LETTER

Hand-Dorsa Vein Recognition Based on Task-Specific Cross-Convolutional-Layer Pooling

Jun WANG[†], **Yulian LI**[†], ***Nonmembers*, *and* Zaiyu PAN**[†a)], ***Student Member***

SUMMARY Hand-dorsa vein recognition is solved based on the convolutional activations of the pre-trained deep convolutional neural network (DCNN). In specific, a novel task-specific cross-convolutional-layer pooling is proposed to obtain the more representative and discriminative feature representation. Rigorous experiments on the self-established database achieves the state-of-the-art recognition result, which demonstrates the effectiveness of the proposed model.

key words: *hand-dorsa vein recognition, pre-trained DCNN, task-specific cross-convolutional-layer pooling*

1. Introduction

Biometric identification techniques which contain fingerprint recognition [1], palmprint recognition [2], face recognition [3] and vein recognition [4], have become a powerful alternative due to their high security. Among these, vein recognition is becoming one of the most popular biometric authentication due to its lives detection, anti-counterfeit and easy acceptability. Hand-dorsa vein recognition systems generally consist of three parts such as image capture and preprocessing, feature extraction and classification. The feature extraction methods whose design is regarded as the most important part in vein recognition systems, have been widely concerned in the last few years. Currently, most of current researches mainly concentrate on designing the effective handcrafted feature extraction methods for vein recognition systems. However, it is difficult to build a more robust and discriminative vein recognition system by handcrafted feature representation methods due to their insufficient representation ability.

Deep convolutional neural network (DCNN) has successfully been applied to some large-scale image tasks due to its outstanding feature representation ability, which also obtains excellent recognition results. However, the performance of DCNN model extremely dependents on the number of training samples to some extent, resulting in the fact that it cannot achieve satisfactory results on some small-scale image recognition tasks such as hand-dorsa vein recognition. To utilize the excellent feature representation capacity of DCNN model for hand-dorsa vein recognition, a novel task-specific cross-convolutional-layer pooling is proposed, which effectively addresses the problem that it is difficult to apply DCNN model to hand-dorsa vein recognition due to the lack of training database. The framework of the proposed model is shown in Fig. 1.

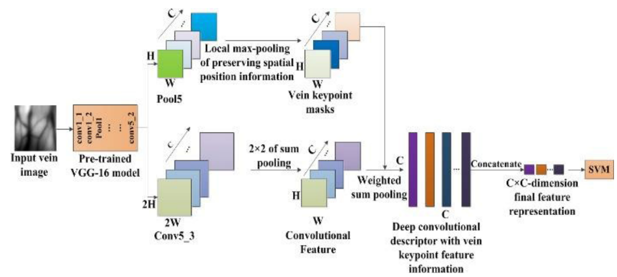

Fig. 1 The framework of the proposed model.

2. Task-Specific Cross-Convolutional-Layer Pooling

Convolutional features of a pre-trained DCNN model are adopted as a feature representation for classification, which also achieves outstanding performance. However, directly utilizing the convolutional features as feature descriptors for hand-dorsa vein cannot obtain acceptable recognition results due to the fact that the feature maps of convolutional layer contain more background information. To better analyze the traits of convolutional features with vein information, we randomly select four vein images on our lab-made database and extract their last convolutional features by the pre-trained DCNN model. The visualization results of feature maps of convolutional layer (pool5 layer) based on vein information are shown in Fig. 2. It can be seen from Fig. 2 that the single cell in the pool5 layer can correspond to two kinds of vein regions including vein region and non-vein region in original input vein images, which decreases the feature descriptor capacity of convolutional activations. Therefore, if the feature maps of one convolutional layer is directly adopted as the mask to aggregate feature maps of other convolutional layers, it cannot obtain the useful and discriminative deep convolutional descriptors for vein recognition. How to remove the non-vein information in convolutional features to explore its potential feature descriptor ability remains a key issue. In addition, during the visualization process of feature maps, we find the fact that the strongest parts of the strong responses in the feature maps of pool5 layer generally correspond to ending point or crossing point, and the strongest parts of the weak responses

Manuscript received June 6, 2019.
Manuscript revised August 18, 2019.
Manuscript publicized September 9, 2019.

[†]The authors are with School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China.
a) E-mail: pzycumt@163.com

DOI: 10.1587/transinf.2019EDL8119

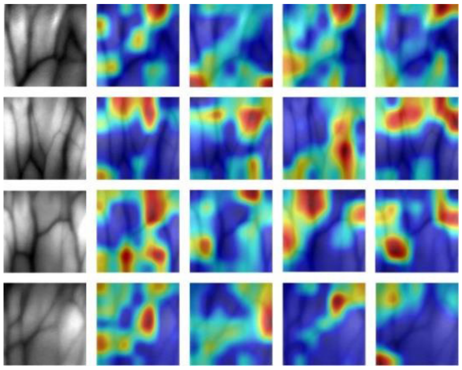

Fig. 2 The visualization results of feature maps of pool5 layer.

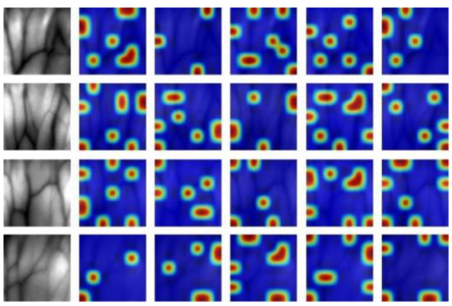

Fig. 3 The visualization results of vein key-point masks.

in the feature maps of pool5 layer generally also correspond to end point or cross point. Inspired by the fact, a novel Preserving Spatial Position of Local Max-Pooling (PSP-LMP) is proposed to acquire the key-point features of feature maps of convolutional layer.

Given the feature maps of pool5 layer $X \in R^{H\times W\times C}$, the vein key-point masks M can be obtained by conduct a 3×3 of PSP-LMP on the feature maps of pool5 layer. In the neighbourhood of 3×3, our proposed PSP-LMP can be represented as follow:

$$M_{3\times 3}(\mathrm{i,j}) = \begin{cases} 1 & if\ X_{3\times 3}(\mathrm{i,j}) == Tmax \\ 0 & else \end{cases} \tag{1}$$

where $X_{3\times 3}$ is the neighborhood of 3×3 in one feature map of pool5 layer, $M_{3\times 3}$ is the partial mask generated by applying the proposed method in the neighborhood of 3×3 and $Tmax$ is maximum of the neighborhood of 3×3.

In order to verify the effect of the proposed PSP-LMP in localizing vein key-point feature of feature maps of convolutional layer based on vein information. We randomly select four vein images on our lab-made hand-dorsa vein database and visualize the obtained vein key-point masks. The visualization results are shown in Fig. 3. It can be seen from Fig. 3 that utilizing the PSP-LMP on the feature maps of pool5 layer can accurately localize vein key-point features, which also demonstrates the effectiveness of the proposed PSP-LMP method.

After the PSP-LMP is applied on the feature maps of pool5 layer, the newly obtained feature maps of pool5 layer, which are named as vein key-point masks in this paper, do not cover non-vein information. Compare with the original feature maps of pool5 layer, vein key-point masks are more representative and discriminative. Thus, utilizing vein key-point masks to aggregate the feature maps of other convolutional layers can obtain more representative and discriminative deep convolutional descriptor. The deep convolutional descriptor with vein key-point information can be acquired as:

$$f_n = \sum_{i}^{H} \sum_{j}^{W} D_n(i, j) M_n(i, j) \tag{2}$$

Where D_n is n-th feature map of other convolutional layer ($n \in \{1, \ldots, C\}$); M_n is the vein key-point mask of n-th feature map and f_n is n-th deep convolutional descriptor with vein key-point information. It should be noted that (i, j) is a particular cell ($i \in \{1, \ldots, H\}$, $j \in \{1, \ldots, W\}$). The detailed process of aggregating deep convolutional feature based on vein key-point mask is shown in Fig. 4. The final feature representation is acquired by concatenating all selective deep convolutional descriptors which are generated by applying vein key-point masks to aggregate feature maps of other convolutional layers.

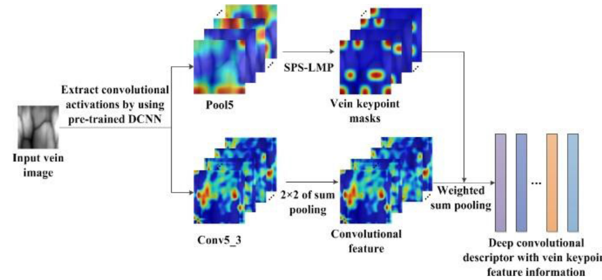

Fig. 4 The detailed process of aggregating deep convolutional feature based on vein ley-point masks.

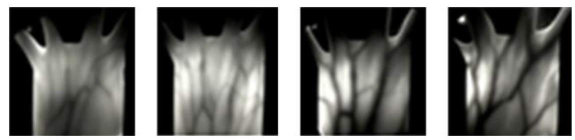

Fig. 5 Samples of lab-made database.

3. Lab-Made Hand-Dorsa Vein Database

To obtain persuasive and satisfactory classification result, a comprehensive hand-dorsa vein database is built containing 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 the fullest of the dorsal vein information, we set the size of the images as 460×680 with extremely high-quality. Figure 5 shows some samples of home-made database. The ROI extraction process [5] specifically designed for this database is conducted followed by the grey and size normalization. Note that the size of vein ROI images is 181×181.

4. Experiments and Analysis

In this part, rigorous comparison experiments on lab-made hand-dorsa vein database are designed to comprehensively evaluate the performance of the proposed model. In our experiments, we adopt VGG-16 model [6] trained on ImageNet database as the pre-trained DCNN model to extract deep convolutional features. The size of input vein ROI image is 224×224, which is generated by utilizing the "imresize" operation of MATLAB on the original vein ROI images. Due to the fact the size of obtained feature representation is too large, PCA is utilized to reduce the dimension of input data of SVM for speeding up its training procedure. In the training process of SVM, the 200×5 vein images which are captured in the first session are adopted as training samples, and the 200×5 vein images which are captured in the second session are used as test samples. In addition, the configuration of SVM used in our experiments is that its kernel function adopt the radial basis function and its penalty parameter as well as gamma are respectively set as 128 and 0.0078.

4.1 Performance Evaluation of Task-Specific Cross-Convolutional-Layer Pooling

In this section, two experiments are conducted on our lab-made vein database to evaluate the effectiveness of the proposed task-specific cross-convolutional-layer pooling model. Different cross-convolutional-layer pooling methods with or without vein key-point mask is analysed to verify the performance of our proposed model in the first experiment. It should be noted that the feature maps of pool5 layer are used to generate the vein key-point mask and the feature maps of conv5_1, con5_2 and conv5_3 layer are regarded as convolutional features which are aggregated by the obtained vein key-point mask. The comparison experiment results of different cross-convolutional-layer pooling methods with or without vein key-point mask for vein recognition are shown in Table 1.

It can be concluded from Table 1 that the recognition results achieved by different cross-convolutional-layer pooling methods with vein key-point mask are always better than the recognition results obtained by different cross-convolutional-layer pooling approaches without vein key-point mask. In addition, the best recognition results with 97.36% are generated by utilizing this cross-convolutional-layer pooling method that vein key-point mask obtained by applying the proposed PSP-LMP on the feature maps of pool5 layer is adopted to aggregate the feature maps of cov5_2 layer, which also indicates the fact that the proposed PSP-LMP method can effectively remove the non-vein information of convolutional features.

The design of the second experiment aims to evaluate the performance of the proposed aggregation method for convolutional feature. Therefore, to fully verify the advantage of the proposed model, we select the several most commonly used encoding and aggregations approaches for convolutional features such as max-pooling, average-pooling, FV, VLAD and SCDA [7] as comparison algorithms. It should be noted that the feature maps of pool5 layer are adopted as convolutional features for comparison model, and the feature maps of conv5_2 and pool5 layer are regarded as convolutional features for cross-convolutional-layer pooling with vein key-point mask. The recognition results of different aggregation methods for convolutional feature are illustrated in Table 2. It can be observed from Table 2 that the proposed task-specific cross-convolutional-layer pooling for vein recognition can achieve the best performance compared with other aggregation methods, which demonstrates the advantage of our proposed model.

Table 1 The evaluation results of different cross-convolutional-layer pooling with or without vein key-point mask for vein recognition.

Convolutional activation	Aggregation methods	Recognition rate (%)
Conv5_1 Pool5	Cross-convolutional-layer pooling	91.97
	Cross-convolutional-layer pooling with vein key-point mask	97.21
Conv5_2 Pool5	Cross-convolutional-layer pooling	92.35
	Cross-convolutional-layer pooling with vein key-point mask	97.36
Conv5_3 Pool5	Cross-convolutional-layer pooling	91.67
	Cross-convolutional-layer pooling with vein key-point mask	96.94

Table 2 The recognition results of different aggregation methods.

Convolutional activation	Aggregation methods	Recognition rate (%)
Pool5	Max-pooling	88.46
	Average-pooing	91.52
	FV	81.08
	VLAD	87.27
	SCDA	93.19
Conv5_2 Pool5	Proposed	97.36

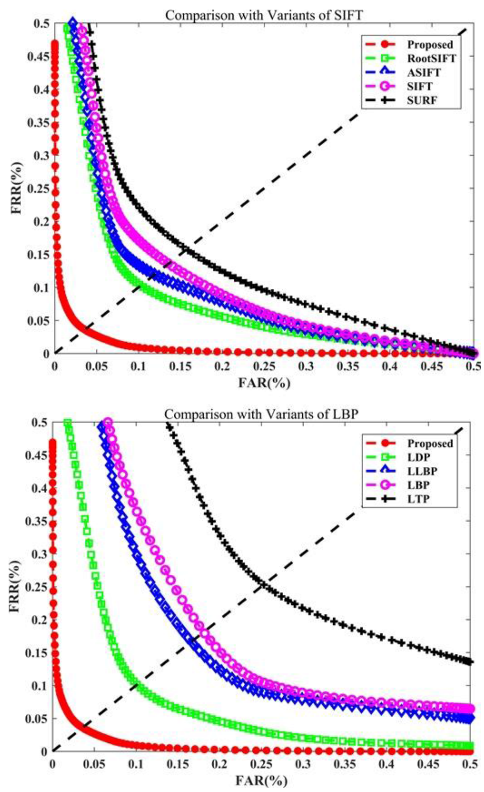

Fig. 6 Comparison of ROC curves between the proposed model and representative state-of-the-art handcrafted methods.

4.2 Comparison with State of the Art Models

Two kinds of representative hand-crafted feature extraction algorithms are used as reference: The one is the local invariant feature model including SIFT, SURF, RootSIFT, ASIFT, 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 and its variants including LDP, LTP, and LLBP, and such model is widely applied for vein based identification application for its efficiency, and it also provides competitive recognition results. The overall performance comparison is illustrated in Fig. 6.

Judging from EER result of verification with the lab-made database, it can be concluded that the proposed model performs far better than the LIF models with EER as 0.042% 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 discriminative feature representation.

5. Conclusions

This paper propose a novel task-specific cross-convolutional-layer pooling model for hand-dorsa vein recognition, which fully utilizes the outstanding feature representation ability of a pre-trained DCNN model. First, the pre-trained DCNN model such as VGG-16 are used to extract convolutional features of input vein images. Then, the PSP-LMP is proposed to generate vein key-point mask by localizing the vein key-point information of feature maps of pool5 layer. Next, vein key-point mask is utilized to aggregate the feature maps of conv5_2 layer for obtaining more discriminative and useful deep convolutional features for vein recognition. State-of-the-art vein recognition results demonstrate the effectiveness of our proposed task-specific cross-convolutional-pooling model. In the future, we will design new attention mechanism to aggregate deep convolutional features and obtain more discriminative and richer convolutional features for vein recognition.

Acknowledgments

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

References

[1] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature matching for partial face recognition," IEEE Trans. Image Process., vol.28, no.2, pp.791–802, 2019.
[2] S. Zhang, H. Wang, W. Huang, and C. Zhang, "Combining modified LBP and weighted SRC for palmprint recognition," Signal, Image and Video Processing, vol.12, no.6, pp.1035–1042, 2018.
[3] C. Lin and A. Kumar, "Contactless and partial 3D fingerprint recognition using multi-view deep representation," Pattern Recognition, vol.83, pp.314–327, 2018.
[4] 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, July 2017.
[5] J. Wang, G. Wang, M. Li, K. Wang, and H. Tian, "Hand vein recognition based on improved template matching," Int. J. Bioautomation, vol.18, no.4, pp.337–348, Dec. 2014.
[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv: 1409.1556, 2014.
[7] X.-S. Wei, J.-H. Luo, and J. Wu, "Selective convolutional descriptor aggregation for fine-grained image retrieval," IEEE Trans. Image Process., vol.26, no.6, pp.2868–2881, 2017.