Abstract—Fatigue driving is a main cause for road traffic accidents. In case an alarm can be sent to drivers one second before the potential traffic accidents, 90% similar traffic accidents can be avoided. Facial expression recognition of drivers is an important component of fatigue driving pre-warning system. For the continuous video images obtained, facial features of eyes, nostril and mouth are firstly acquired via horizontal projection curve to construct the facial feature triangle. Then potential rectangular region of face tracking is generated via rigid constraints. Within the rectangular region of feature tracking, two-dimensional Gabor kernel function is chosen to construct 48 optimal filters, obtain 48 eigenvalues, and carry out training via SVM. Experiment result shows that the recognition rate of the algorithm can be as high as 94.8% for the most common neutral expression. The average recognition rate can be as high as 93.9% for the seven expressions, namely, happy, surprise, sad, angry, fear, and disgust.

Index Terms—Fatigue Driving; Gabor Filtering; Kernel Function; FTGSVM

I. INTRODUCTION

As per the traffic accident statistics and analysis of various countries of Europe and America, of all traffic accidents, 80%~90% are caused by human factors. According to the comprehensive investigation by Indiana University on the causes of traffic accidents, about 85% of them are related to drivers, those caused by vehicle reasons account for about 10%, while those caused by environmental factors only account for about 5%. Of all human factors of drivers that lead to road traffic accidents, fatigue driving is one of the main causes [1] [2]. Relevant data research shows that, if an alarm can be sent for drivers one second before the potential traffic accidents, 90% similar traffic accidents can be avoided [3] [4].

Drivers would always be inattentive and even dozing because of a long time of fatigue driving or tedious objectives they see. When these phenomena occur, the facial expression of drivers would also change obviously [5] [6].

In recent years, expression recognition has attracted more and more attention from scientific and technical workers, and has also obtained some achievements. However, as human face is a rigid body, and facial expression is greatly influenced by age, sex, race, hair, ornaments and illumination, etc., there are few accurate and effective recognition algorithms by far. During the research process on the problem, the expression feature extraction of the early phase is a very vital step, which can provide probability for the real-time processing of the system, and also provide guarantee and high recognition rate for follow-up classification and recognition.

Typical face expression recognition (FER) includes there phases, namely, face detection, expression feature extraction and expression classification. FER algorithm can be divided into two categories, e.g. multiple images sequence and single image. In the paper, the face expressions of multiple continuous images is mainly recognized.

Face feature extraction is mainly divided into two categories: based on geometrical features and based on apparent features. The method based on geometrical features takes advantage of the shape and position of face components (including: mouth, eyes and eyebrow, etc), and extracts features points from face components to constitute eigenvector, thus, representing the geometrical features of face. While the method based on apparent features can extract eigenvectors from the whole face and also specific facial regions, such as Gabor wavelet. The research practice of face recognition shows that: under the premise that it is difficult to obtain the three-dimensional face shapes, it is an appropriate choice to extract multi-direction and multi-scale Gabor wavelet from the image data.

Gabor wavelet transformation can extract features of facial expressions and extract expression eigenvectors via Gabor wavelet transformation with expression sub-regions, thus, constructing expression elastography. The result shows that Gabor wavelet can effectively extract features related to expressions, and can efficiently shield the impacts of illumination variations and individual feature differences [7].

The two feature extraction methods of geometrical features and Gabor wavelet are applied at the same time [8], 34 reference points are marked on the face, and then geometric coordinates of the 34 reference points and 612 Gabor wavelet coefficients are extracted separately as eigenvectors. The experiment result shows that Gabor wavelet can achieve better recognition results than geometrical features.

Zhang Z. et al. has compared various feature extraction methods on the basis of image sequence differences. For the classification of 12 sport units (6 for the upper half face and 6 for the lower half face), Gabor wavelet and
ICA can achieve the best recognition effect. The experiment result shows that it is very important to adopt local filtering, high-frequency and independent statistics for the classification of facial units [9].

As facial images have the facial features of “three stops and five eyes”, Zhu J.X. et al. has marked the sampling sites of 8 rows and 6 columns, extracted Gabor wavelet coefficients of these sampling sites to constitute eigenvectors, then, adopted Adaboost algorithm to reduce the dimensionality of eigenvectors, and finally used the classifier constituted of SVM and k-nearest neighbor for classifying, thus, good classification effect has been achieved [10].

The research object of the facial expression recognition algorithm for FTGSVM fatigue driving that is proposed in the paper is the continuous video facial images, which blends the two facial feature extraction methods based on geometrical features and based on apparent features perfectly. The face detection region has been positioned as per the facial feature triangle constituted of the five basic features of nose, eyebrow, mouth, eyes and ears. Firstly, horizontal projection curve function is adopted for face images to obtain the gray level integral projection drawing of images to be recognized, the feature triangle is applied for the facial features on the projection drawing to generate the triangular region of face recognition, an isosceles triangle is used for the front and a right triangle is used for the side. Then, Gabor wavelet transformation is used in the face regions obtained to extract multi-direction and multi-scale expression features, 48 optimal feature points are chosen to constitute expression eigenvectors, and input into SVM classifier to make training and recognition on the facial expression of drivers, and the recognition effect can reach as high as 94.8%. Finally, detailed analysis and comparison are made on FTGSVM and SVM [11] algorithms in the paper, which prove the validity of the algorithm in the paper.

The constructed expression recognition system is shown in Figure 1.

![Figure 1. Expression recognition system](image)

**II. PREPROCESSING OF VIDEO IMAGES**

An important information content of human face is “three divisions vertically and five divisions transversally”, which represents the position of five sense organs. Specifically, the eyes shall be in the middle of the head, and their length as well as the distance between each other shall be basically equivalent to the width of the nose. The face contour of human can be divided into three parts from the upper side to the lower side, that is, the part from hairline to open wiring, that from open wiring to nose baseline, and that from nose baseline to cheek baseline. The position diagram of five sense organs of human face is shown in Figure 2.

![Figure 2. Position diagram of five sense organs of human face](image)

**A. Horizontal Integral Projection Curve Function**

The preprocessing of video images includes the division of video image subregions and the normalization processing of video images. The former divides the subregions that are most closely related to face from the video images, while the latter includes gray balance and dimension normalization of images. The quality of image preprocessing directly influences the effect and calculated quantity of video image feature extraction.

The geometric normalization processing of fatigue driving firstly makes horizontal integral projection for face, sets up the judgment rules as per the position relation between various minimum values of the projection curve, and determines the horizontal position of eyes, mouth, nostril and eyebrow on the face and the vertical distance between various organs; determines the eye area and mouth area as per the horizontal position and the vertical distance of eyes of mouth on the face, divides these regions to construct the feature triangle and generates the facial processing region; thus, making further feature extraction.

The author of geometric normalization processing method made detailed statement in document [11]. The horizontal integral projection curve function $PH(y)$ of the region can be represented into:

$$PH(y) = \frac{1}{x_2 - x_1} \sum_{i}^{n} I(x, y), y \in (y_1, y_2)$$  \hspace{1cm} (1)$$

Fig. 3 is the face gray level integral project picture obtained from horizontal projection curve function.

![Figure 3. Face gray level integral projection picture](image)

**B. Face Region Tracking in the Video Image Based on Feature Triangle**

For the front face images, the center of the two eyes and the center of the mouth area constitute the isosceles triangle $ijk$ as shown in Fig. 3. We adopt the strategy from the left to the right and from the bottom to the top to...
regulate $i < j < k$ and the isosceles triangle must satisfy the following rigid constraint conditions:

1. $\text{abs}(D(i,j) - D(j,k)) < 0.25 \max(D(i,j), D(j,k))$;
2. $\text{abs}(D(i,j) - D(i,k)) < 0.25 \max(D(i,j), D(i,k))$.

Of which, abs represents the absolute value, $D(i,j)$ represents the distance between the left eye to the middle of the mouth; $D(i,k)$ represents the distance between two eyes; $D(j,k)$ represents the distance between the middle of the mouth to the right eye.

![Isosceles triangle $ijk$ constructed from the front face](image1)

Figure 4. Isosceles triangle $ijk$ constructed from the front face

In fact, the actual region of face includes eyebrow, two eyes, mouth and partial area under the mouth. As per the rectangle tracking region generated from the isosceles triangle constructed before, we only need to calculate the coordinates of four points. Suppose the coordinates of three points of the isosceles triangle $ijk$ are $(X_i, Y_i), (X_j, Y_j), (X_k, Y_k)$, the rectangle region coordinate calculated as per the isosceles triangle in Fig.5 satisfies:

$$
\begin{align*}
X_1 &= X_4 = X_i - 1/3D(i,k) \\
X_2 &= X_3 = X_i + 1/3D(i,k) \\
Y_1 &= Y_2 = Y_i + 1/3D(i,k) \\
Y_3 &= Y_4 = Y_i - 1/3D(i,k)
\end{align*}
$$

(2)

Side image can be divided into left side image and right side image. First of all, the right side image is described in detail. Right triangle $ijk(30^\circ, 60^\circ, 90^\circ)$ constructed by the side image, in the right triangle, $i$ is the right ear hole, $j$ is the right eye center, $k$ is the mouth region. Due to the error of the image itself, error of 25% is to be tolerated. We adopt the strategy from left to right, from bottom to top to regulate $i < j < k$; and Isosceles Triangle $ijk$ must meet rigid constraint conditions as follows:

3. $\text{abs}(D(i,k) - D(j,k)) < 0.60D(i,k)$ and $\text{abs}(D(i,k) - D(j,k)) > 0.40D(i,k)$.

In accordance with the error to be tolerated, 75%–125% of the $D(i,j)$ is to form an arc, this arc is

![Rectangle region coordinate calculated as per the isosceles triangle](image2)

Figure 5. Rectangle region coordinate calculated as per the isosceles triangle

![Right triangle $ijk(30^\circ, 60^\circ, 90^\circ)$ construct from the side image](image3)

Figure 6. Right triangle $ijk(30^\circ, 60^\circ, 90^\circ)$ construct from the side image

According to the rectangular tracking area generated by Right Triangle $ijk$ constructed above, it only needs to calculate the coordinates of four points. Also suppose the coordinates of three points of Right Triangle $ijk$ to be $(X_i', Y_i'), (X_j', Y_j'), (X_k', Y_k')$, as shown in Fig. 7. Rectangle coordinates calculated according to the right triangle should satisfy the following conditions:

$$
\begin{align*}
X_1 &= X_4 = X_i' - 1/6D(i,j) \\
X_2 &= X_3 = X_i' + 1/2D(i,j) \\
Y_1 &= Y_2 = Y_i' + 1/4D(i,j) \\
Y_3 &= Y_4 = Y_i' - 1.0D(i,j)
\end{align*}
$$

(3)

![Coordinates of potential rectangular tracking area calculated according to the right triangle](image4)

Figure 7. Coordinates of potential rectangular tracking area calculated according to the right triangle

The structure of the left side image is similar to that of the right side image, in order to save space, this paper will not discuss it in detail.

The right eye and mouth area $j$ in the positive image are given, according to the rigid constraint conditions satisfied by the isosceles triangle

4. $\text{abs}(D(i,k) - D(i,j)) > 0.13D(i,k)$ and $\text{abs}(D(i,k) - D(i,j)) < 0.19D(i,k)$;
5. $\text{abs}(D(i,j) - D(j,k)) < 0.44D(i,k)$ and $\text{abs}(D(i,j) - D(j,k)) > 0.29D(i,k)$.

in which, abs represents the absolute value, $D(i,j)$ represents the distance between right eye and mouth center, $D(i,k)$ represents the distance between right ear hole and the mouth region, $D(j,k)$ represents the distance between right eye and mouth center, the coefficients are calculated according to the triangle and errors.
concentrated in the right side area of Straight Line $ij$, $k$ is found to construct isosceles triangle, the face potential detection region is rapidly generated. The Rapid detection of feature triangle in the side image is similar to that of the above.

The influence of the background on the fast detection based on characteristic triangle is greatly reduced.

If the Euclidean distance between the centers of block $i$ (right eye) and block $j$ (mouth) is already known, then block center $k$ (left eye) should locate in the area of 75%–125% of Euclidean distance between the centers of block $i$ (right eye) and block $j$ (mouth), which will form a circle.

The search area is only limited in the dark area instead of the whole area of the image as shown in Fig.9. In this way, the triangle based segmentation process can reduce the background part of a cluttered image up to 97%. This process significantly speeds up the subsequent face detection procedure because only 3–9% regions of the original image are left for further processing.

After dividing the subregions related to face from video image, it is necessary to make gray level equalization and dimension normalization processing for the sub-images. The purpose of gray equalization is to eliminate the influences of illumination variation, and eliminate the skin color differences of different races, although the expression features that Gabor wavelet transformation extracts is not sensitive to the illumination variation, the gray level equalization processing can optimize the feature extraction result. And adjust the image mean value and variance of the image via amending the gray level histogram of video sub-images, to complete the equalization processing of images.

III. FEATURE EXTRACTION OF GABOR FILTERING

Two-dimensional Gabor wavelet transformation is an important tool for the time-frequency domain to make signal analysis and processing, and its transformation coefficient has good visual features and biology background, thus, it is widely used in image processing, mode recognition, etc.

Gabor filter of different parameters can capture local feature information in images, which is corresponding to different spatial frequencies, spatial positions and directions. Because of the features of Gabor filter, it is not sensitive to changes to the brightness and face postures, thus, Gabor filter is widely used for image coding, handwritten numeral recognition, face recognition and edge detection, etc.

Then, make two-dimensional Gabor wavelet transformation in the rectangle feature region obtained via the features on the above, to obtain the expression features of the face region of drivers.

Two-dimensional Gabor wavelet kernal function is:

$$\psi_j(k, x) = \frac{k^2}{\sigma^2} \exp(-\frac{k^2}{2\sigma^2})[\exp(k^2 \cdot x) - \exp(-\frac{\sigma^2}{2})]$$ (4)

Of which, $k^2$ ensures that filters of different frequency bandwidths have almost the same energy, $\exp(-\frac{\sigma^2}{2})$ is to offset the DC shunt of the images, thus, the filter will not be sensitive to the overall lamination. Its advantage is that it can maintain the relation of spatial relation while allowing to describe the spatial frequency structure.

$$k = 2^{\frac{\phi}{\pi}} \cdot \mu \cdot \frac{\pi}{4}$$

$\phi$ constitutes different wavelets for sides with different values, and 4 changes to size and directions are adopted in the paper.

$$k = 2^{\frac{\phi}{\pi}} \cdot \mu \cdot \frac{\pi}{4}$$ (v = 1, 2, 3, 4)

Six directions $\phi: 0, 6/8\pi, 12/8\pi$ with interval of $\pi/8$. $\sigma$ is the length of the filter, take $\sigma = \pi$, suppose it is 1 time frequency. Input the image $I(x, y)$ and the wavelet for convolution, then,

$$g(k, x) = \iint I(x, y)\psi_j(k, x)dx dy$$ (6)

Of which, $g(k, x)$ is the amplitude. Thus, the number of Gabor filter is 48, which constitute a group of optimal filters that represent the target features. These filters constitute wavelet subspace, project the images to the wavelet subspace to get the coefficient of the wavelet, and extract the mean value and variance to represent the statistical features of the facial expression images of drivers.
IV. STRUCTURE OF SVM CLASSIFIER

SVM is a new mode recognition method that is developed on the basis of statistical learning theory, which has a lot of special advantages in solving small sample, and non-linear machine high-dimensional mode recognition problem, and can be promoted and applied in other machine learning problems [3], such as function fitting.

Suppose the linear dividable sample set \((x_i, y_i), i = 1, ..., n, x \in R^p, y \in \{+1, -1\}\) is the category symbol, the general form of linear discriminant function of N-dimensional space is \(g(x) = \alpha \cdot x + b\), the classification side equation is \(\omega \cdot x + b = 0\). The classification side is the optimal hyperplane, and the heterogeneous vector that is the closest to the plane is the so-called Support Vector, the distance between the Support Vector and the hyperplane is at the maximum (e.g. edge maximization), and one group of Support Vectors can only determine one hyperplane. The problem for looking for hyperplane is changed into the problem to solve the following quadratic programming:

\[
\Phi(\alpha) = \frac{1}{2} (\alpha \cdot \alpha)
\]

(7)

The constraint condition is the in equation:

\[
y_i[(\alpha \cdot x_i) + b] \geq 1 \quad i = 1, 2, ..., n
\]

(8)

Under the linear condition, the optimal solution to Formula (9) is the saddle point of the following Lagrange function:

\[
L(\alpha, a, b) = \frac{1}{2} \|x\|_2^2 - \sum_{i=1}^{n} \alpha_i [y_i (\alpha \cdot x_i) + b - 1]
\]

(9)

At the time, the optimal hyperplane is the linear combination of vectors in the training set:

\[
\omega = \sum_{i=1}^{n} \alpha_i y_i x_i
\]

(10)

The support vector that has nonzero coefficient \(\alpha_i\), in the expansion equation of \(\omega\) is the vector which makes formula (10) workable. The classification plane of the quadratic programming problem is:

\[
f(x) = \text{sgn} \left( \sum_{i=1}^{n} \alpha_i y_i x_i \cdot x + b \right)
\]

(11)

For the linear impartibility, SVM introduces the slack variable and penalty factor, thus, the target function is changed into:

\[
\Phi(\alpha, \xi) = \frac{1}{2} (\alpha \cdot \alpha) + C \left( \sum_{i=1}^{n} \xi_i \right)
\]

(12)

SVM is similar to the neural network in the form. Input the linear combination of the intermediate node, and every intermediate node is corresponding to a support vector. There is only dot product calculation between vectors, the complex calculation in the high-dimensional feature space can be avoided in case kernal function is adopted, and different kernal functions shall form different algorithms (that is, different VSMs). Currently, major kernel functions include:

- **Polynomial kernal function:** \(k(x, x_i) = (x \cdot x_i + 1)^d\), obtain d-order polynomial classifier:

\[
k(x, y) = \exp\left[- \frac{\|x - y\|^2}{\sigma^2}\right],
\]

the difference between the classifier obtained and traditional RBF is that the center of every basis function is corresponding to one support vector, they and the output weight are determined automatically as per the algorithm.

- **Sigmoid kernal function:** \(k(x, x_i) = \tanh(\gamma (x \cdot x_i) + c)\)

While people’s analysis on facial expression can be traced back to the 20th century. Ekman and Friesen [12] put forward 6 basic facial expressions as per the distribution and movement of facial muscle: happy, surprise, sad, angry, fear, disgust, as shown in Fig.10.

![Figure 10. Six categories of facial expression](image)

During the driving process of drivers, the frequencies of these expressions are not always the same, for example, drivers seldom drive under the emotion of sad and angry, in most cases, it is the neutral expression indicated on the above.

During the research process, we choose RBF and the training algorithm that adopts the Rowley [13] algorithm of VSM, and other two kernal functions will be continued in the follow-up research.

V. EXPERIMENT RESULT AND ANALYSIS

Seven expressions (including the neutral expression) are considered in the experiment, namely, neutral, happy, sad, angry, fear, surprise and disgust. Facial expression data comes from JAFFE (Japanese Female Facial Expression) data base [13] of Japanese Kyushu University, Yale data base and self-built expression data base (only for test), JAFFE data base contains 213 images, and seven expressions; Yale data base contains 165 images, the expressions of 15 people, and 11 images every people; the self-built expression data base contains 200 images. Thus, there are totally 578 expression images.

The experiment is divided into two groups, one group (278 images) is used for network training, another group is used for test, every expression image obtains the structural feature triangles of the eyes, nostril and mouth, including the isosceles triangle and direct triangle, then constructs 48 optimal filters as per the two-dimensional Gabor kernal function, and obtains 48 facial feature points within the feature tracking rectangle region to
obtain the eigenvalue. Every region goes through Gabor wavelet transformation at 4 scales and towards 6 directions as the eigenvectors of the facial image, and then is input to SVM training or recognition.

All expression images go through Gabor wavelet feature extraction before training and testing, expression classification testing result is shown in Table 1, in the test set, there are 45 images for all other six expressions, except that there are 30 images for the neutral expression. Of the 300 test sets, 277 can be recognized correctly, 23 are recognized wrongly, and the average correct recognition rate is 93.9%.

Figure 11(b) is the result which is classified by FTGSVM. There are some of datas which are regarded as noise points wrongly in Figure 11(a), and the classification and recognition are quite clear in Figure 11(b).

In the process of classification and recognition of facial expression, this paper uses the FTGSVM method which is combined with feature triangle region. Under the same condition, SVM [11] method is used to test the classification and recognition of facial expression. The comparative results which are used different methods are expressed in Table 2, and the result is showed that the method of this paper raised gets the good result.

Table 1

<table>
<thead>
<tr>
<th>Expression category</th>
<th>Happy</th>
<th>Surprise</th>
<th>Sad</th>
<th>Angry</th>
<th>Fear</th>
<th>Disgust</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition rate</td>
<td>95.6%</td>
<td>95.6%</td>
<td>93.1%</td>
<td>93.3%</td>
<td>89.9%</td>
<td>94.1%</td>
<td>94.8%</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Recognition Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVC [11]</td>
<td>92.6</td>
</tr>
<tr>
<td>FTGSVM</td>
<td>93.9</td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS

The paper constructs 48 eigenvectors of the optimal filters via the two-dimensional Gabor kernel function in the feature tracking rectangle region, solves the facial expression recognition problems via the strong generalization performance of SVM, and makes experiments for the training algorithm and RBF kernel function by means of the Rowley [13] algorithm of SVM, and the experiment result shows that the method has good performances.

The facial recognition problem of drivers in the intelligent monitoring system includes 3D modeling, face dynamic tracking, expression synthesis, and face recognition. Two-dimensional recognition has the advantages of being simple and rapid, but is restricted by the angle of the camera, while 3D method can accurately express the physical space in the unrestricted complex human movement judgment. However, 3D tracking, three-dimensional facial expression synthesis and 3D recognition are only at the beginning stage, and there are a lot of things that need to be researched further.

ACKNOWLEDGMENTS

This research is partially supported by the Science and Technology Planning Project of Hunan Province (No. 2012GK3096) and Scientific Research Fund of Hunan Provincial Education Department (No. 12C1158 NO. 13C613 and NO. 13C614).

REFERENCES

Machines with Binary Tree Architecture for Multi-Class
Support Vector data for sentence boundary detection in speech,
Extended Set of Haar-like Features for Rapid Object
display.
Z. Zhu, Q. Ji. Eye and gaze tracking for interactive graphic
recognition,
W. Gao, H. Jin. Analysis and Recognition of Facial
image.

Zhu Jianxiang, Su guangda, Li yingchun. Facial Expression Features
Extraction Based on Gabor Wavelet Transformation.


Qiufen Yang received her B.Eng. and M.Eng. degrees degree from Nanjing University, Jiangsu, China and National Defense Science and Technology University, ChangSha, Hunan, China in 1996 and 2004, respectively. She is currently pursuing her Ph.D degree in the School of software, Central South University, Changsha, Hunan, China. She is now an associate professor at Hunan Radio & TV University of science and technology Her current research interests are in the areas of data fusion and computer vision.

Canjun Li received the degree of the B.Eng.(Industrial Automation) and the M.Eng. (Control Science and Engineering) from Zhengzhou University, Henan, China and Central South University, Changsha, China in 2000 and 2007, respectively. She is a visiting scholar at Central South University, Changsha, Hunan, China. Her current research interests are in the areas of modeling and control of complex system.

Zhenjun Li received the degree of the B.Eng. ( Electronic Engineering) and the M.Eng. (Electronic Science and technology) from Xiangtan University, Hunan, China and Hunan University, Changsha, China in1999 and 2009, respectively. His current research interests are in the areas of the key technology of mobile Internet.