is our belief that dynamic features which capture the temporal pattern should be the optimal features to describe fatigue just as facial expression recognition. To account for the temporal aspect of human fatigue, Ji *et al.* [9] introduced a probabilistic framework based on dynamic Bayesian networks for modeling and real-time inferring human fatigue. The dynamic fatigue model integrates fatigue evidences spatially and temporally. However, in summary, there is little research in extracting dynamic features from image sequences for fatigue detection. Achieving high accuracy in fatigue detection is still a challenge due to the complexity and variation of facial dynamics.

In this paper, to account for the temporal characteristic of human fatigue, we propose a novel dynamic features using feature-level fusion to detect fatigue from image sequences. Fig.1 illustrates the framework of the proposed approach. First, after each image in a face image sequence is preprocessed by face detection, geometric normalization and cropping, Gabor filters are used to extract multi-scale and multi-orientation features from each image in the sequence. Then these features are merged according to a fusion rule to produce a single feature. To get the dynamic features of human fatigue, the fused image sequence is divided into rectangle region sequences as dynamic units, and a histogram of each dynamic unit is computed and concatenated as dynamic features. Finally, weak classifiers are constructed on the histogram features and AdaBoost is applied to select the most discriminating dynamic features and build a strong classifier for fatigue detection.
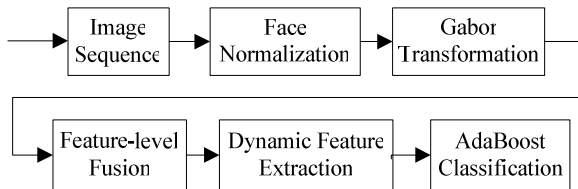

Fig. 1. Framework of the proposed approach

The remainder of this paper is organized as follows. Section 2 introduces Gabor transformation on image sequences. Dynamic feature extraction is showed in section 3. In section 4, feature selection and classifier learning are described. Finally, experimental results and conclusions are presented in section 5 and 6.

2   Multi-scale and Multi-orientation Representation

We exploit Gabor filters to decompose the input face image sequence and get the multi-scale and multi-orientation representation for the sequence after original face images are preprocessed and normalized in a predefined way. Face images in our experiments are normalized with a size of 64*64.

2.1   Gabor Filters

Gabor filters [10] share many properties with the receptive fields of simple cells in the mammalian visual cortex. Due to their biological relevance and computational properties, they have been successfully used in a variety of image processing applications such as face recognition. Gabor filters can be defined as follows:

$$\varphi_j\left(\vec{x}\right) = k_j^2 \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right) \exp\left(i\vec{k}_j \vec{x}\right)$$

Where, $\vec{k}_j = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos\phi_u \\ k_v \sin\phi_u \end{pmatrix}$, $k_v = 2^{-\frac{v+2}{2}}\pi$, $\phi_u = u\frac{\pi}{K}$, and v determines the frequencies of the Gabor filters, u determines the orientations of the Gabor filters. For an image with the gray level distribution of $L\left(\vec{x}\right)$, at a given point $\vec{x} = (x, y)$, The Gabor representation of the image is the convolution of the image with a family of Gabor filters defined as follows:

$$J_j\left(\vec{x}\right) = \int L\left(\vec{x}\right) \varphi_j\left(\vec{x} - \vec{x}'\right) d^2 \vec{x}'$$

In our experiments, we adopt a discrete set of Gabor filters which comprise 5 spatial frequencies, i.e., scales, and 8 distinct orientations. We set $K=8$, $v \in \{0, 1, 2, 3, 4\}$, $u \in \{0, 1, 2, 3, 4, 5, 6, 7\}$.

2.2   Representation for an Image Sequence

The representation for a face image sequence can be derived by convoluting it with the multi-scale and multi-orientation Gabor filters. Each image in the sequence is convolved with the 40 Gabor filters (5 scales and 8 orientations) to generate the Gabor features. Thus, for each pixel position in the face image, 40 complex values can be calculated as Gabor features.

Given one image sequence I with n images, we label each image with I_i, where i is the index of the image. We label the Gabor 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 filter scale, u is corresponding to the Gabor filter orientation. Based on each image $G_{i,u,v}(x, y)$, we obtain a multi-scale and multi-orientation representation G as $\{G_{i,u,v}(x, y) : i \in (0, ..., n-1), u \in (0, ..., 7), v \in (0, ..., 4)\}$ for the image sequence I. Fig. 2 illustrates the representation (magnitudes of the Gabor features) of a face image.
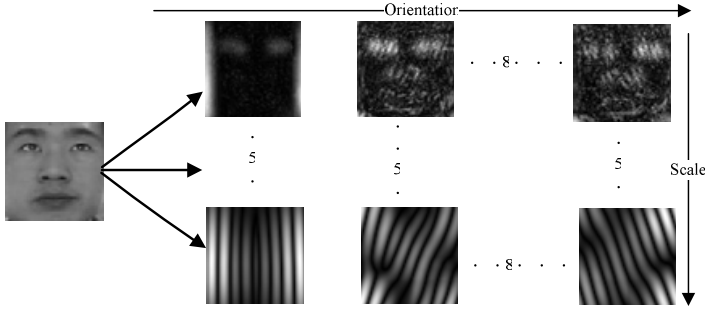
Fig. 2. Multi-scale and multi-orientation representation of a face image

3 Dynamic Feature Extraction Based on Feature-Level Fusion

To obtain the dynamic features of human fatigue, multi-orientation features are merged according to a fusion rule to produce a single feature, and the fused image sequence is divided into region sequences as dynamic units. Then, a histogram of each dynamic unit is computed and concatenated as dynamic features.
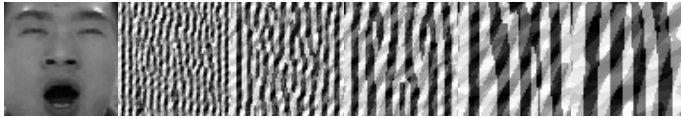
3.1 Feature-Level Fusion

Fig. 3. Original image and multi-scale fused images (real part)

In order to extract the local orientation information, the multi-orientation features are fused into a single code. Gabor filters that we adopted include 8 distinct orientations. 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. We threshold Gabor features to *{0, 1}* codes by the following formula:

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

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

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

There are a total of 256 possible values and each code value represents a type of local orientation. Decimal forms of the fused codes are computed at each scale. Finally, we get 5-scale fused images for one image (Fig.3).

3.2 Dynamic Feature Extraction

To obtain the dynamic features, the fused image sequence is further divided into small non-overlapping rectangle region sequences to enhance its shape information, from which the local histograms are extracted and concatenated into a single extended histogram to capture the temporal information (Fig. 4).

In the experiments, we divide each fused image into 8*8 regions. Each rectangle region can be denoted as $R_{i,v,r}(x,y)(i \in (0,...,n-1), v \in (0,...,4), r \in (0,...,63))$. The same regions in the fused image sequence are concatenated into a dynamic unit (region sequence) $S_{v,r}(x,y)$ as $\{R_{0,v,r}(x,y), R_{1,v,r}(x,y),..., R_{n-1,v,r}(x,y)\}$. A histogram of a dynamic unit can be defined as:

$$h_{i,v,r} = \sum_{x,y} I\{S_{v,r}(x,y) = i)\}, i = 0,...,255, \text{ where } I\{A\} = \begin{cases} 1, A \text{ is true} \\ 0, A \text{ is false} \end{cases}$$
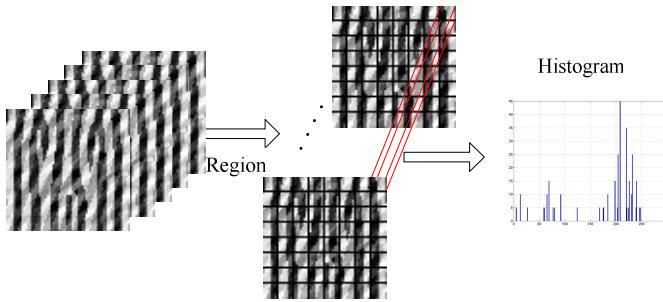

Histogram

Fig. 4. Flowchart of dynamic feature extraction

There are 256 bins for one dynamic unit. Each histogram bin is the number of oc-currences of the corresponding code in a dynamic unit. The resulting histograms on each dynamic unit are combined yielding the extended histogram for an image se-quence. The histogram to descript an image sequence can be defined as

$$H = \{h_{i,v,r} : i \in (0,...,255), v \in (0,...,4), r \in (0,...,63)\}$$

With this histogram, we effectively have a dynamic description of a face image se-quence on spatial and temporal levels.

4 Statistical Learning of Best Features and Classifiers

The number of dynamic features of an image sequence is 256*64*5=81920, which is too high dimensional for fast extraction and accurate classification. We use AdaBoost both to select a small set of features and train the classifier.

4.1 Weak Classifiers

The weak classifiers are the basis of AdaBoost. In our case, the weak classifiers are decision trees based on the histogram features. Decision tree is a tree graph, with

leaves representing the classification results and nodes representing some predicates. Branches of the tree are marked with true or false.

4.2 AdaBoost Learning

AdaBoost [8, 11] 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. The final hypothesis of AdaBoost algorithm is

$$H\left(x\right) = \begin{cases} 1 & if \ \sum_{t=1}^{T} \alpha_t h_t\left(x\right) \ \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

The strong classifier is a linear combination of the T weak classifiers. The AdaBoost learning procedure is aimed to derive α_t and $h_t(x)$.

5 Experimental Results

5.1 Database

The test sets in a lot of researches are often very small with only several subjects. To test the fatigue detection methods, we built a fatigue face database. We used web cameras to catch videos of about forty persons. The videos of each person last several hours which are caught indoors without directions to the subjects.

Fig. 5. Examples of (a) normal and (b) fatigue image sequences in the database

Totally, we got about 50GB videos in AVI format compressed by MPEG-4. We selected thirty subjects' face fatigue videos from the original ones. Then we 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 normal image sequences and 10 fatigue image sequences. There are 5 images with a resolution of 320*240 in each image sequence (Fig. 5).

5.2 Baselines

We present two statistical learning methods, one based on LDA (PCA+LDA classifier) and one on HMM (PCA+HMM classifier) as baselines. PCA+HMM employs

HMM to classify the sequence of PCA coefficients extracted from a face image sequence. PCA+LDA employs LDA to classify the PCA coefficients of face images. 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, PCA coefficients from an image sequence are combined into a feature vector and classified by LDA.

5.3   Results

In our experiments, we implemented Real AdaBoost scheme. The proposed approach was applied to the Gabor real and imaginary parts. The best performance with decision trees of different numbers of splits are showed in Fig. 6. We can see that the proposed approach gets the highest correct rate when the decision trees have one split. Then, when the number of splits increases, the performance decreases.
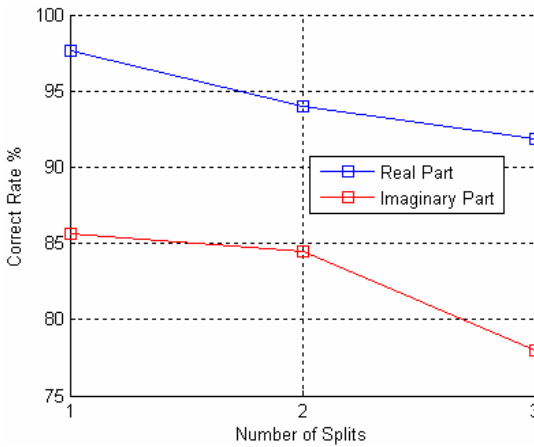

Fig. 6. Best performances with different numbers of splits

Table 1. Confusion matrix for fatigue detection

Method		Normal	Fatigue
PCA+HMM	Normal	74.33%	25.67%
	Fatigue	23.33%	76.67%
PCA+LDA	Normal	78.27%	21.73%
(Single)	Fatigue	18.73%	81.27%
PCA+LDA	Normal	82.00%	18.00%
(Sequence)	Fatigue	17.67%	82.33%
Our Method	Normal	97.33%	2.67%
(Real)	Fatigue	2.00%	98.00%
Our Method	Normal	79.67%	20.33%
(Imaginary)	Fatigue	8.33%	91.67%

The recognition results of all the classifiers are listed in table 1. The average correct rates are 75.50%, 79.77%, 82.17%, 97.67% (real), and 85.67% (imaginary) respectively. The proposed approach with the Gabor real part achieves a very encouraging recognition rate which is much better than the other methods. The correct rate of normal sequences is 97.33% while that of fatigue is 98.00%. PCA+LDA (sequence) is better than PCA+LDA (single) in performance. This may be account for the dynamic characteristic in the sequence. The performance of PCA+HMM is poor. The reason may be that HMM needs enough training samples.

6 Conclusions

A novel dynamic feature using feature-level fusion is presented to account for the spatial and temporal aspects of human fatigue in image sequences. AdaBoost algorithm is used to extract the most discriminative features and construct a strong classifier for fatigue detection. The proposed approach is validated in a real-life fatigue environment with 600 image sequences from thirty people. Experiment results show the validity of the proposed approach, and an encouraging correct rate is achieved. In addition, this approach can be easily extended to video based facial expression recognition. Future efforts will be focused on how to combine the dynamic features from Gabor real parts and imaginary parts to get better performance.

References

1. Wang, R., Guo, K., Shi, S., Chu, J.: A Monitoring Method of Driver Fatigue Behavior Based on Machine Vision. In: Proceedings on intelligent Vehicles Symposium, pp. 110–113 (2003)
2. Miyakawa, T., Takano, H., Nakamura, K.: Development of Non-contact Real-time Blink Detection System for Doze Alarm. In: SICE, vol. 2, pp. 1626–1631 (2004)
3. Dong, W., Wu, X.: Fatigue Detection Based on the Distance of Eyelid. In: IEEE International Workshop on VLSI Design and Video Technology, pp. 365–368 (2005)
4. Wang, R., Guo, L., Tong, B., Jin, L.: Monitoring Mouth Movement for Driver Fatigue or Distraction with One Camera. In: The 7th International IEEE Conference on Intelligent Transportation Systems, pp. 314–319 (2004)
5. Wang, T., Shi, P.: Yawning Detection for Determining Driver Drowsiness. In: IEEE International Workshop on VLSI Design and Video Technology, pp. 373–376 (2005)
6. Bassili, J.: Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face. J. Journal of Personality and Social Psychology 37, 2049–2059 (1979)
7. Zhao, G., Pietikainen, M.: Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. J. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6), 915–928 (2007)

8. Yang, P., Liu, Q., Metaxas, D.N.: Boosting Coded Dynamic Features for Facial Action Units and Facial Expression Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2007)
9. Ji, Q., Lan, P., Looney, C.: A Probabilistic Framework for Modeling and Real-time Monitoring Human Fatigue. J. IEEE Transactions on Systems, Man and Cybernetics, Part A 36(5), 862–875 (2006)
10. Wiskott, L., Fellous, J.M., Kuiger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. J. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)
11. Freund, Y., Schapire, R.E.: A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. J. Journal of Computer and System Sciences 55(1), 119–139 (1997)