Computers, Materials & Continua
DOI:10.32604/cmc.2021.017527
Article

Tech Science Press

Research on Face Anti-Spoofing Algorithm Based on Image Fusion

Pingping Yu[1], Jiayu Wang[1], Ning Cao[2,*] and Heiner Dintera[3]

[1]School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, 050000, China
[2]School of Internet of Things and Software Technology, Wuxi Vocational College of Science and Technology, Wuxi, 214028, China
[3]German-Russian Institute of Advanced Technologies, Karan, 420126, Russia
[*]Corresponding Author: Ning Cao. Email: ning.cao2008@hotmail.com
Received: 01 February 2021; Accepted: 05 March 2021

Abstract: Along with the rapid development of biometric authentication technology, face recognition has been commercially used in many industries in recent years. However, it cannot be ignored that face recognition-based authentication techniques can be easily spoofed using various types of attacks such photographs, videos or forged 3D masks. In order to solve this problem, this work proposed a face anti-fraud algorithm based on the fusion of thermal infrared images and visible light images. The normal temperature distribution of the human face is stable and characteristic, and the important physiological information of the human body can be observed by the infrared thermal images. Therefore, based on the thermal infrared image, the pixel value of the pulse sensitive area of the human face is collected, and the human heart rate signal is detected to distinguish between real faces and spoofing faces. In order to better obtain the texture features of the face, an image fusion algorithm based on DTCWT and the improved Roberts algorithm is proposed. Firstly, DTCWT is used to decompose the thermal infrared image and visible light image of the face to obtain high-and low-frequency subbands. Then, the method based on region energy and the improved Roberts algorithm are then used to fuse the coefficients of the high- and low-frequency subbands. Finally, the DTCWT inverse transform is used to obtain the fused image containing the facial texture features. Face recognition is carried out on the fused image to realize identity authentication. Experimental results show that this algorithm can effectively resist attacks from photos, videos or masks. Compared with the use of visible light images alone for face recognition, this algorithm has higher recognition accuracy and better robustness.

Keywords: Anti-spoofing; infrared thermal images; image fusion; heart rate detection

1 Introduction

With the development of biometrics, face recognition plays a pivotal role in applications such as identity recognition systems, criminal justice database systems, and public surveillance systems.

The subsequent face spoofing attacks have also increased sharply. Attackers often use photos, videos, 3D modeling, masks, and other methods to imitate real human faces and obtain system access authorization for illegal intrusion and face recognition. This poses a serious threat to the security of the face recognition system. It is a necessary research topic to accurately judge the authenticity of human faces and identify human facial information to resist these complex and diverse deception attacks [1].

Face anti-spoofing detection systems are mainly divided into three categories: systems based on specific equipment, systems based on human-computer interaction, and systems based on pure algorithms.

There are relatively many methods based on human-computer interaction to prevent spoofing attacks. For example, Alsufyani et al. [2] used the random movement of infrared light to track the relative movement trajectory of the human eye. Singh et al. [3] and Pan et al. [4] proposed to detect the user's blinking and lip movements to resist people face spoofing attack; Tirunagari et al. [5] used dynamic correlation models to preprocess the video to extract texture features. The disadvantage of this type of detection method which seeks user cooperation is that it takes too long to detect the user's specified action, and it needs the user to request coordinated action. This will affect the user's experience.

Relying on pure algorithms for rapid detection and resolution through user videos or images is also a research hotspot. Wen et al. [6] proposed an algorithm combining image deformation analysis features; Pinto et al. [7] proposed a method of visual frequency analysis to detect video Face attacks; Määttä et al. [8] used LBP features to complete the detection of spoofing attacks; Alhassan et al. [9] combined DMD, LBP, and SVM to perform a liveness test score. Li et al. [10] proposed a face recognition algorithm based on LBP-EHMM; Wild et al. [11] proposed a detection algorithm based on bagging strategy; Pinto et al. [12] proposed a face activity detection method based on visual rhythm analysis. This type of texture feature-based detection method is based on gray-scale image extraction. The extracted features are not comprehensive enough, which affects the final detection result and has limited accuracy. Lee et al. [13] identify real faces from photos by analyzing data. Zhang et al. [14] used an adaptive ellipse fitting method to roughly determine the face area. Then, the study performed AdaBoost-based classification according to face template matching and face skin color distribution statistics, and finally detected facial occlusion. Xia et al. [15] proposed face occlusion detection based on a convolutional neural network. The network model was trained through a large number of occlusion samples. The image to be detected was input to the network, and the result of detecting whether the left and right eyes, nose and mouth were occluded was directly outputted. Kim et al. [16] proposed a face activity detection method for face spoofing attacks on mobile phones. According to the difference in the diffusion speed of reflected light from fake photos and live images, a real-time live detection based on the diffusion speed of reflected light from a single image was proposed. They used the following method: i) introduce the total variation flow to obtain the diffusion speed; ii) use the different diffusion speeds of the reflected light from the active skin and the fake face to distinguish whether there is activity; iii) use the LSP code to extract the speed feature vector on the reflected light diffusion speed distribution map, and iv) use the SVM classifier to determine whether the image comes from a living human face.

Bao et al. [17] proposed a face anti-spoofing detection algorithm that used fusion color texture features using the difference in color features and detailed texture features between real faces and spoofing attack images. The algorithm mainly used infrared at night and lacked spectrum collection color information. Li et al. [18] proposed a face anti-spoofing method based on P-CNN

and ELM to detect 2D spoofing attacks. Combining traditional digital images, Zhang et al. [19] proposed a forensic algorithm for face photos and video spoofing attacks based on the recursive elimination of color texture Markov features and support vector machine features. However, the algorithm could not detect whether the face uses 3D means such as silicone masks. Relying on pure algorithms for anti-spoofing detection of faces, the complexity of the algorithm is relatively high. There are certain restrictions on the detection environment and imaging CMOS cameras, which leads to an increase in the algorithm complexity of the entire face recognition process.

Seeking user cooperation and relying on pure algorithms are not ideal detection methods. With the popularity of infrared cameras, depth cameras and other equipments and the reduction of costs, anti-spoofing algorithms based on specific equipment have become the mainstream. Sun et al. [20] fused the human eye features of color and infrared to detect whether the driver is tired; Wang et al. [21] proposed a three-dimensional face recognition method with elastic matching of the radial curve of the face, but the user experience was lacking.

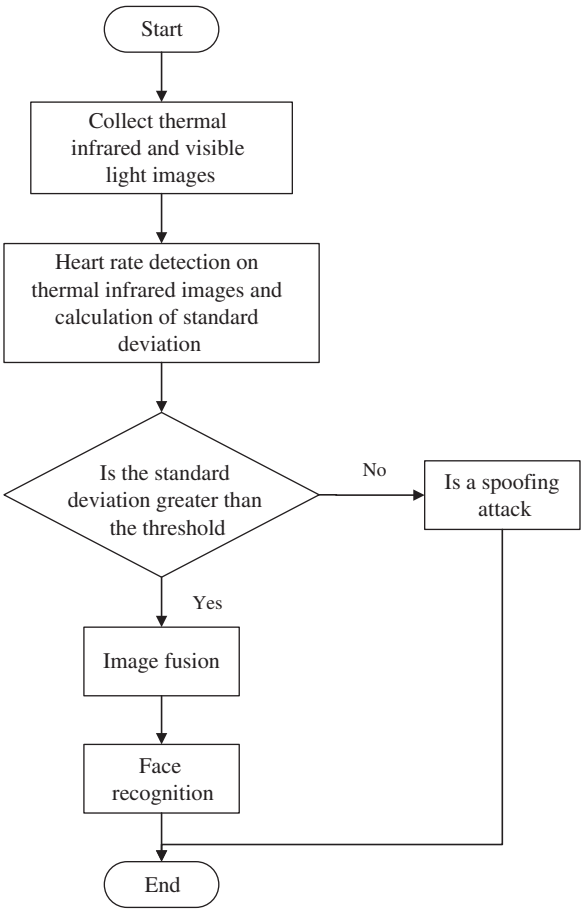

Figure 1: General flow chart of the face anti-spoofing system

This paper proposes a face anti-fraud algorithm based on the fusion of thermal infrared images and visible light images. By detecting the pulse sensitive area in the infrared thermal image, the grayscale value signal of the image is statistically analyzed, and the heart rate waveform is

calculated to distinguish real and fake faces. For faces, we use the method based on dual-tree complex wavelet transform (DTCWT) and improved Roberts operator to fuse the images, and finally identify the identity. The overall flow chart is shown in Fig. 1.

The rest of this article is arranged as follows. The second part introduces real and fake face detection method based on life information analysis. The third part introduces the image fusion based on DTCWT and the improved Roberts algorithm. The fourth part introduces face recognition and gives the results of this method.

2 Real and Fake Face Detection Method Based on Life Information Analysis

The difference between a real face and a deceptive face is that a real face has some vital information, such as capillaries and pulse. The surface of the human body can radiate infrared thermal energy, and the energy is mainly concentrated in the infrared band with a wavelength of 9.312~9.464 μm [22]. The normal temperature distribution of the human face is stable and characteristic, and the important physiological information of the human body can be observed by using the infrared thermal image. Since there are a large number of capillaries in the face, the blood in the blood vessels changes with the beating of the heart. When the heart contracts, the blood increases and the heat radiation energy increases.When the heart relaxes, the blood decreases and the heat radiation energy decreases. Therefore, the pixel value of the infrared thermal image reflected on the infrared thermal imager will fluctuate with the heart's beating, while the thermal infrared image of the spoofing face does not have this phenomenon.

2.1 Heart Rate Signal Detection

Before signal processing, we perform grayscale value processing on the facial thermal image video taken by the infrared thermal imager, and the conversion formula is

$$G_d = 0.299R + 0.587G + 0.114B \tag{1}$$

where G_d is the grayscale value after conversion, and R, G and B are the red, green, and blue component values of the pixel before conversion.

We perform face detection on the input test video and select a region of interest (ROI) to reduce the interference of light changes caused by other factors. Since the forehead part of the gray-scale thermal image of the human face is most sensitive to pulse beats, the forehead part of the grayscale thermal image is selected as ROI for processing in this article. It is mainly divided into two steps:

1) Set ROI template for grayscale thermal image.

In the first few frames of the gray-scale image sequence, we select a relatively clear facial image. We set the distance between the center points of the pupils of the two eyes to $4d$. A $3d \times 3d/2$ rectangle (r_1) in the center of the forehead is selected d above the straight line from the pupils, as shown in Fig. 2.

2) After the ROI template is obtained, extract the sensitive areas of each frame of the gray image sequence.

Due to the large noise interference of individual frames, it is impossible to ensure that the r_1 region information of each frame image is complete and effective. Therefore, the sensitive area of each frame of the image needs to be selected according to the template r_1. In each frame of the gray image sequence, a rectangular block R_n of size $5d \times 3d$ containing the ROI template r_1 in the middle is found, as shown in the dotted area in Fig. 2.

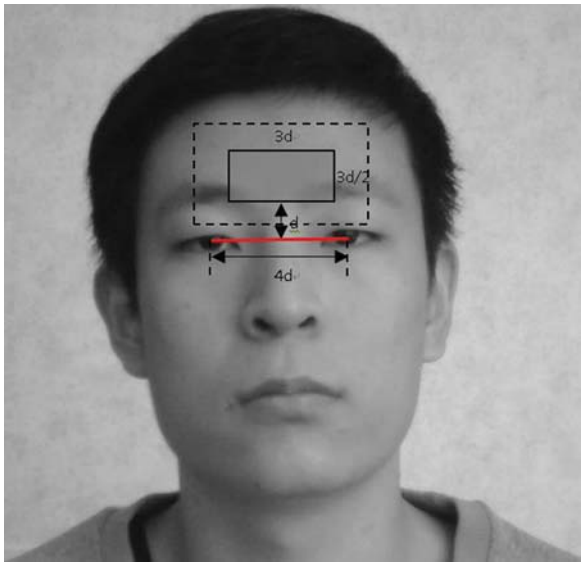

Figure 2: ROI template and rectangular area

The normalized cross-correlation function is used in the R_n area to examine the matching of each $3d \times 3d/2$ candidate block r_n with the ROI template r_1 to ensure that each frame image can select the most accurate sensitive area of grayscale value. The normalized cross-correlation function formula is obtained as follows

$$\gamma(u, v) = \frac{\sum_{x,y}\left[f(x, y) - \overline{f}_{u,v}\right]\left[t(x-u, y-v) - \overline{t}\right]}{\left\{\sum_{x,y}\left[f(x, y) - \overline{f}_{u,v}\right]^2 \cdot \sum_{x,y}\left[t(x-u, y-v) - \overline{t}\right]^2\right\}^{1/2}} \tag{2}$$

Among them, μ refers to the correlation interval on the abscissa axis of the grayscale image, υ refers to the correlation interval on the ordinate axis, $f(x, y)$ represents the candidate block r_n, $t(x, y)$ represents the pixel value of the ROI template r_1, $\overline{f}_{\mu,\upsilon}$ and $\overline{t}$ are the mean values of the pixel values of the candidate block r_n and the ROI template r_1, respectively.

The obtained normalized cross-correlation coefficients of each candidate block r_n are selected, and the block r_n with the largest absolute value and exceeding the specified threshold is used as the sensitive area of the frame. The grayscale value of the sensitive area of each frame of the gray image sequence is averaged to obtain g_n. These average values are arranged in chronological order to get the grayscale value waveform, which is the reflection of the heart rate signal. A set of grayscale value waveforms is shown in Fig. 3.

With the acquisition of a set of heart rates, the average value is calculated and stored in the array. Finally, the calculated variance is compared with the threshold to get the result, which is used to judge whether the given test video is a spoofing attack. Then, we enter the face recognition process on the premise that it is a real face.

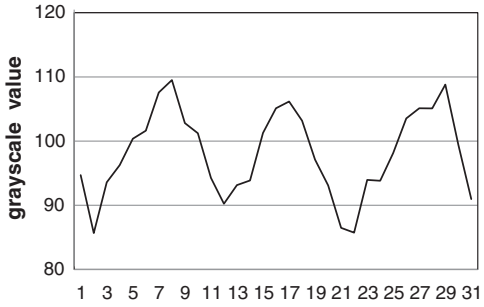

Figure 3: Grayscale value waveform of real faces

2.2 Anti-Spoofing Attack Method

The experiment uses an infrared thermal imager of model G100EX, the temperature measurement range is −40°C to 1500°C, the temperature resolution is 0.04 at 30°C, the pixel is 320 (H) × 240 (V), and the response wavelength is 8∼14 μm.

In a room with a temperature of 22°C, thermal infrared video collection was performed on the experimenter himself, the paper photos of the experimenter, the electronic video of the experimenter, and the mask-wearing experimenter. Considering that the attacker may simulate the mask to the temperature of the human body to attack, the temperature of the mask is increased to approximately 37 degrees for shooting. The collection methods and the presentation results are shown in Fig. 4.

(a) (b) (c) (d)

Figure 4: Collected experimental data. ((a) A group of real face images; (b) a group of printed face images; (c) a group of electronic videos; (d) a group of faceswearing masks)

According to the abovementioned algorithm, the grayscale value waveform acquisition is performed for the three kinds of deception attacks, and the results are shown in Fig. 5.

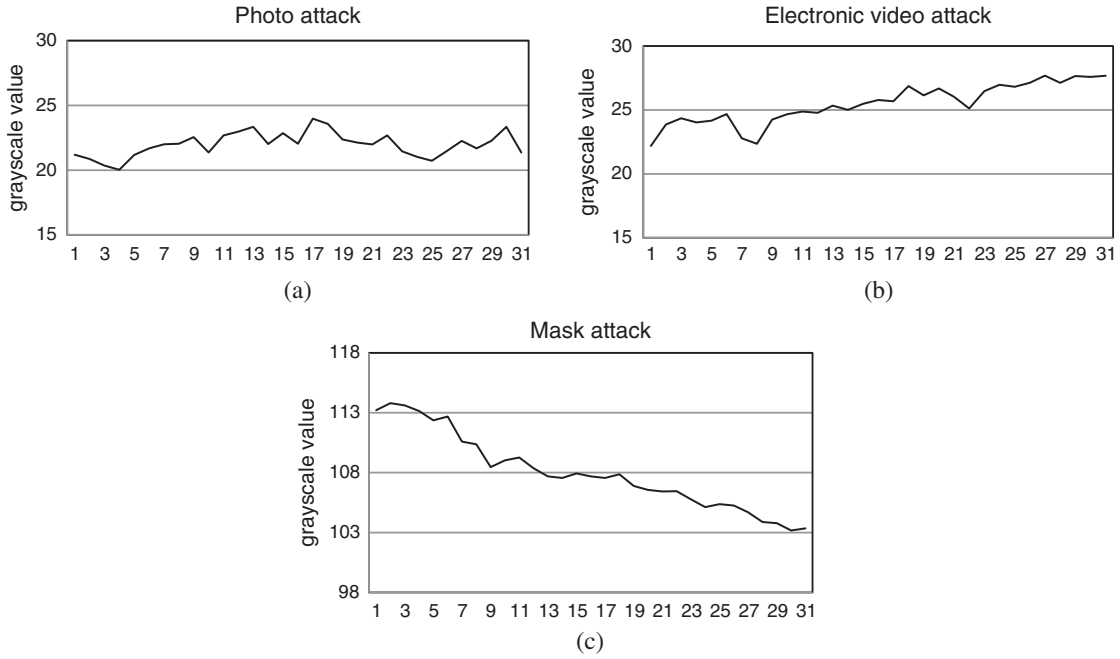

Figure 5: The grayscale value waveforms of various attacks. (a) Grayscale value waveform of electronic photo attack; (b) grayscale value waveform of electronic video attack; (c) grayscale value waveform of mask attack

From the waveform of Fig. 3, it can be seen that the grayscale value of the real face is between 80 and 120, and the fluctuation range is large. The grayscale value of the photo attack and the video attack is quite different from the grayscale value of the real face. It can be clearly distinguished. The grayscale value obtained after wearing the mask is slightly different from the grayscale value of the real face, which is difficult to distinguish directly, but the waveforms of the three deception attacks are relatively stable compared with the waveforms of the real face. Therefore, we use the standard deviation to further calculate the stability of the waveform to determine whether it is a mask attack. The standard deviation formula is obtained as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(g_i - \mu)^2} \tag{3}$$

where N is the number of frames of the image, g_i is the grayscale value of the sensitive area of each frame, and μ is the average value of the grayscale value of the sensitive area of the N frames.

The results are shown in Tab. 1. The standard deviation of real faces is generally above 6, and the standard deviation of photos, videos and masked faces is below 4. Therefore, the threshold for distinguishing true and false faces can be set between 4 and Between 6.

Table 1: The average and standard deviation of grayscale values of real and fake faces

Data group	Average grayscale value of the photo face	Average grayscale value of the video face	Average grayscale value of the mask face	Average grayscale value of the real face
1	21.191	22.196	113.214	94.691
2	20.865	23.865	113.799	85.682
3	20.356	24.356	113.612	93.584
4	20.034	24.031	113.147	96.250
5	20.168	24.168	112.368	100.358
6	21.999	24.686	112.687	101.632
⋮	⋮	⋮	⋮	⋮
Standard deviation	0.9258	1.5353	3.1228	6.6534

A face detection database is constructed which contains 30 real faces, 30 photo faces, 30 electronic video faces and 30 mask faces. The method proposed in this paper is verified by experiments on this database. The results are shown in Tab. 2, which shows that the method proposed in this paper can effectively solve the problem of real face detection.

Table 2: Spoofing attack test results

Experimental data	Number of real faces	Number of spoofing attacks	Number of correct identification	The detection accuracy of real face and deception attack
120	30	90	117	97.50%

3 Image Fusion Based on DTCWT and Improved Roberts Algorithm

The infrared thermal image contains the temperature information of the human body surface, which can be used to distinguish true and false faces through heart rate detection, but the lack of detailed information such as contour texture makes it impossible to recognize the identity of real faces. Visible light images contain rich detailed information, but they have low anti-interference ability under the influence of light, and important face information is often lost during face recognition. The thermal infrared image is not affected by light, and has good anti-interference ability and camouflage recognition ability. The fusion of the thermal infrared image and the visible light image can not only retain the rich detailed information in the visible light image and the temperature information in the infrared thermal image but also make up for the lack of light interference characteristic information [23]. Therefore, this paper uses the dual-tree complex wavelet transform to decompose the visible light image and the infrared thermal image, and obtain the high-frequency and low-frequency subband components of the same size as the source image. The low-frequency subband uses the method based on regional energy to fuse, and the edge enhancement method based on the improved Roberts operator is used for fusion of the high-frequency sub-band. Finally the dual-tree complex wavelet inverse transform is used to obtain the final fused image. The block diagram of the fusion algorithm is shown in the Fig. 6.

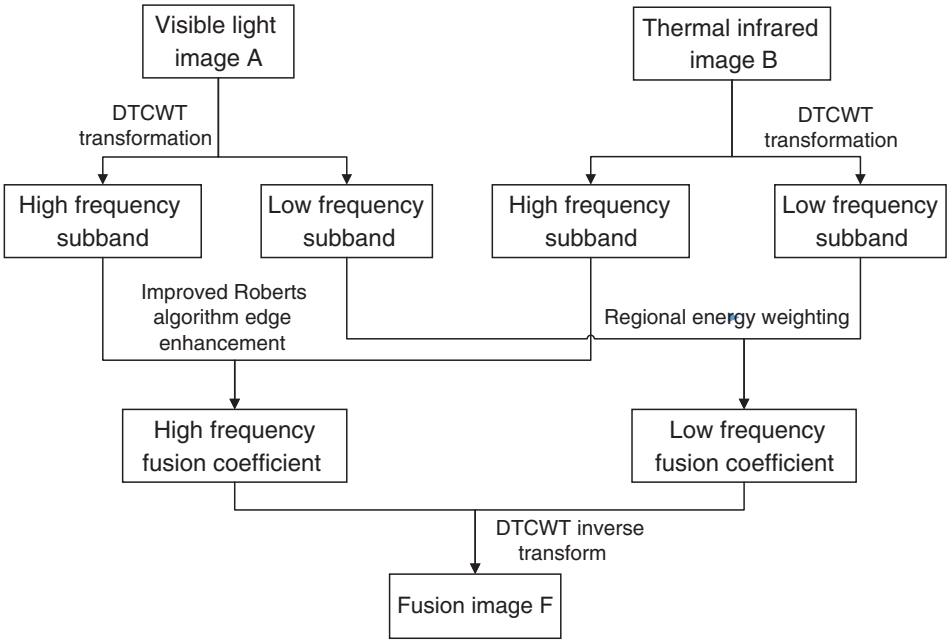

Figure 6: Block diagram of the fusion algorithm

3.1 Improved Roberts Algorithm

3.1.1 Improved Roberts Operator

Roberts operator uses the difference between adjacent pixels in the diagonal direction (45°, 135° direction) in the 2×2 area to approximate the gradient amplitude for edge detection, and the magnitude of the gradient $\mathrm{R}(x, y)$ of a certain point $f(x, y)$ on the image is defined as follows:

$$\mathrm{R}(x, y) = \sqrt{[f(x+1, y+1) - f(x, y)]^2 + [f(x+1, y) - f(x, y+1)]^2} \tag{4}$$

where we elect the threshold t and, when $\mathrm{R}(x, y) > t$, the pixel point $f(x, y)$ is determined to be an edge point. The traditional Roberts operator only calculates the information of 4 pixels in the diagonal direction, ignoring the pixel information in the vertical and horizontal directions. It is easy to cause missing edge pixels, and the threshold needs to be set manually, which has limitations [24].

Given the shortcomings of the traditional Roberts algorithm, this article considers adding vertical and horizontal direction information on the basis of the traditional Roberts operator (as shown in the Fig. 7). Use the template in four directions of 0°, 45°, 90°, 135° in the 3×3 field. The template performs convolution operations on pixels.

The difference in the four directions is obtained as follows:

$$\begin{cases} f_0 = f(i, j+1) - f(i, j-1) \\ f_{45} = f(i-1, j+1) - f(i+1, j-1) \\ f_{90} = f(i-1, j) - f(i+1, j) \\ f_{135} = f(i-1, j-1) - f(i+1, j+1) \end{cases} \tag{5}$$

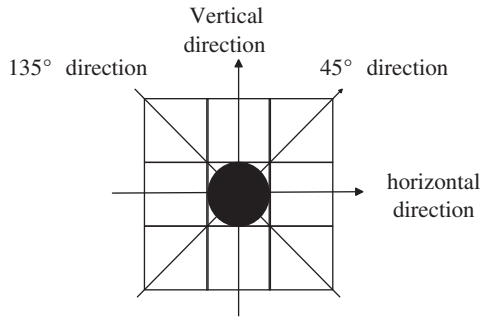

Figure 7: Calculation improvement of the gradient amplitude of Roberts operator

Their corresponding convolution operators are obtained as follows:

$$f_0: \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad f_{45}: \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$f_{90}: \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \quad f_{135}: \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

The improved Roberts operator considers the neighborhood information of pixels in 8 directions, which makes the edge extraction information more complete.

3.1.2 Median Filter Denoising

Although the improved Roberts operator can effectively extract the edge information, the noise generated under the interference of the complex environment in the infrared image and the visible light image will affect the gradient amplitude of pixel value, resulting in the extraction of the false edge formed by the noise. Therefore, we need to denoise the image. Median filtering can protect the edges of the signal from being blurred while filtering out noise. The algorithm is relatively simple and efficient. The two-dimensional median filter expression is obtained as follows:

$$g(x, y) = med\{f(x-k, y-l), (k, l \in W)\} \tag{7}$$

Among them, $f(x, y)$ is the initial image, $g(x, y)$ is the filtered image, and W is the two-dimensional template. A template of $3 * 3$ area is used here.

3.1.3 Threshold Segmentation Based on Otus

Based on the image denoising process, the improved Roberts operator is used for edge extraction. It is necessary to set the threshold t, and determine the edge point when the pixel point (x, y) is greater than t. The selection of the t value is particularly important. The efficiency of the threshold setting is low, and the adaptive ability is poor. So this paper adopts the method of maximum between-class variance (Otus) for threshold segmentation.

Suppose an image with a gray level of L, the range of L is $(0, 1, \ldots, L-1)$, use n_i to represent the number of pixels with a gray level of n, and N to represent the total number of pixels, then, we obtain the following expression:

$$N = n_0 + n_1 + \ldots + n_L = \sum_{i=0}^{L-1} n_i \tag{8}$$

Let $p(i)$ be the probability that a pixel with gray level i appears:

$$p(i) = \frac{n_i}{N} \tag{9}$$

Set the initial threshold t to divide the image into two parts A and B, where the grayscale range of A is $(0, 1, \ldots, t)$, and the grayscale range of B is $(t+1, t+2, \ldots, L-1)$, then the probability of A and B is obtained as follows:

$$P_A(t) = \sum_{i=0}^{t} p_i, \quad P_b(t) = \sum_{i=t+1}^{L-1} p_i \tag{10}$$

The gray average value of the two parts A and B is obtained as follows:

$$\mu_A(t) = \frac{\sum_{i=0}^{t} ip_i}{P_A(t)}, \quad \mu_A(t) = \frac{\sum_{i=t+1}^{L-1} ip_i}{P_B(t)} \tag{11}$$

Then, the between-class variance in the two parts A and B is obtained as follows:

$$d(t) = P_A P_B (\mu_A - \mu_B)^2 \tag{12}$$

When the between-class variance $d(t)$ is the largest, t is the optimal threshold.

3.2 Fusion Strategy

3.2.1 Dual-Tree Complex Wavelet Transform

The dual-tree complex wavelet transform (DTCWT) is composed of two parallel real wavelet transforms, using different low-pass and high-pass filters, each group of decomposition and reconstruction processes is carried out separately, and there is no interaction between data [25].

If $f(t)$ is the image input signal, $s^r(t)$ and $s^l(t)$ are the wavelet functions of the real and imaginary parts, respectively, and $h^r(n)$ and $h^l(n)$ are the real scaling function of the part and the imaginary part, then the wavelet coefficient $W_j^r(k)$ and the scaling coefficient $C_j^r(k)$ of the real part transformation are obtained as follows:

$$D_j^r(k) = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} f(t)\, s^r\left(2^j t - k\right) dt, \quad j = 1, 2, \ldots, J \tag{13}$$

$$C_j^r(k) = 2^{\frac{J}{2}} \int_{-\infty}^{\infty} f(t)\, h^r\left(2^j t - k\right) dt \tag{14}$$

J represents the maximum number of decomposition layers. Similarly, the wavelet coefficient $W_j^l(k)$ and the scale coefficient $C_j^l(k)$ of the imaginary part are obtained as follows:

$$D_j^l(k) = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} f(t)\, s^l\left(2^j t - k\right) dt, \quad j = 1, 2, \ldots, J \tag{15}$$

$$C_j^l(k) = 2^{\frac{J}{2}} \int_{-\infty}^{\infty} f(t)\, h^l\left(2^j t - k\right) dt \tag{16}$$

The final DTCWT output complete wavelet coefficient $W_j(k)$ and scale function $C_j(k)$ are obtained as follows:

$$D_j(k) = D_j^r(k) + iD_j^l(k) \tag{17}$$

$$C_j(k) = C_j^r(k) + iC_j^l(k) \tag{18}$$

The wavelet coefficients and scale coefficients obtained by the above decomposition obtained as follows:

$$D_j(t) = 2^{\frac{i}{2}} \lambda_i \sum_{n \in Z} \left[D_j^r(n)\, s^r\left(2^j t - n\right) + D_j^l(n)\, s^i\left(2^j t - n\right) \right] \tag{19}$$

$$C_j(t) = 2^{\frac{i}{2}} \lambda_{L+1} \sum_{n \in Z} \left[C_j^r(n)\, s^r\left(2^j t - n\right) + C_j^l(n)\, s^i\left(2^j t - n\right) \right] \tag{20}$$

where λ_i is the scale selection coefficient, the value range is 0 or 1, and the reconstructed signal $f^*(t)$ is obtained as follows:

$$f^*(t) = \sum_{j=1}^{L} W_j(t) + C_j(t) \tag{21}$$

3.2.2 Low-Frequency Subband Fusion Strategy

The low-frequency subband part of the image represents the energy distribution of most of the background of the image. In this paper, the weighting method based on regional energy is used to determine the fusion coefficient of the low-frequency subband. The specific fusion steps are obtained as follows.

Step 1: Calculate the regional energy of the low-frequency subband coefficients after DTCWT decomposition.

$$E_F(x, y) = \frac{1}{(2s+1)^2} \sum_{i=-s}^{s} \sum_{j=-s}^{s} L(x+i, y+j)^2 \tag{22}$$

where $E_F(x, y)$ represents the average energy in the image F_l within the neighborhood window of $(2s+1) \times (2s+1)$ centered on the point (x, y); s usually takes 1, 2, 3; $\mathrm{L}(x+i, y+j)$ represents the low-pass subband coefficients after image decomposition.

Step 2: Calculate the weight.

$$\omega = \frac{E_A(x, y)}{E_A(x, y) + E_B(x, y)} \tag{23}$$

Step 3: Calculate the low-frequency subband fusion coefficient:

$$f_L^F(x, y) = \omega f_L^A(x, y) + (1 - \omega) f_L^B(x, y) \tag{24}$$

3.2.3 High-Frequency Subband Fusion Strategy

The high-frequency subband of the image reflect most of the details of edges, textures, contours, etc. The traditional fusion rule of taking the absolute value is susceptible to noise and the fusion effect is low. Moreover, due to the light and other factors, a part of the edge information of the visible light image may be lost, resulting in the loss of local information. Therefore, this paper proposes an edge-enhanced fusion rule based on the improved Roberts operator. The specific fusion algorithm steps are as follows.

Step 1: Perform median filter processing on the high-frequency subband image to remove noise.

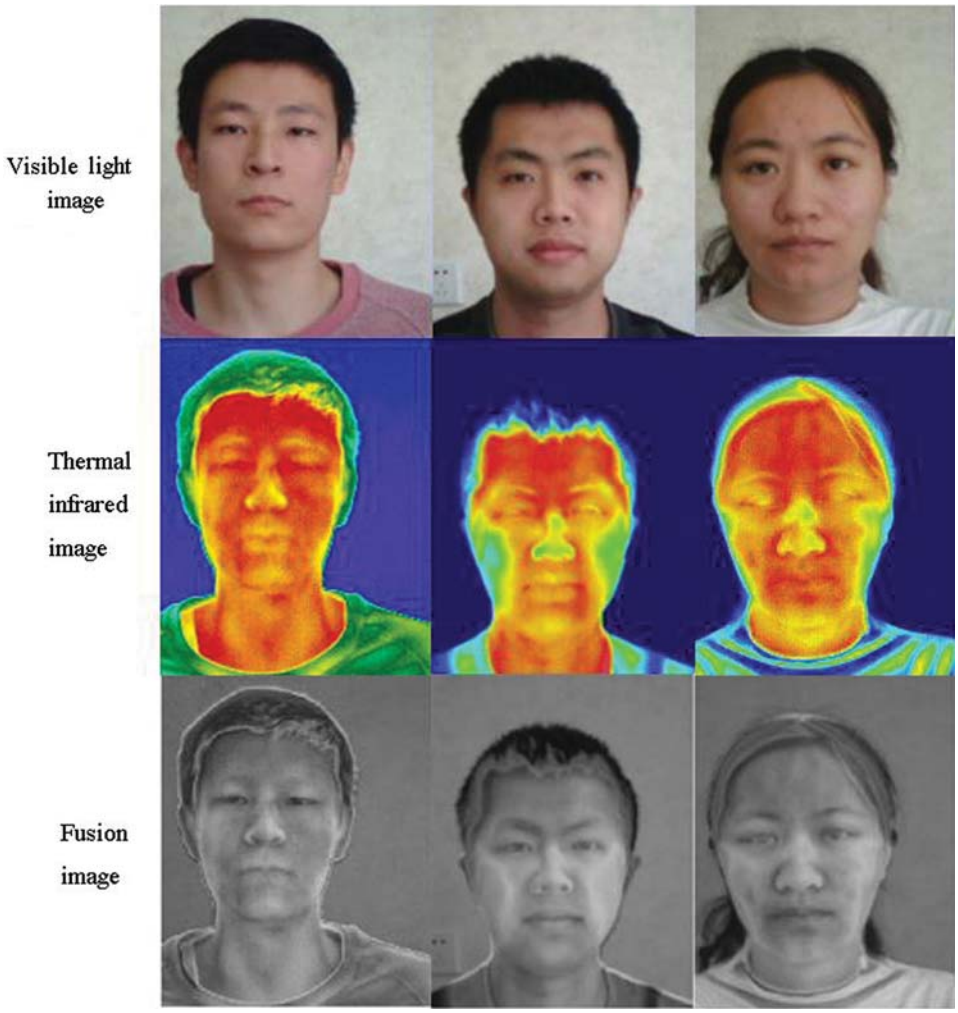

Figure 8: Fused image of real human faces

Step 2: Perform edge information extraction on the high-frequency subband image after denoising. The edge information is extracted according to the above mentioned improved Roberts algorithm. Perform convolution operations based on templates in the directions of 0°, 45°, 90°, 135°, etc. to obtain the gradient value of each pixel. Get $R_A(x, y)$ and $R_B(x, y)$.

Step 3: Use the Otus threshold segmentation method to obtain the best thresholds t_A and t_B.

Step 4: Obtain the high-frequency subband fusion coefficient $f_H^F(x, y)$ as follows:

$$f_H^F(x,y)=\begin{cases} f_H^A(x,y), & R_A(x,y)\geq t_A \boxplus R_B(x,y)\leq t_B \\ f_H^B(x,y), & R_A(x,y)\leq t_A \boxplus R_B(x,y)\geq t_B \\ \max\left\{f_H^A(x,y), f_H^B(x,y)\right\}, & R_A(x,y)\geq t_A \boxplus R_B(x,y)\geq t_B \\ \dfrac{f_H^A(x,y)}{f_H^A(x,y)+f_H^B(x,y)}\times f_H^A(x,y)+\dfrac{f_H^B(x,y)}{f_H^A(x,y)+f_H^B(x,y)}\times f_H^B(x,y), & R_A(x,y)\leq t_A \boxplus R_B(x,y)\leq t_B \end{cases} \tag{25}$$

The low-frequency subband fusion coefficient $f_L^F(x, y)$ and the high-frequency subband fusion coefficient $f_H^F(x, y)$ are inversely transformed by DTCWT to obtain the final fused image F.

3.2.4 Image Fusion Experiment

The real face that has been successfully verified is fused using the above algorithm, and the resulting fused image is shown in Fig. 8. Face recognition is carried out on the fused image, and the identity information of the tested person is verified to realize the face anti-spoofing.

Table 3: Comparison of face recognition results between visible light image and fusion image

Recognition methods	LBP (%)	HOG (%)	LBP + HOG (%)	MB-LBP (%)	LBP + MB-LBP (%)
Recognition rate of the visible light image under normal environment	88.5	90.3	93.3	92.6	94.4
Recognition rate of the fusion image under normal environment	89.9	91.4	93.7	92.5	94.6
Recognition rate of the visible light image with light changes	83.3%	84.8	86.7	86.1	90.1
Recognition rate of the fusion image with illumination transformation	88.4	90.9	92.6	91.3	93.8

4 Test Results and Analysis

In order to verify the advantages of fusion images in face recognition, this paper collects human face information from 100 people. The visible image and thermal infrared image of human faces are collected under normal light conditions and large light changing conditions. Fusion images are obtained by using the above method, and the visible light image and the fusion image database of the human face are established. The visible light image and the fusion

image under different lighting conditions are respectively applied to different face recognition n algorithms (LBP [26], HOG [26], LBP + HOG [26], MB-LBP [27], LBP + MB-LBP [27]) to obtain the face recognition rate, as shown in Tab. 3. From the results, it can be seen that there is not much difference between the recognition rate of visible image and fused image in the same face recognition algorithm under normal illumination; but in the case of large changes in light, the recognition rate of fusion image is significantly higher than that of visible image in the same face recognition algorithm. Therefore, face recognition based on fusion image has strong robustness to illumination changes.

5 Conclusion

This paper proposes an algorithm to resist facial spoofing attacks. Using thermal infrared images, the pixel values of real faces and fake faces of legitimate users are collected, and heart rate signals are detected to distinguish true and false faces. An image fusion algorithm based on DTCWT is proposed to decompose the visible light image and thermal infrared image of real human face. The obtained high-frequency sub-band uses the method based on regional energy for coefficient fusion, and the low-frequency subband uses the improved Roberts algorithm for coefficient fusion. Then use the DTCWT inverse transform to obtain a fusion image containing facial texture features. Different face recognition algorithms are used to verify the recognition rate visible light images and fusion images. The results show that the face recognition algorithm based on fusion images has a higher recognition rate. It can be seen that the algorithm proposed in this paper can effectively reduce the impact of illumination changes on face recognition results in practical application scenes. Combined with heart rate signal detection can effectively distinguish the real faces and spoofing attack face, so as to improve the security of face anti- spoofing system.

Funding Statement: This research was funded by the Hebei Science and Technology Support Program Project (Grant No. 19273703D), and the Hebei Higher Education Science and Technology Research Project (Grant No. ZD2020318).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Zhang, K. Zeng and J. Wang, "A survey on face anti-spoofing algorithms," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 1, pp. 21–34, 2020.

[2] N. Alsufyani, A. Ali, S. Hoque and F. Deravi, "Biometric presentation attack detection using gaze alignment," in *IEEE 4th Int. Conf. on Identity, Security, and Behavior Analysis*, Singapore, pp. 1–8, 2018.

[3] A. K. Singh, P. Joshi and G. C. Nandi, "Face recognition with liveness detection using eye and mouth movement," in *Int. Conf. on Signal Propagation and Computer Technology*, Ajmer, India, pp. 592–597, 2014.

[4] G. Pan, L. Sun, Z. H. Wu and Y. M. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommunication Systems*, vol. 47, no. 3, pp. 215–225, 2011.

[5] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki *et al.,* "Detection of face spoofing using visual dynamics," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 762–777, 2015.

[6] D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.

[7] A. Pinto, H. Pedrini, W. R. Schwartz and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726–4740, 2015.

[8] J. Määttä, A. Hadid and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[9] A. K. Alhassan and A. A. Alfaki, "Color and texture fusion-based method for content-based image retrieval," in *Int. Conf. on Communication, Control, Computing and Electronics Engineering*, Khartoum, pp. 1–6, 2017.

[10] T. Li, L. Y. Wang, Y. Chen, Y. R. Ren, L. Wang *et al.,* "A face recognition algorithm based on LBP-EHMM," *Journal on Artificial Intelligence*, vol. 1, no. 2, pp. 61–68, 2019.

[11] P. Wild, P. Radu, L. Chen and J. Ferryman, "Robust multimodal face and fingerprint fusion in the presence of spoofing attacks," *Pattern Recognition*, vol. 50, no. C, pp. 17–25, 2016.

[12] A. Pinto, W. R. Schwartz, H. Pedrini and A. d. R. Rocha , "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 1025–1038, 2015.

[13] B. K. Lee and Y. S. Lee, "Distinction between real faces and photos by analysis of face data," *Intelligent Automation & Soft Computing*, vol. 26, no. 1, pp. 133–139, 2020.

[14] X. Zhang, L. Zhou, T. Zhang and J. Yang, "A novel efficient method for abnormal face detection in ATM," in *Int. Conf. on Audio, Language and Image Processing*, Shanghai, China, pp. 695–700, 2014.

[15] Y. Xia, B. Zhang and F. Coenen, "Face occlusion detection based on multi-task convolution neural network," in *2015 12th Int. Conf. on Fuzzy Systems and Knowledge Discovery*, Zhangjiajie, China, pp. 375–379, 2015.

[16] W. Kim, S. Suh and J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2456–2465, 2015.

[17] X. A. Bao, X. D. Lin, N. Zhang, L. Xu and B. Wu, "Face anti-spoofing algorithm using color texture features," *Computer Science*, vol. 46, no. 10, pp. 180–185, 2019.

[18] B. Li, B. L. Wang, L. You and M. Yang, "A face anti-spoofing method using parallel convolutional neural networks," *Small Microcomputer System*, vol. 38, no. 10, pp. 2187–2191, 2017.

[19] L. Zhang, F. Peng, L. Qin and M. Long, "Face spoofing detection based on color texture Markov feature and support vector machine recursive feature elimination," *Journal of Visual Communication and Image Representation*, vol. 41, no. 5, pp. 56–69, 2018.

[20] Y. Sun, P. Yan, Z. Li, J. Zou and D. Hong, "Driver fatigue detection system based on colored and infrared eye features fusion," *Computers, Materials & Continua*, vol. 63, no. 3, pp. 1563–1574, 2020.

[21] Z. Z. Wang, X. Zhang, P. P. Yu, W. J. Duan, D. J. Zhu *et al.,* "A new face recognition method for intelligent security," *Applied Sciences*, vol. 10, no. 3, pp. 852, 2020.

[22] Y. Q. Min, W. W. Wan and Y. Yu, "Non-contact face detection based on G-channel heart rate changes," *Computer Applications and Software*, vol. 36, no. 9, pp. 192–197, 2019.

[23] H. Liu and X. Zhou, "Multi-focus image region fusion and registration algorithm with multi-scale wavelet," *Intelligent Automation & Soft Computing*, vol. 26, no. 6, pp. 1493–1501, 2020.

[24] F. C. Wang, M. Zhang and L. M. Gong, "Improved roberts image edge detection algorithm," *Journal of Detection and Control*, vol. 38, no. 2, pp. 88–92, 2016.

[25] G. C. Zhang, J. F. Su and M. X. Tuo, "Infrared and visible light image fusion algorithm in DTCWT domain," *Computer Engineering and Science*, vol. 42, no. 7, pp. 1226–1233, 2020.

[26] Y. Wan, "Fusion with layered feature of LBP and HOG for face recognition," *Journal of Computer-Aided Design and Computer Graphics*, vol. 27, no. 4, pp. 640–650, 2015.

[27] B. Liu, Q. Mi and Y. Xu, "LBP and MB-LBP weighted fusion of face recognition," *Computer Engineering and Design*, vol. 39, no. 2, pp. 551–556, 2018.