Palm Vein Identification Based on Vein Segmentation and Triplet Loss Function

Denis Trofimov [1] and Elena Pavelyeva [1]

[1] *Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskiye Gory 1-52, Moscow, 119991, Russia*

Abstract

In this article the new neural network algorithm for palm vein identification using the triplet loss function is proposed. The neural network model is based on the VGG16 architecture. The similarity learning problem instead of the classification problem is considered. The number of image classes is assumed to be unknown so at the output of the neural network the feature vector is obtained, and then for the pair of palm vein images the distance between them is calculated. Minimization of triplet loss function while training leads to the decrease in distances between the images of the same class, while the distances between the images of different classes increase. The neural network was trained using preprocessed and segmented images from CASIA multi-spectral palmprint image database. The use of segmentation information for palm vein recognition improves the recognition results. Experimental results demonstrate the effectiveness of the proposed method. The value of EER=0.0084 is obtained.

Keywords

Biometrics, palm vein identification, triplet loss, neural network, image segmentation

1. Introduction

Every person has unique biometric characteristics [1]. Some of them are obtained from birth, for example, DNA, fingerprints, iris; others are acquired over time and can change throughout life: gait, voice, signature, etc. Each biometric characteristic has its advantages and disadvantages.

Palm vein recognition is a young and modern technology for human identification. Vein blood contains deoxygenated hemoglobin that absorbs near infrared light, so vein images are obtained in near infrared illumination. Palm veins are usually not visible to others and vein patterns are quite unique, so they can be used for human identification with low risk of biometric information fake or theft.

Currently, the problem of palm vein identification is usually solved based on the use of neural networks, although there are classical mathematical methods for palm vein biometry. For example, in [2] the palm vein identification method based on Gabor wavelet features is proposed. The method consists of five steps: image acquisition, region of interest detection, image preprocessing, features extraction, and feature matching. In [3] authors propose an end-to-end deep CNN framework, called PVSNet, where an encoder-decoder network is used to learn generative domain-specific features followed by a Siamese network in which convolutional layers are pre-trained in an unsupervised fashion as an autoencoder. This model is trained with the triplet loss function that is modified for learning feature embeddings by minimizing the distance between the embedding-pairs from the same subject and maximizing the distance with those from different subjects, with a margin. Authors of [4] propose a palm vein recognition system that combines two approaches using a decision-level fusion strategy. The first approach employs Binarized Statistical Image Features (BSIF) descriptor method on five overlapping sub-regions of palm vein images and the second approach uses a convolutional neural networks (CNN) model on each palm vein image.

Neural network based methods for biometric identification usually solve the classification problem with a preliminary known number of classes. The problem with this approach is the need to retrain the network if a new image class is added. Therefore, the development of neural networks that have feature vectors in neural network output is a promising approach for biometric identification. Then the distance between any two images based on the feature vectors can be calculated, thus, the problem is reduced to the classical formulation for biometry.

In this article the CNN-based method for palm vein identification using the triplet loss function is proposed. The neural network was trained using preprocessed and segmented images from CASIA multi-spectral palmprint image database [5]. The method of CNN training using segmented images of the initial dataset is widely used to improve the performance of networks [6, 7]. The palm vein image segmentation is proposed in [8]. The vein structure is detected using unsupervised learning approach based on W-Net architecture, that ties together into a single autoencoder two fully convolutional neural network architectures, each similar to the U-Net. Then segmentation results are improved using principal curvatures technique. Some vein points with highest maximum principal curvature values are selected, and the other vein points are found by moving from starting points along the direction of minimum principal curvature. The final vein image segmentation results are obtained by intersection of the principal curvatures-based and neural network-based segmentations. The use of segmentation information for palm vein recognition allows us to improve the recognition results.

2. Triplet loss function

The goal of triplet loss function is to provide that two images of the same class have their embeddings close together in the embedding space and two images from different classes have their embeddings far away. Given two images of one class and one image from the other class, the intraclass distance should be less than the interclass distance by some margin (Fig. 1, Fig. 2).

The triplet loss function [9] is defined on triplets of the form $\left(x_i^a, x_i^p, x_i^n\right)$, where x_i^a is an image of a certain class (anchor), x_i^p is an image of the same class (positive example), x_i^n is an image of a class other than the class of images x_i^a and x_i^p (negative example). The elements of the triplet are fed to the input of the neural network mapping $f(x)$, then the everywhere differentiable loss function is calculated

$$L = max\left(0, \; \left\|f(x_i^a) - f\left(x_i^p\right)\right\|_2 - \|f(x_i^a) - f(x_i^n)\|_2 + \alpha\right), \tag{1}$$

where $\| \quad \|_2$ is the $L_2 - norm$, α is a positive parameter. As a result of minimizing such loss function, the mapping $f(x)$ has to acquire the following properties: each anchor mapping $f(x_i^a)$ becomes closer to all mappings of positive examples $f\left(x_i^p\right)$ than to any mapping of negative examples $f(x_i^n)$ (see Fig. 2).
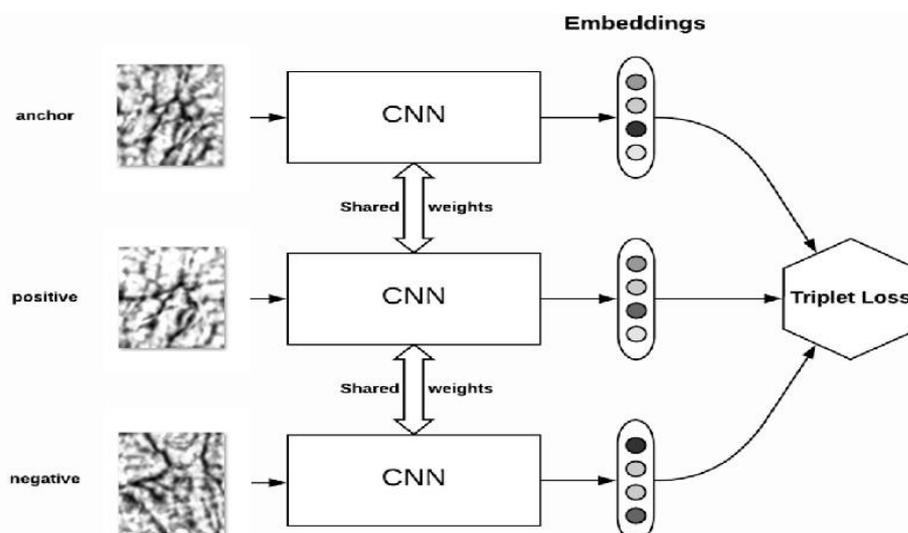

Figure 1: Triplet loss on two positive images of the same class and one negative image from another class
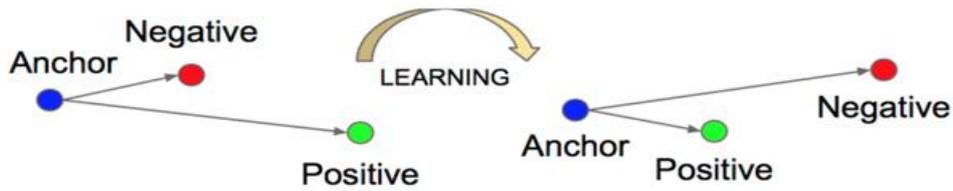
Figure 2: Training scheme using triplet loss

To calculate the loss function, we need firstly to select the triplets. The overall number of triplets has a cubic dependence on the number of images, so it is extremely difficult to iterate through all triplets in practice. Moreover, this search does not make sense, because among all possible triplets there are a large number of trivial ones for which

$$\left\|f(x_i^a) - f(x_i^p)\right\|_2 < \left\|f(x_i^a) - f(x_i^n)\right\|_2$$

so the loss function (1) is zero. Using them in training only slows down calculations and does not make significant changes to the network weights. There are several strategies for selecting triplets:

- Offline strategy: Triplets of examples are generated at the beginning of each epoch. This strategy is inefficient due to the large number of trivial triplets and complicates the training of the network.
- Online strategy: N images are used for each iteration of the network training. Then, for each such iteration, N^3 triplets are created. The relevant ones are selected from them, that is, triplets consisting of two images of one class and images from any other class.

Then there are two ways:

- Batch-all: We select the average loss value among triplets in which the distance between the anchor and the negative example is closer than between the anchor and the positive example, and triplets for which the distance difference between the anchor and the positive example and the anchor and the negative example does not exceed the coefficient α.
- Batch-hard: For each of the anchors in the iteration, we find a positive example with the maximum the distance to the anchor, and we find a negative example with the minimum distance to the anchor.

In this paper, the online batch-hard strategy was chosen.

3. Preprocessed and Segmented Dataset

A preprocessed set of images from the CASIA multi-spectral palmprint image database [5] was selected as the training and testing data set [10, 11]. In total, this set contains 200 different classes (images of the right and left hands of 100 people) with 5 to 6 images of the size of 128×128 pixels in each class (see Fig. 3). The preprocessing process is described in [8]. At first, the contrast-limited adaptive histogram equalization (CLAHE) is used to increase the image contrast. Then the non-local means (NLM) algorithm is used to reduce noise. NLM blurs both noise and veins so CLAHE is applied again.

The segmentation process is described in [11]. The vein structure is detected using unsupervised learning approach based on W-Net [12] architecture, that ties together into a single autoencoder two fully convolutional neural network architectures, each similar to the U-Net. Then segmentation results are improved using principal curvatures technique. Some vein points with highest maximum principal curvature values are selected, and the other vein points are found by moving from starting points along the direction of minimum principal curvature. The final vein image segmentation results are obtained by intersection of the principal curvatures-based and neural network-based segmentations.
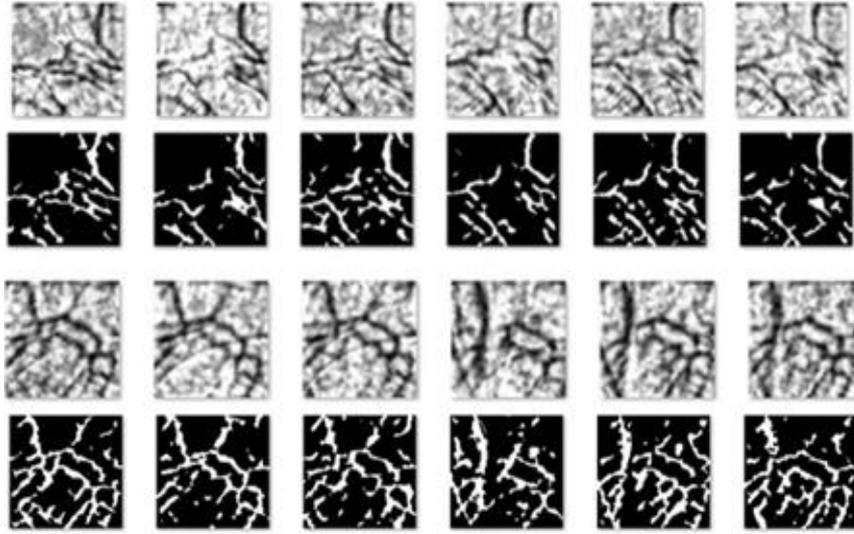
Figure 3: Preprocessed and segmented images

4. Neural network model for palm vein identification problem

The following neural network model based on the VGG16 network was selected for palm vein recognition (see Fig. 4). The size of the original images (128×128) is smaller than the size of the images in [13] (224×224), so the neural network architecture is modified. The input of the neural network receives one or two grayscale images with a size of 128×128 pixels (Fig. 4). We use one input image to obtain the recognition results using only preprocessed vein images, while we use two input images for recognition using preprocessed and segmented images. Each module consists of two convolutional layers of size 3×3 (Fig. 4). The modules are connected via 2×2 layers of max-pulling. The last layer of the neural network is a 78-component feature vector (the dimension of the output vector was found empirically).
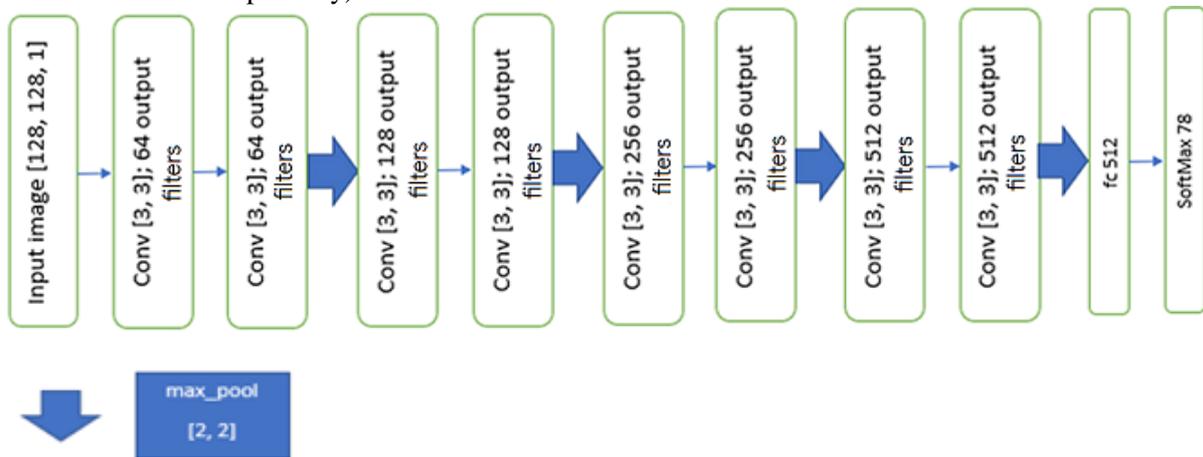

Figure 4: The neural network model

5. Neural Network training

Neural network was trained over 100 epochs, in each of which the best version of the neural network model was selected using the loss function. 826 images were used for training (413 images of the preprocessed dataset and 413 segmented images of the same classes (see Fig. 3). The loss function

$$\sum_{i=1}^{K} \max(0, \; \|f(A_i) - f(P_i)\|^2 - \|f(A_i) - f(N_i)\|^2 + \alpha)$$

is used, where A, P, N – anchor, positive and negative examples, respectively, of α - indent, $K = 100$ - the total number of triplets. Fig. 5 shows the dependence of the loss function on the network training epoch. The parameter value $\alpha = 0.5$ was found empirically.
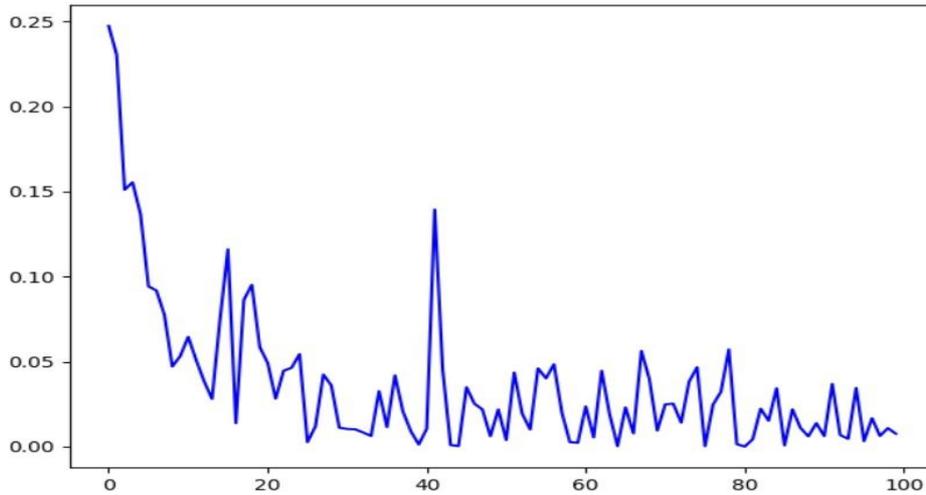

Figure 5: Dependence of the loss function on the network training epoch. The abscissa axis shows the epoch number, the ordinate axis – the loss function

6. Experimental results

To test the developed method, the FAR, FRR and EER values were used.

False Acceptance Rate (FAR) - the probability that the system incorrectly matches the input pattern to a non-matching template in the database. It measures the percent of invalid inputs that are incorrectly accepted.

False Rejection Rate (FRR) - the probability that the system fails to detect a match between the input pattern and a matching template in the database. It measures the percent of valid inputs that are incorrectly rejected.

Equal Error Rate (EER) is the rate at which FAR and FRR are equal. In general, the lower EER, the most accurate biometric system.

After training and testing the network, the following results are obtained (Fig 6).
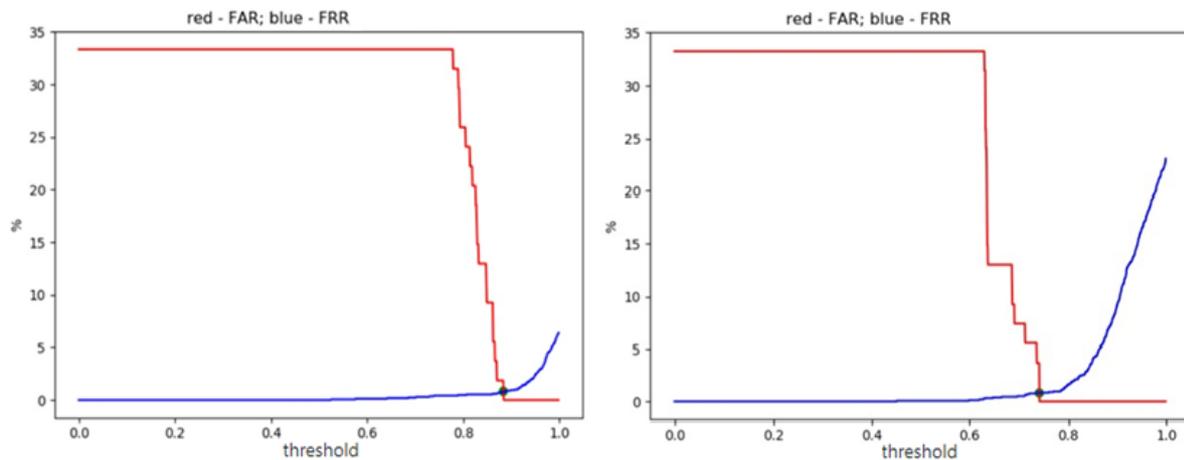

Figure 6: FAR, FRR, EER visualization for a network trained on preprocessed images (left), preprocessed and segmented images (right). EER=0.0095 (left), EER=0. 0084 (right)

It is shown that the use of segmentation information for palm vein recognition improves the recognition results. The EER value 0.0084 is obtained.

A comparison of EERs derived from different approaches for CASIA database is given in Table 1. It shows the good performance of the proposed method.

Table 1
Summary of related approaches for palm vein verification using CASIA database

Algorithm	EER
Zhang H., Hu D. [14]	0.0182
Jhong S. Y. et al. [15]	0.0346
Raghavendra R., Busch C. [16]	0.1010 ± 0.0102
Zhong D. et al. [17]	0.000222
Proposed algorithm	0.0084

The proposed algorithm has an EER worse than the algorithm [17]. However, in [17] the authors have the classification problem with a predetermined number of image classes. For the identification problem with an unknown number of image classes, the proposed algorithm has shown a good result.

7. Conclusion

The new neural network algorithm for palm vein image identification using the triplet loss function is proposed. The preprocessed and segmented datasets of CASIA multi-spectral palmprint image database are used. The use of segmented data allows us to improve the recognition results. The value of EER = 0.0084 is obtained.

8. References

[1] A. K. Jain, R. Bolle, S. Pankanti, Biometrics: personal identification in networked society, Springer Science & Business Media, Vol. 479, 2006.
[2] R. Wang, G. Wang, Z. Chen, Z. Zeng, Y. Wang, A palm vein identification system based on Gabor wavelet features, Neural Computing and Applications 24(1) (2014) 161-168.
[3] D. Thapar, G. Jaswal, A. Nigam, V. Kanhangad, PVSNet: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features, in: 2019 IEEE 5th international conference on identity, security, and behavior analysis (ISBA), 2019, pp. 1-8.
[4] F. O. Babalola, Y. Bitirim, Ö. Toygar, Palm vein recognition through fusion of texture-based and CNN-based methods, Signal, Image and Video Processing 15(3) (2021) 459-466.
[5] CASIA Multi-Spectral Palmprint Image Database, URL: http://biometrics.idealtest.org/.
[6] A. Galdran, A. Anjos, J. Dolz, H. Chakor, H. Lombaert, I. B. Ayed, The little w-net that could: state-of-the-art retinal vessel segmentation with minimalistic models, arXiv preprint arXiv:2009.01907, 2020.
[7] R. Tobji, W. Di, N. Ayoub, FMnet: iris segmentation and recognition by using fully and multi-scale CNN for biometric security, Applied Sciences 9(10) (2019) 1-17.
[8] E. I. Safronova, E. A. Pavelyeva, Palm Vein Recognition Algorithm using Multilobe Differential Filters, in: Int. Conference on Computer Graphics and Vision GraphiCon, Vol. 29, 2019, pp. 117-121.
[9] X. Dong, J. Shen, Triplet loss in siamese network for object tracking, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 459-474.
[10] C. L. Liu, F. Yin, D. H. Wang, Q. F. Wang, CASIA online and offline Chinese handwriting databases, in: 2011 International Conference on Document Analysis and Recognition, 2011, pp. 37-41.
[11] E. Safronova, E. Pavelyeva, Unsupervised Palm Vein Image Segmentation, CEUR Workshop Proceedings, 2020, Vol. 2744, Paper 40, pp. 1-12.
[12] X. Xia, B. Kulis, W-net: A deep model for fully unsupervised image segmentation, arXiv preprint arXiv:1711.08506, 2017.
[13] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255.

[14] H. Zhang, D. Hu, A palm vein recognition system, in: 2010 International Conference on Intelligent Computation Technology and Automation, Vol. 1, 2010, pp. 285-288.

[15] S. Y. Jhong, P. Y. Tseng, N. Siriphockpirom, C. H. Hsia, M. S. Huang, K. L. Hua, Y. Y. Chen, An automated biometric identification system using CNN-based palm vein recognition, in: 2020 International Conference on Advanced Robotics and Intelligent Systems (ARIS), 2020, pp. 1-6.

[16] R. Raghavendra, C. Busch, Novel image fusion scheme based on dependency measure for robust multispectral palmprint recognition, Pattern recognition 47(6) (2014) 2205-2221.

[17] D. Zhong, S. Liu, W. Wang, X. Du, Palm vein recognition with deep hashing network, in: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2018, pp. 38-49.