HIGH ACCURACY RECOGNITION BIOMETRICS BASED ON FINGER VEIN SCREENING SENSOR

Khamis A. Zidan [1]**, Shereen S. Jumaa** [2]
[1] Al- Iraqia University, Baghdad, Iraq
[2] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
khamis-zidan@aliraqia.edu.iq [1], shireen.sadiq81@gmail.com [2]

Abstract- With the great development in information technology, the commonly used biometrics have become very easy to hack such as face, voice and fingerprint etc. This has led to the discovery of new horizons in this field. One of the safest biometrics of today is finger vein- But this technique faces some unique difficulties, the most commonly used being that the vein pattern is difficult to remove because finger vein images are always low in quality, significantly hampered the feature extraction and classification stages. For this purpose, professional algorithms must be considered with the conventional hardware for capturing finger- vein images using Red Surface Mounted Diode (SMD) led. For capturing images, Canon 750D camera with micro lens is used. For high quality images the integrated micro lens is used, and with some adjustments it can also obtain finger print. Features extraction was used by a combination of Hierarchical Centroid and Histogram of Gradients. Results were evaluated with K Nearest Neighbor and Deep Neural Networks using 6 fold stratified cross validation. Results displayed improvement as compared to three latest benchmarks in this field that used 6- fold validation (5- samples for training and one sample for test) for SDUMLA- HMT. The work novelty is owing to the hardware design of the sensor within the finger- vein recognition system to obtain, simultaneously, highly secured recognition with low computation time, finger vein and finger print at low cost, unlimited users for one device and open source.

I. INTRODUCTION

Today the Information security is being controlled by the use of biometric technology- the technique of identifying a person based on biological features such as physical or behavioral characteristics [1] Physical biometrics include iris, hand geometry, face, fingerprint and behavioral biometrics include gait, voice, signature, keystroke pattern etc. Most of these features could be vulnerable to hack attempts [2], [3] and for this reason had led to the suggestion of more secure biometrics such as vein patterns within finger [4] hand [5] and palm [6] .These are nearly no way to reach without the user's consent and hence harder to fake. Finger vein recognition is considered as one of the harder tasks because of low image quality, uneven- illumination and differences in temperature. Finger vein identification systems are also secure against spoof attacks [7], [8], [9] and [10]. So, FVR continue to need more fast and proficient methods. The typical framework of the biometric finger vein recognition (FVR) system is presented in Fig. 1.

II. DATABASE

It is a fact commonly agreed [46], [13] that the hardest finger vein database to preprocess is SDUMLA- HMT database as shown in Fig. 2. The reasons for this non- ideality are many such as low quality of images, Finger area is very small as compared to total image size [14] and Images vary in rotation, translation and shift [15]. In getting of images, a number of people who participate is one hundred and six (106). The format of images of vein is bmp of each person, images of both two hands were taken. Data of three fingers were collected i.e. index, middle and ring fingers for each hand [16] and for each finger, 6 samples were obtained. For this purpose, Joint Lab for Intelligent Computing and Intelligent Systems of

Wuhan University made a device that was used. In this work finger- vein recognition system is modified by developing a new method for image pick up in the hardware of the device. The images are processed by some of enhancement stages, these stages not only have low computational complexity but also high distinguishing power. To evaluate this work, the collected database by proposed device have been used. A number of people who take part is one hundred (100). The format of images of vein is .jpg. Of each person, images of both two hands were taken. Samples were collected as shown in Fig. 3 and Table I.
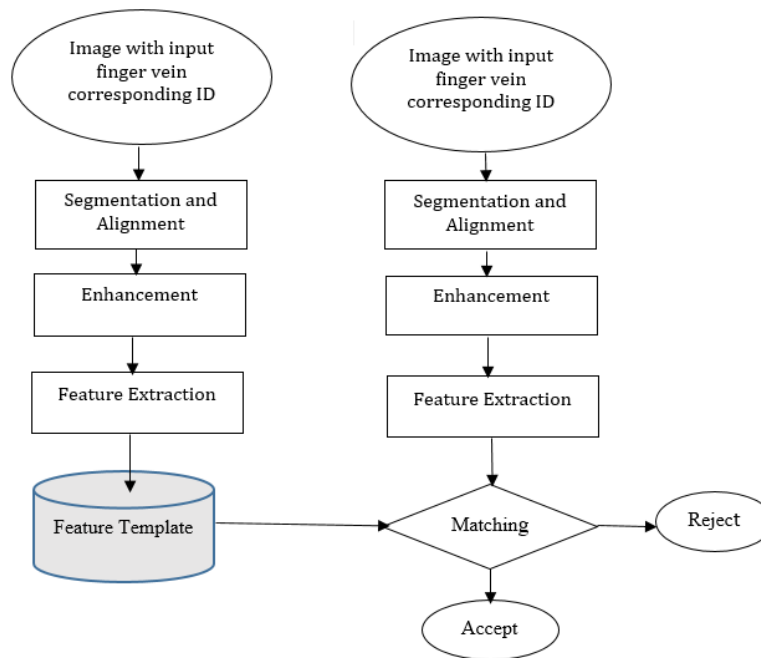

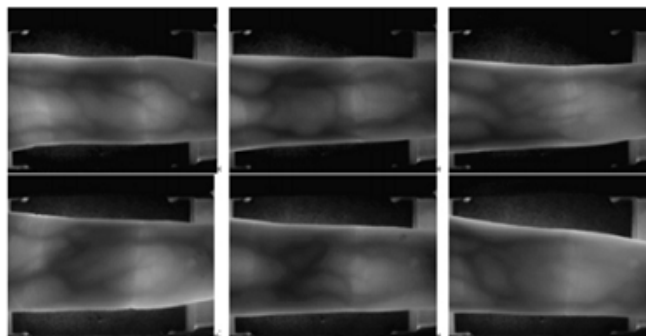Figure 1: Finger- vein recognition process [11]

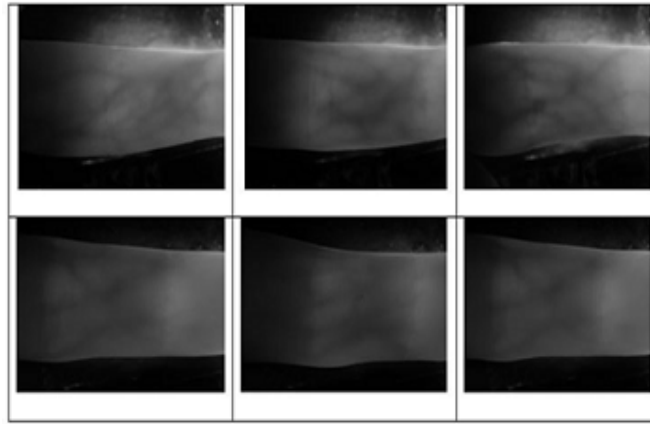Figure 2: SDUMLA- HMT database sample data

Khamis A. Zidan , Shereen S. Jumaa

Figure 3: Database sample data

TABLE I
DATABASE OF A FINGER VEIN

Parameter	Information
No of persons	100
Hands per person	2
Fingers per hand	3
Images per finger	6
Image resolution	910×603
Image format	jpg

III. FINGER VEIN SENSOR

In this work the proposed hardware for getting finger- vein images is modified by using red Surface Mounted Diode (SMD) led on both sides, instead of near- infrared led. For getting images, Canon 750D camera is used with micro lens. To obtain better quality images, the integrated micro lens was used with some adjustments it can also capture finger print. Mobile camera can be used instead of Canon camera for image capturing. The SMD led that are used, have better illumination. The proposed device was used to collect the database of this work. The goal of this work is to increase accuracy and efficiency of finger- vein recognition by reducing the processing time. To achieve this goal, the hardware of finger- vein recognition device is modified by using SMD led instead of near-infrared led, the modified system improves images quality by generating high quality images. Finger vein sensor as shown in Fig. 4.
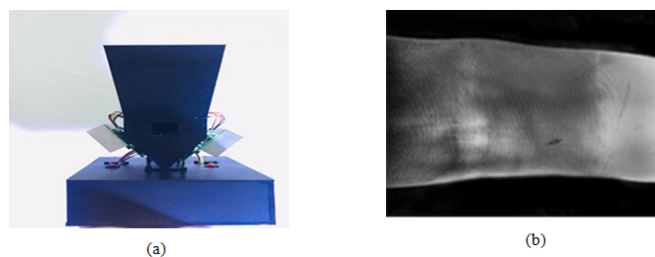

(a)   (b)

Figure 4: (a) Image capturing device (b) Sample of finger- vein image

IV. PROCESSING AND ENHANCEMENT IMAGES

The image acquired is preprocessed in different stages. The proposed method is evaluated as following:

A. Region of interest

Detect the region of interest (ROI) of the finger vein image must be identified and separated from background of image. The mask filter is used to do this localization, which follows the procedure as [17], [18]. The database image is then cropped by keeping the portion of only finger region as shown in Fig. 5.
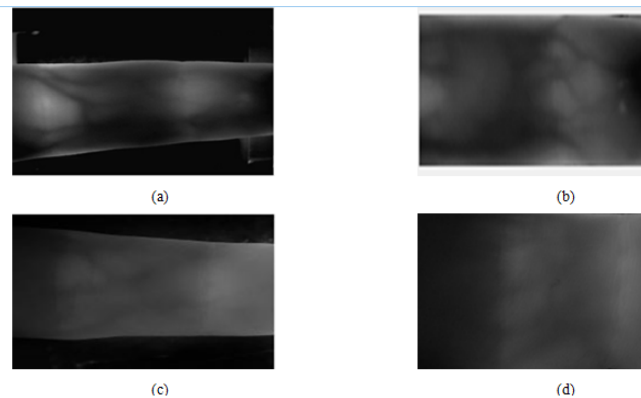

Figure 5: (a) SDUMLA- HMT database image, (b) SDUMLA- HMT cropped finger, (c) The collected data by proposed device, (d) Cropped finger for collected data

B. Gamma correction

To enhance the brightness of the cropped finger, gamma correction is used [19], [20].The enhancement steps is done by manipulating nonlinearity pixel intensities. It increases the quality of image efficiently by guiding the average brightness of an image in the direction desired. The results of gamma correction as shown in the images in Fig. 6.
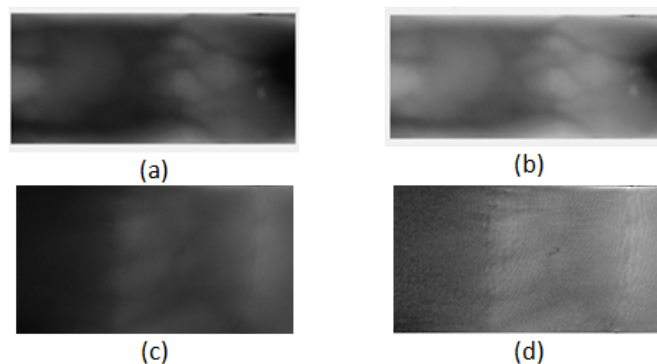

Figure 6: (a)An image before gamma correction,(b) An image after gamma correction, (c) An image before gamma correction,(d) An image after gamma correction

C. *Fuzzy histogram equalization*

To get higher quality images, fuzzy histogram equalization is used to increase the enhancement of veins illumination. Dynamic fuzzy histogram equalization is used because it does not affect the brightness and it introduces the histogram image in a manner that in mapping of peaks does not appear again. The values of histogram are spread in a uniform manner between valley similar areas of two serial peaks [21]. In the Fig. 7 The image after fuzzy histogram equalization and image histogram.
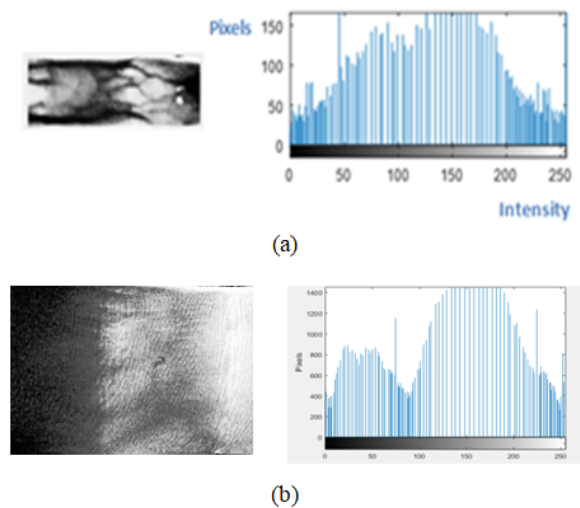

Figure 7: (a) SDUMLA- HMT image after histogram equalization with image histogram, (b) The collected image after histogram equalization with image histogram

D. *Median filtering*

To remove the impulsive and salt- and- pepper noise, Median filtering is used on the image. It smoothly removes noise and keeps the edge- information. The strong point of Median filter depends upon window size [22], [23], [24]. shown that in Fig. 8.

V. FEATURE EXTRACTION

To extract the information and features from images, histogram of oriented gradient (HOG) and hierarchical centroid shape descriptor methods are used.

1) Histogram of Oriented Gradient Feature extraction

In tough visual situation, Histogram of Oriented Gradient (HOG) feature extraction is extensively used due to its ability to do well in those situations [24]. It extracts the information about target edge distribution and is a useful tool for shape identification [25], [26], [27], [28]. Shift and rotations of object do not affect this low- level feature descriptor [29]. Its algorithm works as follows:

- Image is separated into small sections which are called " cells".

- Manifestations of edge alignments in neighboring local section [30] are calculated for each cell (for ' signed gradients' , histogram channels are spread evenly over 0- 360 and for ' unsigned' over 0- 180) .

- For scaling histogram counts, lighting effect is fixed. To do so, add energy measure for each local histogram over the larger connected sections.

- Feature vector, in the end, is accumulated by seeing all orientation counts in the calculated dominant direction of gradients.

The above stated steps therefore normalize the block- cells whose joint histograms establish the HOG feature vector of the complete image. Scaling and rotation do not affect the algorithm because only the key points are used for scale-space extreme a detection. The gradient directions for a sample image patch as shown in Fig. 9 Thus, HOG descriptor primarily comprises of features about histograms of slopes of object edges of an image. HOG feature descriptor can be defined by the equation [31]:

$$F_{HOG} = [F_1, F_2, F_3, ..., F_K] \tag{1}$$

Where: $K =$ the total number of overlapping blocks in an image F_i is the normalized block vector of i-th block:

$$F_i = [h_{1,i}, h_{2,i}, h_{3,i}, ..., h_{C*B,i}] \tag{2}$$

Where: $C =$ total number of cells in a block.

$B =$ total number of bins in a cell.

Bins create a cell and cells make up a block. For a sample image of size 32×32, this procedure can be demonstrated as shown in Fig. 10

2) Hierarchical Centroid Shape Descriptor (HCSD) HCSD [32], [33] is fixed in kd- tree procedure decomposition [34]. It is a feature descriptor mainly for shapes.Centroid coordinates extracted from a binary image create this feature descriptor. It works by iteratively disintegrating an image into sub blocks of pixels by means of the Kd- tree algorithm. It comprises of data regarding centroid coordinates of local sections [35]. If d denotes the depth of the features extraction procedure, then the length of this descriptor is specified by $2 \times (2d - 2)$. Assume I is the $M \times N$ binary image. Its foreground is denoted by I_{fg} and background by I_{bg}. HCSD is then created with the following stages:

- For input I, find its transpose: I^T,

- Find centroid $C(x_c, y_c)$ at the root level for every individual input as equations below:

$$x_c = \frac{m_{10}}{m_{00}} \tag{3}$$

$$y_c = \frac{m_{01}}{m_{00}} \tag{4}$$

Where: $m_{10} =$ first order moment along the x-axis

$m_{01} =$ the first order moment along the y-axis

$m_{00} =$ area of I_{fg}

Raw moment (also recognized as moment of order $(p + q)$) for a $2D$ continuous function $f(x, y)$ is specified by Eq. 5

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \tag{5}$$

The raw moments m_{pq} of a digital image with pixel intensities $I(i, j)$ are specified by Eq. 6

$$m_{pq} = \sum_{i=0}^{M} \sum_{j=0}^{N} i^p j^q I(i, j) \tag{6}$$

- According to their centers of gravity ($x = x_c, y = y_c$) are Iteratively decomposed the image into two sub images until desired depth of breakdown is achieved. At each next stage, it is needed to change the axis of coordinates secured.
- As next step, attained vector is needed to scale over a range of -0.5 to 0.5. Here, 0 is for the root- level centroid. Negative range represents tree decomposition of left sides and positive range represents right side.
- Lastly, extracted features of the images I and I^T are combined. K- tree decomposition structure as shown in Fig. 11

Both of SDUMLA- HMT Database Image and collected database by proposed device are evaluated with K Nearest Neighbor and Deep Neural Networks using 6 fold stratified cross validation. The more details for mechanical operation of KNN and DNN in [36], [37], [38], [39].
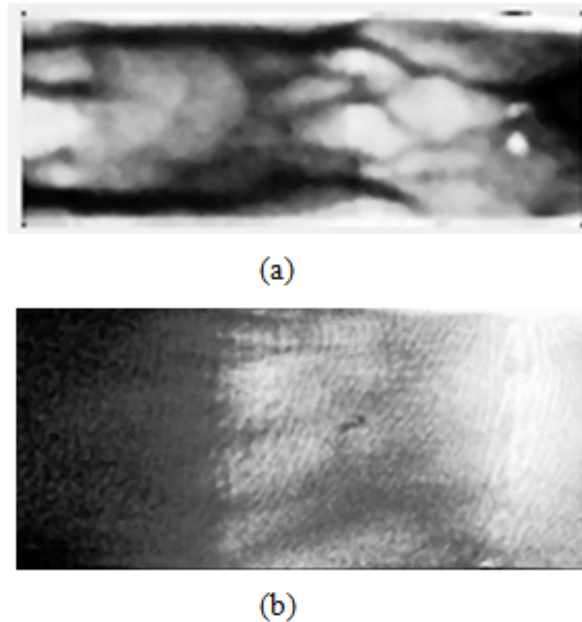

Figure 8: (a) SDUMLA- HMT image after median filter enhancement, (b) The collected image after median filter enhancement
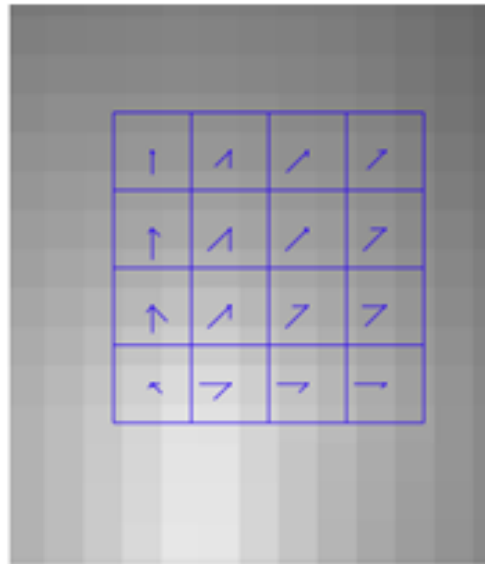
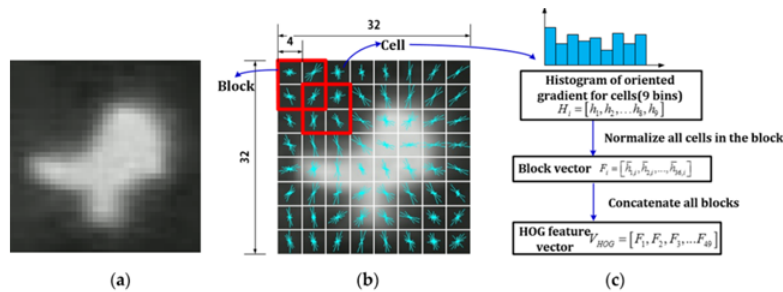Figure 9: Sample HOG descriptors, each cell of the patch shows the orientation of the gradients

Figure 10: (a) Sample image (b) Visualization of gradients. Each block (red outline) contains four cells (white outline) (c) Process of generating HOG feature vector [31]
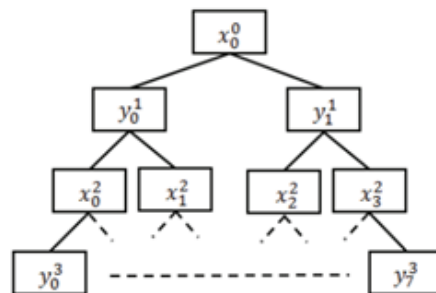

Figure 11: K- tree decomposition structure [34]

Khamis A. Zidan , Shereen S. Jumaa

VI. CLASSIFICATION

1) K- nearest neighbors.

   K- nearest neighbor classifier, also called KNN [40] is a primitive but immensely favored classification method. The label of K closest samples is assigned to a test feature vector [41], [42]. It is a non- parametric passive algorithm [43].

   - It is termed " non- parametric" as it does not cause any inherent assumptions about the placement of data into classes.
   - It is termed " passive" because it does not utilize training data for generalization. It forecasts test data based on all the training data.

2) Deep neural networks The process of learning data representations with many levels of abstraction through a set of computational models based on numerous processing layers is called " Deep Learning" (see Fig. 12) [44], [45].
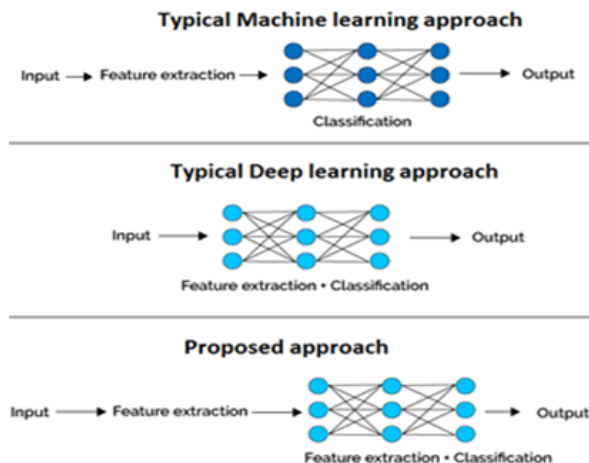

Figure 12: Deep neural network

VII. RESULTS

To evaluate the proposed system the mix finger type and hand type was tested for each feature extraction method and also for each classifier. In the experiment results, it was found that the middle finger is the best of all for the considered database for both SDUMLA- HMT database and collected data by proposed device. Results of the proposed algorithm for recognition and verification for middle finger of right hand, left hand and both hands are collected in the Tables (II, III, IV) . For KNN, results are given with HOG, HC and with the feature vector of HOG and HC. For DNN, dimensions of combined HOG and HC were first reduced (to be equal to 15) using Generalized Discriminant Analysis (GDA) [46] , before using them for classification. The cause for this is that data samples per class were limited and DNN must work best on wide-ranging of data whose feature length is not more than the total available samples. Results of the proposed algorithm with DNN for recognition and verification for middle finger of right hand, left hand are collected in the Tables (V, VI) and KNN results for both.

TABLE II
RESULTS OF KNN WITH RIGHT HAND, MIDDLE FINGER

Method	Accuracy (%)		EER (%)	
	SDUMLA- HMT database	The collected data by proposed device	SDUMLA- HMT database	The collected data by proposed device
Proposed approach using HOG	99.0566	100	0.0089	0
Proposed approach using HC	99.0566	95.238	0.0089	0.226
Proposed approach using HOG + HC	99.0566	100	0.0089	0

TABLE III
R. RESULTS OF KNN WITH LEFT HAND, MIDDLE FINGER

Method	Accuracy (%)		EER (%)	
	SDUMLA- HMT database	The collected data by proposed device	SDUMLA- HMT database	The collected data by proposed device
Proposed approach using HOG	97.1698	80.952	0.0267	0.907
Proposed approach using HC	93.3962	90.476	0.0623	0.453
Proposed approach using HOG + HC	97.1698	80.952	0.0267	0.907

TABLE IV
RESULTS OF KNN WITH BOTH RIGHT AND LEFT HAND, MIDDLE FINGER

Method	Accuracy (%)		EER (%)	
	SDUMLA- HMT database	The collected data by proposed device	SDUMLA- HMT database	The collected data by proposed device
Proposed approach using HOG	97.1698	90.476	0.0148	0.453
Proposed approach using HC	95.7547	92.857	0.0192	0.340
Proposed approach using HOG + HC	97.1698	90.476	0.0148	0.453

TABLE V
RESULTS OF DNN WITH RIGHT HAND, MIDDLE FINGER

Method	Accuracy (%)		EER (%)	
	SDUMLA- HMT database	The collected data by proposed device	SDUMLA- HMT database	The collected data by proposed device
Proposed approach using dimension reduced HOG + HC	89.6226	80	0.93895	0.975

TABLE VI
R. RESULTS OF DNN WITH LEFT HAND, MIDDLE FINGER

Method	Accuracy (%)		EER (%)	
	SDUMLA- HMT database	The collected data by proposed device	SDUMLA- HMT database	The collected data by proposed device
Proposed approach using dimension reduced HOG + HC	87.7358	60	0.93005	0.995

VIII. CONCLUSIONS AND FUTURE WORK

In this work, it introduces new methods of finger vein enhancement for recognition and verification purposes using two type of database, the first one is the SDUMLA database and the second is the data that collected by proposed hardware device. An extensive evaluation of results is done with KNN. KNN proved to be much more accurate in its predictions on test data. The results obtained from this research to compare the accuracy and reduced processing time for finger-vein

recognition for both database with proposed system. The approach developed in this work has optimum results as compared to previous ones especially when we used the collected data for right hand and middle finger, maximum accuracy 100% was reached with an EER of 0.For the SDUMLA database, a maximum accuracy of 97.6415% was reached with an EER of 0.0037. based on the analysis of the 106 persons present in the dataset, it was found that the most discriminative finger is the middle finger of right hand which gives above 99% accuracy. For future work, The developed system may be optimized using some optimization methods, furthermore, using features vein and finger print in classification and recognition approach. Convolution Neural Networks may efficiently be used to check performance using transfer learning techniques.

REFERENCES

[1] Syazana- Itqan, S. K. and N. Saad, " A review of Finger- Vein Biometrics Identification Approaches" , Indian Journal of Science and Technology, 2016.

[2] S. Marcel, M. S. Nixon and S. Z. Li, " Handbook of Biometric AntiSpoofing: Trusted Biometrics Under Spoofing Attacks" , Springer Publishing Company, Incorporated, 2014.

[3] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz and H. Pedrini, " Deep Representations for Iris, Face, and Fingerprint Spoofing Detection" , IEEE Transactions on Information Forensics and Security, 2015.

[4] A. Kumar and Y. Zhou, " Human Identification Using Finger Images, " IEEE Transactions on Image Processing, 2011.

[5] W. Xuebing, Z. Jiangwei and L. Xuezhang, " Research on Enhancing Human Finger Vein Pattern Characteristics" , Asia- Pacific Conference on Power Electronics and Design, 2010.

[6] A. R. S. K. Syazana- Itqan, N. M. Saad, N. A. Hamid and W. H. B. M. Saad, " A Review of Finger- Vein Biometrics Identification Approaches" , Indian Journal Of Science and Technology, 2016.

[7] J. Yang and Y. S. G. Jia, " Finger- Vein Image Matching Based on Adaptive Curve Transformation" , Pattern Recognition, 2017.

[8] H. Zheng, Q. Xu, Y. Ye and W. Li, " Effects of Meteorological Factors on Finger Vein Recognition" , in 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA) , 2017.

[9] H. Ding, " Anti- Spoofing a Finger Vascular Recognition Device with Pulse Detection" , 2016.

[10] X. Xi, L. Yang and Y. Yin, " Learning Discriminative /binary Codes for Finger Vein Recognition" , Pattern Recognition, 2017.

[11] X. M. X. Y. Yin, " Science China Information Sciences" , Finger vein recognition based on deformation information, 2018.

[12] Y. Yin and L. L. X. Sun, " SDUMLA- HMT: A Multimodal Biometric Database" , in Chinese Conference on Biometric Recognition, 2011.

[13] G. Y. Y. Y. a. X. X. Lu Yang, " Finger Vein Recognition with Anatomy Structure Analysis" , IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2017.

[14] Y. Yang, G. Yang and S. Wang, " Finger Vein Recognition Based on Multi- Instance" , International Journal of Digital Content Technology and its Applications, 2012.

[15] J. M. S. a. K. R. P. Wan Kimm " Multimodal Biometric Recognition Based on Convolutional Neural Network by the Fusion of Finger- Vein and Finger Shape Using Near- Infrared (NIR) Camera Sensor" , Sensors MDPI, 2018.

[16] K. Akintoye, M. S. M. Rahim and A. H. Abdullah, " Enhancement of Finger Vein Image Using Multifiltering Algorithm" , ARPN Journal of Engineering and Applied Sciences, 2018.

[17] H. C. L, K. R. P and Eui Chul Lee, " Finger Vein Recognition Using Minutia- Based Alignment and Local Binary Pattern- Based Feature Extraction" , 2009.

[18] S. A. Amiri and H. Hassanpour, "A Preprocessing Approach For Image Analysis Using Gamma Correction" ,I nternational Journal of Computer Applications, 2012.

[19] K. Somasundaram and P. Kalavathi, " Medical Image Contrast Enhancement Based on Gamma Correction" , International Journal of Knowledge Management & e- Learning, 2011.

[20] D. Sheet, H. Garud, A. Suveer, M. Mahadevappa and J. Chatterjee, " Brightness Preserving Dynamic Fuzzy Histogram Equalization" , IEEE Transactions on Consumer Electronics, 2010.

[21] H. Hassanpour, N. Samadiani and S. Salehi, " Using morphological transforms to enhance the contrast of medical images" , The Egyptian Journal of Radiology and Nuclear Medicine, 2015.

[22] M. Nagu and N. Shanker, " Image De- Noising by Using Median Filter and Weiner Filter" , International Journal of Innovative Research in Computer and Communication Engineering, 2014.

[23] A. Makandar and B. Halalli, " Breast Cancer Image Enhancement using Median Filter and Clahe" , International Journal of Scientific & Engineering Research, 2015.

[24] S. A. Korkmaz, A. Akcicek, H. Binol and M. F. Korkmaz, " Recognition of the Stomach Cancer Images with Probabilistic HOG Feature Vector Histograms by Using HOG Features" , in 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY) , 2017.

[25] J. Ouyang, " Combining Extreme Learning Machine, RF and HOG for Feature Extraction" , in IEEE Third International Conference on Multimedia Big Data (BigMM) , 2017.

[26] J. Chen, D. Zhou, Y. Wang, H. Fu and M. Wang, " Image Feature Extraction Based on HOG and Its Application to Fault Diagnosis for Rotating Machinery" , Journal of Intelligent & Fuzzy Systems, 2018.

[27] C. Q. Lai and S. S. Teoh, " Efficiency Improvement in the Extraction of Histogram Oriented Gradient Feature for Human Detection Using Selective Histogram Bins and PCA" , in 9th International Conference on Robotic, Vision, Signal Processing and Power Applications , 2016.

[28] L. Yang, G. Yang, X. Xi, X. Meng, C. Zhang and Y. Yin, " Tri- Branch Vein Structure Assisted Finger Vein Recognition" , IEEE Access , vol. 5, 2017.

[29] O. Deniz, G. Bueno, J. Salido and F. D. l. Torre, " Face Recognition Using Histograms of Oriented Gradients" , Pattern Recognition Letters , 2011.

[30] Y. Ma, X. Wu, G. Yu, Y. Xu and Y. Wang, " Pedestrian Detection and Tracking from Low- Resolution Unmanned Aerial Vehicle Thermal Imagery" , Sensors, 2016.

[31] S. Armon, " Handwriting Recognition and Fast Retrieval for Hebrew Historical Manuscripts" , Hebrew University of Jerusalem, 2011.

[32] A. T. K. W. A. Sexton, " Font Recognition Using Shape- Based Quad- Tree and Kd- Tree Decomposition" , in Proceedings Of The Joint Conference On Information Sciences, 2000.

[33] E. Ilunga- Mbuyamba, J. G. Avina- Cervantes, D. Lindner and J. Guerrero- Turrubiate, " Automatic Brain Tumor Tissue Detection Based on Hierarchical Centroid Shape Descriptor in Tl- Weighted MR Images" , in 2016 International Conference on Electronics, Communications and Computers (CONIELECOMP) , 2016.

[34] L. K. J. H. K, H. J. Wahyono, " Similarity Based Classification of 2- d Shape Using Centroid- Based Tree-Structured Descriptor" , Modern Advances in Applied Intelligence, Springer International Publishing, 2014.

[35] K.Zidan and Sh.Jumaa ," Finger Vein Recognition Using Two Parallel Enhancement Approaches Based Fuzzy Histogram Equalization" , PEN, vol. 7,2019.

[36] K.Zidan and Sh.Jumaa ," A Highly- Verified Biometric Recognition System Using An Ultra- Speed Specifically- Developed Finger Vein Sensor" , PEN, vol. 7,2019.

[37] P. Sun, W. Feng, R. Han, S. Yan and Y. Wen, " Optimizing Network Performance for Distributed DNN Training on GPU Clusters: ImageNet/ AlexNet Training in 1.5 Minutes" , Distributed, Parallel, and Cluster Computing, 2019.

[38] M. Jamal, A. Arun, RossaErik and M.Shapirob, " On Automated Source Selection for Transfer Learning in Convolutional Neural Networks" , Pattern recognition, 2018.

[39] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, " Efficient KNN Classification with Different Numbers of Nearest Neighbors" , IEEE Transactions on Neural Networks and Learning Systems , 2018.

[40] A. Z. A. Zainuddin, W. Mansor, L. Y. Khuan and Z. Mahmoodin, " Classification of EEG Signal from Capable Dyslexic and Normal Children Using KNN" , American Scientific Publishers, 2018.

[41] M. S. Sarma, Y. Srinivas, M. Abhiram, L. Ullala and M. S. Prasanthi, " Insider Threat Detection with Face Recognition and KNN User Classification" , in IEEE International Conference on Cloud Computing in Emerging Markets (CCEM) , 2017.

[42] R. S. P. S.Ponmani, " Classification Algorithms in Data Mining - A Survey" , International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) , 2017.

[43] L. B. R. F. L. T. M. P. Du Tran, " Learning Spatiotemporal Features with 3D Convolutional Networks" , Computer Vision and Pattern Recognition, 2014.

[44] W. Liu, W. Li, L. Sun, L. Zhang and P. Chen, " Finger Vein Recognition Based on Deep Learning" , in 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA) , 2017.

[45] F. Bahmaninezhad and J. H. Hansen, " Generalized Discriminant Analysis (GDA) for Improved i- Vector Based Speaker Recognition" , Interspeech, 2016.

[46] Y. Ait- Sahalia and D. Xiu, " Principal Component Analysis of High- Frequency Data" , Journal of the American Statistical Association , 2017.