ISSN: 2664-9640

Research on non-interactive live face detection method

Yong Deng, Zhao Yang

School of Southwest Petroleum University, Chengdu 610500, China

Abstract

In this paper, the fusion color space is used instead of a single color space, and it is proved that the color information of the fusion color space can improve the accuracy of the face live detection method. The improvement of the LBP feature operator is carried out. After the feature extraction and classification of the fused color and texture features are carried out by using the support vector machine (SVM), the face living body detection can be effectively realized. The Farneback optical flow algorithm is used for calculation, and then processed through the convolutional neural network, and the two-classification of real and fake faces can be realized. This paper designs and conducts relevant experiments to verify and analyze, and has good performance in two public standard databases Replay-Attack and CASIA. At the same time, the method in this paper shows a good accuracy rate on the self-built data set, which proves the feasibility of the method in this paper in the actual situation.

Keywords

live face detection; local binary; optical flow method; convolutional neural network.

1. Introduction

1.1. Live Face Detection Database

To evaluate the accuracy, running speed and other performance indicators of a live face detection method, it is necessary to use a complete and reliable live face detection database. At present, the implementation of identity authentication technology can be roughly divided into the following three ways: (1) based on password information; (2) based on biometrics; (3) based on trusted objects. Among them, biometric authentication technology is becoming more and more popular.

Table 1 Biometric

Feature	Universality	Uniqueness	Stability	Collectability	Performance	Safety
Face	high	high	medium	high	low	low
Fingerprint	medium	high	high	medium	high	low
Handprint	medium	medium	medium	high	medium	medium
Iris	high	high	high	medium	high	high
Retina	high	high	medium	low	high	high
Sign	low	low	low	high	low	low
Timbre	medium	low	low	medium	low	low

It is easy to see from Table 1 that compared with other biological features human faces have many advantages, such as high collectability, strong universality, and good stability. Relying on

ISSN: 2664-9640

the widespread use of these biometric identification technologies, the work efficiency of all walks of life has been promoted, and the production cost has been greatly reduced.

In order to evaluate the accuracy, running speed and other performance indicators of a live face detection method, it is necessary to use a complete and reliable live face detection database.

The basic information of three widely used face live detection databases is given in Table 2:

Table 2 Database information statistics

Name database	Type of data	Number	Data
NUAA	image	15	12641
CASIA	video	50	600
Replay-Attack	video	50	1200

The Reply-Attack database is similar to the CASIA[2] database, and the database samples are very rich. However, the Reply-Attack database better restores the actual application scene, because it uses a variety of ways to reproduce face images. There are also two attack methods during simultaneous collection, which are divided into handheld device shooting and fixed device shooting. Although the number of people participating in the two methods of collection is different, each person shoots multiple times during the shooting. Therefore, the face image sample they are very rich and can simulate the actual situation to a great extent.

1.2. AdaBoost Face Detection Algorithm

The live face detection system consists of two parts, namely face detection and live body detection. Therefore, before performing live face detection, the collected image must be processed to determine whether the current image contains face information. When face information is detected in the image, the image needs to be preprocessed. Locate the face position and eliminate background interference. A face detection algorithm is required. Generally speaking, face detection algorithms are based on the feature points of the face to locate the face, usually in the form of a rectangular frame to frame the face. Although everyone's face structure is the same, that is, they are composed of facial features, but different people will have different expressions, and each person's degree of fatness and thinness is also different. In addition, the light in different environments also has. However, these factors will affect the accuracy of the face detection algorithm. In this paper, the current mainstream face detection algorithm based on the AdaBoost[4] framework with high accuracy will be used to detect and locate the face in the image, so that it can be used in the following experiments.

2. Live Face Detection Based On Color Texture

2.1. Color Space Contrast

After the channels of the RGB, YCbCr and HSV color spaces are separated, the texture feature extraction can be performed in the three spaces respectively. The specific texture feature types are: LBP feature, WLBP feature, CoALBP feature and LDP feature. After the features of SVM are trained and classified, the detection accuracy of the methods in different color spaces can be compared, so as to determine which color space is more suitable for face living detection. The specific experimental results are as follows:

ISSN: 2664-9640

Table 3 Replay-Attack database experimental results EER (%)

Color texture features	Gray	RGB	HSV	YCbCr
LBP	20.0	7.9	8.2	3.7
WLBP	17.2	9.8	6.8	5.4
CoALBP	13.5	6.7	4.1	2.7
LDP	25.1	10.1	7.9	6.5

Table 4 CASIA database experimental results EER (%)

Color texture features	Gray	RGB	HSV	YCbCr
LBP	23.1	21.6	14.1	16.7
WLBP	19.2	13.9	11.2	9.6
CoALBP	15.1	11.6	5.9	10.1
LDP	23.9	14.7	8.1	12.8

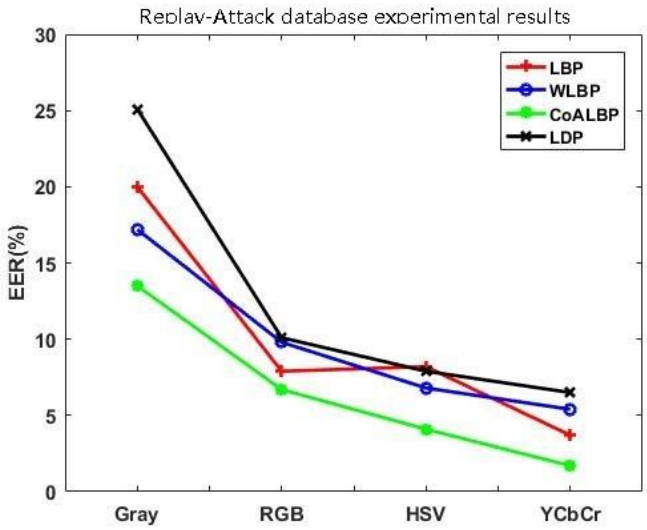

Fig. 1 Replay-Attack database experiment result

ISSN: 2664-9640

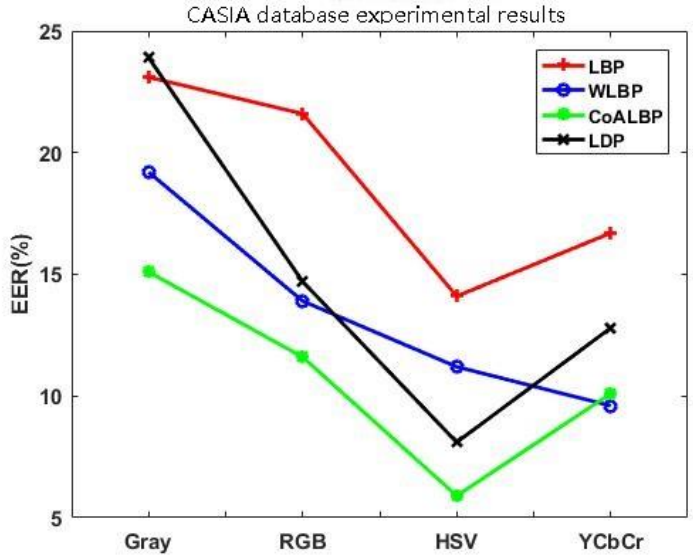

Fig. 2 CASIA database experiment results

As can be seen from Fig. 1 and Fig. 2 and Table 3 and Table 4, the experimental results of different methods in different color spaces are very different, and the four feature extraction operators are in the gray color space has the worst performance. In the Replay-Attack database, the detection accuracy of the Gray color space is the best about 13.5%, while in the HSV color space and the YCbCr color space, the worst detection results are only about 7.9%. In the CASIA database, the Gray color space and the RGB color space perform poorly. The best detection accuracy in the Gray color space is about 23.9%, and the best detection accuracy in the RGB color space is about 21.6%, while in the HSV color space and In the YCbCr color space, the worst detection result is only about 14.1%. For the four texture features, in the Replay-attack database, the CoALBP texture feature and the WLBP texture feature have the best results, and both texture features achieve the best detection results in the YCbCr color space, 1.7% and 5.4%, respectively about. At the same time, although the LDP feature has the worst performance in the Gray color space by about 25.1%, in the other three color spaces, except for the YCbCr color space, which has a large gap with the CoALBP feature, the difference is about 4.8%. In the RGB and HSV color spaces, it is not far from the WLBP feature and CoALBP feature, and it is also an efficient operator, and its worst effect is about 10.1% in the RGB color space. In the CASIA database, the best experimental result is the CoALBP feature, which has a detection effect of about 5.9% in the HSV color space. For WLBP and LDP operators, the experimental results on RGB, HSV and YCbCr color spaces are similar. But the performance of LBP features is the worst in all four color spaces. The above experiments show that in the feature extraction of face live detection, HSV and YCbCr have greater advantages than RGB, and are more suitable as color spaces for face live detection and extraction. At the same time, in these two color spaces, WLBP features, LDP features, CoALBP features have good integration with HSV color space and YCbCr color space. Compared with LBP features, their effects can provide more effective information. Effectively improve the accuracy of the algorithm.

2.2. Fusion color texture feature comparison

Conduct texture feature fusion experiments, namely LBP+WLBP, LBP+CoALBP, LBP+LDP, WLBP+CoALBP, WLBP+LDP and CoALBP+LDP. An optimal fusion texture feature is selected by comparing the experimental results. This experiment is still carried out on the Replay-Attack and CASIA databases, and the experimental results are shown below.

ISSN: 2664-9640

Table 5 Replay-attack database experimental results EER (%)

Color texture features	HSV	YCbCr	HSV+YCbCr
LBP+WLBP	7.2	6.1	5.3
LBP+CoALBP	5.5	2.3	2.9
LBP+LDP	10.4	5.7	3.5
WLBP+CoALBP	4.1	2.1	1.8
WLBP+LDP	5.2	4.3	4.1
CoALBP+LDP	4.0	2.7	2.1

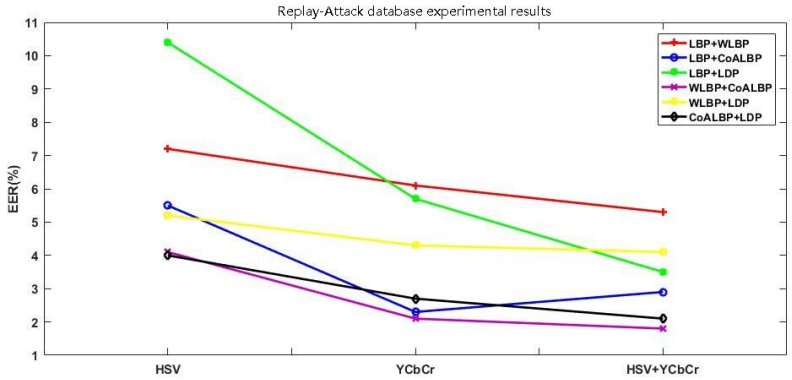

Fig. 3 Replay-Attack database experiment result

As can be seen from Table 5 and Fig. 3, the WLBP+CoALBP texture feature combination and the CoALBP+LDP texture feature combination have the best effect. The detection accuracy of WLBP+CoALBP texture feature combination in HSV color space, YCbCr color space and HSV+YCbCr combined color space is about 4.1%, 2.1% and 1.8%, respectively. The detection accuracy of CoALBP+LDP texture feature combination in HSV color space, YCbCr color space and HSV+YCbCr combined color space is about 4.0%, 2.7% and 2.1%, respectively. In the selection of color space, for all texture feature combinations, the overall performance in fusion color space is better than that of single color space. The detection accuracy is about 10.4% and 5.7%, respectively, but in the HSV+YCbCr combined color space, the detection accuracy becomes 3.5%, an increase of 6.9% and 2.2%, respectively, which proves that the information of the fusion color space is more effective.

Table 6 CASIA database experimental results EER (%)

Color texture features	HSV	YCbCr	HSV+YCbCr
LBP+WLBP	10.7	9.9	6.3
LBP+CoALBP	11.3	9.1	5.7
LBP+LDP	12.6	10.3	7.5
WLBP+CoALBP	3.9	7.2	2.8
WLBP+LDP	11.4	10.6	6.8
CoALBP+LDP	4.2	8.7	3.5

ISSN: 2664-9640

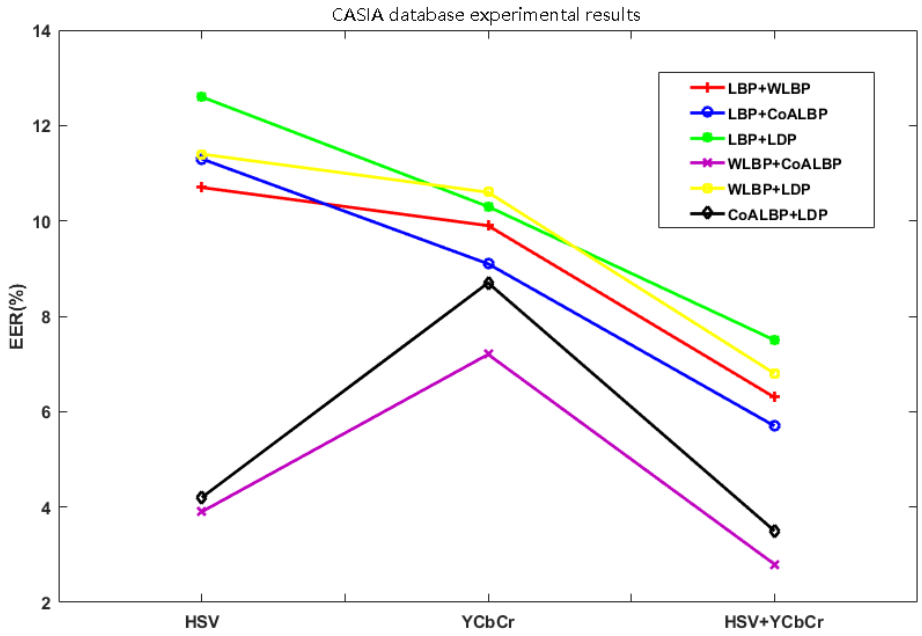

Fig. 4 CASIA database experiment results

As can be seen from Table 6 and Fig. 4, in the experimental results of the CASIA database, the texture feature combination of WLBP+CoALBP has the best performance, and it has the best performance in the three color spaces, which are about 3.9 %, 7.2% and 2.8%. After comparing this result with the performance of the rest of the texture feature combinations in the three color spaces, it can be found that the combination of WLBP+CoALBP texture features has the worst effect in the HSV color space than the LBP+LDP The texture feature combination is improved by 8.7%, the WLBP+LDP combination with the worst effect is improved by 3.4% in the YCbCr color space, and the LBP+LDP combination with the worst effect is improved by 4.7% in the HSV+YCbCr combination color space. At the same time, it can be seen that among the three color spaces, the HSV+YCbCr combination color space has the best combination with each group of texture features. In all experimental data, its detection effect on HSV+YCbCr achieves the best detection effect of about 2.8 %. At the same time, compared with the experimental results of the combined color space and the single color space, the worst improvement effect is 0.7%, and the best improvement effect reaches 5.6%, which shows the effectiveness of the fusion color texture feature method.

At present, most of the live face detection algorithms are carried out on the CASIA and Replay-Attack databases. In order to reflect the effectiveness of the method in this paper, the method in this paper is compared with the existing live face detection methods. The comparison results are shown in Table 7, this paper adopts the color texture feature method combining the fusion color space of HSV+YCbCr and the fusion texture feature of WLBP+CoALBP.

Table 7 Comparison with previous methods EER (%)

Method	Replay-Attack	CASIA
Motion	-	26.6
LBP	20.0	23.1
IDA	7.4	12.9
LBP-TOP	7.9	7.6
DMD	5.3	21.8
The method of this paper	1.8	2.8

ISSN: 2664-9640

As can be seen from Table 7, the method proposed in this paper has different degrees of improvement in the detection results of the CASIA and Replay-Attack databases compared with the previous methods. In the Replay-Attack data, this paper proposes Compared with the DMD method, the accuracy of the proposed method is improved by about 3.5%, and in the CASIA database, the method proposed in this paper also has a 3.6% improvement compared with the FLSB algorithm that performed better in the past.

2.3. SVM

SVM[5][6] is a kind of supervised learning, and the training process is the process of continuously finding the optimal objective function through multiple sets of data. Support vector machines can be divided into the following three types, from simple to complex:

(1) Linearly separable support vector machine;

(2) Linear support vector machine;

(3) Nonlinear support vector machine.

Since there are many kinds of actual situations, it is necessary to make a reasonable choice according to different sample sets. The commonly used kernel functions in support vector machines are shown in Table 8 below.

Table 8 Kernel functions commonly used in SVM

Name	Expression	Parameter
Linear Kernel	$k(x_i, x_j) = x_i^T x_j$	
polynomial kernel	$k(x_i, x_j) = (x_i^T x_j)^d$	d ≥1 bit polynomial
Gaussian kernel	$k(x_i, x_j) = \exp(-\frac{\|\|x_i - x_j\|\|^2}{2\sigma^2})$	$\sigma > 0$ is the bandwidth of the Gaussian kernel
Laplace nucleus	$k(x_i, x_j) = \exp(-\frac{\|\|x_i - x_j\|\|}{\sigma})$	$\sigma > 0$
Sigmoid nucleus	$k(x_i, x_j) = \tanh(\beta x_i^T + \theta)$	$\beta > 0, \theta < 0$,tanhis the hyperbolic tangent function

3. Live Face Detection Based On Optical Flow Information

3.1. The Basic Principle And Selection Of Optical Flow Method

3.1.1. Optical flow field

In the three-dimensional space, the concept of motion field is introduced in order to better describe the motion state when the object is moving. In the two-dimensional plane space, images of consecutive frames are often used to represent the motion of the object. These images have different Grayscale distribution difference [. Similarly, in order to describe the motion of objects on a two-dimensional plane, an optical flow field is introduced. The optical flow field is the projection of the motion of an object on a two-dimensional plane. When the object moves, its image on the two-dimensional plane will also follow the movement, then the pixels in the image will also follow the movement. In such a continuous change, the gray value of the pixel will change accordingly. The process of this change is It can be represented by the optical flow field. When the motion interval of the pixel points is very small, the optical flow field can be regarded as the instantaneous velocity field of the pixel points in motion, which contains the vector motion information of all the pixel points in an image.

ISSN: 2664-9640

3.1.2. Farneback dense optical flow algorithm

The Farneback[7] dense optical flow method is an optical flow field formed by analyzing the displacement of all pixels in the image to obtain the optical flow trajectory of all pixels. The prerequisites for Farneback's dense optical flow method are also constant brightness, small motion, and spatial consistency. When the object moves, its representation (picture) on the two-dimensional plane will also follow the movement, then the pixels in the picture will also follow the movement. In such a continuous picture, the gray value of the pixel will change. The process of this change It can be represented by the optical flow field. When only the middle part of the image is studied, the optical flow research method at this time is called the sparse optical flow method. When all the pixels in the image are studied, the research method at this time is called the dense optical flow method.

The principle of the Farneback dense optical flow method is a polynomial expansion, which approximates the neighboring pixels of two adjacent frames of images, and solves the pixel displacement field by studying the change process of the polynomial.

3.2. Convolutional Neural Network

The network structure of the convolutional neural network[8][9] mainly consists of three parts: the convolutional layer computing layer, the pooling layer and the fully connected layer.

The mathematical calculation formula of the convolution kernel is shown in the following formula (1):

$$\begin{aligned} Z^{l+1}(i,j) &= [Z^l \otimes \omega^{l+1}](i,j) + b \\ &= \sum_{k=1}^{k_l} \sum_{x=1}^{f} \sum_{y=1}^{f} \left[z_k^l(s_0 i + x, s_0 j + y) \omega_k^{l+1}(x,y) \right] + b \end{aligned} \tag{1}$$

In formula (1), l is the number of layers, Z^l is the input layer of the l layer, Z^{l+1} is the output layer, which is the feature layer, b represents the bias, and $k_l$is the lth the number of feature map channels of the layer, f, $s_0$and p are the convolution kernel parameters, which are the convolution size, convolution stride and padding dimension, respectively, $(x,y) \in (0,1,\cdots,L_{l+1})$, $L_{l+1}$is the size of the output feature layer Z^{l+1}, the calculation formula is as follows:

$$L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1 \tag{2}$$

The structure of the convolutional neural network designed in this paper is shown in Fig. 5.

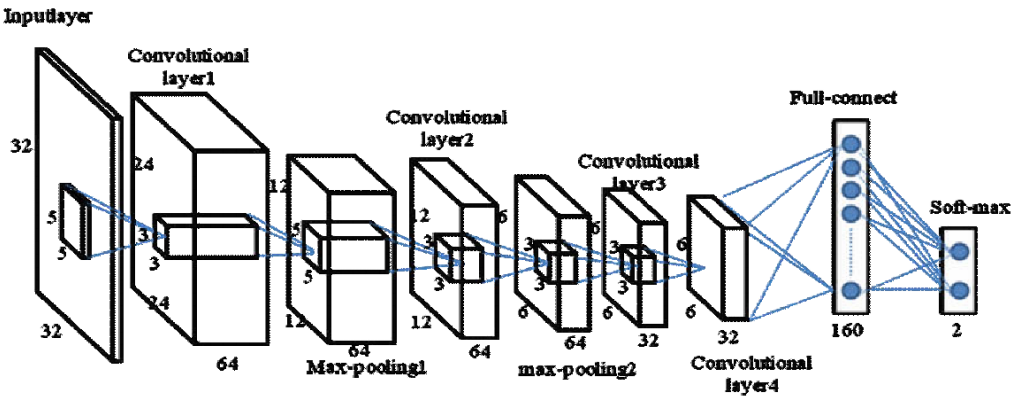

Fig. 5 Convolutional Neural Network Structure

The computation of the forward-propagating convolutional layer is:

ISSN: 2664-9640

$$y_j = \max\{0, \sum_i W_{i,j} * x_i + b_j\} \quad (3)$$

In formula (3), $x_i$represents the *ith* input, y_i represents the *jth* output, and $W_{i,j}$ represents the convolution kernel between the *ith* input and the *jth* output, that is, the weight parameter, the symbol * Represents the convolution operation. b_j is the bias of the *jth* output. The activation function used in the hidden layer is the ReLU function, and its expression is shown in formula (4), which is similar to the general functions $f(x) = \tanh(x)$ and $(x) = (1 + e^{-x})^{-1}$, the ReLU function has faster convergence speed and better fitting ability.

$$f(x) = \max(0, x) \quad (4)$$

The calculation formula of the Max-Pooling layer is as follows:

$$y_j = \max_{k \in D}\{x_i^K\} \quad (5)$$

In formula (5), *D* represents the *ith* input non-overlapping local region. y_j is the maximum value in *D*.

The fully connected layer and the fourth convolutional layer are fully connected, and the calculation of the fully connected layer is expressed as:

$$y_j = \max\{0, \sum_i x_i \times W_{i,j} + b_j\} \quad (6)$$

In formula (6), x_i represents the *ith* output of the fourth convolutional layer, and y_j represents the *jth* output of the fully connected layer.

The Soft-max layer outputs *n* values, which represent the probability distribution of *n* classes. Because live face detection is a binary classification problem, the output of the convolutional neural network is set to two categories, and the calculation of the Soft-max layer is expressed as：

$$y_j = \frac{e^{y_j'}}{\sum_j e^{y_j'}} \quad (7)$$

$$y_j' = \sum_{j=1}^{n} x_i \times W_{i,j} + b_j \quad (8)$$

In the formula (8), x_i represents the 160-dimensional output of the fully connected layer, and $y_j$represents the output of the soft-max layer. This article is set to two categories, so *j*=2.

4. Summary

It is confirmed that the fusion color space is superior to a single color space in color feature extraction; a fusion feature combining CoALBP features and WLBP features is proposed, and combined with the color information of HSV color space and YCbCr color space, and then extracted. A fused color texture feature combining color information and texture feature information. The Farneback optical flow algorithm is used to extract the pixels of each frame of the video stream, and the optical flow map is obtained by calculating the displacement of the pixels in the picture. Finally, the convolutional neural network is used to process the optical flow map to obtain the classification result. It has better performance on the same database, and the method proposed in this paper also has a certain accuracy rate on the self-built data set, which proves the feasibility of the method in the actual situation.

ISSN: 2664-9640

References

[1]. Kollreider K, Fronthaler H. Real-time face detection and motion analysis with application in ”liveness” assessment[J]. IEEE Transactions on information forensics and security, 2007, 2(3):548-558.

[2]. Zhang Z, Yan J, Liu S. A face anti-spoofing database with diverse attacks[C]. Iapr international conference on biometrics, IEEE, 2012:26-31.

[3]. Chingovska I, Anjos A, Marcel S. On the effectiveness of local binary patterns in face anti-spoofing[C]. Biometrics special interest group, IEEE, 2012:1-7.

[4]. Paul Viola, Michael J Jones. Robust real-time face detection[J]. International journal of computer vision, 2004, 57(2):137-154..

[5]. Adankon M, Cheriet M. Support vector machine[J]. Computer science, 2002, 1(4):1-28.

[6]. Soualhi Abdenour, Medjaher Kamal, Zerhouni Noureddine. Bearing health monitoring based on hilbert–huang transform, Support vector machine and regression[J]. IEEE Transactions on instrumentation and measurement, 2014, 64(1):52-62.

[7]. Gunnar Farnebck. Two-Frame motion estimation based on polynomial expansion[J]. Lecture notes in computer science, 2003:363-370.

[8]. LeCun Y, Boser B, Denker J S, Henderson D, Howard R E. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4):541-551.

[9]. HE KaiMing, Zhang XiangYu, Ren ShaoQing, Sun Jian. Deep residual learning for image recognition[C]. Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 2016: 770-778.