Received February 5, 2018, accepted March 5, 2018, date of publication March 13, 2018, date of current version May 2, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2815540

Multimodal Feature-Level Fusion for Biometrics Identification System on IoMT Platform

YANG XIN[1], LINGSHUANG KONG[2], ZHI LIU[2], (Member, IEEE), CHUNHUA WANG[3], HONGLIANG ZHU[1], MINGCHENG GAO[1], CHENSU ZHAO[1], AND XIAOKE XU[4]

[1]Centre of Information Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]School of Information Science and Engineering, Shandong University, Jinan 250100, China
[3]China Changfeng Science Technology Industry Group Corporation, Beijing 100854, China
[4]College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China

Corresponding author: Zhi Liu (liuzhi@sdu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61603073, in part by the National Key Research and Development Program of China under Grant 2017YFB0802300, in part by the Applied Sci-Tech Research and Development Special Fund Program of Guangdong Province under Grant 2015B010131007, in part by the Shandong Provincial Natural Science Foundation, China under Grant ZR2012FZ005, in part by the Key Research and Development Plan of Shandong Province under Grant 2017CXGC1503, and in part by the Fundamental Research Funds of Shandong University under Grant 2015JC038.

ABSTRACT Biometric systems have been actively emerging in various industries in the past few years and continue to provide higher-security features for access control systems. Many types of unimodal biometric systems have been developed. However, these systems are only capable of providing low- to mid-range security features. Thus, for higher-security features, the combination of two or more unimodal biometrics (multiple modalities) is required. In this paper, we propose a multimodal biometric system for person recognition using face, fingerprint, and finger vein images. Addressing this problem, we propose an efficient matching algorithm that is based on secondary calculation of the Fisher vector and uses three biometric modalities: face, fingerprint, and finger vein. The three modalities are combined and fusion is performed at the feature level. Furthermore, based on the method of feature fusion, the paper studies the fake feature which appears in the practical scene. The liveness detection is append to the system, detect the picture is real or fake based on DCT, then remove the fake picture to reduce the influence of accuracy rate, and increase the robust of system. The experimental results showed that the designed framework can achieve an excellent recognition rate and provide higher security than a unimodal biometric-based system, which are very important for a IoMT platform.

INDEX TERMS Multi-model fusion, fisher vector, liveness detection, personal identification, IoMT.

I. INTRODUCTION

Biometric identification systems have been widely used since their reliability was proven. They exhibit many advantages compared to other traditional identification systems, such as keys and passwords that are not easy to falsify or lose. Although biometric techniques seem to be very powerful, currently, they are incapable of guaranteeing a high recognition rate with unimodal biometric systems that are based on a unique biometric signature or resource. In addition, these systems are often affected by the following problems [1], [2]: noise that is generated by the sensor, non-universality, lack of individuality, lack of invariant representation and sensitivity to attacks. Especially, for some internet of medical things (IoMT) platform, the high performance of identification and liveness detection is very important. Biometrics system with multimodal fusion can be applied to address this problem and can be implemented on the IoMT platform.

Recently, multiple researchers have considered the fusion of two or more biometrics for improving recognition performance and authentic security [3]–[5]. Because of its convenience and the low cost of acquisition, multimodal biometric recognition, which mainly relies on the fusion of multiple features that are captured by sensors, has become more popular. Such systems are classified into the following two groups: (1) systems that are based on single-spectrum images, such as fingerprint and palmprint [6]–[9], palmprint and hand shape, palm vein and hand geometry, finger vein and finger shape, and gait and body structure; and (2) systems that are based on multi-spectrum images [10], [11], such as fingerprint and finger vein, finger vein and finger dorsal texture, palmprint and palm vein, and visible face image and infrared face image. However, it is difficult to isolate the necessary features in single-spectrum images and collecting multi-spectrum images requires an expensive device.

In this paper, we propose multimodal biometric recognition based on the fusion of the face, fingerprint and finger vein. The fusion is based on the following considerations: (i) Face, fingerprint and finger vein recognition are adscititious [12]–[15]. Fingerprint and face recognition have high recognition rates, but they are sensitive to the quality of the collected images, which can be low because fingerprints are often damaged and faces can be forged. Different from fingerprint and face recognition, finger vein recognition makes use of features that are inside finger rather than on the finger surface. Thus, finger vein images are difficult to forge. However, the features that are extracted from finger veins are not as precise as those from fingerprints. In view of the advantages and disadvantages, fusing faces and fingerprints with finger veins can compensate effectively for the lack of a single biometric feature [16], [17]. (ii) An integrated device can be designed. Both fingerprint and finger vein recognition systems obtain data from fingers. Thus, an integrated device can be designed that avoids high costs and complex collection using multiple capture devices [18].

This paper is organized as follows: In section 2, we will review fusion levels of multimodal biometrics. In section 3, we will introduce methods of fingerprint recognition, finger vein recognition, and face recognition, our feature-level fusion methods and liveness detection, followed by the experimental evaluations in section 4. The conclusions will be presented in section 5.

II. FUSION LEVEL OF MULTIMODAL BIOMETRICS

In a multimodal biometric system that is based on multiple biometric traits, various levels of fusion are possible: fusion at the feature level, matching score level, and decision level.

A. FEATURE-LEVEL FUSION

Feature-level fusion is the fusion of feature vectors that are obtained from several feature sources: (i) feature vectors that are obtained from different sensors based on a single biometric; (ii) feature vectors that are obtained from different entities based on a single biometric, such as iris feature vectors that are obtained from left and right eyes; and (iii) feature vectors that are obtained from multiple biometric traits.

Feature vectors of different types that are compatible can be combined to form high-dimensional feature vectors. For example, Ryan and Connaughton [8] fused face features with iris features, Angadi and Hatture [13] proposed the fusion of hand geometry features and palmprint features. They trained on face images with high resolution and gait energy images by Principal Components Analysis (PCA) and Multiple Discriminate Analysis (MDA), respectively, to obtain feature vectors. Then, they combined fused them into a single vector after normalization. However, it is difficult to consolidate information at the feature level because feature sets from different biometric modalities may be neither accessible nor compatible [19].

B. SCORE LEVEL FUSION

Score-level fusion is fusion on the matching score level. For this reason, it is also called matching-level fusion. Different matching scores that are obtained by different classifiers or from different biometrics can be fused for matching at this level. Fusion at the matching level can be approached in two distinct ways [17]: as a classification problem and as an information combination problem. In the classification approach, a feature vector is reconstructed using matching scores that are output based on individual matches. Then, these feature vectors are classified into "Accept" (genuine user) or "Reject" (impostor) categories. In the information combination approach, individual matching scores are fused to generate a single scalar score that is used to make the final decision. Note that the individual matching score should be normalized to a uniform field before fusion.

Kabir *et al.* [20] proposed the fusion of face profile and gait at the matching score level. They obtained side face features (EFSI) and gait features (GEI) from a video. Then, they fused them at the matching level using three strategies. Kabir *et al.* [21] proposed matching-level fusion of fingerprint verification using a minutiae matching algorithm and a texture matching algorithm.

C. DECISION LEVEL FUSION

Fusion at the decision level refers to the fusion of matching results that are outputted from different sets of matches for making the final decision. Several fusion strategies are used for decision-level fusion, such as majority voting, Bayesian inference, weighted voting based on Dempster-Shafer theory [18], and AND or OR logical rules.

In theory, the earlier information is combined, the better the achieved results. However, there are many obstacles to achieving satisfactory results from feature-level fusion in practice. The main difficulties are as follows: (i) feature spaces of different biometric features are unknown in most cases, (ii) data sets at the feature level may be incompatible, and (iii) connecting two feature vectors may lead to dimension redundancy due to the substantial increase in size of the fused feature vectors. Consequently, few fusion studies have been conducted at the feature level [22]. Most fusion studies are carried out at the matching level or the decision level [23].

III. PROPOSED METHOD

The fusion of face, fingerprint and finger vein in this paper is performed at the feature level. A block diagram of our fusion method is shown in Fig 1. It mainly consists of four components: face recognition, fingerprint recognition, finger vein recognition, and feature-level fusion.

(i) The image acquisition device is used to collect the user's face, fingerprint and finger vein images, and the images are preprocessed. (ii) The PCA feature of the face image is taken as Feature 1; the breakpoint and the bifurcation point features

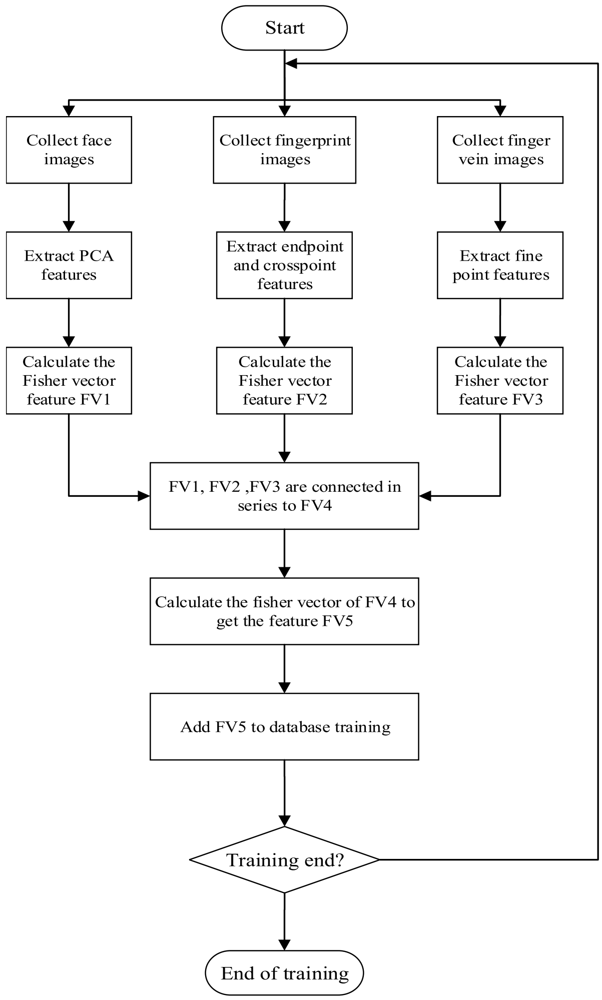


FIGURE 1. **Method flow diagram.**

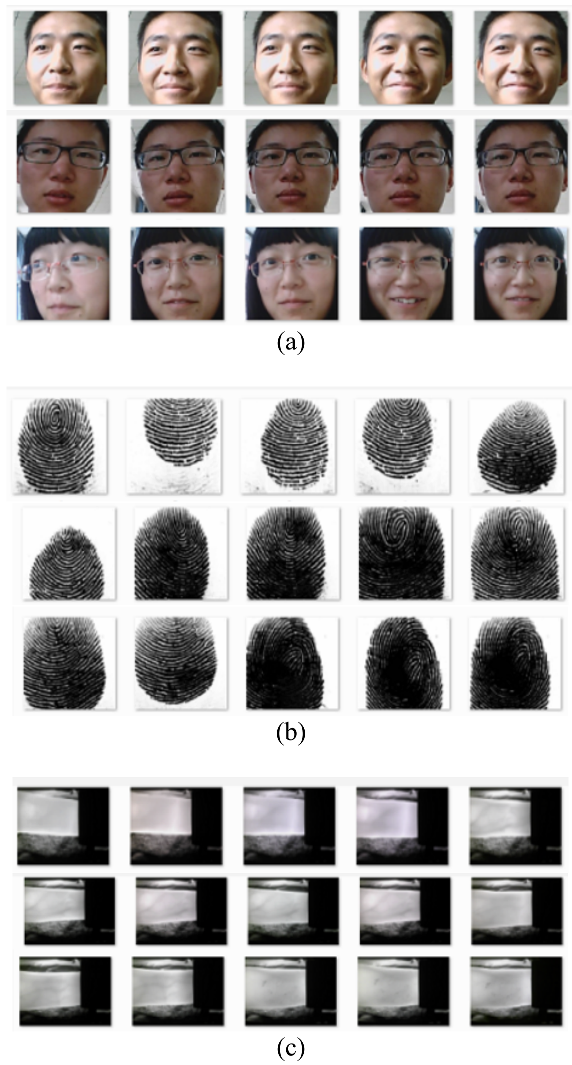


FIGURE 2. **original samples.**

of the extracted fingerprint image are called Feature 2; the fine node of the extracted vein is taken as Feature 3. (iii) The three features of the Fisher vector that were obtained in the second step are recorded as FV1, FV2 and FV3, respectively. (iv) FV1, FV2, and FV3 are concatenated to obtain a new eigenvector and the Fisher vector is calculated for this vector; the newly obtained Fisher vector is denoted as FV4. (v) FV4 is sent to the classifier to be trained to generate the training feature library.

A. FINGERPRINT RECOGNITION

Fingerprint recognition mainly consists of image preprocessing, feature extraction, and matching. To extract fingerprint minutiae accurately, we perform a series of preprocessing steps, including calculation of the orientation field, enhancement, and fingerprint thinning. For fingerprint segmentation, this paper uses the variance method, which is the simplest method, while methods from the literature are used for the calculation of the fingerprint orientation field, enhancement, and finger thinning.

Fingerprint matching based on the point pattern can be approached in two stages: One is preprocessing, including image enhancement, image binarization, refinement, etc. The other is extraction of the detail points and removal of the pseudo-detail points.

In the second stage, the generalization node is extracted, and a 3∗3 template is used to confirm the type and location of the detail points. The template is shown in Fig 2 below. Where N is the test point, N1, N2, ..., N8 is in order, with a pixel value of 0 or 1.

N4	N3	N2
N5	N	N1
N6	N7	N8

The intersection number of the measured point N is:

$$C_N = \frac{1}{2}\sum_{k=1}^{8} |R(k+1) - R(k), R(9) = R(1)| \tag{1}$$

When C_N is 1, it is judged to be a bifurcation point when it is determined as the ridge line endpoint and C_N is 3.

Nodes of fingerprint image are extracted the characteristics because of the influence of noise or itself quality, makes these features are not ideal endpoint and bifurcation point, especially image edge also can produce false endpoint due to image segmentation. Step is to remove the false feature points, iterate through all the feature points, first remove the boundary points, then along the ridge line tracking, set a certain step length, if in this step, find another feature points, then to determine the point of burr or short ridge also remove, and then find the bifurcation point, if within a certain neighborhood, and found another branch point, then to filter it. The remaining feature points are retained as the feature point sets we are going to use.

In this paper, the algorithm based on point mode is used to match, which is to judge the distance between two nodes by using the geometric relation, to measure similarity by matching scores, and to set threshold value to judge.

B. FINGER VEIN RECOGNITION

Similar to fingerprint recognition, finger vein recognition consists of image preprocessing, feature extraction and matching. Image normalization, enhancement, and segmentation are all image preprocessing tasks.

Refers to the finger vein image acquisition quality is very important, a good image quality will be the follow-up algorithm and has a great influence on the final accuracy, so a simple and effective acquisition equipment is very important.

(i) Selection of light source. The wavelength of incident light is generally selected between 720 and 1100 nm, and a clearer vein can be collected. In general, we use infrared emitting diode (LED) to make the image clearer.

(ii) Selection of imaging sensors. CCD is a commonly used photosensitive device. It has high signal-to-noise ratio, sensitivity and good dynamic range. However, due to the complex production process and low production, the cost is high and the device is expensive. However, CMOS devices can be mass-produced and have lower price and better imaging quality. Therefore, it is more appropriate to choose CMOS in finger vein acquisition equipment.

In addition to the above mentioned light source and imaging sensors, also need to filter to eliminate the outside the disturbance of visible light, and the need to design a simple finger positioning device, to a fixed finger position, solve the problem of image registration.

At the time of acquisition, the light intensity distributions of different finger vein images are different. Thus, normalization of image size is necessary [24]. We use a linear gray-scale adjustment of the gray normalization method, as follows:

$$p(i,j) = \frac{p'(i,j) - \min}{\max - \min} \times 255 \tag{2}$$

The ridge wave filter is used to enhance each finger vein image, thereby reducing noise; its form is as follows:

$$G(x,y,\theta,\mu,\sigma) = \frac{1}{2x\sigma^2}\exp\{-\frac{x^2+y^2}{2\sigma^2}\} \times \exp\{2\pi i(\mu x\cos\theta + \mu y\sin\theta)\} \tag{3}$$

where, μ is the frequency, θ is direction of the filter and σ is the standard deviation.

For image segmentation, we use the Niblack method. Its basic strategy is to apply function (4) on a certain image area to obtain the threshold and then perform binarization.

$$T(x,y) = m(x,y) + k \times s(x,y) \tag{4}$$

$$m(x,y) = \frac{1}{r\times r}\sum_{i=x-\frac{r}{2}}^{x+\frac{r}{2}}\sum_{j=y-\frac{r}{2}}^{y+\frac{r}{2}} f(i,j) \tag{5}$$

$$s(x,y) = \sqrt{\frac{1}{r\times r}\sum_{i=x-\frac{r}{2}}^{x+\frac{r}{2}}\sum_{j=y-\frac{r}{2}}^{y+\frac{r}{2}} f^2(i,j)} \tag{6}$$

where (x, y) is the pixel, $m(x, y)$ is mean of the pixel, $s(x, y)$ is the variance, k is the correction factor, $f(I, j)$ is the gray value, $T(x, y)$ is the resulting threshold.

The feature extraction of venous features is characterized by detail points, which not only occupy small space, but also describe it well. Its specific process is, first of all, refers to the vein image skeleton do processing, through to the shape of an eight pixels neighborhood minutiae to determine the endpoint or bifurcation points, for a $3 * 3$ area.

P1	P2	P3
P8	P0	P4
P7	P6	p5

If the value of P0 is 1, Ntrans represents the number of alternation between 0 and 1 from P1 $\sim$ P8. When N$\geq$6, it is considered as the intersection, and the expression is as follows:

$$N_{trans} = \sum_{i=1}^{8}|p_{i+1} - p_i|, \quad p_9 = p_1 \tag{7}$$

Similarly, this method is also used to detect endpoints.

C. FACE RECOGNITION

Face recognition mainly includes four steps: face detection, image preprocessing, feature extraction and identification. The purpose of face detection is to detect the effective face part from the input image, and extract the face part from the background independently.

The preprocessing of image mainly includes: denoising, normalization of gray scale, normalization of scale and Angle processing. The goal of scale normalization is to unify the face images into the same size, so that the positions of the eyes and nose are basically unified. The purpose of the normalization of gray scale is to eliminate the influence of illumination on image and improve the system recognition rate by means of ray compensation or image enhancement.

The feature extraction method can be divided into two types, including knowledge based and statistical learning methods. Feature extraction based on the knowledge of the shape of the main organ according to one's face, and the distance between the parts as feature information of recognize faces, namely geometric features, is the most traditional feature extraction methods, but is usually combined with other algorithm can achieve better results. At present, the extraction method based on statistical learning is the main method of feature extraction in face recognition and can achieve better results.

We use the PCA algorithm to extract feature images, which uses dimensionality reduction to transform multidimensional data into processed low-dimensional data [25]. The following steps are carried out: (i) Image denoising, grayscale normalization, scale normalization and angle processing are performed. (ii) The N images that are collected are denoted as $X_1, X_2, X_3, \ldots, X_n$, and the average image is expressed as follows:

$$X_{ave} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{8}$$

The difference between each image and the average image is:

$$X_i' = X_i - X_{ave} \tag{9}$$

The covariance matrix is:

$$C = \frac{1}{N}\sum_{i=1}^{N} X_i'(X_i')^T \tag{10}$$

It can be seen from the formula that the obtained coordinate system is composed of the corresponding vector of nonzero eigenvalues of covariance. (iii) The resulting feature vector is expressed as a low dimension of the face image.

D. FEATURE LEVEL FUSION

Features level fusion by extracting multiple biological effective vector, the feature information of different modal vector to a certain way together, forming a new feature information vector, to incorporate the new vector as a biometric system input, and then compared with the characteristic vector of training, features matching, eventually determine the results.

In this paper, the three-dimensional features of the face, fingerprint and finger vein are extracted, and their dimensions are different [26], [27]. The traditional feature fusion approach cannot express the intrinsic relationship among the features. To solve this problem, this paper uses the scheme of calculating the Fisher vector and the Gaussian mixture model to unify the dimensions of the different features and obtain a more accurate and effective description of the original image.

We try to solve the problem of the above-mentioned two aspects, one is by solving fisher vector ideas fully excavate internal relations between different characteristics, through the characteristics of deep excavation on the description of the target more accurately and more effectively; Secondly, through the establishment of the Gaussian mixture model to unify the Gaussian model, mean and variance of the weight, solved the problem of the different feature dimensions.

The local feature set of the input image is recorded as $X = \{x_t, t = 1, 2 \ldots, T\}$, $x_t \in \mathrm{R}^D$, where D is the dimension of the local feature or the dimension after reduction. Assuming that each local feature x is independent and follows the same distribution, the formula can be obtained as follows:

$$G_\lambda^X = \sum\nolimits_{t=1}^{T} \nabla_\lambda \log p_\lambda(x_t) \tag{11}$$

$$g_\lambda^X = \sum\nolimits_{t=1}^{T} L_\lambda \nabla_\lambda \log p_\lambda(x_t) \tag{12}$$

That is, when the local features are all independent, their Fisher vector is:

$$x_t \to \varphi_{FK}(x_t) = L_\lambda \nabla_\lambda \log p_\lambda(x_t) \tag{13}$$

The Gaussian Mixture Model (GMM) can characterize the probability density function. Let the set of GMM parameters be $\lambda = \{\omega_k, \mu_k, \varepsilon_k, k = 0, 1, \ldots, K\}$, where ω_k, μ_k and ε_k are the kth weight, mean and covariance, respectively. There are:

$$p_\lambda(x) = \sum\nolimits_{k=1}^{K} \omega_k p_k(x) \tag{14}$$

where p_k represents the kth Gaussian unit:

$$p_k(x) = \frac{1}{(2\pi)^{D/2}\left|\sum k\right|^{1/2}} \exp\{-\frac{1}{2}(x-\mu_k)' \sum_k^{-1} x - \mu_k\} \tag{15}$$

This paper improves the concept of the Fisher vector and the depth of the different features. Furthermore, the accuracy of feature fusion recognition algorithm is improved.

E. LIVENESS DETECTION ALGORITHM

The recognition accuracy of a multi-feature fusion recognition system is always affected by external factors and feature forgery has become the main bottleneck in the development and application of feature recognition systems. Cameras and cell phones are used to capture fake pictures. Due to this situation, in liveness detection is very important. Through the early identification of real pictures and fake pictures, one can effectively eliminate fake features to greatly enhance the accuracy of identification.

Liveness detection algorithms are mainly divided into three categories, namely, Discrete Cosine Transform (DCT) [28], HSV color spatial histogram, and detection algorithms based on SVD and HSV histogram. First of all, the detection algorithm based on DCT coefficient is the extraction of DCT coefficient in the image, and it is used as a feature for living detection. For an image, it contains high frequency characteristics and low frequency characteristics. The low-frequency features are mainly the general outline of the image, while the high frequency features mainly refer to the details and edges of the image. The DCT coefficient mainly refers to the low-frequency characteristics of the image. By setting the high frequency component of the image to zero, the compression and coding of the image are realized. There is a certain correlation between the pixels in the image, and the correlation can be removed by DCT transformation. DCT can use the real - real cosine function as the transformation kernel to make the orthogonal transformation of the image. A small part of the acquired characteristic data can be combined with Fourier transform, which is characterized by high efficiency and fast speed.

The DCT transform sets the high-frequency coefficients of the image to zero and uses the real cosine function as the kernel to process the image. The obtained characteristic data exist in the small part of the DCT transform, which can express the image well. The formula for calculating the DCT transform is as follows:

$$F(u,v)=\frac{2C(u)C(v)}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\frac{(2x+1)u\pi}{2N}\times\cos\frac{(2y+1)v\pi}{2N} \tag{16}$$

where $u, v = 0, 1, \ldots, N-1, x, y = 0, 1, \ldots, N-1$.

$$C(u), C(v)=\left\{\begin{matrix}\frac{1}{\sqrt{2}}, u, v=0\\ 1, u, v=1, 2, \ldots, N-1\end{matrix}\right\} \tag{17}$$

In the feature matrix that is obtained after DCT transformation, the values of different positions represent different information. The upper-left corner of the matrix represents the low-frequency information of the image. The lower-right corner of the matrix represents the high-frequency information. For a feature, the low-frequency coefficients represent the general contour information of the target and the high-frequency coefficients represent the details of the target information.

The reverse formula is needed to restore the image from DCT image to the original image, as shown below:

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)a(u)a(v)\cos\left(\frac{(2x+1)u\pi}{2M}\right)\times\cos\left(\frac{(2x+1)u\pi}{2N}\right) \tag{18}$$

Among them, $x = 0, \ldots, M-1, y = 0, \ldots, N-1$.

IV. MATERIALS AND EXPERIMENTAL RESULTS

In this section, we design a multimodal biological feature fusion system, including data collection, feature extraction, feature fusion, feature training, and testing. In this system, the proposed feature fusion algorithm is compared with the classic feature series and the cascade algorithm. The experimental results are analyzed and the corresponding conclusions are drawn.

A. DATABASE

Face recognition mainly includes four steps: face detection, For the database, we acquired face, fingerprint and finger vein images from 50 students. Each feature corresponded to 5 images.

In the experiment, we selected 4 images of each feature as original training samples and the remaining images served as test samples. To obtain improved recognition accuracy, the set of training samples was expanded by using both the original training samples and mirror images of the original training samples.

As shown in Fig. 2. (a) Face image samples of users were collected from multiple angles. (b) Fingerprint samples were collected in the middle, upper, lower, left, and right areas of users' fingerprints. (c) Finger vein samples were collected in the middle, upper, lower, left, and right areas of the users' finger veins.

B. EXPERIMENT ON FEATURE FUSION SYSTEM

After the establishment of the database, it is necessary to extract and train on the images in the database. For each user, five Fisher vectors are obtained by extracting and merging their face, fingerprint and finger vein feature vectors. The five Fisher vectors are input into the classifier for training and the user's training feature library is obtained. In the testing stage, the face, fingerprint and finger vein image feature information of the test sample is extracted and feature fusion is carried out. Then, the fused features are input into the classification scheme for matching and the recognition result is obtained.

In the framework of the abovementioned multi-feature fusion recognition system, first, this paper carries out a simulation experiment with the proposed Fisher vector secondary feature fusion algorithm. The Gaussian distributions in Gaussian Mixture Model (GMM) will produce the mean, variance and weight parameters in the model Impact, and then affect the fisher vector dimension, and different dimensions of the fisher vector will be multi-feature fusion recognition results have different effects. Therefore, this simulation comparison experiment is carried out using two strategies: different Fisher vector dimensions are tested to determine the accuracy of the trend, and different classifiers, namely, the mainstream k-Nearest Neighbor (kNN), Support Vector Machine (SVM) and Bayes classifiers [29], [30], are tested to determine which classifier is best suited for Fisher vectors.

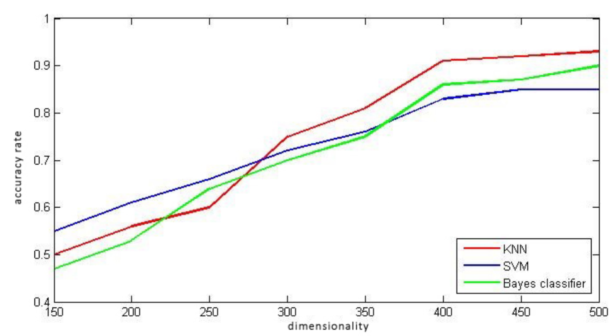


FIGURE 3. The biometric modalities recognition results of fisher vector on kNN, SVM, Bayes classifier.

Fig. 3 presents the biometric modality recognition results of three filters, which are obtained by using the Fisher vector. With the increase of the Fisher vector dimension, the recognition accuracy of the multi-feature fusion recognition algorithm increases. However, because the image pixel information is limited, it is difficult to improve the

TABLE 1. Experimental results on the database use expanded training samples.

Modalities	15 USERS	20 USERS	50 USERS
Face	72.9%	63.3%	52.0%
Figerprint	70.4%	60.0%	46.0%
Figer vein	68.2%	56.6%	42.0%
Serial feature fusion	85.0%	73.3%	62.0%
Parallel feature fusion	83.7%	76.6%	66.0%
secondary calculating FV feature fusion	93.3%	90.0%	88.0%

TABLE 2. Experimental results on the database use expanded training samples.

scene	MULTI-FEATURE FUSION RECOGNITION SYSTEM	MULTI-FEATURE FUSION RECOGNITION SYSTEM BASED ON BIOASSAY
Scene 1	90.0%	90.0%
Scene 2	70.0%	76.7%
Scene 3	66.7%	80.0%
Scene 4	73.3%	83.3%
Scene 5	16.7%	56.7%
Scene 6	20.0%	60.0%
Scene 7	23.3%	63.3%

recognition accuracy after reaching a certain dimension. Moreover, as the dimension increases, the time that is required for multi-feature fusion recognition increases step by step. According to the Fig 3, the kNN classifier outperforms the other two classifiers. Therefore, in this paper, in the multimodal biometric fusion system, kNN is selected as the classifier for feature library training and recognition.

After determining the Fisher vector dimension and classifier, to evaluate the accuracy and robustness of the new algorithm, we simulated and compared the classical serial feature fusion algorithm, the parallel feature fusion algorithm and the proposed Fisher vector secondary feature fusion algorithm. The experiment uses devices to collect the samples from 15, 20, and 30 users to determine the accuracies of the three algorithms, the results are shown in Table 1.

C. EXPERIMENT ON FEATURE FUSION SYSTEM BASED ON LIVENESS DETECTION

The experimental database consists of two parts: the student image dataset and the fake dataset. Mobile phones were used to take pictures for the fake dataset. The database is used for training and testing. To fully evaluate the accuracy of the Fisher vector fusion system based on bioassay, the following seven scenarios are considered:

(1) real face + real fingerprint + real finger vein;
(2) fake face + real fingerprint + real finger vein;
(3) real face + fake fingerprint + real finger vein;
(4) real face + real fingerprint + fake finger vein;
(5) fake face + fake fingerprint + real finger vein;
(6) fake face + real fingerprint + fake finger vein;
(7) real face + fake fingerprint + fake finger vein.

The corresponding image information of these seven scenarios is input into the multi-feature fusion recognition system based on bioassay. To verify that the system can overcome the problems of camouflage and forgery, this experiment will be added to the multi-feature fusion recognition system in *B* experiments as a reference test.

According to Table 2, with the input of a fake image, the recognition accuracy of the multi-feature fusion system is substantially reduced. The effect of the extracted image on the accuracy of the whole recognition system is very large.

For the feature fusion algorithm, because it needs to fuse a variety of features to generate new features, fake features will affect the accuracy of the fused features. Therefore, it is necessary to detect the fake features and remove them before feature fusion to ensure the accuracy of the fused feature. It is concluded from Table 2 that the multi-feature fusion recognition system using real characteristic detection can well remove the fake features and reduce the influence of forged features on the recognition system.

V. CONCLUSION

Multi-biometric identification systems improve the security and accuracy of authentication, which can provide more security for IoMT platform. In this paper, we propose an efficient matching algorithm that is based on secondary calculation of the Fisher vector and uses three biometric modalities: face, fingerprint and finger vein. In our scheme, the three modalities are combined and fusion is performed at the feature level. Experimental results indicate that the proposed method achieves an excellent recognition rate and provides higher security than unimodal biometric-based systems. Moreover, the use of the DCT algorithm for bioassay results in higher accuracy and effectively improves the anti-forgery capability in person identification. This work gives an effective approach for improving the security of IoMT platform.

ACKNOWLEDGMENT

Yang Xin and Lingshuang Kong contribute equally to this paper.

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.
[2] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security Privacy*, vol. 1, no. 2, pp. 33–42, Mar. 2003.
[3] C. Severance, "Anil Jain: 25 years of biometric recognition," *Computer*, vol. 48, no. 8, pp. 8–10, Aug. 2015.
[4] W. S. Zheng, Z. Sun, Y. Wang, X. Chen, P. C. Yuen, and J. Lai, *Biometric Recognition* (Lecture Notes in Computer Science), vol. 32, no. 3. 2017, p. 104.
[5] D. Martinho-Corbishley, M. S. Nixon, and J. N. Carter, "Soft biometric recognition from comparative crowdsourced annotations," in *Proc. Int. Conf. Imag. Crime Prevention Detection*, Jul. 2015, pp. 1–6.
[6] G. Shah, S. Shirke, S. Sawant, and Y. H. Dandawate, "Palm vein pattern-based biometric recognition system," *Int. J. Comput. Appl. Technol.*, vol. 51, no. 2, pp. 105–111, 2015.

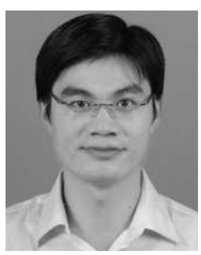
[7] P. R. Nalla and A. Kumar, "Toward more accurate iris recognition using cross-spectral matching," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 208–221, Jan. 2017.
[8] A. Lumini and L. Nanni, "Overview of the combination of biometric matchers," *Inf. Fusion*, vol. 33, no. 2, pp. 71–85, 2017.
[9] N. Lalithamani and M. Sabrigiriraj, "Dual encryption algorithm to improve security in hand vein and palm vein-based biometric recognition," *J. Med. Imag. Health Inform.*, vol. 5, no. 3, pp. 545–551, 2015.
[10] M. I. Gofman, S. Mitra, T. H. K. Cheng, and N. T. Smith, "Multimodal biometrics for enhanced mobile device security," *Commun. ACM*, vol. 59, no. 4, pp. 58–65, 2016.
[11] W. Yang, J. Hu, S. Wang, and C. Chen, "Mutual dependency of features in multimodal biometric systems," *Electron. Lett.*, vol. 51, no. 3, pp. 234–235, 2015.
[12] B. Biggio, G. Fumera, P. Russu, L. Didaci, and F. Roli, "Adversarial biometric recognition: A review on biometric system security from the adversarial machine-learning perspective," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 31–41, Sep. 2015.
[13] S. A. Angadi and S. M. Hatture, "Biometric person identification system: A multimodal approach employing spectral graph characteristics of hand geometry and palmprint," *Int. J. Intell. Syst. Technol. Appl.*, vol. 8, no. 3, pp. 48–58, 2016.
[14] M. A. Haque, K. Nasrollahi, and T. B. Moeslund, "Heartbeat signal from facial video for biometric recognition," in *Proc. Scandin. Conf. Image Anal.*, 2015, pp. 165–174.
[15] E. Maiorana, D. La Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 163–175, Jan. 2016.
[16] M. Gudavalli, D. S. Kumar, and S. V. Raju, "A multibiometric fingerprint recognition system based on the fusion of minutiae and ridges," in *Proc. 49th Annu. Conv. Comput. Soc. India (CSI) Emerg. ICT Bridging Future*, 2015, pp. 231–237.
[17] E. Piciucco, E. Maiorana, and P. Campisi, "Biometric fusion for palm-vein-based recognition systems," in *Proc. Int. Tyrrhenian Workshop Digit. Commun.*, 2017, pp. 18–28.
[18] S. N. Garg, R. Vig, and S. Gupta, "Multimodal biometric system based on decision level fusion," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst.*, Oct. 2017, pp. 753–758.
[19] B. Szymanik and M. Woloszyn, "Magnetic and infrared thermography methods in detection of antipersonnel landmines," *COMPEL, Int. J. Comput. Math. Electr.*, vol. 35, no. 4, pp. 1323–1337, 2016.
[20] W. Kabir, M. O. Ahmad, and M. N. S. Swamy, "A new anchored normalization technique for score-level fusion in multimodal biometrie systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2016, pp. 93–96.
[21] W. Kabir, M. O. Ahmad, and M. N. S. Swamy, "Score reliability based weighting technique for score-level fusion in multi-biometric systems," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2016, pp. 1–7.
[22] X. Yan, W. Kang, F. Deng, and Q. Wu, "Palm vein recognition based on multi-sampling and feature-level fusion," *Neurocomputing*, vol. 151, pp. 798–807, Mar. 2015.
[23] M. Soltane and M. Bakhti, "Soft decision level fusion approach to a combined behavioral speech-signature biometrics verification," *Int. J. Signal Process., Image Process. Pattern Recognit.*, vol. 6, no. 1, pp. 75–90, 2013.
[24] X. Wei, S. L. Phung, A. Bouzerdoum, and A. Bermak, "Invariant image recognition under projective deformations: An image normalization approach," in *Proc. Vis. Commun. Image Process.*, Dec. 2016, pp. 1–4.
[25] J. Feng, H. Xu, and S. Yan, "Online robust PCA via stochastic optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 404–412.
[26] C. Rathgeb, M. Gomez-Barrero, C. Busch, J. Galbally, and J. Fierrez, "Towards cancelable multi-biometrics based on bloom filters: A case study on feature level fusion of face and iris," in *Proc. Int. Workshop Biometrics Forensics*, Mar. 2015, pp. 1–6.
[27] S. Kanade, D. Petrovskadelacrétaz, and B. Dorizzi, "Obtaining cryptographic keys using feature level fusion of iris and face biometrics for secure user authentication," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 138–145.
[28] J. P. Maheshwari, M. Kumar, G. Mathur, R. P. Yadav, and R. K. Kakerda, "Robust digital image watermarking using DCT based pyramid transform via image compression," in *Proc. Int. Conf. Commun. Signal Process.*, Apr. 2015, pp. 1059–1063.
[29] H. Wu, Z. Deng, B. Zhang, Q. Liu, and J. Chen, "Classifier model based on machine learning algorithms: Application to differential diagnosis of suspicious thyroid nodules via sonography," *AJR Amer. J. Roentgenol.*, vol. 207, no. 4, pp. 1–6, 2016.
[30] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013.

YANG XIN received the Ph.D. degree in information security from the Beijing University of Posts and Telecommunications (BUPT). He is currently an Associate Professor with BUPT, meanwhile, he serves as the Vice Director of the Beijing Engineering Laboratory, Cloud Security and Information Security Center, BUPT. His current research interests include network security and big data security.

LINGSHUANG KONG is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong University. His research field includes pattern recognition, machine learning, and deep learning.

ZHI LIU received the Ph.D. degree from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, in 2008. He is currently an Associate Professor with the School of Information Science and Engineering, Shandong University. He is also the Head of the Intelligent Information Processing Group. His current research interests are in applications of computational intelligence to linked multicomponent big data systems, medical image in the neurosciences, multimodal human computer interaction, remote sensing image processing, content based image retrieval, semantic modeling, data processing, classification, and data mining.

CHUNHUA WANG is currently a Senior Engineer with the China Changfeng Science Technology Industry Group Corporation.



HONGLIANG ZHU received the Ph.D. degree in information security from the Beijing University of Posts and Telecommunications (BUPT). He is currently a Lecturer with BUPT, meanwhile, he serves as the Vice Director of the Beijing Engineering Laboratory, Cloud Security and Information Security Center, BUPT. His current research interests include network security and big data security.

MINGCHENG GAO received the master's degree in electronics and communication engineering from Shangdong University in 2011. He is currently pursuing the Ph.D. degree in information security with the Beijing University of Posts and Telecommunications. His main research interests include information security, user cross-domain behavior analysis, and network security.

CHENSU ZHAO is currently pursuing the Ph.D. degree in information security with the Beijing University of Posts and Telecommunications. His main research interests include information security, user cross-domain behavior analysis, and network security.

XIAOKE XU received the Ph.D. degree from the College of Information and Communication Engineering, Dalian Maritime University, in 2008. He was the Post-Doctoral Fellow with Hong Kong Polytechnic University and a Visiting Scholar with the City University of Hong Kong. He is currently a Professor with the College of Information and Communication Engineering, Dalian Minzu University. His current research interests are in complex networks, big data of social networks, nonlinear time series analysis, and data mining.