INTEGRATION OF DISCRIMINATIVE FEATURES AND SIMILARITY-PRESERVING ENCODING FOR FINGER VEIN IMAGE RETRIEVAL

Kuikui Wang[1], *Lu Yang*[2], *Gongping Yang*[1*], *Yilong Yin*[1]

[1]School of Computer Science and Technology, Shandong University, Jinan, P.R. China
[2]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, P.R. China
[*]Corresponding author: gpyang@sdu.edu.cn

ABSTRACT

Although some image retrieval methods were proposed to accelerate finger vein recognition, the insufficient feature (*e*.g., the number of vein point) and unfavorable encoding (*e*.g., predefined threshold based binarization) limited retrieval performance largely. In view of this problem, we develop a new retrieval framework, based on the integration of discriminative texture features and similarity-preserving binary codes. In detail, the vector and scalar features, measuring the gray level, gray difference, and gray gathering of image patch, are both used to represent finger vein image. And to improve the retrieval efficiency, the high-dimensional decimal features are further encoded into the compact binary patterns by principal component analysis (PCA) and similarity-preserving iterative quantization (ITQ). Experimental results on one large finger vein database prove that the proposed method can powerfully improve the retrieval accuracy and efficiency.

Index Terms— Finger vein recognition, image retrieval, statistic texture feature, similarity-preserving encoding.

1. INTRODUCTION

The recognition accuracy and efficiency are two core problems in biometrics. Specially, in finger vein recognition, lots of efforts have been made for improving the accuracy, and some promising results have been reported [1, 2]. But the efficiency problem has not received enough attention, and very limited works tried to deal with it by image classification [3, 4, 5] or image retrieval [6, 7]. By the former kind of methods, one kth (k is the number of image classes) of all enrolled images were matched with the testing image on the average, but by the latter kind of methods, only top t candidates will be selected from the enrolled database and further matched. Therefore, image retrieval may be a better way to narrow down the matching space and therefore increase the recognition speed.

Only two image retrieval methods, *i*.e., locality sensitive hashing (LSH) based method [6] and vocabulary tree (VT) based method [7], were proposed in finger vein recognition. In the first method, the number of vein points in each image patch was binarized by a predefined threshold to obtain the retrieval code. And, based on the retrieval code, LSH was used to select the candidates for the testing image. In the second one, image patch was represented by LBP histogram, and the binary path of each patch in searching its nearest leaf node of a pre-built vocabulary tree was used as retrieval code. Different from the above hashing based candidate selection, the full matching between the testing image and all enrolled images was performed for searching candidates in this method.

Unfortunately, the image features in two current methods were straightforward or insufficient, which may not powerfully represent finger vein image. What's worse, the discrimination of above features was likely discounted in binary encoding. More specifically, it is obvious that the binarization is not a strong encoding method. And in search path based encoding, the node, nearest to the testing patch among all nodes in one layer of VT, is labeled with 1, but there is still difference between the node and the testing patch, therefore causing the quantization loss. Further, the loss is accumulated layer by layer. In summary, the insufficient feature and weak encoding make the retrieval performance unsatisfactory.

Hence, we attempt to improve the retrieval performance by exploring more discriminative feature and similarity-preserving encoding. Inherited from VT based method [7], the statistic texture feature is used. Concretely, in addition to LBP histogram [8, 9], three scalar features (*i*.e., mean LBP, mean gray value and image entropy) are employed to comprehensively represent finger vein image. Besides, considering that high-dimensional decimal features are unfavorable to quick retrieval, PCA [10] is used for compacting image features, and similarity-preserving ITQ [11, 12] is further employed to encode the compact features. Last, the binary codes are used in full matching for candidate selection. The experiments on a large finger vein database prove the advantage of our retrieval framework on both accuracy

The work is supported by National Science Foundation of China under Grant Nos. 61173069, 61472226 and 61573219.

and efficiency.

The rest of our paper is organized as follows. Section 2 describes the proposed finger vein retrieval framework in detail. And the experiments are shown in Section 3. Finally, the paper is concluded in Section 4.

2. FINGER VEIN RETRIEVAL FRAMEWORK

This section presents our retrieval framework, mainly including statistic texture feature extraction, feature compacting, and similarity-preserving encoding. In detail, a set of statistic texture features are firstly extracted from each image patch, and the features of all patches in one image are concatenated to represent this image. Next, the high-dimensional feature is compacted by PCA to mine its principal components. Lastly, the compact decimal feature is encoded by ITQ.

2.1. Statistic Texture Feature Extraction

One vector feature, *i*.e., LBP histogram, and three scalar features, *i*.e., mean LBP, mean gray value and image entropy, are given here. Supposing there are n non-overlapped patches in one image, denoted by $o_i, i = 1, 2, ..., n$, and the points in the patch o_i are denoted by $p_{i,j}, j = 1, 2, ..., m$. In this paper, we set n=96, and m=64. Here, take the feature extraction of the patch o_i as an example.

Vector Feature: LBP Histogram. In histogram construction, the decimal LBP of the current point $p_{i,j}$ is firstly calculated as followings [8, 9]:

$$LBP_{p_{i,j}} = \sum_{g=0}^{G-1} s(p_{i,g} - p_{i,j})2^g \tag{1}$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

Eight neighbors around the point $p_{i,j}$ are involved, *i*.e., G=8. And then all points in the patch o_i are mapped into the corresponding bins of LBP histogram according to its decimal LBP value. Formally, the histogram of image patch o_i can be defined as:

$$H_i = [h_i^0, h_i^1, ..., h_i^b, ..., h_i^{B-1}] \tag{3}$$

$$h_i^b = \sum_{j=1}^{m} I\{(b-1) * g + 1 \leq LBP_{p_{i,j}} \leq b * g\} \tag{4}$$

$$I\{x\} = \begin{cases} 1, if\ x = true \\ 0, otherwise \end{cases} \tag{5}$$

in which B is the number of histogram bins and g is the interval of decimal values in two neighbor bins. In this paper, we set B=16, g=256/16 (Note that the decimal value of the binary number with eight bits varies from 0 to 255).

LBP histogram reflects the gray difference of adjacent points. In finger vein image, the gray variation on vein growth direction is much smaller than it on the perpendicular direction to the vein growth. So, LBP histogram can explore the distribution of finger vein pattern to some extent.

Scalar Features: Mean LBP, Mean Gray Value and Image Entropy. First, mean LBP measures the mean gray difference of all points and their neighbors in one image patch, which can be calculated using the formula:

$$MLBP_i = \frac{\sum_{j=1}^{m} LBP_{p_{i,j}}}{m}, \tag{6}$$

The second feature is the mean gray value of all points in one image patch, and it can be computed as:

$$MG_i = \frac{\sum_{j=1}^{m} p_{i,j}}{m}, \tag{7}$$

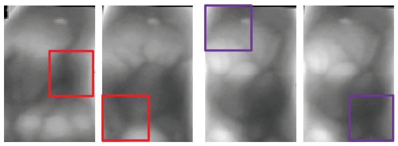
It reflects the gray level of an image patch. As vein point generally has lower gray value than non-vein point in finger vein image, the more vein points one image patch has, the lower value this feature has. Third, image entropy measures the gray gathering of one image patch. The appearance probability of each gray value in [0, 255] is computed for the patch and labeled by $q_v, v = 0, 1, .., 255$. And based on the probability the entropy is defined as:

$$IE_i = -\sum_{v=0}^{255} n_v q_v \log_2(q_v), \tag{8}$$

in which n_v is the number of points with gray value v. The formula can be seen as the weighted image entropy. For one image patch, which only contains vein points or non-vein points, its gray value gathers in a small range and this feature has low value. Oppositely, for the patch filled with both vein points and non-vein points, the gray value scatters in a relatively big range and this feature has high value.

The reasons for choosing these features are summarized. LBP, as finger vein feature descriptor, has achieved good recognition performance, which indicates that it is a good descriptor for finger vein image. And, considering LBP histogram only extracts local feature of an image patch, three global scalar features are employed to assist it. Moreover, the used scalar features are complementary. Such as, two image patches have same gray level, but gray variations between neighboring points may be different, shown in Fig.1(a). And, two patches, with similar image entropies, may have different gray variations, shown in Fig.1(b).

By concatenating the features of all patches in one image, it will be a 1,824-dimensional decimal feature, *i*.e., 96∗16-dimensional vector feature denoted by $V = \{H_i | i = 1, 2, ..., n\}$ and 96∗3-dimensional scalar feature denoted by $S = \{MLBP_i, MG_i, IE_i | i = 1, 2, ..., n\}$. So, to improve retrieval speed, it is essential to compact these features and further convert them into binary pattern.



(a) (b)

Fig. 1. (a) Two patches with similar mean gray values (*i*.e., 116 and 112) but different mean LBP values (*i*.e., 178 and 123); (b) Two patches with similar image entropies (*i*.e., 66 and 65) but different mean LBP values (*i*.e., 174 and 90).

2.2. Feature Compacting

Here, principal component analysis (PCA) [10] is employed to separately compact the vector feature and scalar feature. We take the compact of vector feature as an example. Assuming there are N training images, and their vector features are denoted by $V_i, i = 1, 2, ..., N$. The covariance matrix of these training features can be computed as follows:

$$C = \frac{1}{N} \sum_{i=1}^{N} (V_i - \overline{V})^T (V_i - \overline{V}) \tag{9}$$

The top k eigenvectors of the matrix C are used as PCA basis W. So, the vector feature of each image can be compacted by:

$$CV_i = W * V_i \tag{10}$$

Similarly, we can obtain the compacted scalar feature. In this paper, we set k=64 for both vector feature and scalar feature.

2.3. Similarity-preserved Encoding

In this section, the compacted vector feature and scalar feature are separately encoded into binary pattern by similarity-preserved iterative quantization (ITQ) [12]. This method can preserve the similarity between the decimal feature and the encoded binary pattern by rotating the feature. As we know, the rotation of a matrix can be achieved by the multiplication between the matrix and an orthogonal matrix.

We take the encoding of vector feature as an example. Assuming the orthogonal matrix and the encoded binary pattern are separately denoted by R and B. We can obtain the binary vector feature by minimizing the following quantization loss:

$$Q(B, R) = ||B - CV * R||_F^2 \tag{11}$$

where $CV = [CV_1, CV_2, ..., CV_N]$. The alternating minimization algorithm [12] is used to solve the problem. Similarly, we can get the binary scalar feature. And two kinds of binary features are combined in series as retrieval codes.

By compacting and encoding, the 1,824-dimensional decimal feature is transformed into the 128-dimensional binary code, which will speed up the image retrieval largely. And, the similarity between the decimal feature and the binary code is preserved in encoding, therefore minimizing the quantization loss.

Table 1. Retrieval accuracies (%) of different features.

t	Vector feature	Scalar feature	Fusion of them
1	68.36	56.91	82.75
5	81.42	74.98	90.04
10	85.82	81.46	92.17
20	89.71	87.16	94.39

3. EXPERIMENTS

3.1. Experimental Settings

To assess the proposed method, a fused finger vein database [7] is used, which was built by fusing four public databases, *i*.e., POLYU [1], SDUMLA [13], MMCBNU [14] and FVUSM [15]. There are totally 2,040 fingers, and each finger has 6 images. The region of interest (ROI) of each image was extracted and normalized into 96*64 pixels [1, 16, 17].

Hamming distance based full matching between the testing binary code and all enrolled binary codes is performed for candidate selection. As the classical image retrieval works in biometrics [18, 19], the retrieval accuracy and the penetration rate are used as benchmarks to evaluate all involved methods. The former benchmark is defined as the proportion of testing images, whose genuine enrolled images appear in their candidates. And the latter one measures the number of retrieved enrolled images relative to it of all enrolled images. All experiments are conducted in MATLAB on a PC with Intel i7-4790 3.60GHz CPU and 8G memory.

3.2. Experimental Results

Firstly, to test the complementarity of multiple features we used, the retrieval methods, with different features but same feature compacting, encoding and candidate selection, are compared. In this experiment, we use the first image of each finger as the enrolled sample and others as testing samples. The number of returned candidates for each testing sample, *i*.e., t, varies from 1 to 20, and the retrieval accuracies with all these values are given in Table 1. And the curves between the retrieval accuracy and the penetration rate are shown in Fig. 2.

The experimental results show that, no matter how many candidates are returned, the proposed vector feature-scalar feature fused retrieval method achieves the best accuracy. The main reason is that the local vector feature and global scalar features are complementarity and therefore their fusion increases the discrimination of retrieval feature. And it can be also seen that, the vector feature outperforms the scalar feature when the penetration rate is lower than 4%, especially

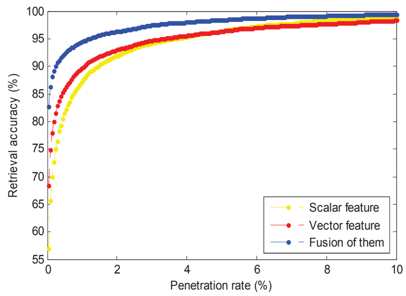


Fig. 2. Retrieval accuracy-penetration rate curves of different features.

Table 2. Retrieval accuracies (%) of different methods.

t	LSH based method [6]	VT based method [7]	Proposed method
1	68.04	84.07	86.34
5	77.08	88.99	92.04
10	79.82	90.98	93.86
20	82.58	92.83	95.39

when only top t candidates are returned. It may be chiefly attributed to the higher discrimination of the multi-dimensional local vector feature than three global scalar features.

Secondly, we make a comparison between the proposed retrieval method and two existing methods, *i*.e., LSH based method [6] and VT based method [7]. As first two images of each finger are used to build vocabulary tree in VT based method, the third image of each finger is used as the enrolled sample, and others are used as the testing samples. The retrieval accuracies with top t candidates are given in Table 2 for all involved methods, and the retrieval accuracy-penetration rate curves of all methods are shown in Fig. 3.

These results show that, our method outperforms two existing methods. There are two underlying reasons for the best retrieval accuracy. First, although three involved methods all employ the statistic feature, the fusion of multiple features in our method has higher discrimination than the single feature in other methods. In detail, the vector and scalar features measure the gray level, gray variation, and gray gathering of one image, which explore more discriminative information than the number of vein points [6] and LBP histogram [7]. Second, the encoding method we used suppresses the quantization loss, and keeps similarity between the decimal feature and the binary code. But, in thresholding based encoding [6], the numerical difference between the number of vein points and the threshold is ignored. And the quantization loss in searching path based encoding [7] is accumulated layer by layer, causing large loss eventually. In a word, the best performance of our method is attributed to the discriminative features and the similarity-preserving encoding both.

Lastly, the time costs of three retrieval methods are evaluated.

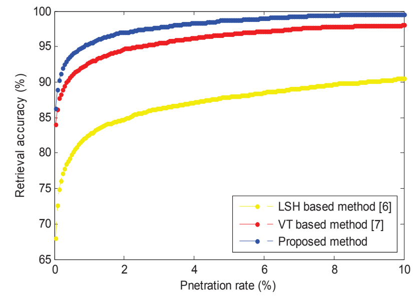


Fig. 3. Retrieval accuracy-penetration rate curves of different methods.

Table 3. Comparison of all methods about time cost (s).

Method	Feature extraction	Encoding	Candidate selection	Total time
LSH based method [6]	0.9616	3.77e-04	0.0089	0.9709
VT based method [7]	0.0158	0.0175	0.0233	0.0566
Proposed method	0.0212	0.0189	8.27e-04	0.0409

In this experiment, one testing image of each finger is used, and the mean time cost per image is reported. The time costs of all compared methods are given in Table 3. The results show that the proposed method spends less time than two current ones. It is mostly attributed to the quick feature extraction and the feature compacting both. In detail, compared with the proposed method, the iterative feature extraction in LSH based method spends larger time, owing to large iteration number for extracting plentiful vein pattern. And, the dimension of binary code in VT based method is 3,744 (*i*.e., 96 patches*39 bits per patch, as the tree with three layers and three branches is used), but it in the proposed method only is 128. So, higher dimensional feature causes more time cost in VT based method.

4. CONCLUSION

In this paper, we try to enhance finger vein retrieval performance by improving the discrimination of retrieval feature. In the proposed retrieval framework, multiple statistic texture features are fused, and the fused feature is encoded by PCA and similarity-preserving ITQ. These texture features can comprehensively represent one image by measuring its gray level, local gray variation, and global gray gathering. And the similarity between the decimal feature and binary code is preserved in ITQ based encoding, therefore minimizing the quantization loss. The experimental results prove the advantage of our method on retrieval accuracy and efficiency.

5. REFERENCES

[1] Ajay Kumar and Yingbo Zhou, "Human identification using finger images," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2228–2244, 2012.

[2] Xiaoming Xi, Lu Yang, and Yilong Yin, "Learning discriminative binary codes for finger vein recognition," *Pattern Recognition*, 2016.

[3] Dun Tan, Jinfeng Yang, Yihua Shi, and Chenghua Xu, "A hierarchal framework for finger-vein image classification," in *2013 2nd IAPR Asian Conference on Pattern Recognition*. IEEE, 2013, pp. 833–837.

[4] Jayachander Surbiryala, R Raghavendra, and Christoph Busch, "Finger vein indexing based on binary features," in *2015 Colour and Visual Computing Symposium (CVCS)*. IEEE, 2015, pp. 1–6.

[5] R Raghavendra, Jayachander Surbiryala, and Christoph Busch, "An efficient finger vein indexing scheme based on unsupervised clustering," in *2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*. IEEE, 2015, pp. 1–8.

[6] Darun Tang, Beining Huang, Rongfeng Li, and Wenxin Li, "A person retrieval solution using finger vein patterns," in *2010 20th International Conference on Pattern Recognition (ICPR)*. IEEE, 2010, pp. 1306–1309.

[7] Kuikui Wang, Lu Yang, Kun Su, Gongping Yang, and Yilong Yin, "Binary search path of vocabulary tree based finger vein image retrieval," in *2016 International Conference on Biometrics (ICB)*. IEEE, 2016, pp. 1–8.

[8] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[9] Eui Chul Lee, Hyunwoo Jung, and Daeyeoul Kim, "New finger biometric method using near infrared imaging," *Sensors*, vol. 11, no. 3, pp. 2319–2333, 2011.

[10] Jian-Da Wu and Chiung-Tsiung Liu, "Finger-vein pattern identification using principal component analysis and the neural network technique," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5423–5427, 2011.

[11] Yunchao Gong and Svetlana Lazebnik, "Iterative quantization: A procrustean approach to learning binary codes," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 817–824.

[12] Yunchao Gong, Svetlana Lazebnik, Albert Gordo, and Florent Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2916–2929, 2013.

[13] Yilong Yin, Lili Liu, and Xiwei Sun, "Sdumla-hmt: a multimodal biometric database," in *Chinese Conference on Biometric Recognition*. Springer, 2011, pp. 260–268.

[14] Yu Lu, Shan Juan Xie, Sook Yoon, Zhihui Wang, and Dong Sun Park, "An available database for the research of finger vein recognition," in *2013 6th International Congress on Image and Signal Processing (CISP)*. IEEE, 2013, vol. 1, pp. 410–415.

[15] Mohd Shahrimie Mohd Asaari, Shahrel A Suandi, and Bakhtiar Affendi Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3367–3382, 2014.

[16] Lu Yang, Gongping Yang, Yilong Yin, and Rongyang Xiao, "Sliding window-based region of interest extraction for finger vein images," *Sensors*, vol. 13, no. 3, pp. 3799–3815, 2013.

[17] Yu Lu, Shan Juan Xie, Sook Yoon, Jucheng Yang, and Dong Sun Park, "Robust finger vein roi localization based on flexible segmentation," *Sensors*, vol. 13, no. 11, pp. 14339–14366, 2013.

[18] Aglika Gyaourova and Arun Ross, "Index codes for multibiometric pattern retrieval," *IEEE transactions on information forensics and security*, vol. 7, no. 2, pp. 518–529, 2012.

[19] Yijing Su, Jianjiang Feng, and Jie Zhou, "Fingerprint indexing with pose constraint," *Pattern Recognition*, vol. 54, pp. 1–13, 2016.