Face Anti-Spoofing Detection Using Least Square Weight Fusion of Channel-Based Feature Classifiers

Xiaoning Song, Qiqun Wu, Dongjun Yu, Guosheng Hu and Xiaojun Wu

February 18, 2020

FACE ANTI-SPOOFING DETECTION USING LEAST SQUARE WEIGHT FUSION OF CHANNEL-BASED FEATURE CLASSIFIERS

Xiaoning Song[*a],Qiqun Wu[a], Dongjun Yu[b],Guosheng Hu[c], Xiaojun Wu[a]

[a]School of Internet of Things Engineering, Jiangnan University, Wuxi; [b]School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing; [c]Anyvision, Belfast BT39DT,U.K

ABSTRACT

Current face biometric systems are vulnerable to spoofing attack. In order to deal with various spoofing attacks, such as photo attack and video replay attack, a stable anti-spoofing algorithm should be designed. In this paper, we proposed a novel face anti-spoofing detection algorithm using least square weight fusion of channel-based feature classifiers. To this end, we first fuse the color and texture features through information entropy, the spatial and frequency features are then filtered and fused by SVM-RFE feature selection method. In addition, the fusion features of two convolutional neural networks are constructed by autoencoder. Second, for the generated three kinds of fusion features, we adpot AdaBoost, SVM and Randomforest to accomplish the robust classification, respectively. The final goal of the proposed methodis to utilize the least square method to adjust the optimal weights of the obtained three kinds of classification results, by this means, the stable and efficient face anti-spoofing detection result can be achieved. Experimental results conducted on two of the most challenging anti-spoofing datasets, including CASIA FASD and Replay-Attack, demonstrate the effectiveness of the proposed method.

Keywords: Channel feature, Least square, Convolutional neural network, Face anti-spoofing detection

1. INTRODUCTION

With the maturity of biometric recognition technology [1], fingerprint recognition [2], iris recognition [3] and speech recognition technology [4] are gradually applied to security systems in all walks of life. Face recognition has gradually become the mainstream because of its interactive, accessible and high degree of visualization. However, these advantages also bring hidden dangers to the security of the system. As early as 2002, Lisa Thalheim et al. used photos and short videos to detect FaceVACS-Logon face system, which successfully spoofed and passed the identity confirmation [5].This event raises great doubts about the security of face recognition technology, and then, face anti-spoofing, the urgent project, arises at the historic moment.

At present, face spoofing mainly includes the following ways: (1) candid face photos; (2) public face videos on the internet; (3) three-dimensional face model synthetized by computer software; (4) face masks made of plastic and rubber. Although bio-simulation technologies such as 3D printing can be gradually put into use for the spoofing systems, considering the cost of equipment, efficiency, convenience and other factors, the most popular current attack method is to take photos and videos of legal users. Currently, texture features such as local binary pattern (LBP) [6], histograms of oriented gradients (HOG) [7] and Haar feature [8] have been commonly used in face anti-spoofing detection, which can obtain a better performance than those of traditional manifold feature representation algorithms. Subsequently, considering the auxiliary discriminative capability present in color space including RGB, HSV, YCbCr, this explicit statistical diversity of human faces can be included for the task of spoofing detection. With the rise of convolution neural network (CNN) algorithm [9], features extracted for anti-spoofing has been improved significantly. However, the relationship among the channel features obtained from different manifold spaces is hard to be precisely represented by an effective collaborative manner.

In this paper, we take full account of the advantages of various channel features, for robust face anti-spoofing detection. According to the different statistical properties of color, texture, spatial and frequency features and the convolutional features of dataset, we develop a novel face anti-spoofing detection algorithm using least square weight fusion (LSWF) of channel-based feature classifiers. Compared with other traditional methods, the contributions of this study are twofold:

(1) We use different extracted methods to fuse different features. By design, we first fuse the color and texture features through regulating variable weight connection, we then filter the spatial and frequency features with SVM-RFE, and fused the convolution features generated by the two networks with autoencoder.

(2) In order to balance the performance of various channel features, we design a least square weight fusion method to evaluate the anti-spoofing detection result, by autonomically assigning the optimal weights of classification results, for achieving the competitive result.

The rest of the paper is organized as follows: in Section 2, the classical methods of face anti-spoofing detection in recent years are briefly introduced. Details of the proposed new method will be presented in Section 3. The analysis of the experimental results and the comparison with other advanced methods will be taken in Section 4. Finally, some conclusions are drawn in Section 5.

2. RELATED WORK

Currently, the main stream face anti-spoofing method, according to the different features, can be divided in two categories: traditional feature extraction and the convolutional feature extraction based on neural network.

2.1 Face anti-spoofing detection based on traditional features

Face anti-spoofing detection is similar to other image processing methods. Generally, feature extraction is also considered in texture, color and illumination. In the field of face anti-spoofing, researchers combine various features in different ways.In [10], considering the diversity of biological features, the authors proposed a method to integrate iris and face features. They used Log-Babor phase encoding algorithm and Laplacianfaces algorithm to extract the two biological features, and used the maximum and minimum probability mechanism for score-level fusion. Considering the face background space, Lin Sun et al. used Gabor descriptor to extract the ground background features and the face background features, and used the similarity of fusion phase difference to make background contrast, in order to resist the attack of replay [11].Yu Cao et al. put forward the method of combining gray level co-occurrence matrix with wavelet analysis to carry out liveness detection. The energy, entropy, inertia matrix, correlation and high frequency subband coefficients of gray level co-occurrence matrix are used as features, which effectively reduces the computational complexity and improves the detection accuracy [12].Litong Feng et al. of City University of Hong Kong designed a multi-scale shear features and combined it with softmax classifier, which obtained good experimental results [13].LeBing Zhang et al. proposed to distinguish real and false faces by color Markov features, and then eliminate redundancy by recursive iteration to achieve real-time detection [14].Jukka Maatta et al. of the University of Oru in Finland used three scales to extract LBP features from gray-scale face images. After feature cascade, support vector machine were used for classification tests [15].Alireza Sepas-Moghaddam of the University of Lisbon, Portugal, considering the influence of each component in the color space, mapped HSV and YCbCr to hue (H), saturation (S), value (V) and brightness (Y), blue component (Cb), red component (Cr) respectively. After taking LBP features and cascading them, the decision-making fusion is carried out [16].Neha D. Patil et al. proposed a method based on image distortion analysis (IDA) and principal component analysis (PCA), which extract features from four aspects: mirror reflection, blur, color moment and color diversity. Principal component analysis of these cascaded features not only improves robustness, but also speeds up operation efficiency [17].

2.2 Face anti-spoofing detection based on neural network

In recent years, with the upgrading of hardware and the solution of storage problem, the neural network algorithm has been paid more attention. Combining traditional features with neural network has become the mainstream method in the field of face spoofing. Gustavo Botelho de Souza et al. first proposed the concept of LBPnet, and trained the traditional LBP features into the neural network, which provided a new reference for the later feature fusion [18].Yousef Atoum of Michigan State University proposed a double-flow fusion method based on neural network[19],which uses local features to distinguish the depressed areas of spatial faces, uses global depth information to detect whether the input face image has face depth information, and then they made score-level fusion from the two aspects. In 2017, Muhammad Asim et al. proposed to combine the neural network with the traditional Spatial-temporal information for training and extracting features, and achieved good accuracy on several video face datasets [20]. In [21], considering the influence of different color spaces on face spoofing, the author proposed a method of extracting color LBP features from the convolutional features of neural networks, and tested them on RGB, HSV and YCbCr color spaces, which improved the robustness of experiments on different datasets.

3. LEAST SQUARE WEIGHT FUSION OF CHANNEL-BASED FEATURE CLASSIFIERS

The contribution of the proposed method in this paper mainly consists of two parts: (1) extract a number of collaborative image features from different channel spaces and fuse them using different methods. (2) design a least square weight fusion method to obtain the predicted result with high degree of confidence. The schematic diagram of the proposed method is shown as Figure 1.

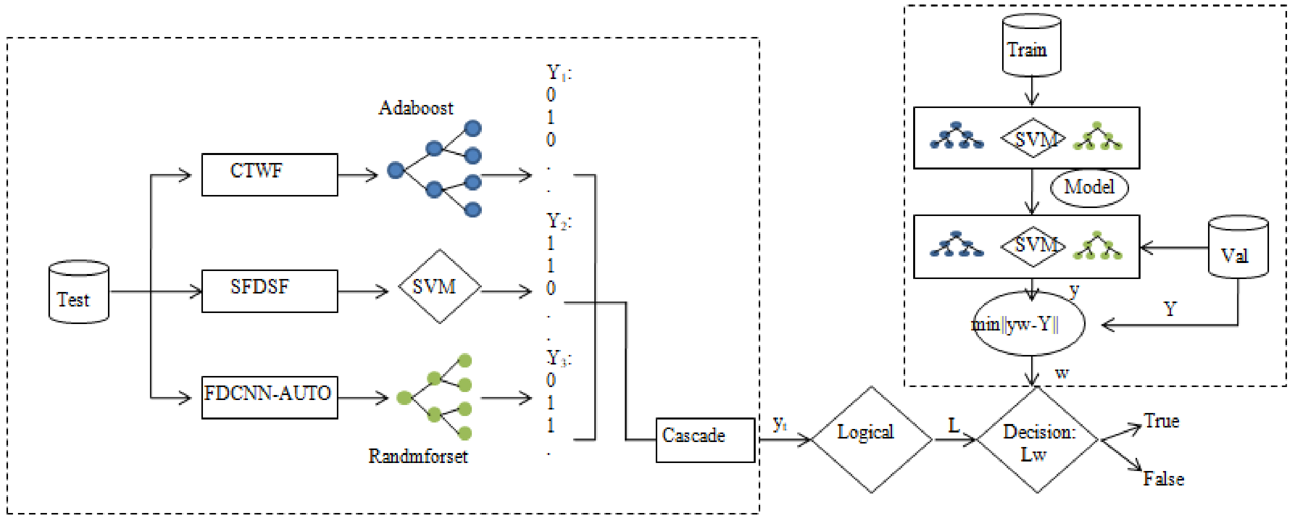

Figure 1. Schematic diagram of the proposed method

In this method, we use three different methods to fuse features. First of all, considering the discrimination effect of color and texture features, the original face image are transformed into different color spaces, the texture features are then extracted from different color components, and histogram components are counted. Second, the technique of information entropy is used to weigh component features. By this means, color texture weighted features (CTWF) are synthesized. Third, considering the spatial and frequency domain, the features of chromaticity moment, shearlet and Haar wavelet are extracted respectively, and the SVM-RFE method is used for feature selection to obtain the spatial-frequency domain selection feature (SFDSF). Due to the significance of feature extraction by neural network, we design a shallow neural network to extract the global face features, and a residual network to extract the local face texture features. The features obtained by the two convolutional neural networks are fused by autoencoder to obtain the self coding features (FDCNN-AUTO).

3.1 Color texture weighted features

In the production process of face video dataset, it involves the different background and illumination changes, which will interfere with the designed algorithm. Therefore, we first crop the video data to reduce the impact of the background image, we then reduce the interference of light changes using gamma correct for each image. The example images via image preprocessing are shown in Figure 2.

.

Figure 2. The example images via image preprocessing

In order to integrate with more edge gradient information, the Roberts operator is modified as shown in Figure 3. By design, the extraction module of the operator is expanded to 3 × 3, we then extract the gradient information of four 2×2 sub modules along the diagonal, the mean value is subsequently taken as the gradient value of the middle pixel. The calculation formula of the improved Roberts operator formula is presented as:

$$\begin{aligned} R_1(x,y) &= \sqrt{(L(x,y)-L(x-1,y-1))^2+(L(x,y-1)-L(x-1,y))^2} \\ R_2(x,y) &= \sqrt{(L(x,y)-L(x-1,y+1))^2+(L(x,y+1)-L(x-1,y))^2} \\ R_3(x,y) &= \sqrt{(L(x+1,y)-L(x,y-1))^2+(L(x+1,y-1)-L(x,y))^2} \\ R_4(x,y) &= \sqrt{(L(x+1,y+1)-L(x,y))^2+(L(x+1,y)-L(x,y+1))^2} \\ R(x,y) &= (R_1(x,y)+R_2(x,y)+R_3(x,y)+R_4(x,y))/4 \end{aligned} \quad (1)$$

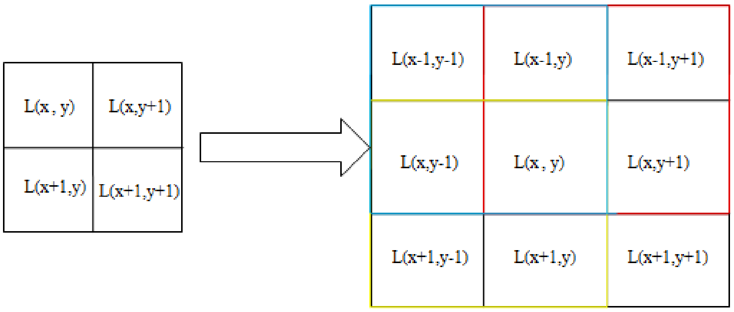

Figure 3. An improved Roberts operator model

The schematic diagram of the construction process of CTWF is illustrated in Figure 4. More specifically, for each original RGB face image, it is first transferred to HSV and YCbCr color spaces, and then three-dimensional face components of each color space can be extracted, respectively. According to the obtained each component of face image, the improved Roberts operator is used to extract significant edge texture features, we then transform six component texture features into histograms, and the weight of histograms can be calculated by means of information entropy. After steps of cascading and fusing, a new color texture weighted feature (CTWF) can be obtained. The obtaining way of color texture fusion considers three aspects of complementarity: the various color feature spaces, the balancing degree of integration of color feature component, and the difference between the illumination and texture transformation.

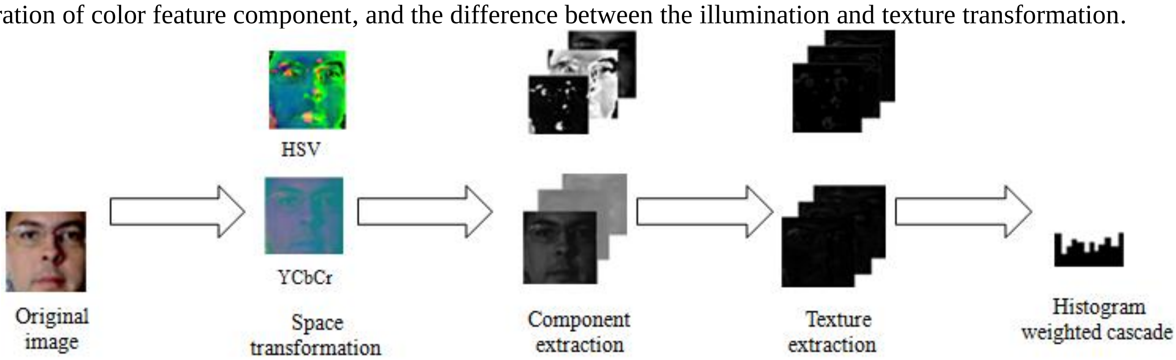

Figure 4. The construction process of CTWF

3.2 Spatial-frequency domain selection features

Chromaticity moments is a simple feature extraction method based on the mathematical calculation method, in which the first-order moments (mean) are used to represent the color brightness, the second-order moments (variance) are for presenting the color distribution range, and the third-order moments (skewness) are available to describe the color distribution and symmetry. In addition, we add the percentage of peak pixel to all pixels and the percentage of valley pixel to all pixels in the channel histogram as well, for highlighting the difference of image color distribution. For a picture with size of m× n, the calculation method of its five components is described as formulas (2), where I (i,j) represents the pixel value of a certain position, by means of calculation the formula, the chromaticity moment feature CMF = [E, D, S, max, min] under a single channel can be obtained. In order to better reflect the color difference, the original RGB face image is transformed into HSV and YCbCr color spaces, the chromaticity moment features under six color components are then calculated, respectively, we finally obtained a total of 5× 6 = 30 dimensional features .

$$\begin{aligned} \mathrm{E} &= \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{I(i,j)}{M\times N} \\ \mathrm{D} &= \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{(I(i,j)-E)^2}{M\times N} \\ \mathrm{S} &= \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{(I(i,j)-E)^3}{M\times N} \\ \max &= \frac{bin_{max}}{M\times N} \\ \min &= \frac{bin_{min}}{M\times N} \end{aligned} \tag{2}$$

Moreover, Shearlet transform has good anisotropy, which is defined in different scales, positions and directions. Shearlet is beneficial to detect direction information, especially sensitive in edge singularities, and which has good geometric properties of multi-dimensional functions. The process of constructing shear wave features is presented in Figure 5. For each input face image, the second level shearlet transform is used for wavelet decomposition, by this means, the high-frequency coefficient and low-frequency coefficient of the image can be simultaneously obtained. Meanwhile, the number of histograms in the eight directions of the coefficients in the two frequency bands are counted, respectively. After regularization, the number of histograms is cascaded to obtain the 2× 8 = 16 dimensional feature vector.

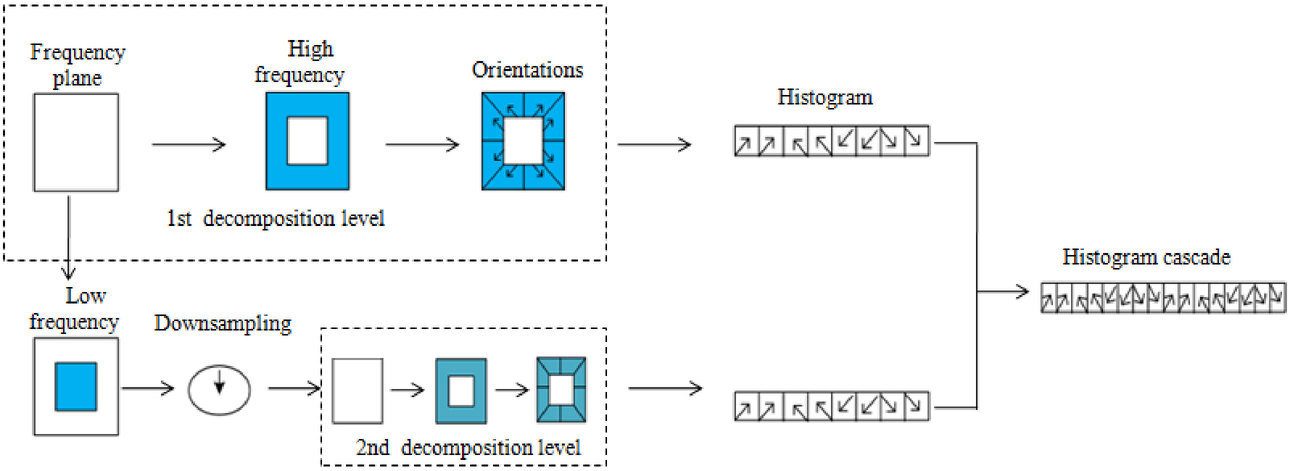

Figure 5. Feature structure of shearlet

Haar wavelet is discontinuous in the time domain, but it can decompose the image quickly, thus it is often used for wavelet decomposition in the field of face recognition to obtain the frequency domain information of face. As shown in Figure 6, for the original face image under the R color channel, after decomposition, a low-frequency approximate image containing the main information and energy of the face and a high-frequency face image with filtering in three directions including horizontal direction (H), vertical direction (V) and diagonal direction (D) can be obtained.

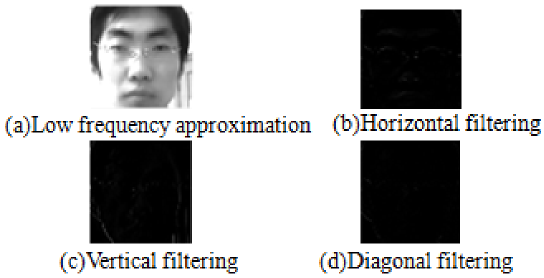

Figure 6. First level Haar filtering

It can be seen from the figure that the horizontal filter and the vertical filter can better obtain the details of the face image. Therefore, we calculate the mean (E) and variance (D) of the horizontal filter and the vertical filter respectively to form the eigenvector. The construction method of Haar wavelet features is shown in Figure 7. In order to obtain more detailed information of human face, we extract three color components (R, G and B) from the original face image, and perform Haar wavelet transform from one to four levels for each color component, we then calculate the mean value, variance of each level of horizontal filter image and vertical filter image, respectively. Finally, we obtain a total of $3 \times 4 \times 2 \times 2 = 48$ dimensional eigenvector.

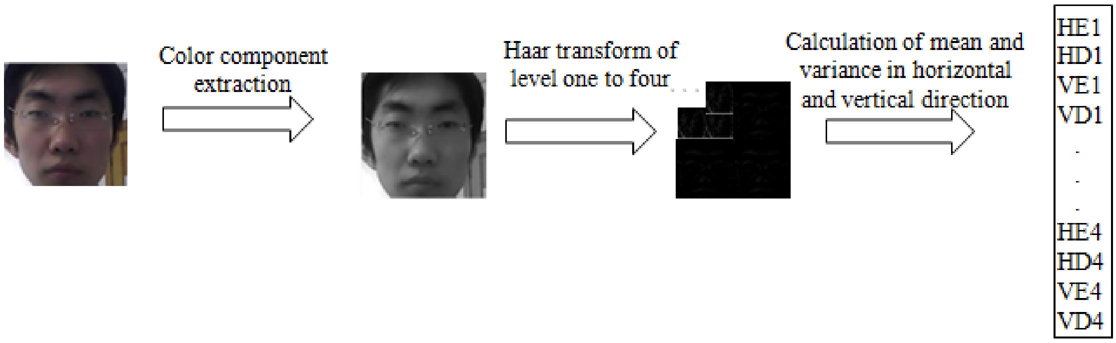

Figure 7. Haar feature extraction

After step of the extraction of three kinds of features, we apply SVM-RFE to perform filter for the obtained three cascading features, by this means, the features with high recognition scores can be reserved. The complete construction algorithm for obtaining spatial -frequency domain selection features is presented in Figure 8.

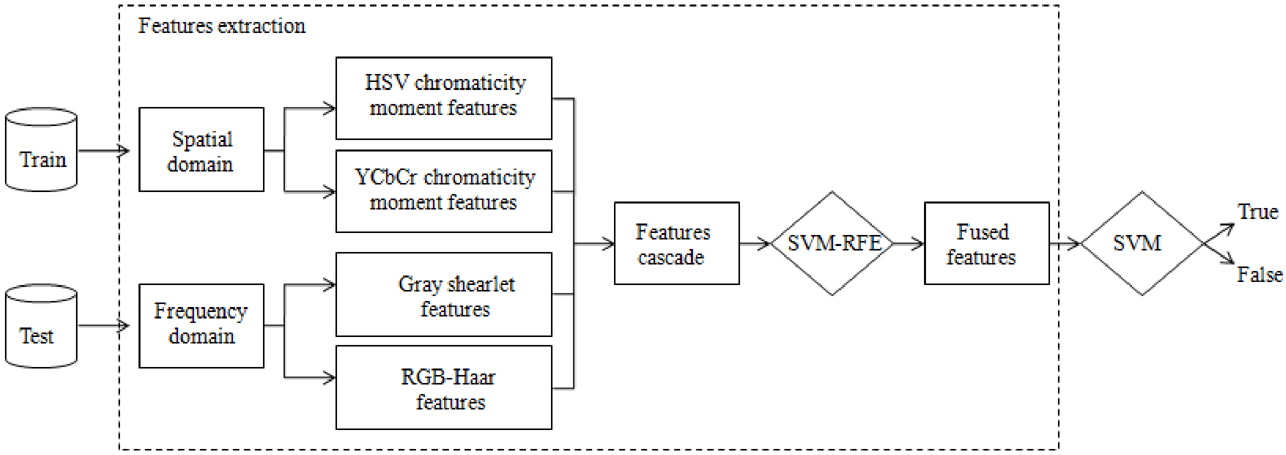

Figure 8. The construction of SFDSF

3.3 Features based on double convolutional neural network and autoencoder

Considering the integrity of RGB face global information and the significance of LBP feature face texture, we design a double network fusion face anti-spoofing detection method. More specifically, on the one hand, a shallow convolution neural network is first designed, in which RGB face is used as an input image for extracting the global features; on the other hand, R, G and B color components are extracted from RGB face image, respectively, on the basis of which the local texture features using LBP method are simultaneously extracted. After that, a multi-core convolution residual network is designed for training, by this means, the full connection layer features from above the two networks are obtained, we then use autoencoder to perform feature fusion, by design, the obtained fused features can be used for face anti-spoofing.

Meanwhile, in order to extract the depth significant convolution feature of human face, we designed two shallow neural networks named Single-Net and LBP-Resnet. Among them, Single-Net consists of three convolution layers, two pooling layers and two full connection layers; LBP-Resnet is a residual network consisting of 3× 3, 5×5 and 1×1 multi-scale convolution cores.

In fact, for obtaining the depth information of human face as completely as possible, the input of Single-Net is only acquired from the RGB face images which have been preprocessed. As shown in Table 1, it includes three convolution layers, to be more specific, conv1 has three convolution cores, each of which has size of 64×64, conv2 has thirty-two convolution cores, each of which has size of 32×32,and conv3 has eight convolution cores, each of which has size of 16×16. After each volume accumulation layer, batchnorm is used for regularization to prevent over fitting. We use the Relu function to activate and prevent the gradient from disappearing. In addition, the convolution kernel of 2×2 is used to maximize the pooling of each pooling layer to reduce the value deviation caused by the parameter error of the convolution layer.

.

Table 1. The structure of Single-Net

Layer	Kernel Size	Stride	Input Maps	Output Maps
Conv1	3×3	1	3×（64×64）	32×（64×64）
Pool1	2×2	2	32×（64×64）	32×（32×32）
Conv2	3×3	1	32×（32×32）	8×（32×32）
Pool2	2×2	2	8×（32×32）	8×（16×16）
Conv3	3×3	1	8×（16×16）	8×（16×16）
Fc1	—	—	2048	32
Fc2	—	—	32	2
Softmax	—	—	2	2

Actually, LBP-Resnet is designed for our method, which is a 45 layers lightweight residual network. In contract to the traditional residual network, in this method, the LBP texture feature extraction layer is added in the network, we also apply the 3D LBP face as the input image of the neural network. As shown in Figure 9, LBP-Resnet is composed of four residual units. Different from the general residual network, in order to better extract the texture features of LBP face images on different scales, the convolution kernels of 3×3, 5×5 and 1×1 scales are used in the same residual unit to extract the input LBP feature images. In addition to assigning 64 convolution kernels for the third residual unit, the number of convolution kernels for the other residual units is set to 32. Meanwhile, after the fourth residual unit, an identity mapping unit is added to supplement the LBP texture features that were lost in the convolution process. Finally, the fully connected layers are also added to the network for transforming the output feature map into one-dimensional feature vector.

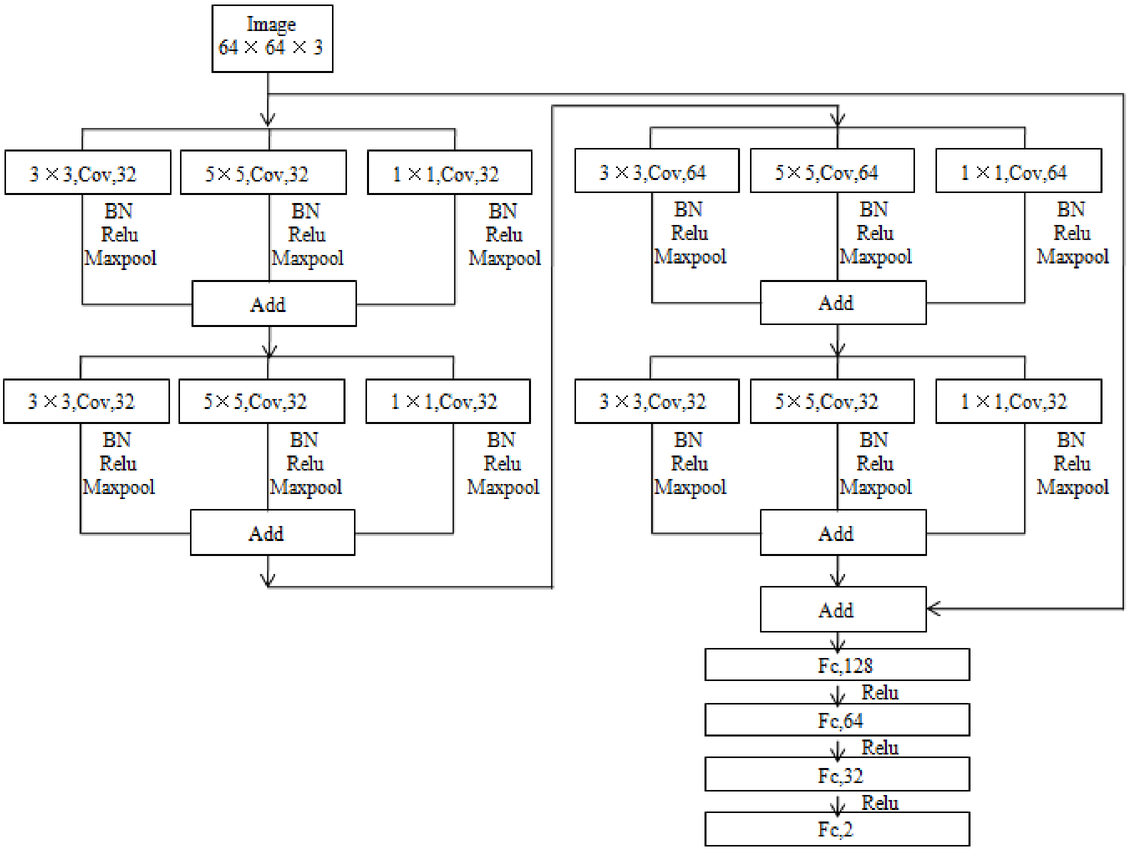

Figure 9. The structure of LBP-Resnet

After these preparations, the construction process of features based on double convolutional neural network and autoencoder (FDCNN-AUTO) is presented in Figure 10. Two shallow networks are for training RGB face map and LBP face map respectively. We take 32-dimensional full connection layer features from the two networks to cascade, and then use autoencoder to fuse the features to get the final FDCNN-AUTO.

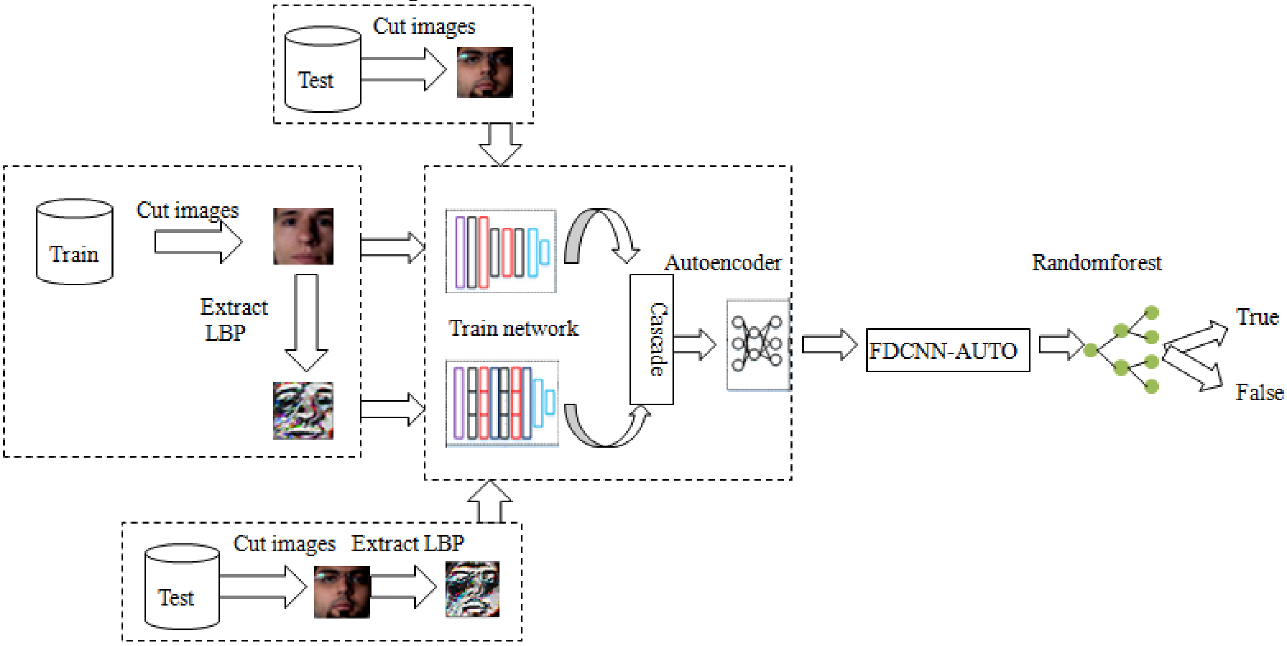

Figure 10. The construction of FDCNN-AUTO

3.4 Decision fusion method based on Least Square

In this section, three fusion features are designed and trained on three classifiers, including: AdaBoost, SVM and Randomforest. After that, the optimal weight distribution of each channel classifier will be autonomically assigned. Specifically, the predicted face labels of each feature are denoted as y_1, y_2 and y_3 respectively. Assuming that $y = [y_1, y_2, y_3]$ stands for face label matrix for three feature prediction, **w** is the weight of three prediction results, the actual label matrix of face image is denoted by **Y**, then the weighting process by least square method can be detailed as:

$$\min||yw\text{-}Y||^2 \tag{3}$$

The specific solutions of the equation can be obtained by:

$$\begin{aligned} &||\mathbf{yw} - \mathbf{Y}|| \\ &= (\mathbf{yw} - \mathbf{Y})^{\mathrm{T}}(\mathbf{yw} - \mathbf{Y}) \\ &= (\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}} - \mathbf{Y}^{\mathrm{T}})(\mathbf{yw} - \mathbf{Y}) \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathbf{yw} - 2\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathbf{Y} + \mathbf{Y}^{\mathrm{T}}\mathbf{Y} \end{aligned} \tag{4}$$

For w in (4), we derive:

$$\begin{aligned} &\frac{\partial(\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathbf{yw} - 2\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathrm{Y} + \mathbf{Y}^{\mathrm{T}}\mathbf{Y})}{\partial\mathbf{w}} \\ &= \frac{\partial(\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathbf{yw} - 2\mathbf{w}^{\mathrm{T}}\mathbf{y}^{\mathrm{T}}\mathbf{Y})}{\partial\mathbf{w}} \\ &= (\mathbf{y}^{\mathrm{T}}\mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{y})\mathbf{w} - 2\mathbf{y}^{\mathrm{T}}\mathbf{Y} \end{aligned} \tag{5}$$

We then set Eq. (5) to be zero for obtaining the weight as below:

$$w = (y^{\mathrm{T}}y)^{-1}y^{\mathrm{T}}Y$$

Therefore, the obtained w is viewed as the weight that minimizes the errors between the predicted results and the real category of face image samples. We then bring a set of optimal weights obtained from the validation set into the classification stage for the robust evaluation.

For the three feature prediction labels $\boldsymbol{y_t} = [y_1, y_2, y_3]$ in the test set, we first perform the operation on exclusive or logical discrimination with the actual test set label **Y** to obtain the discrimination matrix **L**, we then combine with the weight **w** calculated by the least square to make the final decision-making judgment on the true or false faces.

$$\boldsymbol{y_t} = \begin{cases} \mathbf{Y}, \frac{\mathbf{L}\times\mathbf{w}}{3} > \frac{1}{2}\times|\mathbf{w}_{\min}| \\ 1 - \mathbf{Y}, \text{ otherwise} \end{cases} \tag{6}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Experiment setup

Experimental results conducted on two of the most challenging anti-spoofing datasets. For Replay-Attack video dataset [22], we extract one picture every four frames, generating 23251 pictures to be training set, 30646 pictures to be test set and 23156 pictures for validation set. For CASIA FASD video dataset [23], due to the lack of validation set, we extract 20 subsets from 30 test subsets to compose of validation set, thus the designed optimal weight can be calculated by least square method. For CASIA FASD, we extract a face picture every two frames, we then obtain 22631 images in training set, 33041 images in test set and 22027 images in validation set. Note that for the experiment setting, each face image was resized to 128×128. In order to test the effectiveness of the designed algorithm, we use the three indicators of accuracy (ACC), equal error rate (EER), and half total rate (HTER) as the basis for evaluating the effectiveness of the algorithm.

Table 2. Dataset settings

Dataset	Number of videos			Number of pictures		
	Train	Test	Val	Train	Test	Val
Replay-Attack	360	480	360	23251	30646	23156
CASIA FASD	20×12	30×12	20×12	22631	33041	22027

4.2 Experimental results of color texture fusion

Table 3 and table 4 show the experimental results on Replay-Attack and CASIA FASD data sets respectively. It can be seen from the table that in the feature extraction of gradient texture based on six single color components, whether in Replay-Attack or CASIA FASD, the effect of Roberts texture feature extraction based on hue component is better, which also shows that the difference between real face and spoofing face in color hue is relatively obvious after being photographed by camera. After six color component texture features are fused by entropy weighting, the experimental results are improved obviously. The recognition accuracy on Replay-Attack is 95.21%, EER and HTER are reduced to 2.32% and 7.39% respectively. The accuracy on CASIA FASD is increased to 88.56%, EER and HTER are reduced to 7.34% and 13.60% respectively. Experiments show that it is difficult to imitate the color change of face in the light and the texture difference of the real face by photo spoofing and video spoofing.

Table 3. Experimental results of CTWF on Replay-Attack

Replay-Attack	ACC%	EER%	HTER%
H	89.71	12.60	15.77
S	83.28	23.12	25.63
V	80.15	30.32	36.51
Y	83.05	24.20	27.36
Cb	81.26	27.36	33.58
Cr	80.07	33.34	37.28
CTWF	**95.21**	**2.32**	**7.39**

Table 4. Experimental results of CTWF on CASIA FASD

CASIA FASD	ACC%	EER%	HTER%
H	83.25	18.64	20.38
S	76.38	24.88	30.02
V	69.55	33.92	37.11
Y	72.60	30.87	33.37
Cb	74.06	31.31	34.28
Cr	74.89	31.84	33.01
CTWF	**88.56**	**7.34**	**13.60**

4.3 Experimental results of spatial and frequency domain fusion

Table 5 and table 6 show the experimental results on two datasets respectively. It is clear from the table that for a single feature, the chromaticity moment feature has a better identification effect on true and false faces. The recognition rate on Replay-Attack reaches 95.23%, HTER and EER can be reduced to 5.77% and 5.23%.The recognition rate on CASIA FASD is 86.21%, HTER and EER are reduced to 16.80% and 18.15% respectively. After SVM-RFE feature selection, the three fusion features not only reduce the complexity of operation, improve the efficiency, but also improve the recognition rate. The recognition rate on Replay-Attack is 95.71%, on CASIA FASD is 86.90%, which is about 0.5% and 0.3% higher than the best recognition effect of single feature respectively. As a whole, CASIA FASD contains more spoofing modes and more interference from external light than Replay-Attack, so the recognition effect on CASIA FASD data set is worse.

Table 5. Experimental results of SFDSF on Replay-Attack

Replay-Attack	ACC%	EER%	HTER%
Chromaticity moment	95.23	5.23	5.77
Gray-Shearlet	73.54	51.11	49.85
RGB-Haar	90.03	10.78	13.82
SFDSF	**95.71**	**5.15**	**6.06**

Table 6. Experimental results of SFDSF on CASIA FASD

CASIA FASD	ACC%	EER%	HTER%
Chromaticity moment	86.21	18.15	16.80
Gray-Shearlet	65.53	50.39	49.47
RGB-Haar	80.39	22.56	23.68
SFDSF	**86.50**	**15.38**	**16.99**

4.4 Experimental results of double network fusion

Table 7 shows the results of double network fusion test on Replay-Attack and CASIA FASD. In the experiment of Replay-Attack, EER and HTER of self coding fusion on test set are 0.93% and 2.77% respectively. It can be seen that the FDCNN-AUTO method proposed by us has a better recognition effect in the complex face spoofing of high-definition video and light environment changes. On the CASIA FASD, the EER and HTER obtained by FDCNN-AUTO method are 5.06% and 4.41% respectively. In addition, compared with single network and single feature direct classification, the feature fusion method proposed by us has obvious advantages in each parameter index.

Table 7. Experimental results of FDCNN-AUTO

Network/ Parameters	Replay-Attack			CASIA FASD		
	ACC%	EER%	HTER%	ACC%	EER%	HTER%
Single-Net	96.39	4.93	4.92	93.22	7.93	6.71
LBP-Resnet	95.32	5.87	5.54	91.72	8.10	9.74
FDCNN-AUTO	**98.45**	**0.93**	**2.77**	**96.10**	**5.06**	**4.41**

4.5 The results of least square weight fusion and comparison with the state-of-art experiments

Finally, we use the least square weight fusion (LSWF) method to calculate the optimal weight of the classification results, and make decision fusion for the discrimination of the three fusion features, so that the final discrimination results are further improved. Then we compare the final results with some state-of-the-art techniques in recent years, and the detailed results are presented in Table 8. It can be seen from the table that the method using least square weight fusion of channel-based feature classifiers proposed by us have obvious advantages for the identification of true and false faces.

Table 8. Comparison with other state-of-art experiments

Method	Replay-Attack		CASIA FASD	
	EER%	HTER%	EER%	HTER%
DPCNN[24]	2.9	6.1	4.5	
Nonlinear diffusion CNN[25]	-	10	-	-
CNN LBP-TOP[20]	3.33	4.70	8.02	9.94
3DCNN[26]	0.16	0.042	-	11.33
LsCNN[27]	0.33	2.50	4.44	-
Illumination CNN[28]	-	5.50	-	3.88
LiveNet[29]		5.74		4.59
CTWF	**2.32**	**7.93**	**7.46**	**13.97**
SFDSF	**5.15**	**6.06**	**15.38**	**16.99**
FDCNN-AUTO	**0.93**	**2.77**	**5.06**	**4.41**
LSWF	**0.09**	**0.75**	**0.76**	**2.32**

5. CONCLUSION AND FURTHER WORK

In this paper, we propose a novel face anti-spoofing detection algorithm using least square weight fusion (LSWF) of channel-based feature classifiers. Different from general single feature anti-spoofing experiments, we extract a number of image features from different channel spaces, containing the color, texture, spatial domain, frequency domain and convolution features. In addition, for different channel features, we first use three different methods for feature fusion. According to the predicted results, a least square weight fusion method is used latter to assign the optimal weights of classification score fusion. The final result of our experiment is certainly comparative with the state-of-art experiments. Next, we will further improve and study the method of feature fusion.

REFERENCES

[1] Jain, Anil K., Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." IEEE Transactions on circuits and systems for video technology 14.1 (2004): 4-20.
[2] Ali, Mouad MH, et al. "Overview of fingerprint recognition system." 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, 2016.
[3] Wildes, Richard P. "Iris recognition: an emerging biometric technology." Proceedings of the IEEE 85.9 (1997): 1348-1363.
[4] Meng, Jianliang, Junwei Zhang, and Haoquan Zhao. "Overview of the speech recognition technology." 2012 fourth international conference on computational and information sciences. IEEE, 2012.
[5] Thalheim, Lisa. "Body Check-Biometric Access Protection Devices and their Programs Put to the Test." c't Magazine(2004).

[6] Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification." International Conference on Advances in Pattern Recognition. Springer, Berlin, Heidelberg, 2001.

[7] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 1. IEEE, 2005.

[8] Erdem, C. E., et al. "Combining Haar feature and skin color based classifiers for face detection." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.

[9] Sharif Razavian, Ali, et al. "CNN features off-the-shelf: an astounding baseline for recognition." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2014.

[10] Fenghua Wang,Jiuqiang Han,Xianghua Yao.”A multi-feature fusion method based on iris and face” in Journal of Xi'an Jiaotong University,February 2008, pp.133-137.

[11] Lin Sun,Gang Pan.”Real-time detection of video playback counterfeiting attacks in face recognition” in Journal of Circuits and Systems, April2010,pp.39-46.

[12] Yu, C. A. O., T. U. Ling, and Wu Li-fang. "Face Liveness Detection using Gray Level Co-Occurrence Matrix and Wavelets Analysis in Identity Authentication." Journal of Signal Processing 30.7 (2014).

[13] Po, Lai-Man, et al. "Face liveness detection using shearlet-based feature descriptors." Journal of Electronic Imaging 25.4 (2016): 043014.

[14] Zhang, Le-Bing, et al. "Face spoofing detection based on color texture Markov feature and support vector machine recursive feature elimination." Journal of Visual Communication and Image Representation 51 (2018): 56-69.

[15] Määttä, Jukka, Abdenour Hadid, and Matti Pietikäinen. "Face spoofing detection from single images using micro-texture analysis." 2011 international joint conference on Biometrics (IJCB). IEEE, 2011.

[16] Sepas-Moghaddam, Alireza, et al. "Face spoofing detection using a light field imaging framework." IET Biometrics 7.1 (2017): 39-48.

[17] Patil, Neha D., and Sujata V. Kadam. "Face spoof detection techniques: IDA and PCA." 2016 Online International Conference on Green Engineering and Technologies (IC-GET). IEEE, 2016.

[18] De Souza, Gustavo Botelho, et al. "Deep texture features for robust face spoofing detection." IEEE Transactions on Circuits and Systems II: Express Briefs 64.12 (2017): 1397-1401.

[19] Atoum, Yousef, et al. "Face anti-spoofing using patch and depth-based CNNs." 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017.

[20] Asim, Muhammad, Zhu Ming, and Muhammad Yaqoob Javed. "CNN based spatio-temporal feature extraction for face anti-spoofing." 2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2017.

[21] Li, Lei, and Xiaoyi Feng. "Face Anti-spoofing via Deep Local Binary Pattern." Deep Learning in Object Detection and Recognition. Springer, Singapore, 2019. 91-111.

[22] Chingovska, Ivana, André Anjos, and Sébastien Marcel. "On the effectiveness of local binary patterns in face anti-spoofing." 2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG). IEEE, 2012.

[23] Zhang, Zhiwei, et al. "A face antispoofing database with diverse attacks." 2012 5th IAPR international conference on Biometrics (ICB). IEEE, 2012.

[24] Li, Lei, et al. "An original face anti-spoofing approach using partial convolutional neural network." 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2016.

[25] Alotaibi, Aziz, and Ausif Mahmood. "Deep face liveness detection based on nonlinear diffusion using convolution neural network." Signal, Image and Video Processing 11.4 (2017): 713-720.

[26] Gan, Junying, et al. "3D convolutional neural network based on face anti-spoofing." 2017 2nd international conference on multimedia and image processing (ICMIP). IEEE, 2017.

[27] de Souza, Gustavo Botelho, João Paulo Papa, and Aparecido Nilceu Marana. "On the learning of deep local features for robust face spoofing detection." 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2018.

[28] Bresan, Rodrigo, et al. "Facespoof buster: a presentation attack detector based on intrinsic image properties and deep learning." arXiv preprint arXiv:1902.02845 (2019).

[29] Rehman, Yasar Abbas Ur, Lai Man Po, and Mengyang Liu. "LiveNet: Improving features generalization for face liveness detection using convolution neural networks." Expert Systems with Applications 108 (2018): 159-169.