Hand Vein Recognition with Bag of SIFT Feature Model

Fuqiang Li[1, a], Tongzhuang Zhang[2, b*], Yong Liu[3, c]

[1]Xuhai College, China University of Mining and Technology, Xuzhou, 221000, China

[2]School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221000, China

[3]Xuhai College, China University of Mining and Technology, Xuzhou, 221000, China

[a]email: lifqcumt@126.com, [b]email: ztzcumt@163.com, [c]email: liuycumt@126.com

Keywords: Vein Recognition; SIFT; Mismatching; BOSF; SVM

Abstract. SIFT, which is widely used for feature extraction in image recognition task especially when image rotation, translation, uneven illumination occur, has been applied in vein recognition task more widely. However, mismatching between intra-class and inter-class, which is unbearable for personal identification task, is unavoidable under the traditional ratio-based matching framework. To solve such problem, the paper proposes bag of SIFT feature (BOSF) model to realize SIFT feature based matching framework from the perspective of feature coding and classification. Finally, the proposed approach is rigorously evaluated on the self-built database and achieves the state-of-the-art EER (Equal Error Rate) of 0.026%, which demonstrates the effectiveness of the proposed model.

Introduction

Vein recognition, which is not introduced until 1990 by MacGregor and Welford [1], has become one of the most popular biometric identification methods with the advantage of unique, portable and inherent properties when compared with other biometric features such as fingerprints, iris, hand gesture and face. The main characteristics we stress when putting vein recognition into practice not only because the vein patterns of individual is different even between identical twins [2], but also lying in the fact that it is easy-acceptable, anti-counterfeit and also with high recognition rate.

Meeting all the requirements of biometric identification like other personal traits [3], vein recognition is also characterized with high convenience in image acquisition and feature representation which results in wide investigation on vein information research in hand covering palm vein [4-6], dorsal vein [7] [8] [9] and finger vein [10]. And the entire framework for the three kinds of recognition is more or less same, which covers vein image preprocessing, vein image feature extraction and representation, classifier design and vein recognition [11]. The most important and difficult part, in reality, is the feature extraction methods design, many studies conducted recently focuses on how to acquire the salient features more efficiently, and the most promising one is the local invariant feature based model for its robust to rotation, illumination change. The most popular scale-invariant feature representation methods cover SIFT [12], SURF [13], ASIFT [14], RootSIFT [15], all of which covers three main steps including the built of scale-space, extrema detection and generation of descriptors. Pierre et al. [16] tried SIFT for the vein feature representation after the preprocessing, which includes noise removal and contrast enhancement. However, the preprocessing part would result in the loss of useful vein information and the proposed method could not arise as many keypoints as possible. Other researchers [17] [18] [19] tried to extract the SIFT and SURF feature directly from the vein images without binarization process, and the recognition performance was improved. However, there exist the mismatching pairs between detected keypoints which results in high false acceptance rate (FAR) and equal error rate (EER). Kang et al. [20] proposed the hierarchical mismatching removal method based on neighborhood searching and LBP histograms after the DoG-HE enhancement and SIFT keypoints detection to improve the accuracy of feature matching. The analytical disadvantage of the method is the existence of non-vein keypoints on the palm such as the palmprint part and the noise part. It

could be concluded from [20] that there exists mismatching between intra-class and inter-class matching, which is unacceptable for a reliable personal identification system.

To solve such a problem, the paper design bag of SIFT feature (BOSF) model to coding the obtained SIFT feature vector and adopts SVM to classify the coding feature vector instead of matching them directly following [12], and the experimental results demonstrate the efficiency of the proposed BOSF model in solving the mismatching problem.

SIFT extraction of hand vein images

SIFT is of great invariance performance on the change of translation, rotation and affine transformation, and utilizing it as the feature representation method has the advantage of removing the influence of location inconsistency for the same person. Inspiring by the research on [21-22], the paper proposes the model combing SIFT and Bag of Word (BOW) to accomplish the feature representation task and introduces the SVM as the classifier to group the test images into the corresponding ones. The recognition task is conducted at the self-built vein database (as shown in Figure 1) including a total of 500 images covering 50 different samples.

Fig.1. Lab-built Hand-dorsa Vein Database

Before extracting the SIFT feature, it proves necessary to conduct the pre-process procedure regarded to the amounts of background and noise information caused by the system. In the part of dismissing the background information, the threshold set method is adopted to obtain the hand vein information $H(x, y)$, and the specific criterion is as follows:

$$HT(x,y) = \begin{cases} 0, & H(x,y) < T \\ 1, & H(x,y) \geq T \end{cases} \tag{1}$$

The following procedure is that trying the morphological erosion operation on the template image $HT(x,y)$ to cut the edge part of the vein image $HV(x,y)$, which often produces the useless feature points, and then combine the processed image with the original image to get the final ROI vein image:

$$H_{ROI}(x,y) = HV(x,y)H(x,y) \tag{2}$$

Then trying the Gaussian filtering and Median filtering process for decrease the effect caused by the white noise and Salt&Pepper noise. After the process above, it comes to go the patching process on the pre-processed vein image from the training-set database, the size set in the paper is 16×16 and the step set is 8, and then extracting the SIFT feature with each patch to get the final feature collection $X = \{x_1, x_2, \cdots, x_N\}, x_I \in R^{128}$ mixed by the each 128 dimensional descriptors.
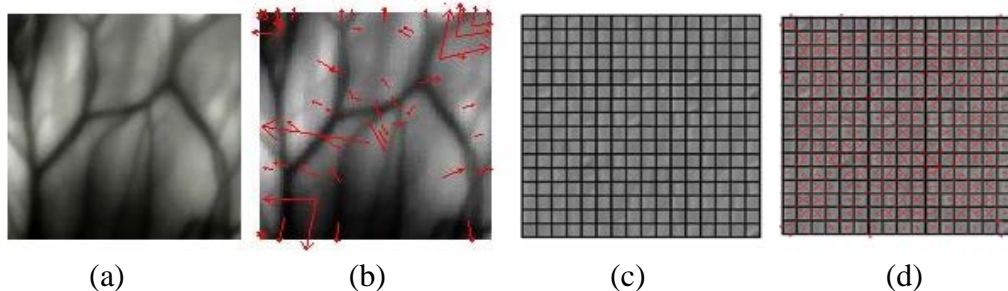

(a)   (b)   (c)   (d)

Fig.2. SIFT points distribution (a).ROI image (b).Corresponding SIFT distribution of ROI (c).Patches of ROI image (d).Corresponding SIFT distribution of Patches

Generation of BOSF Coding Vector

K-means is defined as one of the classical unsupervised clustering algorithms, the main theory of which is that grouping the to-be-dealt samples {x_1,x_2,···,x_N } into K kinds of clusters including {c_1,c_2,···,c_K }. Obviously, the characteristic of the sample distribution is that the similarity, estimated by the Euclidean distance, within class is high while that of different classes is low. The corresponding cost function is expressed as follows:

$$J = \sum_{j=1}^{K} \sum_{i=1}^{N} \left\| x_i^j - c_j \right\|^2 \tag{3}$$

The main procedure of dictionary generation with K-Means is:

(1) Set the initial K, chosen from the input N samples, data-based clustering center;

(2) Computing the distance between every sample and every clustering center, and classifying them into corresponding group according to the principle of minimum distance;

(3) Re-computing the new clustering center with the classified small groups and refresh the original one with it;

(4) Repeat the procedure in (2) and (3) until the generating center keeping the same as the former one.

The paper introduces the K-means based method to construct the bag of words model with the 350 sample images after extracting the SIFT feature and get the final K-centers dictionary $\{c_1, c_2, \cdots, c_K\}, c_i \in R^{128}$.

In the part of feature representation experiment, the paper respectively introduces SPM(Spatial Pyramid Matching) [22]and SCSPM(Linear Spatial Pyramid Matching using Sparse Coding) [23], and the influential shortcoming of the two adopted methods after result analysis is that the SPM model perform bad in the intra-class distinction experiment while for the SCSPM model is constrained by the great cost of time consumption as the result of figuring out the sparse-coding image on each input vein image. The paper makes great effort to introduce the LLC (Locality-constrained Linear Coding) [24] for improving the classical SPM model classification performance. The main modification theory is that: finding the K basis vectors, nearest by the to-be-coded vector X_i, in the feature dictionary, and then conducting the LLC procedure based on the figured K basis vectors.

After obtaining the training feature dictionary $B = [b_1, b_2, \cdots, b_k] \in R^{D \times K}$ by importing K-means, the image patching process is performed on the input vein images for the later SIFT extraction experiment to construct the real vein feature vectors set $X_I = \{x_1, x_2, \cdots, x_m\} \in R^{D \times m}$, then the following formulas are designed to realize the LLC coding experiment:

$$\min_{C} \sum_{i=1}^{m} \| X_i - Bc_i \|^2 + \lambda \| d_i \odot c_i \|^2 \tag{4}$$

$$s.t. \; 1^T c_i = 1, \forall i \tag{5}$$

The importation of $1^T c_i$ enables the LLC coding being resistance to the unexpected translation change when capturing the hand vein image, the sign of $\odot$ indicates the corresponding multiplying operation of input elements, the importation of $d_i = exp\left(\frac{dist(x_i, B)}{\sigma}\right) \in R^K$ is set to control the locality of the feature coding, the sign σ is set to regulate the weight of locality varying with distance. The vital item $\| d_i \odot c_i \|^2$ enables the LLC coding with the following advantages:

(1) Better reconstruction performance. The reconstruction of to-be-coded feature vector is obtained by the nearest basis vectors, which proves to be better than the vector quantization and SCSPM;

(2) Local smoothness. In coding procedure of SCSPM, the learning of overcomplete base is essential resulted by the sparsity. So the choice of reconstruction basis is set different for the similar input to-be-coded vector. While in the coding process of LLC, the similar vein image patches result in the similar coding vectors.

(3) Analytical solution during coding. The SCSPM is come up with great demand for iterative calculation while the LLC coding can derivate the unique the analytical solution:

$$\tilde{c}_i = (C_i + \lambda diag(d)) \backslash 1, \tilde{c}_i = \tilde{c}_i / 1^T \tilde{c}_i \qquad (6)$$

After getting the feature coding vectors $C = \{c_1, c_2, \cdots, c_m\} \in R^{K \times m}$ by LLC method, the paper introduces the 3-levels pyramid models [22] to realize the classification$(4 \times 4, 2 \times 2, 1 \times 1)$, and then grouping the maximum vectors from all the regional coding feature vectors into K dimensional to represent the regional max pooling, the specific expressions are as follow:

$$Pooling_{Max}\left(\begin{bmatrix} 9 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 8 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 7 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 8 \\ 7 \end{bmatrix} \qquad (7)$$

The final coding result for each input vein image is 21504 dimensional vectors.

Results and Discussion

In the part of feature extraction and representation experiment, the proposed method firstly conduct the patching process to get the 16×16 images with the interval of 8 pixels, and then the SIFT feature is extracted to get the 128 dimensional descriptors. Following the SIFT procedure, the corresponding vein images are random grouped into training database with 350 samples and recognition with 150 samples, and the K-means clustering method is conducted on the training database to get the feature dictionary with the dimensionality of 1024. In the final procedure, the LLC coding is performed to get the final feature representation with the dimensionality of 21504 after the 3-levels(4×4, 2×2, 1×1) pyramid pooling process.

It can be figured out that the dimensionality is a little bit high with 21504 after the feature representation experiment, the classification part try the PCA method to decrease the dimensionality to 100(set after amounts of experiments). During the classification procedure, the SVM classifier is firstly trained with the 350 training samples, and the realization of SVM is based on the Liblinear Software [25]. After obtaining the trained model, the classification experiment with the remaining samples is conducted and the specific recognition result is shown in Table 1.

Table 1. Experimental Results with BOSF Model

Database Setup		Methods	Experimental Results	
Training Set	350	BOSF model	RR (%)	99.2
			EER (%)	0.32
Testing Set	150	Traditional SIFT ratio matching	RR (%)	89.6
			EER (%)	5.49

It could be inferred from the results in Table 1 that the proposed BOSF model performs far better that the traditional SIFT with the EER of 0.32% and 5.49%. The analyzed reason lies in the fact that the LLC coding procedure and following PCA processing could pick up the feature vector corresponding to the true-matching keypoints, by which both the possibility of mismatching and total time consumption of matching could be eliminated greatly.

Conclusion

To improve the robustness of hand vein recognition system to rotation, unsteady illuminance, and ensure the advantage of contact-free, the paper focuses on SIFT and try to develop hand vein-based personal identification system with it. However, the disadvantage of mismatching between both intra-class and inter-class with SIFT introduced in [20] is unbearable. To solve such problem brought by ratio matching strategy, feature vector coding followed by SVM classification model is introduced, and the experimental results with high performance with the proposed BOSF model fully demonstrates the efficiency. The BOSF algorithm firstly extract the patch-based SIFT feature, and then the 128-dimensional vectors are coded into 21504-dimensional LLC vector, followed by PCA to obtain the 100-dimensional representation vector, and then the SVM classifier is trained and tested with the lab-made database. The success of the BOSF model lies in the fact that the dictionary learning and coding processing could invert the original SIFT feature vector to another matrix subspace, where the difference of each feature vector is well defined so that the SVM could well classify them into different groups without loss.

Acknowledgement

References

[1] M. J. Morgan, "Features and the primal sketch," Vis. Res., vol. 51, no. 7, pp. 738-753, 2011.

[2] A. Kumar, M. Hanmandlu, and H. M. Gupta, "Online biometric authentication using hand vein patterns," in Proc. IEEE Symp. Comput. Intell. Secur. Defence Appl.,
Ottawa, ON, Canada, 2009, pp. 1-7.

[3] Moreno. P, Marín-Jiménez. M. J, Bernardino. A, Santos-Victor. J and Blanca. N. P. D. L, "A comparative study of local descriptors for object category recognition: sift vs hmax," Lecture Notes in Computer Science, vol. 4477, pp. 515-522.

[4] C. Lin and K. Fan, "Biometric verification using thermal images of palm-dorsa vein patterns," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 2, pp. 199-213, Feb, 2004.

[5] S. Malki and L. Spaanenburg, "Hand veins feature extraction suing DT-CNNs," Proc. SPIE, vol. 6590, May. 2007.

[6] P. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based onSIFT matching," in PROC. IEEE Int. Conf. Biometrics, Alghero, Italy, 2009, pp. 1290-1298.

[7] S. Zhao, Y. Wang, and Y. Wang, "Biometric identification based on low quality hand vein pattern images," in PROC. IEEE Int. Conf. Mach. Learn. Cybern., Kunming, China, 2008, pp. 1172-1177.

[8] A. Kumar and K. Prathyusha, "Personal authentication using hand vein triangulation and knuckle shape," IEEE Trans. Image Process., vol. 18, no. 9, pp. 2127-2136, Sep. 2009.

[9] Y. Wang, K. Li, j. Cui, L. Shark, and M.Varley, "Study of hand-dorsa vein recognition," in Proc. Int. Conf. Intell. Comput., Changsha, China, 2010, pp. 490-498.

[10] N. Miura, A. Nagasaka, and T. Miyatake, "Feature extraction of finger-vein pattern based on repeated line tracking and its application to personal identification," Mach. Vis. Appl., vol. 15, no. 4, pp. 194-203, 2004.

[11] Mikolajczyk, K. Schmid, and C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Anal and Mach Int., vol. 27, no. 10, pp. 1615-1630, Oct. 2005.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision., vol. 60, pp. 91-110, 2004.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. G.Van, "Speeded-up robust features (SURF)," Computer Vision & Image Understanding., vol. 110, no. 3, pp. 346-359, 2008.

[14] J. M. Morel, and G. Yu, "Asift: A new framework for fully affine invariant image comparison," Siam Journal on Imaging Sciences., vol. 2, no. 2, pp. 438-469, 2009.

[15] R. Arandjelovic, and A. Zisserman, "Three things everyone should know to improve object retrieval," in IEEE International Conference on Computer Vision and Pattern Recognition, pp.2911-2918, 2012.

[16] P. O. Ladoux, C. Rosenberger, and B. Dorizzi, "Palm vein verification system based on sift matching," in Advances in Biometrics. Springer Berlin Heidelberg, 2009, pp. 1290-1298.

[17] L. Xiuyuan, L. Tiegen, D. Shichao, H. Jin, and W. Yun, "Fast recognition of hand vein with surf descriptors," Chinese Journal of Scientific Instrument., vol. 32, no. 4, pp.831-836, 2011.

[18] H. Wang, L. Tao, and X. Hu, "Novel algorithm for hand vein recognition based on retinex method and sift feature analysis," Electrical Power Systems & Computers., vol. 99, pp. 559-566, 2011.

[19] H. G. Kim, E. J. Lee, G. J. Yoon, S. D. Yang, E. C. Lee, and M.Y. Sang, "Illumination normalization for sift based finger vein authentication," in Advances in Visual Computing. Springer Berlin Heidelberg, 2012, pp. 21-30.

[20] W. Kang, Y. Liu, Q. Wu, and X. Yue, "Contact-free palm-vein recognition based on local invariant features," PLOS ONE., vol. 9, no. 5, pp. 1-12, 2014.

[21] O. Russakovsky, Y. Lin, K. Yu, et al, "Object-centric spatial pooling for image classification" [C], in European Conference on Computer Vision, 2012.

[22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories" [C], in IEEE Conference on Computer Vision and Pattern Recognition, 2006.

[23] J. Yang, K. Yu, Y. Gong, et al, "Linear spatial pyramid matching using sparse coding for image classification" [C], IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[24] J. Wang, J. Yang, K. Yu, F, "Locality-constrained linear coding for image classification"[C], IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[25] R. Fan, K. Chang, C. Hsieh, et al, "LIBLINEAR: A library for large linear classification"[J], Journal of Machine Learning Research, 9:1871-1874, 2008.