IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 4, 2018, accepted May 19, 2018, date of publication May 23, 2018, date of current version June 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2839720

Spatial Pyramid Pooling of Selective Convolutional Features for Vein Recognition

JUN WANG, ZAIYU PAN, GUOQING WANG, MING LI, AND YULIAN LI
School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China

Corresponding author: Zaiyu Pan (pzycumt@163.com)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 2018XKQYMS26.

ABSTRACT Deep neural network (DNN) has demonstrated astounding performance in large-scale image recognition task, and pre-trained DNN models trained for one task have also been applied to domains different from their original purposes. Following such an idea, a novel hand-dorsa vein recognition model is constructed by adopting DNN pre-trained on a large-scale database as a universal feature descriptor. Unlike most of these studies which adopt activations of the fully connected layer of DNN as the image representation, we adopt convolutional activations as the region representation. However, not local features of all regions are equally important for final classification. Thus, to solve this issue, a novel selective convolutional feature model based on spatial weighting is proposed to acquire more robust and discriminative feature representation. In specific, a spatial weighting scheme is applied on the convolutional activations to weigh the importance of local features at different regions for classification, which further enhances the discriminability of feature representation. Besides, to take full advantage of the spatial information of the convolutional activations, spatial pyramid pooling is introduced to obtain feature representation with rich spatial information. The final image representation is formed by concatenating the local features of different level of the spatial pyramid. Series rigorous experiments on the lab-made database are conducted to evidence the effectiveness and feasibility of the proposed model. What is more, an additional experiment with subset of PolyU database illustrates its generalization ability and robustness.

INDEX TERMS Deep neural network, hand-dorsa vein recognition, selective convolutional features, spatial weighting, spatial pyramid pooling.

I. INTRODUCTION

With the development of technology and the increase of safety awareness, traditional identification authentication techniques including passwords, signature and smart cards are no longer the best choice. Instead, biometric identification techniques focusing on the intrinsic physiological characteristics such as face [1], iris [2], palmprint [3] and hand shape [4] have become more and more popular. Among these, vein recognition is becoming one of the most popular biometric identification techniques due to its liveness detection, anti-counterfeit and easy acceptability. However, traditional vein recognition model is usually designed by the extraction methods of handcrafted feature, due to the insufficient capacity of feature representation, it is difficult to establish a more robust and discriminative vein recognition model. Recently, deep neural network has successfully applied on large-scale image recognition task because of the capacity in learning discriminative and representative features. However, due to the sufficient vein database, it is difficult to effectively train DNN model on small-scale vein database. To utilize the advantage of feature representation of DNN, Wang *et al.* [5] propose a transfer learning model that the coarse-to-fine scheme is leveraged to train task-specific model in a step way such that the inherent correction between the neighboring models could serve as initialization base to relieve the problem of over-fitting. This method can effectively apply DNN model to small-scale vein recognition task and obtain astounding performance. However, due to the complexity of task-specific transfer learning model, it leads to the low practicability of vein recognition system. In this paper, we directly adopt pre-trained DNN model from ImageNet instead of DNN model trained on vein information database as the feature extractor. This method can realize a simple, efficient, yet highly discriminative and robust vein recognition system.

Activations of a DNN pre-trained on a large-scale database [6], [7], such as ImageNet, can be used as a universal image representation and this method has obtained high performance on some image recognition task. Most of current models take activations of the fully connected layer as the image presentation, whereas, the feature maps of convolutional layer are rarely employed, and some researches [8], [9] have also indicated that directly utilizing the activations of convolutional layer as image representation produces inferior performance. However, comparing with the activations of the fully connected layers, the feature maps of convolutional layer cover more spatial information and rich semantic information. Why does this situation happen? We think the background information and noise information covered in the activations of convolutional layer can result in the low discriminability of the final feature representation, and the detailed analysis process is introduced in Section III. Therefore, in this paper, combining with the inherent distribution characteristic of vein information, a novel selective convolutional feature based on spatial weighting is proposed to remove the non-vein information. Besides, to take full advantage of the spatial information of the convolutional activations, spatial pyramid pooling is introduced to obtain the feature representation with rich spatial information.

A. RELATED WORK

Vein recognition, which meets basic requirement of biometric identification systems as other personal traits [10], is becoming more popular because of high convenience in the procedure of vein image acquisition and feature representation. Different kinds of vein information including finger vein [11], [12], palm vein [13]–[15], dorsal vein [16]–[18], forearm vein [19], wrist vein [20], and scleral vein [21], [22], have been investigated for realizing a robust and efficient personal identification system. In general, nearly all the vein recognition systems share the same configuration, which covers vein image acquisition and preprocessing, feature extraction and representation, classifier design and matching [23]. During the procedure of the vein recognition systems, the most important and difficult part is the design of feature extraction methods because of the inherent characteristics of vein image. Most of current studies conducted in this field focus on generating more robust and discriminative feature representation. State-of-the-art techniques described in literature could be summarized into four groups as follows:

Global topology-based methods: The line-like and curve-like models, which cover the maximum curvature point method [24], the mean curvature [25], the Gabor filter [26], the repeated line tracking method [27], principal curvature [28], the multi-scale Gaussian matched filtering and scale production [29], have been successfully adopted for robust and discriminative feature extraction. What's more, other quantized parameters for describing the geometrical distribution, such as the radius, angle, length and minutiae coordinates, are also imported as mutual feature for improving final performance. In the part of classification and identity recognition, the generated feature vectors would be verified with effective distance evaluation methods including the Hamming distance [30]–[32] and the Hausdorff distance [33], [34]. What's more, other effective and specified criterion such as the Phase-Only-Correlation [35] and pixel-matched ratio [36] are also designed for the task-specific feature classification. The main problem, however, is that accurate segmentation for obtaining topological distribution is essential before conducting the proposed feature extraction. Owing to the existence of low contrast distribution and uneven illumination, it is difficult to acquire accurate segmentation. Therefore, these methods don't have robustness and reliability.

Local geometry-based models: Most of the local analysis methods follow the framework of segmenting the network into binarized one to get the high-quality vascular network distribution for later feature extraction such as the positions and angles of short straight line vectors [37], vein knuckle shapes [38], endpoints and crossing points [39] and tri-branch of vein structure [40]. However, the difficulty in accurate vein segmentation usually brings down the distinguishing ability of these methods. Besides, those geometry describing models are also sensitive to scaling, rotation, displacement and other geometric transformation, which are unavoidable for contactless vein acquisition systems.

Local invariant-feature-based models: Regarding the solution for solving the condition of unconstrained geometric transformation, the most referred hand-crafted feature extraction model is the scale-invariant feature description models including the basic one, SIFT [41]–[43] and its variants such as SURF [44], RootSIFT [45] and ASIFT [46]. The feature vectors generation procedures of these models share the same configuration including the built of scale-space, the extrema points detection and descriptors generation. Ladoux *et al.* [15] adopted SIFT for hand vein describing after simple preprocessing, which refers to the procedure of denoising and contrast enhancement (CE). However, it is argued that the preprocessing would result in the vein loss and the low contrast distribution renders great difficulty in generating enough descriptive keypoints. To remove the negative effect of CE, SIFT or SURF [47]–[49] is extracted directly with input without any pre-processing and binarization procedure, and this idea results in considerable performance improvement. However, the existence of mismatching pairs between unpaired keypoints usually brings up the false acceptance rate (FAR) and equal error rate (EER), which is unacceptable for vein recognition.

DNN-based models: Deep Neural Network has acquired excellent performance in image recognition. Based on their powerful capacity for feature representation, some researchers brought them into biometrics. Several deep learning models such as in [50] and [51] have been built for face recognition and have shown great success on the LFW face database. Although the DNN model is used for both face recognition task and other face attributes analysis task, lacking in enough training database like the ImageNet restricts its generalization ability in other computer

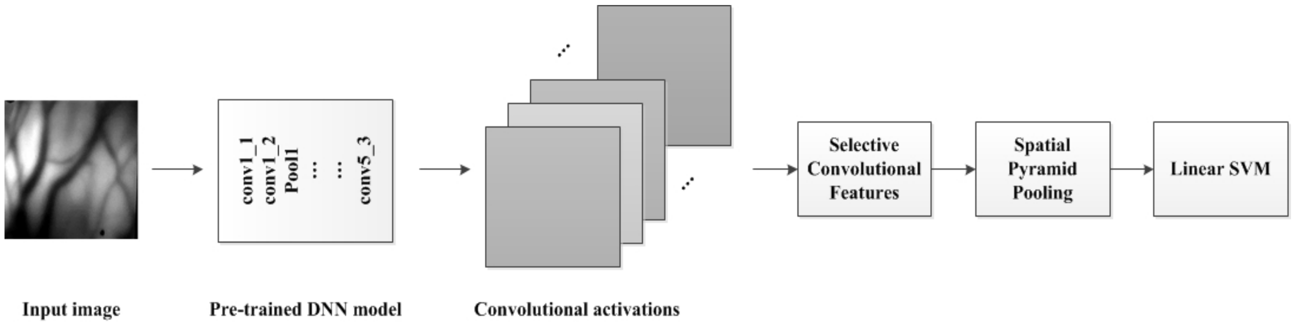


FIGURE 1. The framework of our proposed spatial pyramid pooling of selective convolutional features.

vision tasks like hand vein image analysis, where only a small-scaled lab-made database is available. To employ the discriminative ability of feature representation of DNN models for identity recognition based on vein information, the task-specific transfer learning model based on the coarse-fine strategy [5] is proposed to achieve acceptable results. However, the complexity of this model restricts the practicability under real environment.

B. OUR WORK

In this paper, we adopt the feature maps of convolutional layer instead of the activations of the fully connected layer as the region representation. Because the convolutional activations contain the background information and different local features in the convolutional activations are not equally important for final classification, we must select the discriminative feature and weigh the importance of different local features in the convolutional activations. Therefore, combining with the inherent property of vein information, a novel selective convolutional features model (SCF) based on spatial weighting is proposed to enhance the representative capacity of convolutional activations. In specific, one single cell in the last pool layer corresponds to one local vein patch in the input vein image, and all single cell in the last pool layer form a regular grid of vein patches in the input vein image. Therefore, we weigh the importance of one single cell in the last pool layer by judging the size of vein region in the vein patch corresponded to this single cell. To evaluate the size of vein region in vein patches, a spatial weighting scheme is proposed. Besides, spatial pyramid pooling is introduced to produce the feature representation with rich spatial information.

Fig.1 shows the framework of our proposed spatial pyramid pooling of selective convolutional features (SPP-SCF), which has been evaluated with the lab-made hand vein database for identity recognition. Overall, the contributions of the paper are summarized as follows:

1) We propose the first framework for hand-dorsa vein recognition with the pre-trained DNN model from ImageNet. Comparing with other DNN-based model, on the one hand, our model is simple, efficient and yet achieve acceptable results. On the other hand, our proposed model has the high ability of discriminability and generalization.
2) We adopt the feature maps of convolutional layer instead of the activations of the fully connected layer as the region representation, and this method remains the rich semantic information and spatial information of the input vein image.
3) A task-specific selective convolutional features model based on spatial weighting is proposed to obtain the discriminative convolutional features.
4) Spatial pyramid pooling is introduced to acquire the representative and robust feature representation because it can take better advantage of the spatial information.

The remainder of the paper is organized as follows. The selective convolutional features is described in Section II. Section III introduces the spatial pyramid pooling, and the experimental results and analysis are presented in Section IV. Finally, we summarize the paper and conclude the future work in Conclusions.

II. SELECTIVE CONVOLUTIONAL FEATURES MODEL

Unlike most of current studies that adopt the activations of the fully connected layer of DNN model as the image representation, in this paper, we employ the convolutional activations of DNN model as the region representation. However, directly adopting convolutional activations as the vein patch representation cannot achieve the satisfactory performance. Therefore, to obtain the discriminative convolutional features, a novel task-specific selective convolutional activations model based on spatial weighting is proposed.

A. THE ANALYSIS OF CONVOLUTIONAL FEATURES WITH VEIN INFORMATION

Comparing with the activations of the fully connected layer, the convolutional activations contain rich semantic information and spatial information. However, the convolutional activations also cover the background information and noise information, thus, it leads to the low discriminability of convolutional activations. Fig. 2 shows the vein image's visualized feature maps which are produced by the last pool layer. In this figure, it can be observed that the single cell in the

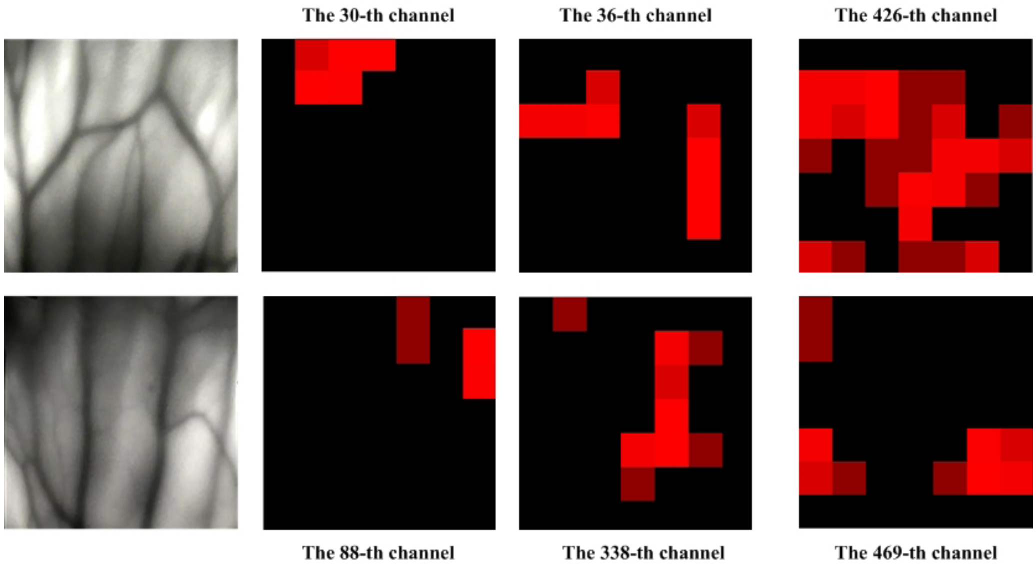


FIGURE 2. Samples of visualized feature maps of vein images.

last pool layer can correspond to three kinds of vein patches including non-vein region, vein region with small amount of vein information and vein region with large amount of vein information. The strongest responses of feature maps in the last pool layer are not always corresponding to the vein region in the input vein image, besides, in the aggregation process of the convolutional activations, local deep features corresponded to the last two kinds of vein patches are served as equally importance for final classification. Thus, it leads to the low discriminability of convolutional activations. The visualized feature maps of the last pool layer also demonstrate the initial idea about the low discriminative of convolutional activations. Therefore, how to remove the non-vein information to full explore its representative and discriminative potential for vein recognition remains a key issue.

To acquire more discriminative convolutional features, we must effectively weigh the significance of each single cell in the last pool layer. Due to the inherent characteristics of vein information, it is difficult to directly remove the non-vein information in the last pool layer by the threshold methods. In the last pool layer, each single cell corresponds to the vein patches of input vein image, thus, we adopt local contrast estimation based spatial weighting to weigh the size of vein region of each vein patch corresponded to each single cell in the last pool layer, which can effectively weigh the importance of each single cells and remove the non-vein information.

B. LOCAL CONTRAST ESTIMATION BASED SPATIAL WEIGHTING

By analyzing convolutional activations with vein information, it can be concluded that single cell in the last pool layer corresponds to three kinds of vein patches covering non-vein region, vein region with small amount of vein information and vein region with large amount of vein information. Therefore, we select discriminative convolutional feature by weighing the size of vein region in vein patch corresponded to single cell in the last pool layer.

The local contrast estimation is defined as the problem of finding the optimal local binary pattern, which equals to finding the best threshold to classify the local region into object part and background part. Driven by this theory and the formulation of LBP [52], we organize the problem as finding the optimal threshold used for local binary pattern coding, and the spatial weight is calculated by measuring the variance distribution between the one with coding image and that of the original image. The detailed procedure for generating weight ω is described as follows.

The *LBP* coding could be viewed as approximating the local pixel intensity distribution in local region by two separated clusters separated by the center pixel (τ). To measure the effect of the binarization, the residual error between the distribution of original image and coding one could be expressed as follows:

$$\epsilon(\tau) = \frac{1}{P}\left\{\sum\nolimits_{i|r_i\geq\tau}(r_i-\mu_0)^2+\sum\nolimits_{i|r_i<\tau}(r_i-\mu_1)^2\right\} \quad (1)$$

where

$$\mu_0 = \frac{1}{P_0}\sum\nolimits_{i|r_i\geq\tau} r_i, \quad P_0=\sum\nolimits_i S(r_i-\tau);$$
$$\mu_1 = \frac{1}{P_1}\sum\nolimits_{i|r_i<\tau} r_i, \quad P_1=\sum\nolimits_i S(\tau-r_i).$$

and $S(x)$ is the signal function.

Here, μ_0 and μ_1 are the mean value of the two separated clusters, and we define σ_w^2 and σ_b^2 as the variance within clusters and between clusters respectively. Based on the definition of Fisher Criterion, the better discrimination of *LBP* indicates the smaller value of $\epsilon(\tau)$, which corresponds to σ_w^2.

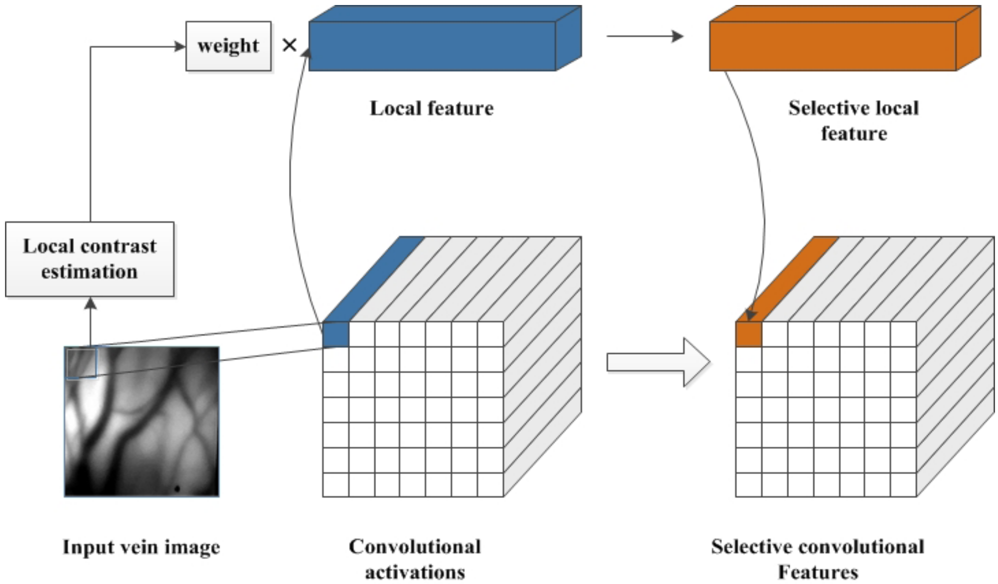


FIGURE 3. The procedure of the proposed selective convolutional features based on spatial weighting.

According to Fisher discriminant criterion [63], minimizing σ_w^2 coincides with maximizing the Fisher discriminant score, and the same as maximization of σ_b^2.

$$\tau^* = arg \max_{\tau \in \{r_i\}_{i=1}^P} \sigma_b^2(\tau) \tag{2}$$

Where

$$\sigma_b^2(\tau) = \frac{P_0}{P}(\mu_0 - \tau)^2 + \frac{P_1}{P}(\mu_1 - \tau)^2 = \frac{P_0 P_1}{P^2}(\mu_0 - \mu_1)^2.$$

With the optimized threshold τ^*, the optimized voting weight could be accordingly determined as follows:

$$w = \sqrt{\frac{\sigma_b^2(\tau^*)}{\sigma^2 + C}} \tag{3}$$

Where C is a small constant designed to avoid numerical instability of w, which refers to coding of non-vein region. And in the vein coding experiment, C is set as 0.01^2. The value distribution of w corresponds to the situation that the higher value is generated from the region with vein whereas the lower one is the non-vein region.

Given the feature maps of the last pool layer $X \in R^{H \times W \times C}$, the selective convolutional features based on spatial weighting can be defined as follows:

$$F = \sum_i^H \sum_j^W X(i,j) w(i,j) \tag{4}$$

Where X (i, j) $\in R^{1 \times C}$ is the local feature at the (i, j)-th location, and $w(i,j)$ generated from Eq. (3) is the size of vein information of vein patch corresponded to the (i, j)-th location. Local convolutional features are weighed to obtain selective convolutional features such that more critical parts such as vein region with large amount of vein information can play a more important role in final classification process, which further enhances the discriminative and representative capacity of final vein feature information. The procedure of our proposed task-specific selective convolutional features model based on spatial weighting is as shown in Fig. 3.

III. SPATIAL PYRAMID POOLING

After obtaining the selective convolutional features, we need consider how to efficiently employ spatial information of convolutional activations. The traditional aggregation methods such as average-pooling, max-pooling and fisher vector cannot realize accept performance because these methods cannot fully employ the spatial information of convolutional activations. To address this problem, spatial pyramid pooling (SPP) [53] is introduced to aggregate the convolutional features and obtain more representative vein feature with rich spatial information.

In our experiment, the pre-trained DNN model is VGG16 [54] trained on ImageNet database, and the size of input vein image is 224×224. Therefore, we can get 7×7×512 convolutional activations, on the one hand, these activations are regarded as 512 feature maps of size 7×7, on the other hand, they are also served as 49 deep local features of 512-dimension. After acquiring the weighted activations of $pool_5$ layer, a 3-level pyramid pooling is conducted. The detailed process is introduced as follows: First, a level 0 pyramid (1×1) is obtained by performing the average pooling with the window size of 7 and the stride of 7. Second, we conduct an average pooling with the window size of 4 and the stride of 3 to acquire the level 1 pyramid (2×2). Finally, an average pooling with the window size of 3 and the stride of 2 is employed to gain the level 2 pyramid (3×3). It should be noted that the outputs of spatial pyramid pooling are kC-dimensional vector, where C is the number of feature

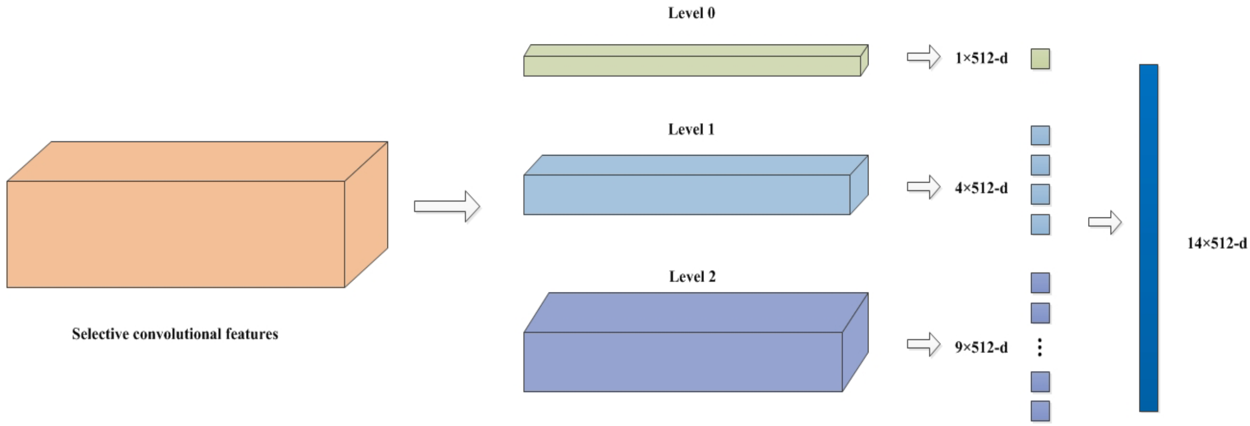


FIGURE 4. The process of spatial pyramid pooling.

TABLE 1. The configuration of 3-level pyramid.

	Level 0	Level 1	Level 2
Input size	7×7×512	7×7×512	7×7×512
Pool	average	average	average
Window size	7	4	3
Stride	7	3	2
Output size	1×1×512	2×2×512	3×3×512

maps in the $pool_5$ layer and k is the number of all single cells in 3-level spatial pyramid. The configuration of 3-level pyramid is as shown in Table 1 and Fig. 4 illustrates the process of SPP.

The proposed SPP-SCF methods is summarized in Algorithm 1.

Algorithm 1 The SPP-SCF Model

Input:

An input vein image I (224×224)

A pre-trained DNN model (VGG-16)

Procedure:

1) Extract deep descriptors $\boldsymbol{X} \in \boldsymbol{R}^{H\times W\times C}$ from I using the pre-trained DNN model.
2) Calculate the weighting (w) of vein patches corresponded to every single cells in the pool5 layer by using the local contrast estimation method.
3) The selective convolutional features (F) is acquired by performing multiplication operation $\boldsymbol{F} = \sum_i^H \sum_j^W \boldsymbol{X}\,(\boldsymbol{i},\boldsymbol{j})\boldsymbol{w}(\boldsymbol{i},\boldsymbol{j})$.
4) Conduct a 3-level spatial pyramid pooling on the selective convolutional features.
5) Concatenate $F_{1\times1}$, $F_{2\times2}$ and $F_{3\times3}$ to form the final feature presentation f which is a 14∗512-d feature vector.

IV. EXPERIMENTS AND DISCUSSION

In this part, rigorous comparison experiments are designed to comprehensively evaluate the performance of the proposed model. On the one hand, experiments with the small-scale lab-made database are conducted firstly to analyze the superiority of the proposed spatial pyramid pooling of selective convolutional features model. On the other hand, to check the generalization ability of the proposed model, additional experiments with the PolyU palmprint database [55] is carried out and state-of-the-art recognition result fully demonstrates the generalization ability and robustness of the proposed model.

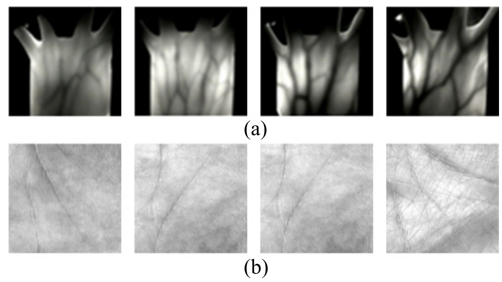


FIGURE 5. Samples of hand-dorsa vein database and PolyU palmprint database. (a) Lab-made hand-dorsa vein database. (b) PolyU palmprint database (NIR part).

A. FORMULATION OF GPP ALGORITHM

Our hand-dorsa vein database has 200 individuals where male and female are respectively 100, and for each person, 10 right hand-dorsa vein images are captured. All hand-dorsa vein images in our database are acquired in two specifically set sessions separated by a time interval of more than 10 days, and at each time, five samples are acquired from each subject at the wavelength of 850nm. To employ the dorsal hand vein information, we set the size of the images as 460∗640 with extremely high quality, and Fig. 5 shows some samples of the lab-made hand-dorsa vein database and the PolyU palmprint database.

B. EXPERIMENT DETAILS

In our experiments, VGG-16 model is employed as the pre-trained DNN model to extract deep convolutional activations.

TABLE 2. **The evaluation results of SCF by using different aggregation methods.**

Methods	Dimension	Recognition rate (%)
$Pool_5$+max-pooling	512	88.54
SCF+max-pooling	512	89.06
$Pool_5$+average-pooling	512	91.15
SCF+average-pooling	512	91.67

Note that the feature maps of pool5 layer in VGG-16 model are regraded as the convolutional activation in our methods. Thus, X is 7×7×512 convolutional activations. The size of input vein ROI image is 224×224, and the single cell in the pool5 layer corresponds to the vein patch in the input vein ROI image that has a size of 32×32. And the selective convolutional features are 49 deep features of 512-d, which are also regarded as the 512 feature maps of size 7×7. A 3-level pyramid pooling is introduced to aggregate the selective convolutional features, and the detailed configuration is as shown in Table 1. It should be noted that the final feature presentation is a 14×512-d feature vector.

C. PERFORMANCE EVALUATION OF SELECTIVE CONVOLUTIONAL FEATURES

In this section, the effectiveness of the proposed selective convolutional features model based on spatial weighting is evaluated on our lab-made hand-dorsa vein database. Due to only focusing on the impacts of the convolutional features, we directly employ the max-and average-pooling methods to aggregate the feature maps of the last pool layer. The experiment results are as shown in Table 2.

It can be concluded from Table 2 that Comparing the original convolutional features, the two kinds of feature aggregation methods applied on selective convolutional features realize the accept performance, which also illustrates the high representative and discriminative of selective convolutional features. At the same time, it is obvious that the initial idea about the low discriminative of convolutional activations is demonstrated in this experiment.

D. PERFORMANCE EVALUATION OF SPATIAL PYRAMID POOLING

In this part, the effectiveness of spatial pyramid pooling is verified on the lab-made hand-dorsa vein database. It should be noted that we adopt the selective convolutional features as the deep descriptors of input vein image in this experiment. Several encoding and pooling methods are introduced to evaluate the advantage of spatial pyramid pooling. This experiment is not only evidencing the effectiveness of spatial pyramid pooling but also illustrating the proposed spatial pyramid pooling of selective convolutional features model. The recognition results are as shown in Table 3.

Judging from the results in Table 3, the SPP method realizes high performance comparing with other encoding and pooling methods, and the state-of-the-art recognition result with 96.35% indicates the effectiveness of SPP and also demonstrates the advantage of spatial pyramid pooling of selective convolutional features model based on spatial weighting.

TABLE 3. **The recognition results with different aggregation methods.**

Methods	Dimension	Recognition rate (%)
SCF+max-pooling	512	89.06
SCF+ average-pooling	512	91.67
SCF+FV	2048	84.17
SCF+SPP	7168	96.35

E. COMPARISON WITH STATE-OF-THE-ART MODELS

After evaluating the effect of the proposed model on the vein recognition task in the above sections, attention is moved to design experiment to demonstrate the advantage of the proposed SPP-SCF over other state-of-the-art feature extraction algorithms (Since that there is no state-of-the-art performance in hand-dorsa vein recognition tasks with the recently popular feature learning algorithms, only the hand-crafted feature extraction algorithms are implemented for performance comparison.), multi-modal experiments in the scenario of verification are designed. In such scenario, the first image is regarded as gallery whereas the remaining images are exploited as probe, and the overall performance comparison is illustrated in Fig. 6.

Two kinds of representative hand-crafted feature extraction algorithms are used as reference: The one is the local invariant feature model including SIFT [40], SURF [43], RootSIFT [44], ASIFT [45], and it has the advantages of being invariant to rotation, translation, scale uncertainty and even uniform illumination, which makes it the best one among all hand-crafted algorithms. The other one is the LBP [52] and its variants including LDP [56], LTP [57], and LLBP [58], and such model is widely applied for vein based identification application for its efficiency, and it also provides competitive recognition results.

Judging from EER result of verification with the lab-made database, it can be concluded that the proposed spatial pyramid pooling of selective convolutional activations model performs far better than the LIF (Local Invariant Feature) models with EER as 0.06% whereas the best of LIF is 0.105% with RootSIFT and the best of LBPs is 0.113% with LDP, and the state-of-the-art vein recognition results fully demonstrate the ability of the proposed model for obtaining complete and discriminative feature representation.

To further demonstrate the effectiveness of the proposed model, we compare the proposed method based on the pre-trained DNN model with the transfer learning model based on the fine-tuning DNN model [5]. Although the EER result as 0.058% achieved the transfer learning model is little better than the proposed model with EER as 0.06%, comparing with the transfer learning model, our model need not the complex process of fine-tuning strategy and can be employed as universal feature representation for vein recognition. Thus, our proposed model can realize the simple, efficient vein recognition system.

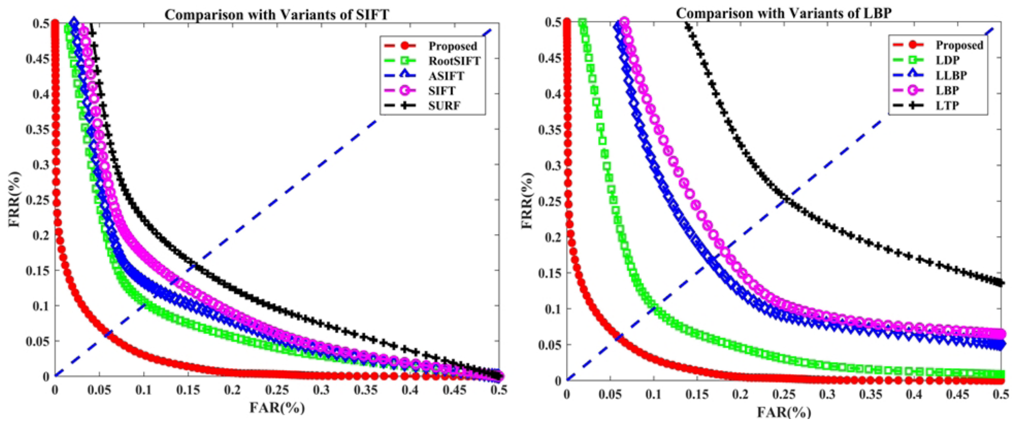


FIGURE 6. Comparison of ROC curves between the proposed feature learning methods and representative handcrafted methods (Left: SIFTs, Right: LBPs).

TABLE 4. Summary of EERs derived from recently published vein recognition models using polyU palmprint database.

Reference	Year	Methodology	Performance(EER)
[59]	2009	Extract palmvein features with multiscale matched filters, and then generate the matching score with ICP algorithm	0.557%
[60]	2011	Adopt NMRT and Hessian phase respectively to realize feature extraction and matching	NMRT: 0.004% Hessian:0.43%
[61]	2013	Combine curvelet transform and Gabor filter for feature extraction and employ score level fusion strategy to obtain matching score for verification	0.1023%
[62]	2014	Employ Gaussian-Random transform to extract the orientation matrix and then compute the principal direction based on the matrix as palm vein feature, and matching between feature bins is used for identification	0.14%
[63]	2017	Extract vein features with quality-specific discriminative LBP algorithm and adopt improved Chi-square for verification	0.079%
Proposed method	2018	Extract vein features with the proposed selective convolutional features of pre-trained DNN model and adopt SVM for classification.	0.068%

F. GENERALIZATION EVALUATION WITH POLYU PALMPRINT DATABASE

The PolyU Multispectral Palmprint Database is employed here to evaluate the generalization ability of the proposed complementary feature encoding, and only the near-infrared images of PolyU database is adopted for verification since the focus of the experiment is palm-vein verification. The proposed system as illustrated in Fig. 1 is accordingly followed to obtain state-of-the-art recognition results, and comparison with other models on PolyU database is as shown in Table 4.

It can be observed that the proposed model could achieve state-of-the-art EER level compared with others, which is served as strong proof for the generalization ability of the model.

V. CONCLUSION

Currently, DNN model has successfully applied to the large-scale image recognition task, however, it is difficult to apply DNN model to the small-scale image recognition task such as vein recognition. If we directly train DNN model on the small-scale database, it can produce the problem of over-fitting in the training process of DNN. Recently, Activations of a DNN model pre-trained on a large-scale database such as ImageNet can be used as a universal image representation and this method has also obtained high performance on some image recognition task. Following this idea, we employ the pre-trained DNN model (VGG-16) from ImageNet instead of the DNN model trained on the small-scale database as the region representation for vein recognition. To obtain more spatial information of vein feature, we adopt the feature maps of fully convolutional layer instead of the activations of the fully connected later as the region representation. However, directly employing the convolutional activations as the region representation cannot achieve the satisfactory recognition result because they cover the background information and noise information. Thus, to remove the non-vein information and fully evaluate the impacts of different local features at different locations, the selective convolutional features model based on spatial weighting is proposed to acquire more representative and discriminative convolutional activations for classification. In specific, one single cell in the last pool layer corresponds to one local vein patch in the input vein image, and all single cell in the last pool layer form a regular grid of vein patches in the input vein image. Therefore, we weigh the importance of one single cell in the last pool layer by judging the size of vein region in the vein patch corresponded to this single cell. To evaluate the size of vein region in the patches, a spatial weighting scheme is proposed. Besides, to efficiently utilizing the spatial information of convolutional activations, a 3-level pyramid pooling is introduced to aggregate the feature maps of pool5 layer.

We also argue that the proposed spatial pyramid pooling of selective convolutional features model based on spatial weighting is also applicable for other computer vision task solved with pre-trained DNN models.

In the future, a DNN model based on selective convolutional features, which is end-to-end trainable, will be discussed for both hand vein recognition task and other image recognition task.

REFERENCES

[1] C.-X. Ren, Z. Lei, D.-Q. Dai, and S. Z. Li, "Enhanced local gradient order features and discriminant analysis for face recognition," *IEEE Trans. Cybern.*, vol. 46, no. 11, pp. 2656–2669, Nov. 2016.

[2] 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.

[3] W. Jia, R.-X. Hu, Y.-K. Lei, Y. Zhao, and J. Gui, "Histogram of oriented lines for palmprint recognition," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 3, pp. 385–395, Mar. 2014.

[4] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, Aug. 2016.

[5] G. Wang, J. Wang, and Z. Pan, "Bimodal vein recognition based on task-specific transfer learning," *IEICE Trans. Inf. Syst.*, vol. E100-D, no. 7, pp. 1538–1541, Jul. 2017.

[6] B.-B. Gao, X.-S. Wei, J. Wu, and W. Lin. (2015). "Deep spatial pyramid: The devil is once again in the details." [Online]. Available: https://arxiv.org/abs/1504.05277

[7] X. S. Wei, J. H. Luo, J. Wu, and Z. H. Zhou, "Selective convolutional descriptor aggregation for fine-grained image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2868–2881, Jun. 2017.

[8] P. Agrawal, R. Girshick, and J. Malik, "Analyzing the performance of multilayer neural networks for object recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 329–344.

[9] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[10] A. Kumar, M. Hanmandlu, and H. M. Gupta, "Online biometric authentication using hand vein patterns," in *Proc. IEEE Symp. Comput. Intell. Secur. Defence Appl.*, Ottawa, ON, Canada, Jul. 2009, pp. 1–7.

[11] L. Yang, G. Yang, Y. Yin, and X. Xi, "Finger vein recognition with anatomy structure analysis," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[12] J. Yang, Y. Shi, and G. Jia, "Finger-vein image matching based on adaptive curve transformation," *Pattern Recognit.*, vol. 66, pp. 34–43, Jun. 2017.

[13] W. Kang and Q. Wu, "Contactless palm vein recognition using a mutual foreground-based local binary pattern," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1974–1985, Nov. 2014.

[14] S. Malki and L. Spaanenburg, "Hand veins feature extraction using DT-CNNS," *Proc. SPIE*, vol. 6590, p. 65900N, May 2007.

[15] P.-O. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based on SIFT matching," in *Proc. Int. Conf. Adv. Biometric*, Alghero, Italy, 2009, pp. 1290–1298.

[16] J. Wang, G. Wang, and M. Zhou, "Bimodal vein data mining via cross-selected-domain knowledge transfer," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 733–744, Mar. 2018.

[17] J. Wang and G. Wang, "Hand-dorsa vein recognition with structure growing guided CNN," *Optik-Int. J. Light Electron Opt.*, vol. 149, pp. 469–477, Nov. 2017.

[18] Y. Wang, K. Li, J. Cui, L.-K. Shark, and M. Varley, "Study of hand-dorsa vein recognition," in *Proc. Int. Conf. Intell. Comput.*, Changsha, China, 2010, pp. 490–498.

[19] H. Zhang, C. Tang, A. W.-K. Kong, and N. Craft, "Matching vein patterns from color images for forensic investigation," in *Proc. IEEE 5th Int. BTAS*, Arlington, VA, USA, Sep. 2012, pp. 77–84.

[20] J. E. S. Pascual, J. Uriarte-Antonio, R. Sanchez-Reillo, and M. G. Lorenz, "Capturing hand or wrist vein images for biometric authentication using low-cost devices," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Darmstadt, Germany, Oct. 2010, pp. 318–322.

[21] Y. Lin, E. Y. Du, Z. Zhou, and N. L. Thomas, "An efficient parallel approach for sclera vein recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 2, pp. 147–157, Feb. 2014.

[22] Z. Zhou, Y. Du, N. Thomas, and E. Delp, "A new human identification method: Sclera recognition," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 3, pp. 571–583, May 2012.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[24] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. ICCV*, vol. 2, Oct. 2005, pp. 1458–1465.

[25] T. Kobayashi, "BFO meets HOG: Feature extraction based on histograms of oriented p.d.f. gradients for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 747–754.

[26] O. Russakovsky, Y. Lin, K. Yu, and L. Fei-Fei, "Object-centric spatial pooling for image classification," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, 2012, pp. 1–15.

[27] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Aisserman, "The devil is in the details: An evaluation of recent feature encoding methods," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 76.1–76.12.

[28] J. H. Choi, W. Song, T. Kim, S.-R. Lee, and H. C. Kim, "Finger vein extraction using gradient normalization and principal curvature," *Proc. SPIE*, vol. 7251, p. 725111, Feb. 2009.

[29] Y.-B. Zhang, Q. Li, J. You, and P. Bhattacharya, "Palm vein extraction and matching for personal authentication," in *Advances in Visual Information Systems*. Berlin, Germany: Springer, 2007, pp. 154–164.

[30] X. Qian, S. Guo, X. Li, F. Zhong, and X. Shao, "Finger-vein recognition based on the score level moment invariants fusion," in *Proc. Int. Conf. Comput. Intell. Softw. Eng. (CiSE)*, Dec. 2009, pp. 1–4.

[31] J.-C. Lee, "A novel biometric system based on palm vein image," *Pattern Recognit. Lett.*, vol. 33, no. 12, pp. 1520–1528, 2012.

[32] W. Yang, X. Huang, F. Zhou, and Q. Liao, "Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion," *Inf. Sci.*, vol. 268, no. 6, pp. 20–32, 2014.

[33] L. Wang, G. Leedham, and S. Gho, "Infrared imaging of hand vein patterns for biometric purposes," *IET Comput. Vis.*, vol. 1, no. 3, pp. 113–122, 2007.

[34] L. Wang, G. Leedham, and D. Siu, "Minutiae feature analysis for infrared hand vein pattern biometrics," *Pattern Recognit.*, vol. 41, no. 3, pp. 920–929, 2008.

[35] J. Yang and Y. Shi, "Finger–vein ROI localization and vein ridge enhancement," *Pattern Recognit. Lett.*, vol. 33, no. 12, pp. 1569–1579, 2012.

[36] W. Song, T. Kim, H. Kim, J. Choi, H. Kong, and S. Lee, "A finger-vein verification system using mean curvature," *Pattern Recognit. Lett.*, vol. 32, no. 11, pp. 1541–1547, 2011.

[37] J. M. Cross and C. L. Smith, "Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification," in *Proc. Int. Carnahan Conf. Secur. Technol.*, Sanderstead, U.K., Oct. 1995, pp. 20–35.

[38] P. MacGregor and R. Welford, "Imaging for security and personnel identification," *Adv. Imag.*, vol. 6, no. 7, pp. 52–56, 1991.

[39] K. Wang, Y. Yuan, Z. Zhang, and D. Zhuang, "Hand vein recognition based on multi supplemental features of multi-classifier fusion decision," in *Proc. Int. Conf. Mechatronics Autom.*, Luoyang, China, Jun. 2006, pp. 1790–1795.

[40] L. Yang, G. Yang, X. Xi, X. Meng, C. Zhang, and Y. Yin, "Tri-branch vein structure assisted finger vein recognition," *IEEE Access*, vol. 5, pp. 21020–21028, 2017.

[41] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[42] D. Huang, R. Zhang, Y. Yin, Y. Wang, and Y. Wang, "Local feature approach to dorsal hand vein recognition by centroid-based circular keypoint grid and fine-grained matching," *Image Vis. Comput.*, vol. 58, pp. 266–277, Feb. 2017.

[43] D. Huang, Y. Tang, Y. Wang, L. Chen, and Y. Wang, "Hand-dorsa vein recognition by matching local features of multisource keypoints," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1823–1837, Sep. 2015.

[44] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[45] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, 2009.

[46] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2911–2918.

[47] L. Xiuyuan, L. Tiegen, D. Shichao, H. Jin, and W. Yun, "Fast recognition of hand vein with surf descriptors," *Chin. J. Sci. Instrum.*, vol. 32, no. 4, pp. 831–836, 2011.

[48] H.-B. Wang, L. Tao, and X.-Y. Hu, "Novel algorithm for hand vein recognition based on retinex method and sift feature analysis," in *Electrical Power Systems and Computers*, vol. 99. Berlin, Germany: Springer, 2011, pp. 559–566.

[49] H.-G. Kim, E. J. Lee, G.-J. Yoon, S.-D. Yang, E. C. Lee, and S. M. Yoon, "Illumination normalization for SIFT based finger vein authentication," in *Advances in Visual Computing*. Berlin, Germany: Springer, 2012, pp. 21–30.

[50] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[51] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2892–2900.

[52] E. C. Lee, H. C. Lee, and K. R. Park, "Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction," *Int. J. Imag. Syst. Technol.*, vol. 19, no. 3, pp. 179–186, 2009.

[53] L. Liu, C. Shen, and A. van den Hengel. (2014). "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification." [Online]. Available: https://arxiv.org/abs/1411.7466

[54] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[55] D. Zhang, W.-K. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, Sep. 2003.

[56] B. J. Kang, K. R. Park, J.-H. Yoo, and J. N. Kim, "Multimodal biometric method that combines veins, prints, and shape of a finger," *Opt. Eng.*, vol. 50, no. 1, p. 017201, 2011.

[57] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 168–182, Jun. 2010.

[58] B. A. Rosdi, W. S. Chai, and S. A. Suandi, "Finger vein recognition using local line binary pattern," *Sensors*, vol. 11, no. 12, pp. 11357–11371, 2011.

[59] H. Chen, G. Lu, and R. Wang, "A new palm vein matching method based on ICP algorithm," in *Proc. Int. Conf. Interact. Sci.*, 2009, pp. 1207–1211.

[60] Y. Zhou and A. Kumar, "Human identification using palm-vein images," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1259–1274, Dec. 2011.

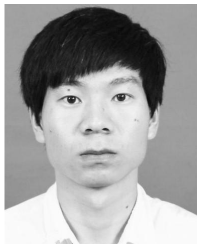
[61] J. Sun and W. Abdulla. "Palm vein recognition by combining curvelet transform and Gabor filter," in *Proc. 8th Chin. Conf. Biometric Recognit.*, Jinan, China, 2013, pp. 314–321.

[62] P. O. Staff, "Palm-vein classification based on principal orientation features," *PLoS ONE*, vol. 9, no. 11, pp. 1–12, 2014.

[63] J. Wang and G. Wang, "Quality-specific hand vein recognition system," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2599–2610, Nov. 2017.

JUN WANG received the B.S., M.S., and Ph.D. degrees in control theory and control engineering from the China University of Mining and Technology, Jiangsu, China, in 2003, 2009, and 2015, respectively.

He is currently an Associate Professor with the School of Information and Electrical Engineering, China University of Mining and Technology, where he is also with the Jiangsu Province Laboratory of Electrical and Automation Engineering for Coal Mining. His current research interests include intelligent control system design, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.



ZAIYU PAN received the B.S. degree in electrical engineering and automation from Chaohu University, Anhui, China, in 2016. He is currently pursuing the M.S. degree in control science and engineering with the China University of Mining and Technology. His current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.

GUOQING WANG received the B.S. and M.S. degrees in information engineering from the China University of Mining and Technology, Xuzhou, Jiangsu, China, in 2014 and 2017 respectively. He is currently pursuing the Ph.D. degree with in the School of Computer Science and Engineering, University of New South Wales, Sydney, Australia.

He joined the Jiangsu Province Laboratory of Electrical and Automation Engineering for Coal Mining, China University of Mining and Technology. His current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.

MING LI received the B.S. and M.S. degrees in automation and the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Xuzhou, Jiangsu, China, in 1983, 1989, and 2002, respectively.

He is currently a Professor with the School of Information and Electrical Engineering, China University of Mining and Technology, where he is also with the Jiangsu Province Laboratory of Electrical and Automation Engineering for Coal Mining. His current research interests include intelligent control system design, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.

YULIAN LI received the B.S. degree from the College of Physical and Electronic Information, Luoyang Normal University, Henan, China, in 2015. She is currently pursuing the M.S. degree in control science and engineering with the China University of Mining and Technology. Her current research interests include image processing, machine learning, computer vision, and pattern recognition, especially on vein recognition and other biometrical identification methods.