Local vein texton learning for finger vein recognition

Lu Yang, Gongping Yang, Yilong Yin*, and Lumei Dong,

School of Computer Science and Technology, Shandong University,
Jinan, 250101, P.R. China
yangluhi@163.com, gpyang@sdu.edu.cn, ylyin@sdu.edu.cn,
dongcomeon@gmail.com

Abstract. In finger vein recognition, the input image is generally labeled in accordance with the nearest enrolled neighbor. However, it is so rigid that it is inadequate for some cases. This paper explores a modified sparse representation method for finger vein recognition. In the method, each block in a finger vein image will be sparsely represented by dictionary textons, not simply labeled by the nearest enrolled block, and the representation coefficients of all blocks are arranged to be a two-dimensional histogram to model the image. As textons is learned from local vein pattern, not global vein pattern. Therefore, for encode global geometric information of finger vein pattern, the representation coefficient histogram is projected to different lines, and then connected in parallel to generate more powerful image features. Extensive experiments on the HKPU finger vein database show the effectiveness of the modified sparse representation method in finger vein recognition.

Keywords: Finger vein recognition, texton learning, linear projection

1 Introduction

Finger vein, a new physiological biometric trait, has been explored for personal identity by research groups in recent years. As near-infrared light can be absorbed by the hemoglobin in finger vein, but can transmit other finger tissues, it is used to capture finger vein pattern, which will be shown as shadow patterns in image [1]. Like other biometric traits, finger vein has several desirable properties, for example, universality, distinctiveness, permanence and acceptability. In addition to, it has other distinct advantages in living body identification, noninvasive and noncontact image capture and spoofing resistance.

In the post, several algorithms for finger vein feature extraction have been proposed, which can be categorized into three groups: (1) Vein pattern based algorithms, for example, repeated line tracking [2], maximum curvature point [3], mean curvature [4] and Gabor filter [5]. These algorithms first segment vein network, and then use the topological structure of vein network as vein feature. However, the segmental results are often unsatisfying in low quality images. (2) Local binary pattern based algorithms, for example, local binary pattern (LBP) [6], local line binary

*Corresponding Author: Yilong Yin, e-mail: ylyin@sdu.edu.cn.

pattern (LLBP) [7] and personalized best bit map (PBBM) [8]. These methods are less practical in finger vein recognition, because they do not distinguish vein area and background area of image in feature extraction. (3) Minutia based algorithm [9]. As the number of minutiae in finger vein is limited, the recognition result of this kind algorithm is defective.

Although some advancements have been made, a phenomenon may be a hindrance to the development of finger vein recognition, that the input finger vein image is generally labeled in accordance with the nearest enrolled neighbor. This kind of nearest neighbor based classification is so rigid that it is inadequate for some cases. For example, if the distances between the input image and two or more enrolled images are very approximate, it is obviously inappropriate to label the input image based on the nearest enrolled image. Hence, rigid nearest neighbor based classification may be not optimal.

Recently, the theory of sparse representation (SR) has been successfully used in pattern recognition. Wright *et al.* [10] introduced a sparse representation framework for robust face recognition, in which the original face samples are used as dictionary to sparsely represent input face image. Yang *et al.* [11] use the metafaces, learned from original face images, to represent input face image under the framework of sparse representation. Xie *et al.* [12] uses textons learned from image patches to model texture images in texture classification. Xin *et al.* [13] introduced sparse representation for finger vein recognition, in which the dimensionality of finger vein image is firstly reduced by sparsity preserving projection (SPP), and then SR is used for finger vein representation in low dimensional space. The method in [13] is based on the sparse property of finger vein image, but only if the number of training images is large enough, images can show sparse property.

In finger vein image, some local vein patterns look similar at a certain extent, and these patterns appear in images at high frequencies, which perhaps can be used as textons. So, like patch-based sparse texton learning [12], in this paper, we try to use the combination of blocks in enrolled images to model the input image. But different from texture image, finger vein patterns have obvious geometric structure. To further encode geometric information of finger vein pattern, we incorporate the idea of linear projection [14] into SR, and generate a series of line features.

In detail, in the new SR method, textons in dictionary are learned from blocks in training images in training process. For an input finger vein image, we first model it by representing each block in it over all learned textons. Thus, for the input image, a two-dimensional SR coefficients histogram is achieved. And then, in order to encode geometric information of finger vein pattern, the two-dimensional histogram is projected to different lines to generate more powerful line features. Last, the line feature will be used in matching. The proposed method will meet two goals in finger vein recognition: 1) the local finger vein pattern can be adequately represented by the dictionary of textons, and 2) the global spatial information of finger vein pattern can be encoded effectively by linear projection.

The rest of the paper is organized as follows. Section 2 reviews the concepts of SR. Section 3 describes the proposed SR method in detail. The experimental results and analysis are shown in section 4, and section 5 concludes this paper.

2 Sparse Representation based Classification

SR based classification takes root in visual perception mechanism, i.e., that visual neurons generate sparse representation for visual perception. SR can be used as a feature descriptor, and its aim is to seek sparse representation for testing signal over training signals.

Assume there are k classes, and each class i has n_i training samples, denoted by $A_i = [s_{i,1}, s_{i,2} \ldots, s_{i,n_i}] \in R^{m\times n_i}$, where $s_{i,j}, j = 1,2,\ldots,n_i$, is an m-dimensional vector stretched from the jth sample of the ith class. Denote by $A = [A_1, A_2, \ldots, A_k] \in R^{m\times n}$ the set of training samples from all classes, where $n = n_1 + n_2 + \cdots + n_k$, is the number of training sample. For a new sample y from the ith class, if it can be approximately coded as a linear combination of all training samples: i.e., $y = A$, where $\alpha = [\alpha_1, \ldots, \alpha_i, \ldots, \alpha_k] = [\alpha_{1,1}, \alpha_{1,2} \ldots, \alpha_{1,n_1} \ldots, \alpha_{k,n_k}]$, meanwhile most coefficients in α are nearly zero and only coefficients in α_i have significant values, we say y has a sparse representation over training set A. The sparse representation based algorithm [10] is summarized as follows.

1. Normalize the columns of A to have unit $l2$-norm.
2. Solve the l1-minimization problem to get representation coefficients of y:
   $\hat{\alpha}_1 = arg\ min_{\alpha,A} ||A\alpha - y||_2 + \lambda||\alpha||_1$.
3. Compute the residuals
   $r_i(y) = ||y - A\delta_i(\hat{\alpha}_i)||_2,\ i = 1,2,\ldots,k$.

where $\delta_i: R^n \to R^n$ is the characteristic function that selects the coefficients associated with the ith class.

4. Output $identity\,(y) = arg\ min_i r_i(y)$.

3 Modified Sparse Representation

We present the modified SR method in this section. It mainly includes three steps: local texton learning, feature descriptor and matching. The textons in dictionary is the fundamental of the proposed method, which is learned from local finger vein pattern. In feature descriptor, SR coefficients of all blocks in an image constitute a two-dimensional representation coefficient histogram, and the histogram is projected onto different lines to generate a series of line features, which is seen as image feature. In matching, in order to overcome the translation and/or rotation in finger vein images, multiple model histograms are built for each class.

3.1 Local Vein Texton Learning

Local vein texton learning is illustrated in Fig. 1, and we will present the learning process at great length in the following.

Before texton learning, all training images are partition into blocks, and some typical blocks are selected and initialized as dictionary texton. In detail, the image, randomly selected from database, is partitioned into nonoverlapping blocks with size of $p \times p$ pixels, and the block is stretched to a w-dimensional vector $(w = p \times p)$. Hence, using all selected training images, we can construct a training block dataset $B = [B_1, B_2, \dots, B_s]$, where $B_i, i = 1,2,\dots,s$, is all blocks in ith image. For an image, $B_i = [b_{11}, b_{12}, \dots, b_{CR}]$, where $b_{cr}, c = 1,2,\dots,C$ and $r = 1,2,\dots,R$, is one block in a training image.

Assume there are 100 images selected from database, and each image is partitioned into 100 blocks, it is totally 10, 000 blocks. If all blocks are used to learning texton, two limitations will be caused: 1) as the number of blocks is so large that learning process is very time-consuming, and 2) as the vein pattern in blocks are redundancy, the learned textons are not representative, which may not represent images well. So, K-means algorithm is used to determine l typical blocks. The mean vectors of all clusters are initialized as dictionary, denoted by $D = [d_1, d_2, \dots, d_l] \epsilon R^{w \times l}$.

After dictionary initialization, sparse representation is used to optimize dictionary D and SR coefficients denoted by $\Lambda = [\beta_1, \beta_2, \dots, \beta_l]$, and the optimization objective can be written as

$$arg\, min_{D,\Lambda}\, ||B - D\Lambda||_F^2 + \lambda ||\Lambda||_1, \tag{1}$$

This joint optimization problem is solved according to the method proposed in [11, 15]. The optimized dictionary D is based on the image block, not the whole image, and its texton is local vein pattern, not whole vein pattern in finger. Therefore, we call this learning process local texton learning.

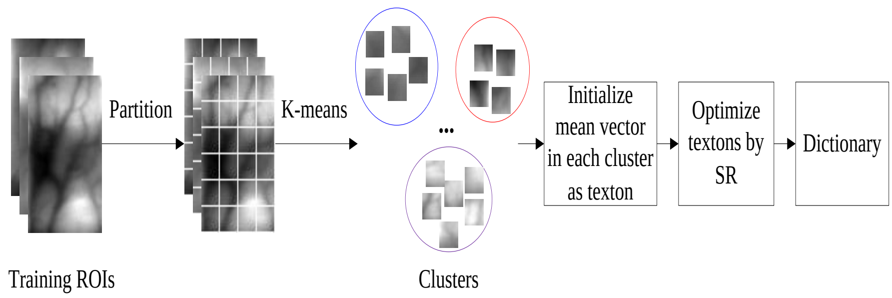

Fig. 1. Dictionary learning. K-means is used to gain typical image blocks and initialize as texton; and SR is used to optimize textons to get the dictionary

3.2 Feature Extraction

Dictionary learning has been introduced in the above section, and in this section we will describe how to use learned dictionary to represent finger vein image. In feature descriptor of finger vein image, we not only seek the effective representation of image block in locality, but also encode the spatial relationship of blocks from global perspective.

For a given finger vein image, we first partition it into nonoverlapping blocks with size of $p \times p$ pixels like training blocks, and then the learned dictionary is used to sparsely represent each block. The following problem is solved to get the SR coefficients:

$$arg\,min_{\beta_{ij}}\ ||b_{ij} - D\beta_{ij}||_2^2 + \lambda||\beta_{ij}||_1, \tag{2}$$

We use the method proposed [15] to solve this problem. Thus, for block b_{ij} at ith row and jth column of an image, we can obtain a coefficient vector β_{ij} saved in sub-histogram h_{ij}. And we arrange the sub-histogram h_{ij} according to the position of the block b_{ij} in image. So, for an image, a two-dimensional SR coefficient histogram can be achieved, shown in Fig. 2.

Although the locality of finger vein image is sought by sparsely representing each block over the learned dictionary, the global geometric information of finger vein pattern, i.e., the spatial relationship of blocks, has not been encoded. Hence, the idea of liner projection is used to encode geometric information of finger vein pattern from global perspective. For the achieved two-dimensional SR coefficient histogram, we project it onto lines with arbitrary angle to get a family of one-dimensional histogram, called line features, which is can depict the geometric information of global finger vein pattern. In finger, vein pattern is always parallel to finger, or grows at 45^0 or 135^0 angle with finger. Hence, projections with 0^0, 45^0, 90^0 and 135^0, are taken to maximize the geometric information of finger vein pattern, receiving four one-dimensional histograms H^0, H^{45}, H^{90} and H^{135}, shown in Fig. 2. Projection is a process of sum calculation for above achieved two-dimensional SR coefficient histogram along certain direction. We take H^{90} as an example to explain the projection process. For each column of above achieved two-dimensional histogram, we calculate the sum of the column as one element in H^{90}. All line features will be connected in parallel to be two-dimensional histogram H as the feature of an image.

3.3 Matching

As the feature of finger vein image is in two-dimensional histogram, Chi-square distance is used to measure the similarity between the input and enrolled images, as following:

$$d_{chi-square}\left(H_i, H_y\right) = \sum \frac{(H_i - H_y)^2}{H_i + H_y}, \tag{3}$$

where $H_i, i = 1,2, \dots, k,$ denotes the model histogram of the enrolled image. Similarly, H_y is the SR coefficient histogram of the input image.

The random of finger placement in image acquisition causes translation or/and rotation in finger vein image. In order to overcome this shortcoming, for each enrolled class we use multiple model histograms calculated from multiple images of this class, and the minimum value of matching scores between multiple model histograms and the input histogram is used.

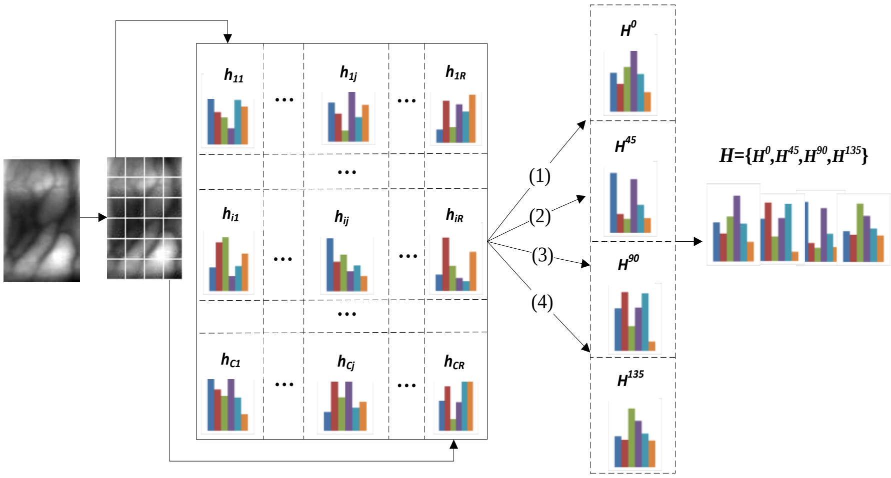

Fig. 2. Feature descriptor. For one image, two-dimensional SR coefficient histogram is first achieved, and then the histogram is projected to four directions, last, features in four direction are connected.

3.4 Discussions

In this section, we discuss the relationship between our proposed sparse representation method and one related work: the existing sparse representation based finger vein recognition [13].

The differences between the existing method and our method mainly focus on three aspects. First, in existing sparse representation based finger vein recognition, the dictionary is learned from whole images, while textons in our dictionary are learned from image blocks. In addition, each image has its class label, but for one image block, it does not have class label, which creates the second difference. In existing method, each class has one sub-dictionary, and the global dictionary consists of all sub-dictionaries. In other word, the existing method learn class-specific dictionary. However, the dictionary in our method has no class label, which is a common dictionary. Last, the third aspect is related to the way of identify. In detail, residual of sparse representation is used to identify the input finger vein image in existing method, but we make use of SR coefficient histogram to perform finger vein recognition.

4 Experimental Results

In this section, we evaluate the effectiveness of our proposed SR method in finger vein recognition.

4.1 Experimental Settings

We perform experiments on one public finger vein database from the Hong Kong Polytechnic University [5]. The HKPU database contains 3, 132 finger vein images of 156 subjects captured in two separate sessions. In each session, each subject provided six finger vein images from the left index finger to the left middle finger, respectively. As only 105 subjects turned up for image acquisition in second session, we use 2, 520 images (i.e., 105 subjects $\times$ 2 fingers $\times$6 images$\times$2 sessions) in our experiment. The training set used to learn dictionary includes one random selected image with 210 classes (In this paper, each finger is seen as one class.). For each class, 8 model histograms are built from 8 images, and remaining 4 images are used to test. The region of interest (ROI) of each image is segmented and normalized into 96$\times$64 pixels using method proposed in [16]. Some typical finger vein images and the corresponding ROIs are shown in Fig. 3.

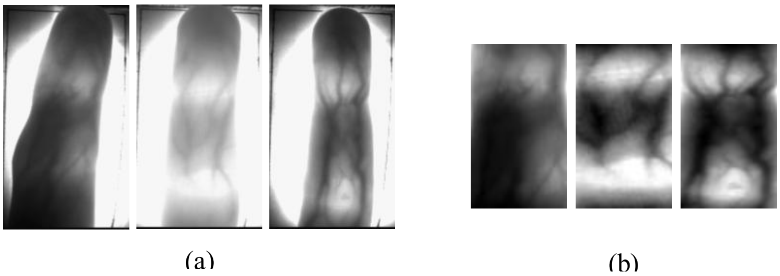

Fig. 3. (a) Some typical finger vein images. (b) The corresponding ROIs.

4.2 Comparison with other methods

In this subsection, we compare the proposed method with four existing methods. In the four methods, one is the existing sparse representation based finger vein recognition [13], and each of three others are selected from each of three feature extraction algorithm groups depicted in Introduction. We fix the parameters as presented in above subsection. Besides, there are some parameters in SR, and we adjust them to get the best performance: λ=0.1, the size of image block is set as 8$\times$8 pixels, and 90 textons are learned. We compare the recognition rates of all methods. The results are shown in Table 1. We can see from the results that the proposed method achieves a better recognition rate than other methods. The reason is that the proposed method seeks a sparse representation for modeling input image over the learned dictionary, and uses representation coefficients to perform classification, which is more flexible than nearest neighbor based classification. In addition, the proposed method encodes both the local vein pattern and the global spatial information, which cannot be done by other methods.

Table 1. Performance comparison of different methods.

Methods	Recognition rate (%)
Mean Curvature [4]	0.9298
LBP [6]	0.9393
Minutia [9]	0.7887
Existing SR method [13]	0.9418
Proposed method	0.9524

4.3 Investigation of parameter values

We study the performance of the proposed method with different parameter values in this subsection. There are three main varying parameters in the proposed method: the size of image block, the number of dictionary textons and the number of model histograms in each class.

First, we set the number of textons to 90, but vary the other parameters. In this experiment, 8×8 pixels, 12×12 pixels and 16×16 pixels are separately assigned to the size of block, and 4, 6, 8 model histograms are built for each class. The evaluation results are given in Table 2. The results show that when the size of block is 8×8 pixels, the proposed method achieves better recognition rate, and the larger size of block does not make the recognition rate increase. Besides, the more model histograms, the higher recognition rate.

Table 2. The recognition rate (%) of different blocks and models.

	8×8 pixels	12×12 pixels	16×16 pixels
4 models	0.8655	0.8452	0.8440
6 models	0.9286	0.9190	0.9131
8 models	0.9524	0.9488	0.9345

Then, we study the influence of the number of dictionary textons on the performance of the proposed method. Here we fix the size of block as 8×8 pixels, but vary the number of dictionary textons and the number of model histograms. The number of dictionary textons varies from 70 to 150, and the number of models in each class varies from 4 to 8. Fig. 4 illustrates the recognition rate of different textons and models. We can clearly see from this figure that when the number of textons is 90, the proposed method achieves best accuracies. After that, with the increase of number of textons, the recognition rate keeps unchanged on the whole. And, the more model histograms, the higher recognition rate, which is same with above experiment.

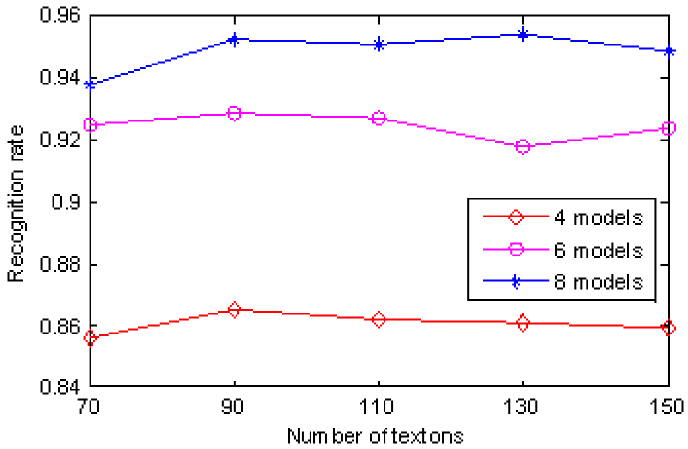

Fig. 4. The recognition rate (%) of different textons and models

5 Conclusions

In this paper, we proposed a modified sparse representation method for finger vein recognition. In the proposed method, textons in dictionary are learned from image blocks, and the image is modeled as distributions over the learned textons. The locality of finger vein pattern is sought by sparsely representing each image block over the learned textons, and the global geometric information of finger vein pattern is also encoded by the liner projections of four directions. Besides, to overcome the translation and/or rotation of finger vein image, we built multiple enrolled models for each class. The proposed method achieves better recognition rate than other methods. However, textons in dictionary are learned from finger vein image, and there are all kinds of noises in image, which are harmful for recognition performance. So, applying the proposed method on feature of finger vein image may be an interesting future issue.

Acknowledgments. This work is supported by National Natural Science Foundation of China under Grant No. 61173069, 61472226 and Shandong Natural Science Funds for Distinguished Young Scholar under Grant No. JQ201316. The authors would particularly like to thank the anonymous reviewers for their helpful suggestions.

References

1. Hashimoto, J.: Finger Vein Authentication Technology and Its Future. In Proceedings of the Symposium on VLSI Circuits Digest of Technical Papers, Honolulu, HI, USA, pages 15–17, 2006.

2. Miura, N., Nagasaka, A., Miyatake, T.: Feature extraction of finger vein patterns based on repeated line tracking and its application to personal identification. Machine Vision and Applications, 15 (4): 194–203, 2004.
3. Miura, N., Nagasaka, A., Miyatake, T.: Extraction of finger-vein patterns using maximum curvature points in image profiles. IEICE TRANSACTIONS on Information and Systems, E90-D (8): 1185–1194, 2007.
4. Song, W., Kim, T., Kim, H.C., Choi, J.H., Kong, H.J., Lee, S.R.: A finger-vein verification system using mean curvature. Pattern Recognition Letter, 32 (11) : 1541-1547,2011.
5. Kumar, A., Zhou, Y. B.: Human identification using finger images. IEEE Transactions on Image Processing, 21 (4): 2228–2244, 2012.
6. Lee, E. C., Jung, H., Kim, D.: New finger biometric method using near infrared imaging. Sensors, 11 (3): 2319–2333, 2011.
7. Rosdi, B.A., Shing, C.W., Suandi, S.A.: Finger vein recognition using local line binary pattern. Sensors, 11(12): 11357-11371, 2011.
8. Yang, G.P., Xi, X.M., Yin, Y.L.: Finger vein recognition based on a personalized best bit map, Sensors 12 (2012): 1738-1757.
9. Yu, C.B., Qin, H.F., Zhang, L., Cui, Y.Z.: Finger-vein image recognition combining modified hausdorff distance with minutiae feature matching. Interdisciplinary Sciences: Computational Life Sciences, 1(4): 280-289, 2009.
10. Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., Ma, Y.: Robust face recognition via sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(2): 210-227, 2009.
11. Yang, M., Zhang, L., Yang, J., Zhang, D.: Metaface learning for sparse representation based face recognition. 17th IEEE International Conference on Image Processing (ICIP), pages 1601-1604, 2010.
12. Xie, J., Zhang, L., You, J., Zhang, D.: Texture classification via patch-based sparse texton learning. 17th IEEE International Conference on Image Processing (ICIP), pages 2737-2740, 2010.
13. Xin, Y., Liu, Z., Zhang, H.X., Zhang, H.: Finger vein verification system based on sparse representation. Applied optics, 51(25): 6252-6258, 2012.
14. Cao, Y., Wang, C.H., Li, Z.W., Zhang, L.Q., Zhang, L.: Spatial-bag-of-features. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3352-3359, 2010.
15. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: A method for large-scale l1-regularized least squares. IEEE Journal on Selected Topics in Signal Processing, 1(4): 606–617, 2007.
16. Yang, L., Yang, G.P., Yin, Y.L., Xiao, R.Y.: Sliding window-based region of interest extraction for finger vein images. Sensors, 13 (3): 3799-3815, 2013.