*International Journal of Latest Research in Engineering and Technology (IJLRET)*
*ISSN: 2454-5031*
*www.ijlret.com || Volume 04 - Issue 11 || November 2018 || PP. 36-40*

# Multi-feature face liveness detection method combining motion information

Changlin LI
*(School of Computer Science and Technology, University of Science and Technology of China,*
*Hefei, Anhui, 230026, P. R. China)*

**Abstract:** Face authentication technology has achieved high stability and efficiency, and play an important role in many fields. Therefore, counterfeiting attacks or replay attacks have been a major threat to the biometric authentication system, and the identification system is vulnerable to various deliberate attacks. This paper proposes a face detection method that combines multi-feature fusion of motion information and texture information. The non-rigid motion analysis extracted the motion information of the facial part. The facial background consistency analysis obtained the correlation between the facial motion and the background motion. The multi-scale analysis obtained the difference of the dynamic texture of the image. Because each feature clearly has its own meaning, the proposed method has stronger generalization ability. Experimental results show that Long-short-term Memory Network algorithm is used to achieve better detection accuracy by using a variety of features.

**Keywords:** face recognition, face liveness detection, Multi-feature fuse, LSTM

## I. INTRODUCTION

Face recognition has been playing an important role in the era of AI, with the promotion and application of brush face payment, the popularity of face card and the release of cell phone face. However, face recognition system is still faced with many threats because of the complexity of the face recognition environment and the continuous improvement of attack technology. The counterfeiting attacks or replay attacks have been a major threat to the biometric authentication system, and the identification system is vulnerable to various deliberate attacks. In order to identify the face image from the camera is a real person, or a photo, the Counterfeit attack detection is essential to identity authentication system. Therefore, the study on the liveness detection of face authentication system to distinguish the real face from photos or video has a very important significance [1].

Research on spoofed face detection was first published by Li in 2004[2]. For different applications, the detection method of face fake attacks is also different. At present, researchers at home and abroad mainly include liveness detection based on facial image attributes, liveness detection based on facial physiological behavior, and liveness detection combined with facial image attributes and facial physiological behavior. Among them, the detection of face image attributes mainly uses three-dimensional depth information, Fourier spectrum, multi-spectral imaging technology to distinguish living faces. Such detection methods have the advantages of simple principle and strong usability, Maatta et al.[3] proposed an analysis method based on multi-scale and multi-region LBP features at the 2011 International Joint Conference on Biometric (IJCB). The method first extracts the multi-scale and multi-region LBP features from the image[4], combines them into a 833-dimensional feature vector, and then feeds them into a binary SVM classifier for classification. The liveness detection of human face physiological behavior mainly judges the living face through human-computer interaction and blinking judgment. This type of detection method has high accuracy but can only be realized in a specific environment, and some liveness face detection methods need device that add characteristics, such as an infrared imaging device. K. Kollreider et al.[5] use the interactive liveness detection method, the system requires the user to make some simple actions, read some of the numbers, etc., and analyze whether the lip movements are consistent to determine whether it is a living face. In addition to the different liveness detection methods described above, there are a wide variety of studies that combine a variety of different methods to improve detection accuracy. Tronci et al.[6] proposed combining motion information (blinking) and texture information to perform liveness detection. There is a certain complementarity between the two mechanisms. After fusion, the detection of the living body becomes more robust.

In this paper, the liveness face detection technology is deeply researched, and the existing algorithms for extracting face features are analyzed, and their advantages and disadvantages are compared. Based on various advanced algorithms, this paper proposes a face detection method that combines multi-feature fusion of motion information and texture information, and conducts experiments on various public databases. The experimental results show that video faces, photo face and real face can be distinguished very effectively.

Finally, we tested the accuracy of different classifiers for feature classification. Experiments show that by using a variety of features, Long-short-term Memory Network algorithm is used to achieve better detection accuracy.

## II. NON-RIGID MOTION ANALYSIS

The essence of facial non-rigid motion analysis is to find non-rigid motion patterns in the local facial region. For a fixed front image, this is quite easy because there is only non-rigid motion. However, in practice, the face may also exhibit rigid motion, such as head shift or rotation. Therefore, it is necessary to extract the non-rigid motion of the face region from the global rigid body motion of the face, where a batch image alignment method is used to separate the non-rigid motion and the rigid motion. Batch image alignment uses a series of rigid transformations to align several images with a fixed image, with residuals being non-rigid motion. Inspired by the literature [7], this paper uses the low rank matrix decomposition method to detect the position of the face in each frame of the video and roughly align according to the position of the detected eyes, and then let them form a matrix $I = [I_1, I_2, \cdots, I_n]$,where $I_i$ is generated by raster scanning the original input two-dimensional image to a one-dimensional image, and n is the number of frames in the video. The basic appearance of the faces in these frames should be highly correlated, even identical. However, considering the presence of both face stiffness and non-rigid motion, the expression of I is as follows:

$$I = I^0 \circ \tau + E \qquad （1）$$

Among them，$I^0$ is a basic low rank matrix, $\tau$ is an affine transformation matrix, which is used to simulate rigid facial motion of facial pose and position change, and E is a sparse residual caused by non-rigid facial motion. Each frame $I_i$has a corresponding transformation matrix $\tau_i$， which is used to convert $I_i^0$ into $I_i$， while a non-rigid motion that cannot be modeled by $\tau_i$ is considered as residual $E_i$。 Our goal is to recover these components and find true non-rigid facial motion information from residual E. Since $I^0$ is basically low rank and E is sparse (face motion always occurs in some local areas, such as eyes and mouth), the problem can be modeled as the following optimization problem:

$$\{I^0, E, \tau\} = argmin\|I^0\|_* + \lambda\|E\|_1, I \circ \tau = I^0 + E \qquad （2）$$

Among them, $\|*\|_*$ is kernel norm, $\|*\|_1$ is $\ell^1$ norm.

For a true face sequence, there are some frames with large sparse residuals in the eye area that correspond to the blink or the mouth area corresponding to the mouth motion. There may be some residuals in the fake face video, however, they have evenly distributed noise throughout the face area, and by defining the area features, the two residuals can be easily distinguished.

Once the sparse non-rigid motion matrix E is obtained, we can use it to extract feature information that distinguishes between false faces and real faces. We mark the areas that need to be focused on as $\Omega_i$。 In the experiment, the area of focus is the eye area, the left eye area is recorded as $\Omega_1$， and the right eye area is recorded as $\Omega_2$。 The facial non-rigid motion feature $s_i, i = 1{,}2$ can be expressed by

$$S_i = \max_{j=1,2,\cdots,n}\left\{\frac{\sum_{(x,y)\in\Omega_i}|E(x,y)|}{\sum_{(x,y)\in\mathrm{Face}_j}|E(x,y)|} \times \frac{\sum_{(x,y)\in\mathrm{Face}_j} 1(x,y)}{\sum_{(x,y)\in\Omega_i} 1(x,y)}\right\} \qquad （3）$$

Where $\sum_{(x,y)\in\Omega_i}|E(x,y)|$ is the residual of the jth frame in the region$\Omega_i$, $\sum_{(x,y)\in\mathrm{Face}_j}|E(x,y)|$is the residual of the jth frame in the face area. $\sum_{(x,y)\in(\cdot)} 1(x,y)$is the area of the area used to eliminate the influence of the regional scale. If the $S_i$ value of a certain face sequence is relatively large, the corresponding face is considered to be a real face, otherwise it is considered to be a non−real face. The feature obtained from the non−rigid motion analysis is $(S_1, S_2)$, because the feature is not only the dimension is small, only two dimensions, and there is better generalization ability, from the training of a small number of samples Can be obtained. By adding additional areas, such as the mouth, the method can also detect other facial non-rigid motions.

## III. FACE BACKGROUND CONSISTENCY ANALYSIS

In practical applications, only non-rigid motion detection is not enough, because some real human faces may only have some small motions that cannot be detected by batch image alignment, so in this section, we consider the facial background consistency feature as complementary to non-rigid motion detection methods.

Facial background consistency is based on the fact that if the target face is real, its motion should be completely independent of the background motion; otherwise the motion relationship between the face and the background cannot be completely independent due to the limitations of the display medium.

First, you need to capture the scene motion. Since accurate motion information is not needed, the background modeling method based on Gaussian Mixture Model (GMM)[8] is used to describe the motion in the scene. The reason for using GMM instead of dense optical flow is that GMM is more efficient and robust to illumination and noise. In GMM, scene motion is described in an indirect but effective manner and works well in the following experiments. After the initialization is completed, the GMM outputs a foreground background binary image $B_i, i = 1, \cdots, n$ of a given frame $I_i$

$$B_i = \begin{cases} 1 & if\ (x, y) \in foreground \\ 0 & ohterwise \end{cases} \tag{4}$$

First, face detection is performed on each frame of image, and the face and background areas of each frame are detected. After each frame passes the face detection, the face and the background area can be roughly separated according to the position of the face. Then, in the face area and the background area, the trend of motion in the area as a function of time is analyzed. If the movement trend of the face area and the background area is very close, or the height is consistent, it can be judged to be a non-real face, and if the consistency is low, the face is independent of the background and can be regarded as a real face. Define $mt_j, j = 1{,}2$ to describe the trend of motion

$$mt_j = \frac{\sum_{(x,y)\in\Omega_j} B_i(x,y)}{\sum_{(x,y)\in\Omega_j} 1(x,y)} \tag{5}$$

Where $\sum_{(x,y)\in\Omega_j} B_i(x,y)$ is the number of foreground pixels, $\sum_{(x,y)\in\Omega_j} 1(x,y)$ is the area of the area $\Omega_j$. In this section, $\Omega_1$represents the face area and $\Omega_2$ represents the background area. The above equation represents the motion of the scene by calculating the motion area of the foreground.

The motion trend consists of a series of vectors, including the face region motion $mt_1(i)$ in each frame, and the background region motion $mt_2(i)$. The distance between $mt_1$ and $mt_2$ is presented as a chi-square distribution, which is used herein as the motion consistency distance D.

$$\mathrm{D} = \sum_{i=1}^{n} \frac{(mt_1(i) - mt_2(i))^2}{mt_1(i) + mt_2(i)} \tag{6}$$

Since the motion in the real face video is independent of the background, the motion consistency distance in the real face video is much larger than the motion consistency distance in the non-real face. The value of D is used as the first dimension feature of the motion consistency of the face background.

The trend of the movement of the face area and the background area in the real face video is very different, and the movement trend of the face area and the background area in the handheld fake face video and the fixed false face video is very similar.

The motion in the hand-held fake face video is global, while the motion of the real face is local, so the total motion trend can be utilized as another detection feature. This motion entropy $me(i), i = 1, 2, \cdots, n$ is defined as follows:

$$me(i) = -p_i \log(p_i) - (1 - p_i)\log(1 - p_i) \tag{7}$$

Where $p_i$ is the total foreground ratio. In some cases $p_i = 0$， then $me(i)$cannot be calculated, so you need to use $p_i = \min\{p_i, 1/\mathrm{n}\}$（n is the total pixel point），smooth it. Next, the distribution histogram of the motion entropy is used to represent the motion entropy distribution of each frame in the video. The motion entropy distribution of real face and false face is obviously different. The motion entropy of false face is close to 0, and the motion entropy of real face video is very large. Motion entropy is the second dimension feature of the consistency of facial background motion.

## IV. LIVENESS FACE DETECTION BASED ON MULTI-FEATURE FUSION

The features proposed in the first two sections and the multi-scale analysis features are complementary in practical applications. Non-rigid motion analysis can classify real faces and fake faces in face videos with facial expressions, while face background consistency analysis can classify real faces and fake faces in videos with complex backgrounds. For background-clean images, facial background consistency analysis is ineffective, but multi-scale analysis is effective because the background has no effect on the method. In the experiments below, we utilized different detection features for different image backgrounds. For video with complex backgrounds, non-rigid motion and facial background consistency analysis are used, while for particularly clean backgrounds, non-rigid motion and multi-scale analysis are used. Since the background conditions can be easily estimated by edge detection, the classification of the background is completely adaptive.

In this paper, we proposed the LSTM network to solve the problem of gradient explosion and disappearance that may appear in tradition RNN. LSTM adds a new state variable to the traditional RNN to record the long-term state of the input raw data, called the unit state. In order to control the cell state of the unit better, LSTM adds three control gate, the input gate, output gate, and forget gate.

After the above process, we can summarize the steps of the LSTM networks:

(1). Input the original face data set, the original data enters the input layer, and after the data is calculated by the excitation function, output the output vector of the input layer .

(2). Take the output result vector of step (1), the output vector of the hidden layer at the previous moment, and the state information vector retained by the cell at the previous moment as input, and input them to the initial node of the hidden layer.

(3). After the input of the initial state, the processing of the input gate, the output gate, the forget gate, and the cell is performed, and the corresponding result data vector is output according to the opening and closing state of each gate.

(4). If it is the last hidden layer, the output result of step (3) is used as the final output result of the hidden layer. If it is not the last layer, the output result of step (3) is used as the data input of the next hidden layer, and continue with calculations.

(5). Take the final output of step (4) as input, input it to the output layer, and output the result vector of the output layer after processing by the hidden neuron inside the output layer.

(6). Compare the final output vector of the output layer with the expected output, and calculate the corresponding bias based on the equation.

(7). The back propagation process begins, and the weight of each gate is constantly updated based on the actual output and the expected result bias.

(8). Repeat step (7) until the actual output result and the expected result bias value meet the expected requirements and reach the ideal bias value, stop the iteration.

## V. EXPERIMENTAL RESULT

In this section, two databases are used to verify the proposed algorithm. The first is the NUAA database, and the second is the Iadiap replay attack database, which is more challenging for the generalization capabilities of the algorithm. Real face video is captured in different backgrounds under different lighting conditions. The fake face video is printed on A4 paper by a color laser printer and held or fixed in front of the acquisition device. In the experiment, 70% of the videos were selected as the training set, and 30% of the videos were used as the test set for the development set and other videos. Experiments were carried out on the NUAA database using Non-Rigid Motion Analysis (NRMA), Facial Background Consistency Analysis (FBCA), Multi-Scale Dynamic Analysis (MSDA) and the proposed method. The results were also compared with the other two algorithms LBP- TOP + SVM [9], MTA + linear kernel[10] were compared, and the test results are shown in Table 1. Experiments were carried out on the PRINT-ATTACK database. The results were also compared with partial least squares (PLS)[11], mixed model (MTA)[8]. The test results are shown in Table 2.

Table 1. Comparison of experimental results on the NUAA database

| Method | Accuracy |
|---|---|
| NRMA | 81.52% |
| FBCA | 90.06 |
| MSDA | 88.74% |
| The proposed method | 92.37% |
| LBP-TOP + SVM[9] | 89.32% |
| MTA+ linear kernel[10] | 88.83% |

Table 2. Comparison of experimental results on the PRINT-ATTACK database

| Method | Background | Accuracy |
|---|---|---|
| NRMA | All background | 90.04% |
| FBCA | Complex background | 97.55% |
| MSDA | Uniform background | 97.62% |
| The proposed method | All background | 100% |
| PLS[11] | All background | 99.375% |
| MTA[8] | All background | 100% |

The experimental results of these algorithms are shown in Tables 1 and 2. As can be seen from the table, each feature can obtain relatively good experimental results alone. The non-rigid motion analysis extracted the motion information of the facial part. The facial background consistency analysis obtained the correlation between the facial motion and the background motion. The multi-resolution analysis examined the difference of the image texture. The experimental results show that the proposed method achieves the best detection accuracy in two standard databases, which verifies the effectiveness of the proposed method.

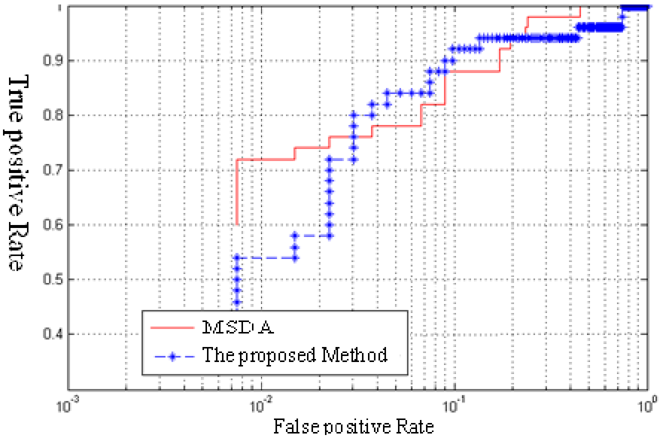

Figure 1. Liveness face detection ROC curve

It can be seen from Figure 1 and the above table that the algorithm achieves better experimental results by fusing various features. Since each feature clearly has its own meaning, the algorithm has stronger generalization ability.

## VI. CONCLUSION

By combining multi-scale dynamic analysis and operational information analysis methods, this paper proposes a more generalized face detection algorithm. Motion information analysis includes non-rigid motion analysis and face background consistency analysis. Because each feature clearly has its own meaning, the algorithm has stronger generalization ability. The experimental results show that by combining various features, Long-short-term Memory Network algorithm is used to achieve better detection accuracy.

## REFERENCES

[1]. Santosh T, Norman P, David W, et al. Detection of Face Spoofing Using Visual Dynamics. *IEEE Transaction on Information forensics and Seturity.10(4),*2015:762-777.

[2]. Li,J.W. Y,H,Wang. T,N,Tan.et al. Live face detection based on the analysis of fourierspectra.*International Society for Optics and Photonics In Defense and Security*,2004:296-303.

[3]. Maatta J, Hadid A, Pietikainen M. Face spoofing detection from single images using micro-texture analysis. *Proceedings of the Biometrics (IJCB), 2011 International Joint Conference on*, 2011: 1 - 7.

[4]. Ojala T, Pietik01inen M, M01enp T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 24(7),*2002: 971 - 987.

[5]. K.Kollreider. Real-Time Face Detection and Motion Analysis With Application in “Liveness” Assessment. *IEEE Transactions on Information Forensics and Security,2(3-2),*2007:548-558

[6]. Tronci R, Muntoni D, Fadda G, et al. Fusion of multiple clues for photo-attack detection in face recognition systems. *Proceedings of the Biometrics, International Joint Conference on*, 2011: 1-6.

[7]. PENG Y, GANESH A, WRIGHT J, et al. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 34(11),* 2012: 33-46.

[8]. STAUFFER C, GRIMSON W E L. Adaptive Background Mixture Models for Real-Time Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999

[9]. PEREIRA T D F, ANJOS A, MARTINO J M D, et al. LBP － TOP Based Countermeasure against Face Spoofing Attacks; *proceedings of the International Conference on Computer Vision*, 2012.

[10]. MAATTA J, HADID A, PIETIKAINEN M. Face spoofing detection from single images using micro-texture analysis; *proceedings of the International Joint Conference on Biometrics*, 2011.

[11]. SCHWARTZ W R, ROCHA A, PEDRINI H. Face spoofing detection through partial least squares and low-level descriptors; *proceedings of the Biometrics (IJCB), 2011 International Joint Conference on*, 2011.