FINGER VEIN VERIFICATION WITH VEIN TEXTONS

LUMEI DONG, GONGPING YANG, YILONG YIN, XIAOMING XI, LU YANG and FEI LIU

School of Computer Science and Technology, Shandong University, Jinan, 250101, P.R. China

Finger vein pattern has become one of the most promising biometric identifiers. In this paper, a robust method based on Bag of Words (BoW) is developed for finger vein verification. Firstly, some robust and discriminative visual words are learned from local base features such as Local Binary Pattern (LBP), Mean Curvature and Webber Local Descriptor (WLD). We name these visual words as Finger Vein Textons (FVTs). Secondly, each image is mapped into a FVTs matrix. Finally, spatial pyramid matching (SPM) method is applied to maintain spatial layout information by representing each image as pyramid histogram which is performed for matching by histogram intersection function. Experimental results show that the proposed method achieves satisfactory performance both on our database and the open PolyU database. In addition, our method also has strong robustness and high accuracy on the self-built rotation and illumination databases.

Keywords: Finger Vein Verification; Bag of Words; Finger Vein Textons; Spatial Pyramid Matching.

1. Introduction

Biometrics or biometric authentication, refers to using distinctive physiological (e.g., fingerprints, faces, iris, and palm geometries) and behavioral (e.g., gaits, voices, typing rhythms) methods to automatically recognize an individual\(^1\)\(^2\). Now it is widely used in everyday life and raises a heated research because of its better security, higher efficiency. Recently, finger vein verification, as one of the most promising biometric techniques, has received considerable attention from researchers due to its advantages over other biometric techniques\(^3\)\(^4\): (1) non-contact; (2) live-body identification; (3) high security; (4) small device size. Now, it is playing a more and more important role in many mission-critical applications such as access control, personal identification, E-passport, etc.

Generally speaking, a typical finger vein verification process includes the following four steps, namely, image capturing, pre-processing, feature extraction and matching. Among these steps, feature extraction is very important. Many researchers are dedicated to studying effective feature extraction methods in finger vein verification. According to whether finger vein network is segmented or not,
these feature extraction methods can be divided into two categories. The first kind of methods are based on the segmented finger vein network. For example, Miura et al.\textsuperscript{5} proposed a repeated line tracking method to extract finger vein features. Song et al.\textsuperscript{6} used the mean curvature (MC) method, which views the vein image as a geometric shape and then finds the valley-like structures with negative mean curvatures. In 2005, a method of using local maximum curvatures across the vein images was developed which significantly improved the robustness of finger vein \textsuperscript{7}. Kumar et al.\textsuperscript{8} investigated a finger vein feature extraction approach based on Gabor filters successfully. These methods can obtain good performance under the assumption that the finger vein networks are segmented properly. However, if the quality of captured images is low, the network may not be segmented properly. Then, the extracted features based on improper network will make the performance of verification degrade dramatically.

To alleviate the above problem, feature extraction without segmentation has been proposed. Some local pattern-based methods such as Local Binary Pattern (LBP)\textsuperscript{9}, Local Line Binary Pattern (LLBP)\textsuperscript{11}, Generalized Local Line Binary Pattern (GLLBP)\textsuperscript{12}, Histogram of Salient Edge Orientation Map (HSEOM)\textsuperscript{13} and personalized best bit map (PBBM)\textsuperscript{14} are applied to extract different effective features respectively. Among these methods, LBP can reflect the texture of the finger vein very well. As a variant of LBP, LLBP demonstrates a better accuracy than LBP. However, LLBP is limited in extracting horizontal and vertical line patterns. As an orientation-selectable LLBP method, GLLBP extends LLBP for line pattern extraction into any orientation and effectually improves the matching accuracy. HSEOM is an orientation-based local descriptor which can efficiently utilize edge and orientation features in finger vein images. PBBM is rooted in LBP and can select the personalized discriminative bits for each subject. Though these methods have attained promising experimental results, extracted feature is based on a pixel which is sensitive to the translation and rotation of finger in image. In addition to local pattern-based methods, some methods like principal component analysis (PCA)\textsuperscript{15}, linear discriminant analysis (LDA)\textsuperscript{18} and two-dimensional principal component analysis ((2D)\textsuperscript{2}PCA)\textsuperscript{17} were used to extract global finger vein feature. Nevertheless, global feature is short of local detailed finger vein information, which is also very important for verification.

Recently, the Bag-of-Words (BoW) method which describes an image as a set of orderless visual words has been successfully applied in texture analysis and visual classification\textsuperscript{19,20,21}. The method can capture a significant proportion of the complex statistics of real images and visual classes in a convenient local form, and it also has good resistance to occlusions, geometric deformations and illumination variations\textsuperscript{22}. Due to these advantages of BoW, several researchers have applied it in biometric recognition. For example, the authors\textsuperscript{26} improved the performance for coarse iris classification using a visual-dictionary-learning algorithm to represent the global characteristics of iris images. Li et al.\textsuperscript{23} proposed a face and express recognition method based on BoW considering holistic and local image features, which
gave excellent face recognition performance under various conditions. A real-time hand gesture detection and recognition system using Bag-of-Features and support vector machine techniques achieved satisfactory performance under variable scale and illumination conditions\textsuperscript{24}. Paul Scovanner et al.\textsuperscript{25} applied a BoW approach to represent videos in action recognition. Inspired by these literatures, we think that the BoW method is an intuitively good candidate to extract robust finger vein feature. During finger vein research work, we observed that the finger vein images exhibit important regularity. Some local finger vein patterns are similar and appear repeatedly in the image. These similar and repetitive local finger vein features can be represented by a visual word, which is called Finger Vein Texton (FVT). Fig. 1 shows some FVTs intuitively.

Based on the above analysis, we think the BoW model which can learn some typical FVTs is suitable for finger vein feature extraction. To better describe the feature of finger vein, the spatial pyramid representation of FVTs which reveals the global spatial layout and the local detail of the finger vein is used to make the feature more robust. In our previous work\textsuperscript{10}, we have proposed a finger vein recognition method based on the spatial pyramid representation of local features such as gray, texture and shape to maintain spatial layout information of finger vein. The spatial pyramid representation is based on the improved BoW method (i.e., spatial pyramid matching (SPM))\textsuperscript{27}. Motivated by the SPM method, a new finger vein verification method which can extract robust finger vein feature with global layout and local details information is proposed. We firstly extract local base feature of finger vein image like LBP\textsuperscript{3}, MC\textsuperscript{6}. Secondly, we construct codebook from some collected finger vein patches which are represented by local base features. The codebook consists of our learned FVTs. Thirdly, the finger vein image is mapped into a FVTs matrix by mapping each patch to the closest FVT. Fourthly, spatial pyramid is constructed by
partitioning the FVTs matrix into a sequence of increasingly fine sub-regions, and the frequency of FVT in each sub-region is calculated to maintain spatial layout information. Lastly, histogram intersection function is used to measure similarities of extracted features. Experimental results show that our method can significantly improve verification performance. In addition, related results also show that the proposed method is robust to rotation and illumination variations of finger vein images, which is very crucial to a practical finger vein verification system.

The rest of this paper is organized as follows. Section 2 presents the technical background briefly. Section 3 describes the details of our proposed method. Experiments and results are presented and discussed in Section 4. Finally, Section 5 concludes this paper.

2. Technical Background

2.1. Bag of Word

The BoW algorithm was initially developed for text information retrieval (IR) and text categorization (TC)\textsuperscript{28}. Recently, it has been successfully applied in texture classification and visual recognition\textsuperscript{19,20,21}. The basic idea of BoW approach is to represent an image using an unordered collection of visual words. In a typical BoW method, some local orderless patches are firstly extracted from a number of images and these local patches are represented by feature descriptors. Secondly, we cluster some collected patches to produce a codebook. Each clusters center is used as a visual word. And the number of the clusters is the codebook size. Thirdly, each patch in an image is assigned to the nearest visual word using an appropriate distance, and finally images can be represented by a visual word histogram reflecting the distribution of all the visual words.

2.2. Spatial Pyramid Matching

Recently, traditional BoW methods have demonstrated impressive performance, but they disregard all the spatial layout information. To overcome the limitation, the spatial pyramid matching (SPM) approach is proposed in\textsuperscript{27}. It works by dividing the image into a sequence of increasingly fine sub-regions and computing histograms of local features found inside each sub-region. In SPM, all feature vectors are quantized into $M$ discrete types. It is assumed that features of the same type can be perceived equivalent to one another. Let us construct $L$-levels spatial pyramid. In level 0, there is only one sub-region, namely, the whole image. In level 1, the image is partitioned into 4 approximate same sub-regions. In level $l$, the image is divided into $(2^l)^2$ approximate same sub-regions. For two different images $X$ and $Y$, $X_m$ and $Y_m$ represent the coordinates of features of type $m$ found in the respective images. Spatial pyramid matching kernel $K$ is defined as

$$K^L(X,Y) = \sum_{m=1}^{M} k^L(X_m,Y_m).$$  \hfill (1)
\[ k^L(X_m, Y_m) = \frac{1}{2^L} I_m^0 + \sum_{i=1}^{L} \frac{1}{2^{L-i+1}} I_m^i. \]  

(2)

\[ I_m^l = I(H_{X}^{l,m}, H_{Y}^{l,m}) = \sum_{i=1}^{D_l} \min(H_{X}^{l,m}(i), H_{Y}^{l,m}(i)). \]  

(3)

Here, \( I_m^l \) represents the number of matches of features of type \( m \) at level \( l \). It is measured by the histogram intersection function. \( D_l \) is the total number of sub-regions at level \( l \). \( H_{X}^{l,m}(i) \) denotes the number of features of type \( m \) appearing in the \( i \)-th sub-region at level \( l \) in image \( X \). In Fig. 2, we show an example of constructing a three-level pyramid intuitively. It can be seen that SPM will reduce to a standard bag of words when \( L \) is 0. When \( L \) is 2, there will be \( 21 \times M \)-dimensional histogram intersections more than \( M \)-dimensional histogram intersections of a standard BoW. However, it is efficient because the histogram vectors are extremely sparse and the computational complexity of the kernel is linear in the number of features.

3. The Proposed Method

In this section, we describe the proposed method for finger vein verification. The vein verification procedure mainly includes preprocessing, feature extraction based on FVTs and matching. Detailed descriptions of these steps are as follows. The
framework of our proposed finger vein verification method is demonstrated in Fig. 3.

Fig. 3. Framework of the proposed method.

3.1. Preprocessing

For obtaining efficient features, it is necessary to processing the acquired finger vein images. The preprocessing operation mainly includes gray processing, Region of Interest (ROI) extraction, size normalization and gray normalization. Image gray processing is used for transforming the original 24-bit color image into an 8-bit gray image to reduce the computational complexity. The ROI of our database can be obtained according to the maximum and minimum abscissa values of the finger contour which is extracted using the edge-detection method with a Sobel operator\textsuperscript{14}, while the ROI of PolyU database can be obtained by the following three primary steps including: (1) detect the skew angle of the image and correct the image; (2) determine the height of the ROI based on the phalangeal joints of the finger, (3) acquire the width of ROI based on internal tangents of fingers edges\textsuperscript{29}. Then, to reduce the influence of personalized factors such as different finger size and changing location, the ROI region is normalized to the same size by using the bilinear interpolation. The size of the normalized ROI is set to be 240 × 160. After size normalization, gray normalization is used to obtain a uniform gray distribution.

3.2. Feature Extraction

In this paper, feature extraction process includes three steps. We firstly extract local base features from the partitioned dense and regular patches. Secondly, a small codebook of visual words is learned, which are called Finger Vein Textons (FVTs). Thirdly, the pyramid histogram of FVTs is compiled to represent the global and local characteristics of finger vein images.
3.2.1. Local base feature extraction

This section briefly introduces three kinds of typical feature descriptors: the MC, the LBP and the webber local descriptor (WLD).

As a typical finger vein feature extraction method based on the segmented finger vein network, MC has achieved good performance in finger vein verification. MC method considers the vein image as a geometric shape and finds the valley-like structures with negative Mean Curvatures. In this paper, the computation of MC descriptor is on a regular dense patch. For each patch, we construct a feature vector by arranging the Mean Curvature values of points in this patch by row.

As a typical finger vein feature extraction method without segmentation, LBP has attained much attention in finger vein recognition. An LBP can be described as an ordered set of binary values determined by comparing the gray values of a center pixel and its 3x3-neighborhood pixels. All binary codes can be concatenated together and converted to a decimal. In this paper, we compute the histogram of LBP in every regular dense patch. Each patch has a 256-dimensional LBP histogram vector which contains the distribution of vein texture information on a patch level.

The WLD descriptor is proposed by Chen et al. which has performed well in texture classification and face detection. It is a simple, powerful and robust local descriptor inspired by Webbers Law. Since WLD can give expression to texture and orientation information, we try to apply it to finger vein verification. In this paper, the histogram of WLD is calculated in every regular dense patch.

3.2.2. Codebook of FVTs Generation

A schematic diagram illustrating the steps of learning the codebook of FVTs is shown in Fig. 4. There are many different clustering techniques for the generation of codebook. In this paper, codebook is calculated using the k-means clustering algorithm. Firstly, the finger vein image is partitioned into dense regular patches
and local base features are extracted on a patch level. Then, we collect a set of local base features from random finger vein images and perform k-means clustering algorithm to obtain the most informative centers, namely, a codebook of FVTs. To improve the generalization capabilities, we choose 100 finger vein images from the database randomly in our experiment. As the center of a learned cluster, a FVT can reflect certain characteristic of finger vein and represent all finger vein features in this cluster.

3.2.3. Pyramid Histogram of FVTs Compilation

The procedure of compiling pyramid histogram of FVTs is shown in Fig. 5. After the construction of codebook, the extracted base feature of a patch in a finger vein image is mapped into the closest FVT according to the similarity between the FVT and the feature. Euclidean distance is used as the similarity measure. So, a finger vein image is represented by a FVTs matrix. Codebook mapping is beneficial and robust to finger vein recognition because it can make an intractable number of distinct possible features categorized into a manageable number of FVTs, reducing the influence of extreme features. After that, we firstly construct spatial pyramid by repeatedly subdividing a FVTs matrix and compute histograms of image features over the resulting sub-regions. And then, we represent finger vein images as concatenated histograms of all sub-regions. Pyramid histogram combines the spatial layout and local details of FVTs. Therefore, it can represent the finger vein more properly. In addition, since spatial pyramid has multiple resolutions in a principled fashion, it is robust to failure at individual levels. To make a trade-off between discriminability and computational complexity, we use three level pyramid histogram representation (i.e., level 0, level 1, level 2) in our method. Pyramid histogram can represent the statistical information of global FVTs in the level 0 and describe local detailed information at the level 1 and 2.

![Fig. 5. The procedure of compiling pyramid histogram of FVTs.](image-url)
3.3. Matching

Now every finger vein image is represented by the pyramid histogram of FVTs. The final step is matching two finger vein images. We use the histogram intersection (Eq. (3)) to measure similarities between the input finger and the registered finger in the database. The value of histogram intersection is compared with a threshold to identify whether the user is accepted or not.

4. Database and Experiment

In this section, we report results on four databases: the first one finger vein database is from our MLA Lab, named MLA database; the second one is constructed by Hong Kong Polytechnic University, named PolyU database; the last two one are the self-built rotation and illumination databases from the MLA database, named rotation database and illumination database respectively.

4.1. The Experimental Databases

The MLA database was collected using the capturing device manufactured by the Joint Lab for Intelligent Computing and Intelligent System of Wuhan University, China. The database was collected from 34 individuals (20 males and 14 females). Each individual participated in two sessions, separated by 20 days. The age of the participants was between 19 and 48 years old, and their occupations included university students, professors, and workers at our school. Each individual provides 4 fingers, namely, left index, left middle, right index, right middle, each of which contributes 30 images. In total, the database contains 4080 (34 subjects × 4 fingers × 30 samples) finger vein images from 136 different fingers. The original spatial resolution of the data is 320 × 240. The size of the region used for feature extraction is reduced to 240 × 160, after ROI extraction and size normalization.

The PolyU finger vein database consists of 6264 images collected from 156 volunteers over a period of 11 months (April 2009–March 2010). The finger images were acquired in two separate sessions. In each session, each of the subjects provided six image samples from the index finger to the middle finger, respectively. Each sample consists of one finger vein image and one finger texture image. However, we only use the finger vein image. In addition, as only 105 subjects turned up for the imaging during the second session, we have used the images acquired in the first session in our experiment. Consequently, there are totally 1872 (156 subjects × 2 fingers × 6 samples) finger vein images from 312 different fingers. After ROI extraction and size normalization, the size of the region used for feature extraction is reduced to 240 × 160.

The rotation database is formed by rotating each preprocessed finger vein image in the MLA database at a random degree which is in the scope of [-10, +10]. Consequently, the rotation database contains 4080 (136 classes × 30 samples) finger
vein images. Some typical finger vein images from the self-built rotation database are shown in Fig. 6.

Since finger vein image is captured by near infrared methods, the quality of acquired vein image is highly sensitive to illumination. To verify the robustness of our method, we choose 70 fingers influenced by illumination from the preprocessed images of the MLA database to construct a illumination database with 2100 (70classes × 30samples) finger vein images. Some typical finger vein images from the self-built illumination database are shown in Fig. 7.

Fig. 6. Sample finger vein images from the self-built rotation database.

Fig. 7. Sample finger vein images from the self-built illumination database.

4.2. The Experiment Settings

All the experiments are implemented in MATLAB, and conducted on a PC with 2.9G CPU and 4.0G memory. In this paper, five experiments are designed to evaluate the proposed method: (a) Experiment 1 evaluates the performance of the proposed method by comparing with corresponding base feature (i.e., LBP, MC and WLD). (b) Experiment 2 is designed to verify the robustness of our method on rotation and illumination databases constructed from the MLA database respectively. (c) Experiment 3 discusses the influence of codebook size and patch size. (d) Experiment 4 is conducted to compare the performances of the proposed method with our previous work. (e) In Experiment 5, we focus on measuring the average processing time of our proposed method.
4.3. **Experiment 1**

In this section, we compare the proposed method with base features on our finger vein database and PolyU database respectively. We choose three typical base features (i.e., MC, LBP and WLD) introduced in section 3.2.1 as base features.

For the MLA database, we use the first 10 samples of each class in the database to generate the class center, which is calculated by averaging the corresponding feature values of the training samples. And the remaining 20 samples of each class are used as test samples. Consequently, there are 2720 (136 × 20) intraclass matchings and 367200 (136 × 20 × 135) interclass matchings in total. Similarly, for the PolyU database, we use the first 3 samples of each class to generate the class center and the remaining 3 samples as test samples. The class center is also calculated by averaging the corresponding feature values of the training samples. Therefore, there are 936 (312 × 3) intraclass matchings and 291096 (312 × 3 × 311) interclass matchings in this database. In this paper, the performance of a system is evaluated by Equal Error Rate (EER), False Rejection Rate (FRR) at zero False Acceptance Rate (FAR), FAR at zero FRR, FRR at 0.001 FAR, FAR at 0.001 FRR. The EER is defined as the error rate when the FRR is equivalent to the FAR. It is suitable to measure the overall performance of biometrics systems because the FRR and FAR are treated equally. In this experiment, for MLA database, we set the sizes of patch for FVT-MC, FVT-LBP and FVT-WLD are 4 × 4 pixels, 8 × 8 pixels and 9 × 9 pixels respectively, while the sizes of codebook are 55, 205 and 105 respectively. For PolyU database, we set the sizes of patch for FVT-MC, FVT-LBP and FVT-WLD are 3 × 3 pixels, 8 × 8 pixels and 9 × 9 pixels respectively, while the sizes of codebook are 55, 205 and 105 respectively. These are the optimal parameters for different base feature on different database which are discussed in section 4.5.

The Receiver Operating Characteristic (ROC) curves of the MLA database and the PolyU database are shown in Fig. 8. The EER, FRR at zero FAR, FAR at 0.001 FRR and FRR at 0.001 FRR values of both databases are listed in Table 1, Table 2 respectively. The results from Fig. 8, Table 1 and Table 2 show that our method achieves much lower EER than the base-feature-based method on both databases. For example, the EERs of FVT-LBP method on the MLA database and the PolyU database are 0.074 and 0.0238 respectively, outperforming the LBP-based method. This is because our method uses the most informative features (FVTs) learned from base features for verification, which reduces the influence of extreme features at certain degree. Besides, the method also maintains the spatial layout information which is very crucial to finger vein verification. These results also indicate that the proposed method is robust and general to different base features and databases.

4.4. **Experiment 2**

In order to verify the robustness of our method, we evaluate the performance of different methods on the self-built rotation and illumination databases respectively.
Fig. 8. ROC curves by different methods. (a) MLA database; (b) PolyU database.

Table 1. The performances by different methods on the MLA database.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FAR at zero FRR</th>
<th>FRR at zero FAR</th>
<th>FAR at 0.001 FRR</th>
<th>FRR at 0.001 FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.1617</td>
<td>0.9972</td>
<td>0.8077</td>
<td>0.9931</td>
<td>0.6143</td>
</tr>
<tr>
<td>LBP</td>
<td>0.0442</td>
<td>0.9592</td>
<td>0.4360</td>
<td>0.5689</td>
<td>0.2129</td>
</tr>
<tr>
<td>WLD</td>
<td>0.0574</td>
<td>0.7632</td>
<td>0.7629</td>
<td>0.5801</td>
<td>0.3566</td>
</tr>
<tr>
<td>FVT-MC</td>
<td>0.0464</td>
<td>0.9955</td>
<td>0.9353</td>
<td>0.5910</td>
<td>0.3706</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>0.0074</td>
<td>0.8305</td>
<td>0.7728</td>
<td>0.0868</td>
<td>0.0676</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>0.0077</td>
<td>0.7812</td>
<td>0.7588</td>
<td>0.1672</td>
<td>0.0375</td>
</tr>
</tbody>
</table>

Table 2. The performances by different methods on the PolyU database.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FAR at zero FRR</th>
<th>FRR at zero FAR</th>
<th>FAR at 0.001 FRR</th>
<th>FRR at 0.001 FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.2484</td>
<td>0.9842</td>
<td>0.9626</td>
<td>0.9839</td>
<td>0.8600</td>
</tr>
<tr>
<td>LBP</td>
<td>0.0846</td>
<td>0.7290</td>
<td>0.7831</td>
<td>0.7285</td>
<td>0.4626</td>
</tr>
<tr>
<td>WLD</td>
<td>0.1358</td>
<td>0.7002</td>
<td>0.9679</td>
<td>0.6996</td>
<td>0.6827</td>
</tr>
<tr>
<td>FVT-MC</td>
<td>0.1037</td>
<td>0.7764</td>
<td>0.9776</td>
<td>0.7756</td>
<td>0.6250</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>0.0238</td>
<td>0.4367</td>
<td>0.6720</td>
<td>0.4551</td>
<td>0.1282</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>0.0310</td>
<td>0.3945</td>
<td>0.9594</td>
<td>0.3936</td>
<td>0.1645</td>
</tr>
</tbody>
</table>

The experimental settings are the same as Experiment 1. The sizes of patch and codebook for different base feature on the two self-built databases are the same as the parameters on MLA database in Experiment 1. The EER, FRR at zero FAR, FAR at zero FRR, FRR at 0.001 FAR and FAR at 0.001 FRR values of different methods on both databases are listed in Table 3 and Table 4 respectively.

From Table 3 and Table 4, we can see that the proposed method achieves lower EER compared with base-feature-based method. It is also worth noting that, as compared to Table 1, the performance of our method decreases much less than that
of base-feature-based method. For example, the EER of FVT-LBP method only increases 0.0033 on the rotation database when comparing to the MLA database, while the EER of LBP increases 0.0248. This indicates that the features extracted by our method have stronger robustness to the images with rotation and illumination. The main reason is that codebook mapping has good resistance to geometric deformations and illumination variations. Codebook mapping could map two slightly different patches which are affected by rotation and illumination into the same FVT. For example, for two finger vein image patches which have a little difference from the same finger, traditional base features like LBP about the two patches may be different but our proposed method will map them into the same FVT. Fig. 9 intuitively shows that FVT's mapping is robust to rotated finger vein image.

Table 3. The performance by different methods on the rotation database.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FAR at zero FRR</th>
<th>FRR at zero FAR</th>
<th>FAR at 0.001 FRR</th>
<th>FRR at 0.001 FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.2635</td>
<td>0.9994</td>
<td>0.9570</td>
<td>0.9937</td>
<td>0.8735</td>
</tr>
<tr>
<td>LBP</td>
<td>0.0690</td>
<td>0.9199</td>
<td>0.8460</td>
<td>0.7521</td>
<td>0.3960</td>
</tr>
<tr>
<td>WLD</td>
<td>0.0881</td>
<td>0.8467</td>
<td>0.9728</td>
<td>0.6141</td>
<td>0.5691</td>
</tr>
<tr>
<td>FVT-MC</td>
<td>0.0781</td>
<td>0.9434</td>
<td>0.9625</td>
<td>0.8332</td>
<td>0.6055</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>0.0107</td>
<td>0.5919</td>
<td>0.6338</td>
<td>0.2316</td>
<td>0.0599</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>0.0129</td>
<td>0.2643</td>
<td>0.5680</td>
<td>0.1660</td>
<td>0.0482</td>
</tr>
</tbody>
</table>

Table 4. The performance by different methods on the illumination database.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FAR at zero FRR</th>
<th>FRR at zero FAR</th>
<th>FAR at 0.001 FRR</th>
<th>FRR at 0.001 FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.1636</td>
<td>0.9940</td>
<td>0.7350</td>
<td>0.9934</td>
<td>0.5029</td>
</tr>
<tr>
<td>LBP</td>
<td>0.0512</td>
<td>0.5784</td>
<td>0.4143</td>
<td>0.5322</td>
<td>0.2293</td>
</tr>
<tr>
<td>WLD</td>
<td>0.0648</td>
<td>0.5976</td>
<td>0.6957</td>
<td>0.4574</td>
<td>0.3571</td>
</tr>
<tr>
<td>FVT-MC</td>
<td>0.0529</td>
<td>0.6044</td>
<td>0.8314</td>
<td>0.5064</td>
<td>0.3650</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>0.0084</td>
<td>0.1156</td>
<td>0.0821</td>
<td>0.0668</td>
<td>0.0193</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>0.0101</td>
<td>0.0996</td>
<td>0.0579</td>
<td>0.0752</td>
<td>0.0221</td>
</tr>
</tbody>
</table>

4.5. **Experiment 3**

In this section, we use the MLA database and the PolyU database to evaluate the influence of codebook size and patch size on our method. In these experiments, we just fix one parameter and vary the other.

Firstly, we test different sizes of patches when extracting local base features. For MLA database and PolyU database, we fix the sizes of codebook for FVT-MC, FVT-LBP and FVT-WLD are 55, 205 and 105 respectively. The left of Fig. 10 shows the performance of different patch sizes on the two databases. We can see that the best results of different features are obtained in different ranges of patch
size. However, the performance is relatively stable within that range. The trends of the same feature on the two databases are similar with the increase of the patch. The patch size around $3 \times 3$ pixels, $7 \times 7$ pixels and $8 \times 8$ pixels have satisfactory results for FVT-MC, FVT-LBP and FVT-WLD respectively. We also find that the performance is not good enough when the patch size is too small. A small patch contains poorer finger vein information so that generated codebook of FVTs cannot be very discriminative. Besides, the cost of extracting local base features will increase when using a small patch. Conversely, if the patch size is too large, the patch may contain richer finger vein information. But, the number of patches will become small and the frequencies that FVTs appear in a vein image will be low. The feature with low frequencies will be less discriminative. Therefore, the discriminability of large patch size is also unsatisfactory.

Secondly, we test different sizes of codebook about the three features with the corresponding optimal patch size. For MLA database, we fix the sizes of patch for FVT-MC, FVT-LBP and FVT-WLD are $4 \times 4$ pixels, $8 \times 8$ pixels and $9 \times 9$ pixels respectively. For PolyU database, we fix the sizes of patch for FVT-MC, FVT-LBP and FVT-WLD are $3 \times 3$ pixels, $8 \times 8$ pixels and $9 \times 9$ pixels respectively. Related results are shown in the right of Fig. 10. We discover that the best performances of different base features are obtained in different ranges of codebook sizes, possibly related to the discriminability of the feature. For example, the performances of FVT-MC are satisfactory around the codebook size of 55. However, for codebook size between 105 and 305, FVT-LBP reports similar satisfactory results. But FVT-WLD has good and stable performance when the codebook size between 55 and 255. We also find that the performance is unsatisfactory when the codebook is too small.
Fig. 10. The effects of using different patch size and codebook size. (a1) (a2) (a3) are the effects of patch size on MLA database and the PolyU database corresponding to FVT-MC, FVT-LBP and FVT-WLD respectively; (b1) (b2) (b3) are the effects of codebook size on MLA database and the PolyU database corresponding to FVT-MC, FVT-LBP and FVT-WLD respectively.

small. This can be explained that the FVT is not very discriminative and some dissimilar finger vein features may be mapped into the same FVT. From the trend
in the right of Fig. 10, we can see that the performance tends to drop when the codebook size becomes too large. Because the FVTs become more discriminative to distinguish irrelevant variations (i.e., noise) and map similar finger vein features to different FVTs. Certainly, the cost of codebook generation, codebook mapping and the matching between two histograms will increase when using a large codebook.

4.6. Experiment 4

In this experiment, we use the PolyU database to compare the performances of the proposed method in this paper with our previous work. The pyramid histogram of gray (PHG), pyramid histogram of texture (PHT) and pyramid histogram of orientation gradient (PHOG) were used to reflect the global and local information of gray characteristic, texture characteristic and shape characteristic of the finger vein in our previous work. The base features that we use in this paper are the same as the previous work. The details of PHG, PHT and PHOG can be found in. In addition, the experimental settings are the same as Experiment 1. The ROC curves of the PolyU database are shown in Fig. 11. The EER, FRR at zero FAR, FAR at zero FRR, FRR at 0.001 FAR and FAR at 0.001 FRR values of different methods are listed in Table 5.

From Fig. 11 and Table 5, we can see that the proposed method achieves lower EER compared with the method in our previous work. Though both of the two methods use spatial pyramid representation of local features to maintain spatial layout information, the local features are quite different. Our new proposed method learns robust and discriminative local features (FVTs) from the traditional base feature in which is sensitive to rotation and illumination. As the center of a learned cluster, a FVT can represent all finger vein features in this cluster. In addition, since the proposed method could map two slightly different patches which are affected by rotation and illumination into the same FVT, it has good resistance to geometric deformations and illumination variations.

Table 5. The performance by different methods on the PolyU database.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FAR at zero FRR</th>
<th>FRR at zero FAR</th>
<th>FAR at 0.001 FRR</th>
<th>FRR at 0.001 FAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHG</td>
<td>0.0996</td>
<td>0.7320</td>
<td>0.7169</td>
<td>0.7319</td>
<td>0.4476</td>
</tr>
<tr>
<td>PHT</td>
<td>0.0534</td>
<td>0.7377</td>
<td>0.5951</td>
<td>0.7376</td>
<td>0.2350</td>
</tr>
<tr>
<td>PHOG</td>
<td>0.1138</td>
<td>0.9943</td>
<td>0.6944</td>
<td>0.9942</td>
<td>0.3964</td>
</tr>
<tr>
<td>FVT-Gray</td>
<td>0.0986</td>
<td>0.8501</td>
<td>0.9156</td>
<td>0.8497</td>
<td>0.5470</td>
</tr>
<tr>
<td>FVT-Texture</td>
<td>0.0300</td>
<td>0.8066</td>
<td>0.7115</td>
<td>0.8060</td>
<td>0.1442</td>
</tr>
<tr>
<td>FVT-OG</td>
<td>0.0865</td>
<td>0.7676</td>
<td>0.9209</td>
<td>0.7664</td>
<td>0.5032</td>
</tr>
</tbody>
</table>

4.7. Experiment 5

To verify that our method can be used in real-time applications, we measure the average processing time of key steps in our method. Table 6 and Table 7 show the
average times required for training, feature extraction and matching on the two databases (i.e., MLA database and PolyU database). The training time refers to generate codebook from random 100 finger vein images. Feature extraction time includes extracting local base feature time, mapping codebook time and compiling pyramid histogram time. The matching time is from the input of two finger vein features to getting the final matching result. From Table 6 and Table 7, FVT-LBP spends the least feature extraction time due to the computational simplicity of LBP. We can also see that the average training time is time-consuming, but the training process can be done off-line. In a word, from Table 6 and Table 7, we can conclude that the proposed method can be used in real-time applications.

Table 6. The average processing time by different methods on the MLA database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Training</th>
<th>Feature Extraction</th>
<th>Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVT-MC</td>
<td>340 s</td>
<td>182 ms</td>
<td>0.073 ms</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>339 s</td>
<td>174 ms</td>
<td>0.106 ms</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>356 s</td>
<td>401 ms</td>
<td>0.076 ms</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper proposes a novel finger vein verification method with vein textons. Our method is able to extract the discriminative and robust features by trained FVTs. Each FVT (cluster center in feature space) can represent some similar vein patterns. Meanwhile, the proposed method also maintains the spatial layout and local details.
information of finger vein by using SPM. To the best of our knowledge, it is the first time to introduce BoW to finger vein verification. Experimental results show that our method can achieve better performance, especially under rotation and illumination conditions. It should be pointed out that the method is general for finger vein verification. Besides LBP, MC and WLD, we can also use other features such as LDP\(^9\), LLBP as the base feature. In the future, we plan to test and verify the effectiveness of other base features. In addition, every FVT is treated as equally important, so considering the weight of FVT is also our future work.

Acknowledgments

The work is supported by National 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 like to thank the anonymous reviewers for their helpful suggestions.

References


<table>
<thead>
<tr>
<th>Method</th>
<th>Training</th>
<th>Feature Extraction</th>
<th>Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVT-MC</td>
<td>240 s</td>
<td>215 ms</td>
<td>0.070 ms</td>
</tr>
<tr>
<td>FVT-LBP</td>
<td>205 s</td>
<td>87 ms</td>
<td>0.094 ms</td>
</tr>
<tr>
<td>FVT-WLD</td>
<td>323 s</td>
<td>715 ms</td>
<td>0.116 ms</td>
</tr>
</tbody>
</table>


31. J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, “WLD:


---

**Biographical Sketch and Photo**

**Lumei Dong** received her BSEE in 2008 from Shandong University. Now she is studying at Shandong University for a Master degree in computer science and technology. Her main research interests are biometrics and machine learning.

**Gongping Yang** received his Ph.D. degree in computer software and theory from Shandong University, China in 2007. Now he is a professor in the School of Computer Science and Technology, Shandong University. His research interests are biometrics, medical image processing, and so forth.

**Yilong Yin** is the director of MLA Lab and a professor of Shandong University. He received his PhD in 2000 from Jilin University. From 2000 to 2002, he worked as a postdoctoral fellow in the Department of Electronic Science and Engineering, Nanjing University. His research interests include machine learning, data mining, computational medicine, and biometrics.

**Xiaoming Xi** received his BSEE in 2006 from Shandong Jiaotong University. Now he is studying at Shandong University for a PhD degree in computer application technology. His main research interests are biometrics and machine learning.
Lu Yang received her BSEE in 2007 from Shandong University of Technology. Now she is studying at Shandong University for a PHD degree in computer application technology. Her main research interests are biometrics and machine learning.

Fei Liu received her B-SEE in 2008 from Shandong University. Now she is studying at Shandong University for a Master degree in computer science and technology. Her main research interests are biometrics and machine learning.