LETTER

Bimodal Vein Recognition Based on Task-Specific Transfer Learning

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

SUMMARY Both gender and identity recognition task with hand vein information is solved based on the proposed cross-selected-domain transfer learning model. State-of-the-art recognition results demonstrate the effectiveness of the proposed model for pattern recognition task, and the capability to avoid over-fitting of fine-tuning DCNN with small-scaled database.
key words: *gender recognition, vein recognition, transfer learning*

1. Introduction

Recent researches on other biometric traits in the field of personal identification including faces [3], and palmprint [4] etc. motivate another research topic "soft biometric", the ability of which is argued and verified in terms of serving as the complementary information for better biometric recognition performance in an unified framework. Apart from enhancing the performance of identity recognition, soft biometric, covering gender, age, height, emotion and other related attributes, are also being investigated widely and deeply because of their potential applications in social interactions, forensics, surveillance, entertainment, criminalistics, and military affairs. Among all these soft biometric characteristics, gender information is one of the most widely researched fields because of its simplicity and practical applications. However, the unconstrained situations such as arbitrary pose, non-uniform illumination and the characteristics of being easy to be copied and destroyed render in unreliable gender classification system with the existing traits. To address such problems and promote the development of gender information based applications, the first hand vein based gender classification model is proposed in this letter. Also this is the first bimodal recognition model with hand vein information for both gender and identity recognition task.

Transfer learning [1], [2] based feature extraction refers to the situation that a state-of-the-art Deep Convolutional Neural Network (DCNN) model trained with large-scaled database can be adopted as universal image feature descriptor, and doing so leads to impressive performance for various of image recognition tasks especially for the condition that the supply of target training data is limited [5]. Traditional transfer learning models usually adopt the activations from a single DCNN layer, usually the second fully-connected layer, as final representation. The performance, however, will be degraded greatly if the pattern distribution between the source training data (domain A) and the target training data (domain B) is greatly different, thus making the idea that adopting state-of-the-art DCNN model (such as VGG [6] and AlexNet [7]) trained on large-scaled dataset for generating discriminative feature representation for small-scaled dataset unreliable and making great similarity between source domain and target domain a necessity for generating effective and discriminative feature representation. To take full advantage of DCNN for generating highly discriminative feature representation so as to improve the vein based gender classification, and also solve the problem of degraded performance caused by great domain difference between hand-dorsa vein image and source training image of DCNN (VGG or AlexNet), the DCNN re-trained with hierarchical transfer learning strategy, instead of adopting the original DCNN as feature extractor directly, is proposed for discriminative feature generation. The model training guided by the "coarse-to-fine" transfer learning strategy is a hierarchical procedure (as shown in Fig. 1) of model re-training with different database, and each neighboring database shares more similar pattern distribution, thus improving the discriminative ability of the resulted feature representation for specified task. Besides, the task of identity recognition is also solved under the "general-to-specific" training framework. What's more, the problem of over-fitting when training deep architecture with small-scaled database is also tackled with the proposed hierarchical knowledge transfer strategy.

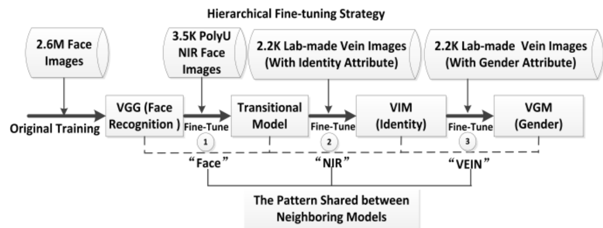


Fig. 1 Proposed hierarchical knowledge transfer procedure.

2. Guided DCNN Re-Training for Feature Extraction

Unlike traditional DCNN based feature extraction model that directly adopting the output of the second fully con-

Manuscript received February 14, 2017.
Manuscript revised March 19, 2017.
Manuscript publicized April 17, 2017.
[†]The authors are with School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116 China.
a) E-mail: wangguoqingcumt@163.com
DOI: 10.1587/transinf.2017EDL8031

nected layer (usually FC7) in the pre-trained model as feature vector, we try to obtain the discriminative and robust feature representation by way of fine-tuning the state-of-the-art DCNN model (the VGG-face [8] model is selected as the basis in this paper and the general diagram for VGG-face could be referenced from [8]) with the lab-made database, and the linear classifier is embedded into the fine-tuning network for obtaining discriminative task-specific feature representation.

It could be observed from Fig. 1 that the fine-tuning procedure is not realized directly from VGG (source) to Identity or Gender (target), but there exists some intermediate models for step-aware transfer learning. An interesting setup driven by numerous of experiments and the conclusion in [9] could be analyzed that each neighboring model shares some common patterns, which in this paper shows as "Face (VGG face to NIR face)-Near Infrared (NIR face to NIR vein)", during transfer learning and such design could well find the similar weights for the same pattern for the purpose of speeding up the convergence of training. On the other hand, such a "coarse-to-fine" (which could also be referenced as "general-to-specific") transfer learning scheme has the advantage of relieving the over-fitting.

The specific transfer learning procedure under the "coarse-to-fine" fine-tuning scheme is as follows: Firstly, VGG face model trained on very large (2.6M) face database is selected as the very beginning model, and then it is fine-tuned on small-scaled PolyU NIR face database with only identical attribute annotation, and the shared pattern of face images could help speed up the convergence by starting the training process from the latent face descriptive parameter space generated by former model, and the model fine-tuned in this step is called FRM. Secondly, we fine-tune the FRM on the small-scaled lab-made hand vein database with annotated identity attribute for purpose of vein recognition, and the VIM model is obtained by training the FRM with new database under fine-tuning scheme. Finally, by exploiting the inherent correlation between identical attribute and gender attribute within vein images, the VGM is obtained by training the VIM with the same database under the "general-to-specific" scheme, and the structure of the output is adjusted according to specific task during each transfer learning step.

The "coarse-to-fine" transfer learning strategy is far better than the randomly initialized parameters because the inherent correlation between neighboring models could on the one hand help obtain better representation and on the other hand speed up the convergence. Furthermore, two sequential training with the same database for different purpose indirectly increase the amount of training database and relieve the problem of over-fitting.

After obtaining robust task-specific feature representation, the simple but effective linear-SVM is adopted as a basis for robust classifier design, and the entire framework of the proposed strategy is illustrated as Fig. 2.

In the gender classification with hand vein information experiment, the linear-SVM is adopted directly as bi-class classifier, and multi-class classifier by combining numbers of bi-class linear-SVM training within grouped samples is realized for identity recognition.

3. Datasets and Experimental Setup

Two main databases in terms of face and hand vein are adopted in the experiments. One is the PolyU NIR-face database for fine-tuning the VGG model to obtain FRM, and it consists of 335 samples imaged under near-infrared condition. Before feeding the face images into DCNN, a robust joint face detection and alignment model [10] based on multitask cascaded CNNs is adopted for region of interest (ROI) cropping followed by normalizing the ROI as 224*224. Another one is the lab-made hand-dorsa vein database containing 98 females and 102 males whose ages vary from 19 to 62. For each sample, 10 hand-dorsa vein images were acquired in two specifically set sessions separated by a time interval of more than 10 days, and at each time, five images were acquired from each subject at the wavelength of 850nm, Fig. 3 shows some samples of male and female in the dataset. To the fullest of the dorsal vein information, we set the size of the images as 460*680 with extremely high-quality. The ROI extraction [11] process specifically designed for this database is conducted followed by the same size normalization with the face database.

Caffe deep learning framework is adopted for fine-tuning DCNN models and extracting the FC7 layer features for SVM training. As the common parameters configuration of all three models, a momentum of 0.9, a weight decay of 0.0005 and total learning iterations of 30000 are used as default value. The basic learning rate for fine-tuning FRM is 0.01 and 0.001 for VIM and VGM, while the learning rate is reduced by polynomial with gamma as 0.1, and also batch size is set as 120.

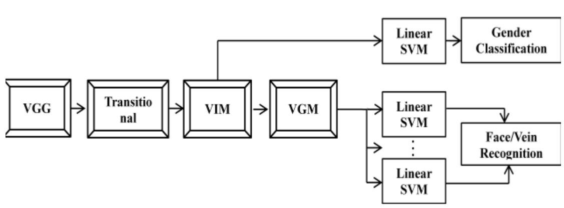


Fig. 2 Identity and gender recognition with the proposed model.

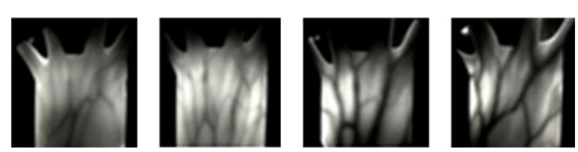


Fig. 3 Samples of Lab-made database (the left two are the female samples and the right two are male ones).

4. Experiments and Analysis

4.1 Effect of Different Fine-Tuning Strategies

In this part, both the identity and gender recognition experiments with the lab-made database are evaluated to demonstrate the effectiveness of the proposed model. What's more, the effect of different fine-tuning strategies is also evaluated to further verify the superiority.

4.1.1 Identity Recognition with Hand Vein Information

In this part, all experiments are conducted under hypothesis that the VIM is the ending part of the model. Different fine-tuning strategies including direct and "coarse-to-fine" fine-tuning are evaluated. The direct one refers to the procedure that fine-tuning VIM directly from VGG face or AlexNet without PolyU NIR face while the other one fine-tunes the VIM from the transitional model fine-tuned from VGG with random face database or PolyU NIR face database under the "coarse-to-fine" scheme, thus resulting in five different models for identity recognition experiment, and the corresponding results with the four models are illustrated in Fig. 4.

It could be concluded from the results in Fig. 4 that even the worst performance with EER as 0.415% fine-tuned directly from AlexNet is competitive with those hand-crafted ones, which demonstrates the ability of DCNN in learning discriminative representation for a pre-defined classification task. On the other hand, the results of the indirect fine-tuned with the transitional process performs a little bit better than the direct one, indicating that training a deep neural network with larger amount of data improves the ability of the model for learning more discriminative representation.

4.1.2 Gender Classification with Hand Vein Information

In this part, four fine-tuning strategies including fine-tuning

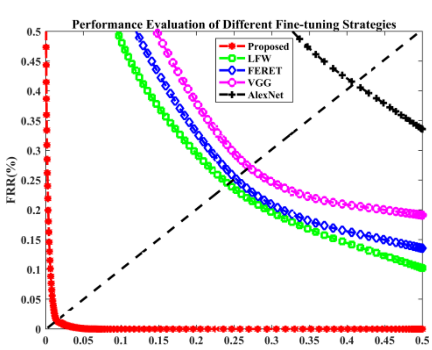


Fig. 4 Comparison of ROC curves with different fine-tuning strategies (VGG and AlexNet represent the two direct fine-tuning models while LFW and FERET represent "coarse-to-fine" fine-tuning models).

the VGM directly from the VGG or AlexNet without transitional parts and the other two refer to removing the VIM or not from the entire framework in Fig. 1. The gender classification results in Table 1 fully demonstrate the effectiveness of the "general-to-specific" fine-tuning strategy, which could also serve as a guideline for solving other computer vision task with transfer learning.

4.2 Comparison with State of the Art Models

In this part, bimodal experiments in terms of both identity and gender recognition with hand vein images are conducted, and also the state-of-the-art hand-crafted feature extraction is also adopted as reference to demonstrate the superiority of the proposed feature learning model.

4.2.1 Identity Recognition with Hand Vein Information

Two kinds of representative hand-crafted feature extraction algorithms are used as reference: The one is the local invariant feature model including SIFT, SURF, ASIFT, RootSIFT, and it has the advantages of being invariant to rotation, translation, scale uncertainty and even nonuniform 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.

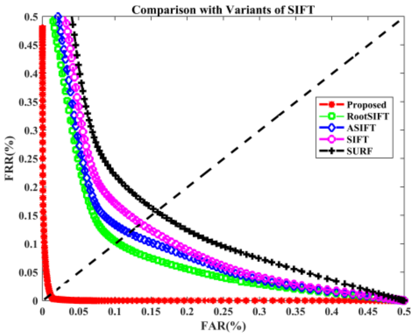
Judging from EER result of verification with the lab-made database, it can be concluded that the proposed feature learning model performs far better than the LIF models with EER as 0.013% whereas the best of LIF is 0.105% with RootSIFT and the best of LBPs is 0.113% with LDP, and the state-of-the-art identity recognition results fully demonstrate the ability of the proposed model for learning the robust and discriminative identity-specific feature.

4.2.2 Gender Classification with Hand Vein Information

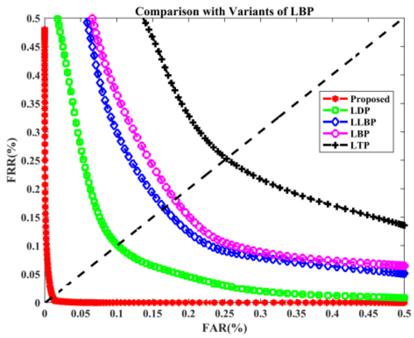
Driven by the general idea to solve an unknown problem that borrowing knowledge from another similar problem, we focus on review the existing framework for gender classification with faces, and then rigorous experiments with the representative of the researched methods for solving gender classification with hand vein information is conducted. Unsatisfied results, however, motivate us to generate the proposed "coarse-to-fine" fine-tuning framework to learn the best gender-specific feature with the DCNN model. The experimental results shown in Table 2 fully demonstrate the effectiveness of the proposed model for gender classification task.

Table 1 Gender classification results with different transfer learning strategies.

Models	AlexNet	VGG	Without VIM	Proposed
Accuracy	16.38%	29.56%	59.41%	**91.60%**



(a)



(b)

Fig. 5 Comparison of ROC curves between the proposed feature learning methods and representative state-of-the-art hand-crafted methods (a: SIFTs, b: LBPs)

Table 2 Gender classification results with different transfer learning strategies.

Original Modality	Model Name	Accuracy
Gender Classification with body	BIF+PCA	81.5±1.5%
Bimodal gender classification with face and corresponding fingerprint	BOW+D-LDA	86.5±2.3%
Gender Classification with face	ICA+LDA	83.5±3.8%
Gender Classification with hand-dorsa vein	Fine-tuned DCNN+SVM	**91.6±1.2%**

5. Conclusions

Gender, one of the most widely researched "soft biometric" attribute, has been widely used for numerous applications. To cope with the latent problem of existing gender classification systems, the first model for designing more robust gender classification system with hand-dorsa vein images is proposed with hierarchical transfer learning strategy, and the identity recognition problem is also solved within the same framework, which is the first model for both gender and identity recognition. State-of-the-art recognition results in both situations fully demonstrate the effectiveness of the proposed model.

We also argue that the proposed hierarchical transfer learning strategy with special database selection for utilizing the inherent correlation within source and target dataset is also applicable for other computer vision task solved with transferable DCNN models.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 61379143 and the Provincial Innovative Program for Graduate Training under Grant SJLX15_0693.

References

[1] L. Zhen, J. Yang, and H. Liu, "Learning from multiple sources via multiple domain relationship," IEICE Trans. Inf. & Syst., vol.E99-D, no.7, pp.1941–1944, July 2016.

[2] Y. Cheng, X. Wang, and G. Cao, "Multi-source tri-training transfer learning," IEICE Trans. Inf. & Syst., vol.E97-D, no.6, pp.1668–1672, June 2014.

[3] A. Moeini and H. Moeini, "Pose-invariant gender classification based on 3D face reconstruction and synthesis from single 2D image." ELECTRON LETT., vol.51, no.10, pp.760–761, May 2015.

[4] 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, March 2014.

[5] S.J. Pan and Q. Yang, "A Survey on Transfer Learning." IEEE Trans. Knowl. Data Eng., vol.22, no.10, pp.1345–1359, Oct. 2010.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Proc. 2015 IEEE Conference on Comput. Vis. Pattern Recognit., pp.1–9, 2015.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," Proc. 25th International Conference on Neural Information Processing Systems, pp.1097–1105, 2012.

[8] O.M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," Proc. British Machine Vision Conference, pp.1–12, 2015.

[9] G. Ozbulak, Y. Aytar, and H.K. Ekenel, "How Transferable Are CNN-Based Features for Age and Gender Classification?" Proc. 2016 International Conference of the Biometrics Special Interest Group, pp.1–6, 2016.

[10] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," International Journal Bioautomation.., vol.23, no.10, pp.1499–1503, Oct. 2016.

[11] 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.