SPECIAL ISSUE PAPER

Single-sensor hand-vein multimodal biometric recognition using multiscale deep pyramidal approach

Shruti Bhilare[1] · Gaurav Jaswal[2] · Vivek Kanhangad[1] · Aditya Nigam[2]

Abstract

Biometrics has emerged as a powerful technology for person authentication in various scenarios including forensic and civilian applications. Deployment of biometric solutions that use cues from multiple modalities enhances the reliability and robustness of authentication necessary to meet the increasingly stringent security requirements. However, there are two drawbacks typically associated with multimodal biometrics. Firstly, the image acquisition process in such systems is not very user-friendly, primarily due to the time and effort required to capture biometric samples belonging to multiple modalities. Secondly, the overall cost is higher as they employ multiple biometric sensors. To overcome these drawbacks, we employ a single NIR sensor-based image acquisition in the proposed approach for hand-vein recognition. From the input hand image, a palm-vein and four finger-vein subimages are extracted. These images are then enhanced by CLAHE and transformed into illumination invariant representation using center-symmetric local binary pattern (CS-LBP). Further, a hierarchical non-rigid matching technique inspired by the architecture of deep convolutional networks is employed for matching the CS-LBP features. Finally, weighted sum rule-based matching score-level fusion is performed to combine the palm-vein and the four finger-vein modalities. A set of rigorous experiments has been performed on an in-house database collected from the left and right hands of 185 subjects and the publicly available CASIA dataset. The proposed approach achieves equal error rates of 0.13% and 1.21%, and rank-1 identification rates of 100% and 100% on the in-house and CASIA datasets, respectively. Additionally, we compare the proposed approach with the state-of-the-art techniques proposed for vascular biometric recognition in the literature. The important findings are (1) the proposed approach outperforms all the existing techniques considered in this study, (2) the fusion of palm-vein and finger-vein modalities consistently leads to better performance for all the feature extraction techniques considered in this work. (3) Furthermore, our experimental results also suggest that considering the constituent palm-vein and finger-vein images instead of the entire hand-vein images achieves better performance.

Keywords Vascular biometrics · Hand-vein recognition · Deep matching and finger-vein recognition

1 Introduction

Nowadays, it is important that reliable security solutions are employed to perform person recognition. The traditional password and token-based recognition approaches have several inadequacies such as the risk of passwords being cracked, forgotten or divulged or the tokens being stolen or lost. Therefore, they have been largely replaced by a more efficient alternative namely, biometric authentication, which uses various physiological or behavioral characteristics of

a person for establishing or verifying the identity. In the literature, several biometric traits have been investigated, including fingerprint, face, ear, iris, palmprint and gait. However, the unimodal biometric solutions have been shown to be quite vulnerable to sensor-level attacks such as presentation attacks, which involve presenting a fake copy of the genuine user's sample to the biometric sensor. On the contrary, multimodal biometric solutions combine information extracted from several biometric traits, thereby exhibiting higher resilience against presentation attacks. Moreover, they also provide better biometric performance and increased population coverage, which are essential for performing large-scale recognition. However, multimodal biometric systems have two major drawbacks: firstly, high overall cost of the system owing to the multiple biometric sensors and sec-

✉ Shruti Bhilare
  phd12110103@iiti.ac.in

1 Indian Institute of Technology Indore, Indore, India

2 Indian Institute of Technology Mandi, Mandi, India

ondly, higher degree of user cooperation required to capture samples belonging to different biometric modalities at the time of authentication (or enrollment). The above-mentioned drawbacks, however, can be overcome to a certain extent if multiple modalities can be acquired from a single body location. For instance, palmprint and palm-vein images can be captured by varying the wavelength of the incident light.

In order to achieve better cost-effectiveness in multimodal biometric recognition systems, it is desirable to reduce the number of sensors employed for image acquisition. This is achievable when the biometric characteristics are in close proximity. For instance, in the case of a hand image, palmprint, fingerprint and hand geometry features can be extracted from an image acquired using a single sensor. Thus, to enhance user-friendliness, while exploiting the benefits of multimodal biometrics, we propose to capture the frontal hand image in such a way that it covers the palm and finger regions. In particular, we extract the vascular biometric characteristics from palm and finger regions in the image.

The rest of the paper is organized as follows: In Sect. 2, we briefly outline the related work. Section 3 describes the proposed approach in detail. Section 4 presents a comprehensive analysis of the experimental results and comparison with other techniques. Finally, Sect. 5 concludes the paper with directions for future work.

2 Literature review

The idea of multimodal biometric systems has received much attention from researchers owing to its enhanced reliability and performance. Unimodal biometric systems often have to deal with issues such as limited population coverage and susceptibility to spoofing. Since multibiometric systems incorporate information from multiple biometric traits, these problems are reduced to a great extent.

Multimodal biometric systems involve the fusion of information from multiple biometric traits to achieve a performance superior as compared to the individual modalities. Multimodal biometric systems vary based on the number of sensors, traits, feature sets and classifiers employed. But often, the term multimodal biometrics is used synonymously with the scenarios in which multiple biometric traits are considered for fusion. One of the earliest works in multimodal biometrics was the fusion between the voice and face biometric traits at the matching score level [1]. Several biometric traits have been explored in the literature; for instance, authors in [2] employ face and ear biometric traits for biometric recognition, while authors in [3] utilize iris, fingerprint and face traits. In majority of the reported works, biometric modalities are obtained from two or more different locations on the human body. Not only does this make the process of enrollment and authentication incon-

venient for the users, it also adds to the cost incurred by the employment of multiple sensors. This makes deployment of such multimodal systems less appealing for civilian applications. Thus, the choice of biometric traits for fusion plays a key role in the usability of the biometric system. For instance, the authors in [4] make use of the finger-vein and fingerprint modalities which are acquired from the frontal finger region making the biometric acquisition process more user-friendly. Similar studies have been conducted for palmprint and palm-vein-based [5], finger-vein and finger dorsal texture-based [6], palm and hand dorsal vein patterns [7,8] and finger-knuckle- and finger-vein-based [9] recognition. Although these approaches require less effort on the users' part, the complexity of the associated hardware is high.

Thus, in order to enhance the user-friendliness and cost-effectiveness while still enjoying the performance benefits of multimodal biometrics, single-sensor multimodal biometrics have been investigated in the past. Researchers have used a single biometric sensor for acquisition of hand images and have extracted features from palmprint, fingerprint and hand geometry biometric traits [10–12]. A more recent study utilizes multiple information from a single image captured using a single-sensor, specifically, a CCTV camera. The information includes gait, head and height for criminal investigation [13]. The hand-based biometric acquisition is prone to presentation attacks which have been shown to be highly successful in spoofing palmprint-based [14], fingerprint-based [15] and hand geometry-based [16] biometric systems. This has led to increased attention towards vascular biometrics. Besides being unique, the subcutaneous vein structures have the added advantage of lying underneath the skin surface, and thereby, making it less vulnerable to spoofing attacks [17]. Majority of the recent works in vascular biometrics explore the line-based and curvature-based information in the vessel structures present in the biometric samples. The authors have demonstrated the effectiveness of these features for palm-vein- [18–22], finger-vein- [23–27] and dorsal hand-vein-based [28,29] biometric systems. However, much recent work has been focused on employing the deep learning techniques in various domains, and the field of biometrics is not an exception. These works have shown that the deep convolutional neural network (DCNN)-based features perform better than the handcrafted features for face [30,31] and finger-vein [32] verification systems. Furthermore, a hierarchical convolutional architecture, similar to that of DCNN, has been employed for improving the matching performance by the authors in [33]. Encouraged by the recognition performance in [33], we adopt the matching technique in the proposed work.
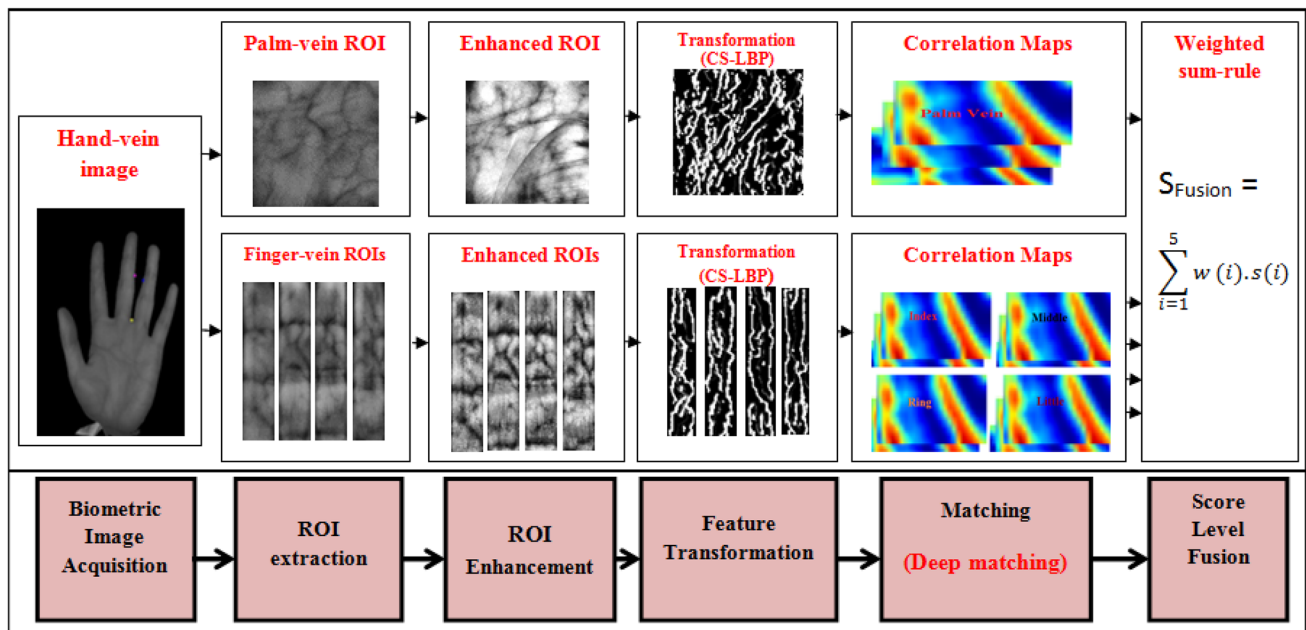
Fig. 1 Overview of the proposed approach

3 Proposed approach

Figure 1 shows an overview of the proposed approach. The hand-vein images are segmented to obtain the palm-vein and the finger-vein regions of interest (ROIs). For majority of the images, ROIs have been extracted automatically. However, in some cases, manual intervention was required. ROI extraction is followed by image enhancement using contrast-limited adaptive histogram equalization (CLAHE). The enhanced images are then transformed using CS-LBP, a computationally efficient and illumination invariant feature extraction technique used in the literature [34]. Further, a hierarchical non-rigid dense matching technique [35] inspired by the functioning of deep convolutional networks has been employed for matching the resulting CS-LBP images. Finally, weighted sum rule-based matching score-level fusion is performed to combine the palm-vein and the four finger-vein modalities.

3.1 Image preprocessing

As part of the preprocessing, the acquired hand-vein image is binarized using graylevel thresholding [36] to segment the hand region. This segmented image is further processed to separate the palm-vein and finger-vein ROIs.
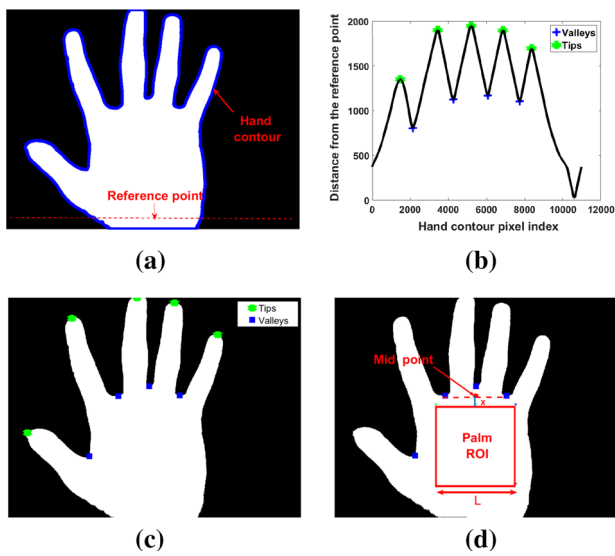
3.1.1 Palm-vein ROI extraction

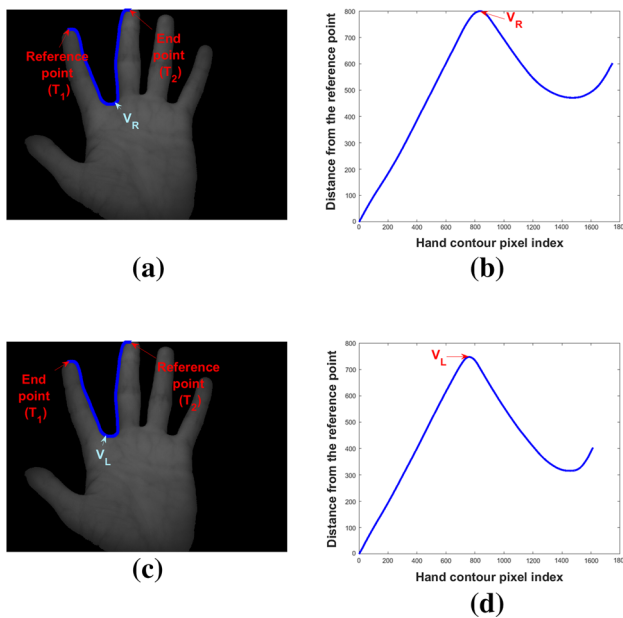In order to extract the palm-vein ROI, we locate the contour of the hand in the binarized image. (The contour is highlighted in blue in Fig. 2a.) A reference point is then selected near the base of the hand. Distance between the reference point and every point on the hand contour are computed and plotted against the hand contour pixel index starting at the reference point. The resulting radial distance function is shown in Fig. 2b. Peaks and valleys on this curve [37] correspond to the finger tips and finger valleys, respectively, as shown in Fig. 2c. Finally, the valley points closer to the index and little fingers are joined and at a perpendicular distance of x from the center of this line, a square region of size $L \times L$ corresponding to the palm-vein ROI is extracted (Fig. 2d).

3.1.2 Finger-vein ROI extraction

In this work, we supplement the palm-vein ROIs with the information extracted from the four finger-vein ROIs to improve the biometric recognition performance. We exclude the thumb since it is challenging to acquire the frontal images of the thumb as well as the four fingers as part of a single image. In order to segment the finger-vein ROIs from the binarized images, three points on each finger contour are identified including the fingertip denoted by T and two points on the base of the finger. As described in the previous section, in the process of palm-vein ROI extraction, the locations of fingertips and finger valleys on the hand contour are identified. However, on close observation, it can be observed that the valley between any two fingers is not a single point. Instead, it is a wider region comprising several points. For instance, Fig. 3a, c shows that the points V_L and V_R both lie in the same valley region between the index and middle
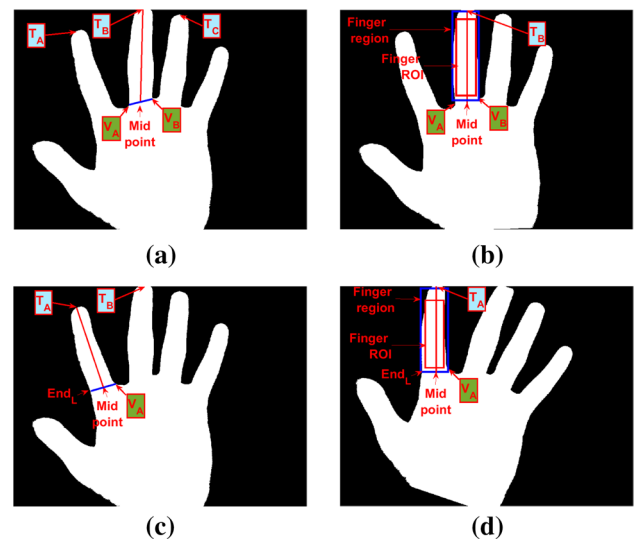
Fig. 2 Palm-vein ROI extraction **a** hand contour localization, **b** plot of radial distance function, **c** finger tips and valleys, **d** palm-vein ROI

Fig. 4 Finger-vein ROI extraction: for inner fingers (first row) and for outer fingers (second row)

fingertip as the end point. The valley point corresponds to the maximum on the radial distance curve between the reference and the end points. Refer Fig. 3 where point T_1 is considered to be the reference point to locate V_R, and similarly, T_2 is considered to be the reference point to locate V_L.

The inner fingers namely, middle and ring have two finger valleys V_A and V_B, to their left and right, respectively. V_A is located using radial distance curve with T_A and T_B as the reference and end points, respectively. Similarly, V_B is located using the points, T_B and T_C as the reference and end points, respectively, as shown in Fig. 4. The line joining the fingertip T_B with the midpoint of the line $V_A V_B$ acts as the finger axis. Figure 4b shows the rotated image with vertical finger axis. The outer blue rectangle represents the finger boundary, while the inner red rectangle corresponds to the ROI which is extracted by iteratively reducing the outer boundary from the lateral and upper ends, until the background is eliminated.

Outer fingers which include index and little fingers, have a single valley associated with them. In order to locate the other end of the finger contour, the contour length l between the valley V_A and the fingertip T_A is measured and a point End_L along the contour at a contour distance of l from T_A is identified, as shown in Fig. 4c. Once the two base points are located, we perform the same set of operations as for inner finger-vein ROI extraction. The finger-vein ROI for index finger is shown in Fig. 4d.

Fig. 3 Finger-vein ROI extraction **a** right valley V_R detection, **b** corresponding radial distance curve, **c** left valley V_L detection and **d** corresponding radial distance curve

fingers. They represent the leftmost and rightmost points of the valley region, respectively. We consider the valley point that lies closer to a finger as its base point, resulting in two base points per finger on its either side.

In order to locate the valley points, we employ the same approach as the one used for palm-vein ROI extraction; however, with two reference points. Specifically, for a region along the hand contour between any two consecutive fingertips T_1 and T_2 (highlighted in blue in Fig. 3a, c). One of the fingertips is considered as the reference point and the other

The extracted palm-vein and finger-vein ROIs are then enhanced using CLAHE and normalized to a size of 150×150 pixels and 150×30 pixels, respectively. Due to the variation in illumination in some of the images, Otsu's thresholding [36] does not find the optimum threshold to segment the hand image from the background. We discuss the treatment of such images in Sect. 4.

3.2 Feature Transformation

The enhanced and normalized palm-vein and finger-vein ROIs contain several curvilinear vessel structures of varying shapes and sizes. In order to effectively characterize these structures, we employ CS-LBP [34], a computationally efficient and illumination invariant feature representation used in the literature for image region description. It captures the image gradient information by comparing the gray levels of pixels in a local image neighborhood. The CS-LBP representation is derived from local binary patterns (LBP) [38] which is computed by assigning a binary code to each pixel of the image by thresholding the N neighborhood pixel intensities with the value of the center pixel and concatenating the N binary values to obtain a binary string. CS-LBP is a more compact form of LBP. Instead of comparing the intensities with the center pixel, pixels located symmetrically about the center are compared and thresholded based on a preset value T, which is typically set to 1% of the range of the pixel intensities. Specifically, we considered a neighborhood of 8 pixels ($N = 8$ and radius $R = 2$) centered at the current pixel. In particular, the difference between the pixel intensities of the ith and $(i + \frac{N}{2})$th pixels, denoted by g_i and $g_{i+(\frac{N}{2})}$, respectively, is considered and a binary code is assigned to each pixel in the enhanced ROI $I(x, y)$ based on Eq. 1. This generates the CS-LBP image which is an encoded template with decimal equivalent values ranging from 0 to 15 and is used for matching. Figure 5 shows the palm-vein and finger-vein ROIs and the corresponding CS-LBP and uniform LBP images for comparison. Although the CS- LBP and the uniform LBP extract texture features, the resultant images indicate that the two descriptors characterize the palm-vein image very differently.

$$CS - LBP_{R,N,T}(x, y) = \sum_{i=0}^{\left(\frac{N}{2}\right)-1} s\left(g_i - g_{i+\left(\frac{N}{2}\right)}\right)2^i,$$

where

$$s(p) = \begin{cases} 1, & \text{if } p > T \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

3.3 Matching

Subsequently, in order to compare the gallery (I_g) and probe (I_p) feature templates obtained in the previous step, we employ a robust image matching technique, known as deep matching [35]. It performs non-rigid dense matching between the images in two stages namely, bottom-up and top-down. The bottom-up stage consists of a pipeline similar to the architecture of DCNNs. In particular, it contains the convolutional layer along with max-pooling, subsampling and rectification layers. However, the weights of the filters used
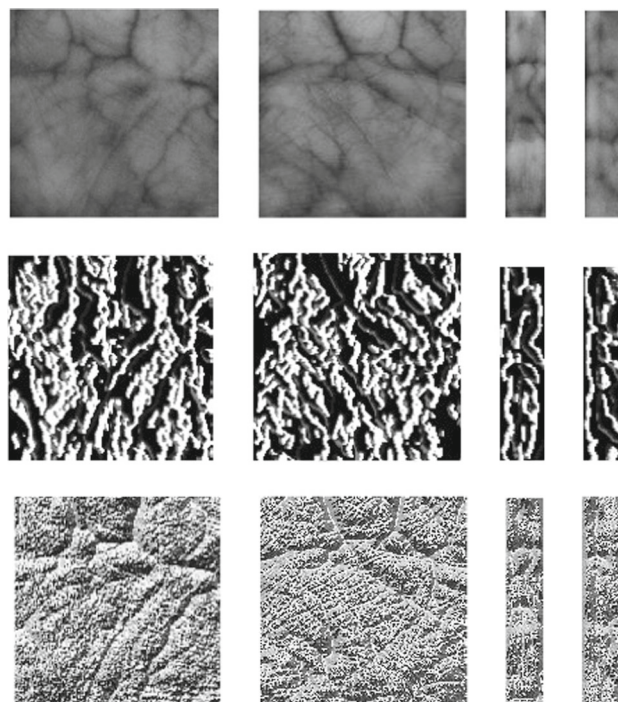

Fig. 5 Illustration of textural differences between CS-LBP and LBP images: palm-vein and finger-vein ROIs (first row), corresponding CS-LBP images (second row) and uniform LBP images (third row)

for convolution operations are not learnt as in DCNNs; instead, the image patches of the first image act as convolution filters for the second image. Specifically, I_g is partitioned into non-overlapping patches of size $n \times n$ where $n = 4$. These patches act as convolution filters which when convolved with I_p result in level-1 correlation maps which form the bottommost layer of the correlation map pyramid. These correlation maps represent the similarity scores between the image patches. Let $C_{N,p}$ denote the correlation maps created using image patches of I_g having size $N \times N$ and centered at p. Further, let $C_{N,p}(p')$ denote the similarity between patch of I_g centered at p with the patch of I_p centered at p'.

Further, in order to perform comparison at multiple image scales, recursive aggregation of correlation maps is performed in a bottom-up manner. A multilevel correlation pyramid is formed as shown in Fig. 6. Specifically, four smaller quadrants in the correlation maps for image patches of size $\frac{N}{2} \times \frac{N}{2}$ in the lower layer are aggregated to form correlation maps corresponding to larger image patches of size $N \times N$. The aggregation involves the operations as such as max-pooling, sub- sampling, shift and nonlinear rectification as in a deep convolutional framework . Furthermore, the four children quadrants ($C_{\frac{N}{2},p_i}$) with centers p_i for $i = 1...4$ are allowed to move in the local neighborhood. Max-pooling and subsampling denoted by $P \circ S$ are performed on the children quadrants to select the best match using kernel size of 3×3 and stride of 2. Subsequently, a translation of 1 pixel
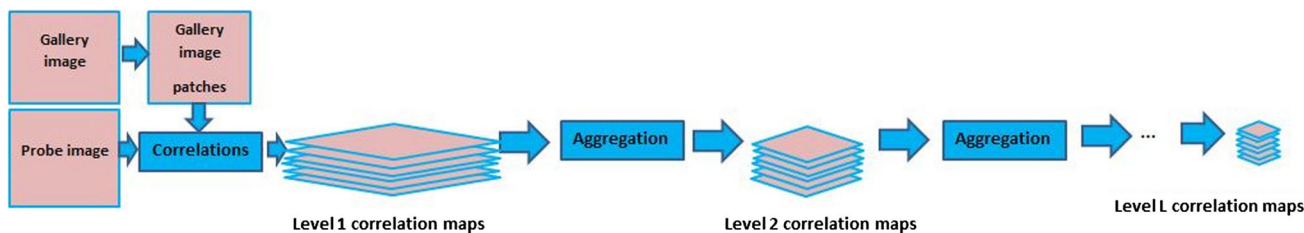
Fig. 6 Bottom-up stage of deep matching inspired by the architecture of DCNNs: building the correlation map pyramid
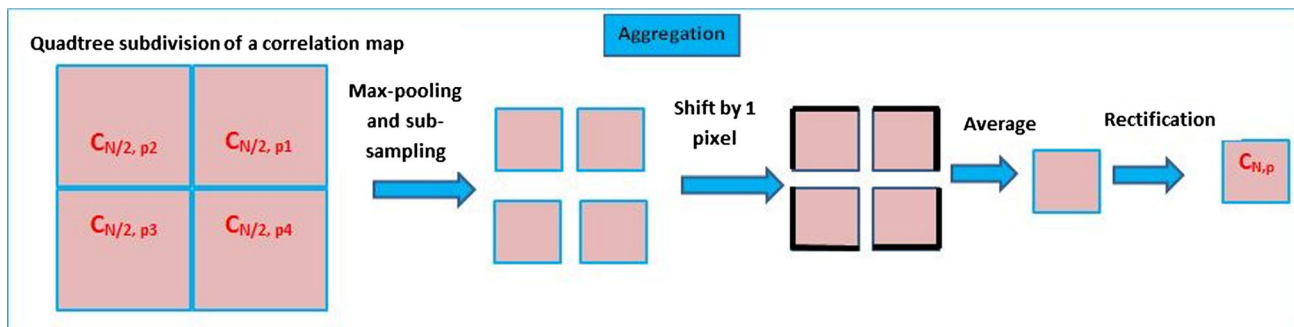

Fig. 7 Aggregation performed at each layer in the bottom-up stage

and averaging is performed followed by nonlinear rectification for contrast enhancement of the correlation maps. The parameters used in this work for all the aggregation tasks are the same as in [35]. Figure 7 shows the sequence of aggregation operations performed on the correlation maps resulting in the correlation map pyramid.

The final matching score between gallery and probe image is computed by backtracking along the correlation map pyramid constructed in the bottom-up manner. This process is shown in Fig. 8. The entry point is determined based on the maxima in the topmost pyramid level. Further, Eq. 2 is employed recursively to undo the aggregation operations to identify the matching patches. Finally, the matching score is defined as the number of matches that fulfill the deformation tolerance constraints, which include scaling factors between $[\frac{1}{2}, \frac{3}{2}]$ and the value of rotation angles between $[-30, 30]$ degrees.

$$d_i = \arg \max_{d \in \{-1,0,1\}^2} C_{\frac{N}{2}, p_i(p'_i + d)} \qquad (2)$$

where d denotes the local neighborhood of p'_i and d_i corresponds to the quadrant with the highest value.

3.4 Score-level fusion

Subsequently, in order to supplement the palm-vein biometric trait with the additional information from finger-vein images, matching score-level fusion is performed using the weighted sum rule. Prior to fusion, the matching scores are normalized using the Tanh normalization [39].

4 Experiments and results

In order to evaluate the effectiveness of the proposed approach, we performed extensive experiments on two datasets, considering verification and identification scenarios. In particular, we have performed three sets of experiments. In the first set of experiments, we have compared the performance of the proposed approach with several existing feature extraction techniques that have been shown to be effective for palm-vein- and finger-vein-based biometric recognition systems in the literature. In the second set of experiments, we assessed the improvement in performance of the palm-vein biometric system on fusion with the finger-vein images at the matching score level. Additionally, in the third set of experiments, we have compared the performance of the proposed approach with a related work that performs matching of the entire hand-vein images [40]. In this section, we describe the datasets used in this work and discuss the results obtained from the three sets of experiments.

4.1 Database description

There are three publicly available datasets for research on palm-vein biometrics namely, The Hong Kong Polytechnic University Multispectral Palmprint Database (PolyU database) [41], the Idiap Research Institute VERA Palmvein Database (VERA palmvein dataset) [42] and CASIA Multispectral Palmprint Image Database (CASIA dataset) [43]. Since our approach utilizes vascular information from the entire hand, we required hand-vein images having full frontal
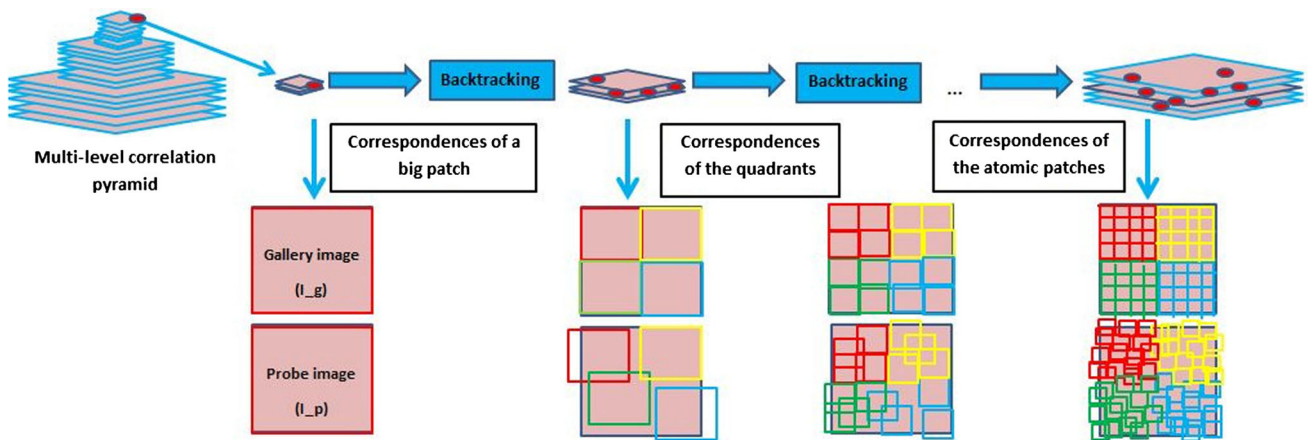
Fig. 8 Top-down stage of deep matching: finding the matching correspondences in the images
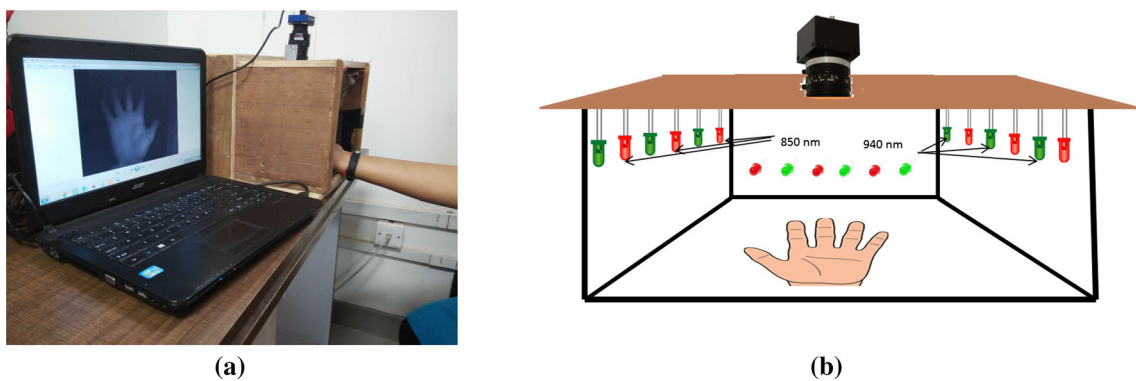

(a) **(b)**

Fig. 9 Imaging setup employed for hand-vein image acquisition

hand coverage with complete palm and finger areas. Since the hand-vein images in the PolyU and VERA palm-vein datasets contain only partial finger-vein areas, these datasets are not suitable for our study. On the other hand, the images in the CASIA dataset cover the whole hand area. Thus, we used the CASIA dataset for evaluating our approach. This dataset consists of 6 images per hand acquired from left and right hands of 100 subjects. In addition to this, we collected a larger in-house dataset (IITI hand-vein dataset), consisting of 6 samples per hand acquired from left and right hands of 185 subjects.

4.1.1 Imaging setup

There are two modes in which NIR illumination can be employed for image acquisition namely, the transmission and the reflection modes. In the transmission mode, NIR light is transmitted through the biometric sample and the biometric capture device is on the other side of the light source, as in most of the finger-vein acquisition devices [43]. On the other hand, in the reflection mode, both the light source and camera are placed on the same side of the biometric sample, as
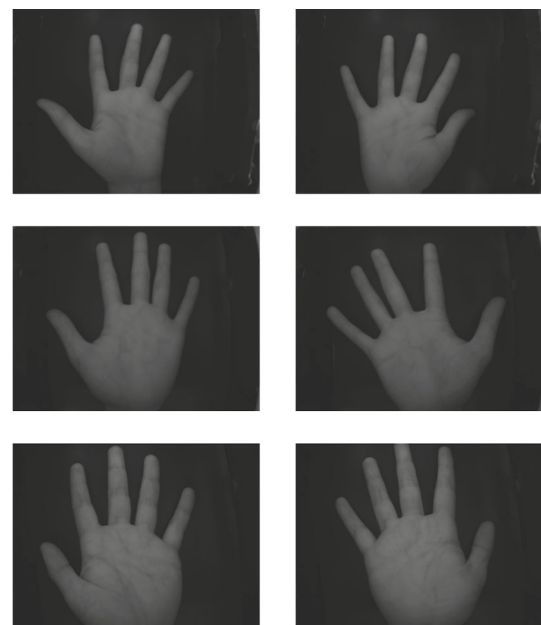

Fig. 10 Hand-vein images of left and right hands of three subjects from the IITI hand-vein dataset
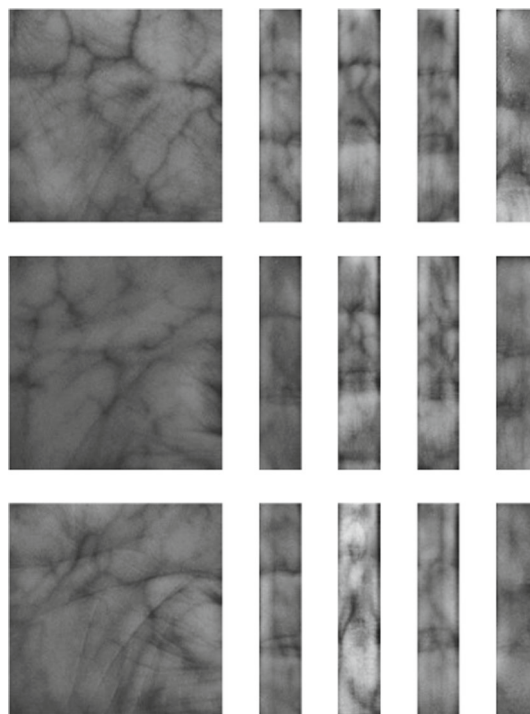
Fig. 11 Two binary images showing the segmentation failure cases in our database

Fig. 12 Palm-vein and finger-vein ROIs extracted from the hand-vein images of 3 subjects from the IITI hand-vein database

in case of palm-vein image acquisition [18]. In this work, we acquired the hand-vein images using the latter scheme (Fig. 9). The imaging setup used for collecting the database consists of a wooden box, covered from all sides except the front side through which the subject places his hand on the floor of the box, as shown in Fig. 9a. An NIR sensor (mvBluefox IGC) mounted on the top of the box was connected to a laptop in order to acquire the images of the hand. Infrared LEDs having peak wavelengths of 850 and 940 nm were placed alternately to form arrays of LEDs. Specifically, three arrays of LEDs were employed. Two arrays were placed on the box ceiling to illuminate the hand, and the third array was placed at a lower elevation on the front wall of the box in order to illuminate the finger regions. The NIR illumination arrangement employed for image acquisition is shown in Fig. 9b. Hand images were collected from the left and right hands of 185 subjects in an unconstrained and contactless manner. Some of the images collected using our setup are shown in Fig. 10. The vein patterns in palm-vein images are denser compared to the finger-vein images and are more clearly visible. Nevertheless, the information extracted from the finger-veins can be employed to enhance the performance of the palm-vein biometric system, as shall be demonstrated through the experiments.

4.1.2 ROI extraction

For some of the hand-vein images in the database, automatic hand segmentation using the threshold determined by Otsu's method did not produce the desired results. It led to either inclusion of parts of the background or exclusion of parts of the hand-vein region, mostly the finger regions (Fig. 11). This may be due to adverse effect of non-uniform illumination in some of the samples. Besides, since the images were acquired in an unconstrained manner, fingers in some of the images were not well separated, leading to incorrect hand contour localization. As a result, subsequent employment of radial distance curve for identifying the finger tips and valleys did not work correctly. Figure 11 shows two examples of failed ROI extraction in which finger-vein regions are not segmented correctly. Overall 4.12% of the total number of images had incorrect hand segmentation. In all such cases,

we extracted the ROIs manually in order to evaluate the performance of the proposed approach on the entire dataset. The ROIs (palm-vein and finger-vein) extracted from the images are shown in Fig. 12.

4.2 Performance evaluation in unimodal framework

In the first set of experiments, we performed a comparative evaluation of the proposed approach with several existing feature extraction techniques that have been shown to be effective for palm-vein and finger-vein biometrics in the literature. In particular, we assessed the performance for palm-vein and finger-vein modalities individually (without score-level fusion). For comparison, we employed neighborhood matching radon transform (NMRT) [20] and Hessian phase [20] approaches, which have yielded promising results for palm-vein biometrics. Additionally, we employed the maximum principal curvature (MPC) [44], wide line descriptor (WLD) [26] and repeated line tracking (RLT) [25] techniques, which have been shown to be effective for finger-vein biometrics. Table 1 shows the values selected for the parameters used in these feature extraction techniques. Figures 13 and 14 show the feature templates extracted from palm-vein and finger-vein ROIs, respectively, using the above-mentioned and CS-LBP feature extraction techniques. These encoded feature templates created using the above-mentioned techniques were used for matching. In order to

Table 1 Parameter settings for feature extraction using the existing techniques

| Technique	Parameter	Value	
		Palm-vein	Finger-vein
MPC [44]	Sigma, σ	3	3
WLD [26]	Radius, r	9	9
	Threshold, t	1	1
	Global threshold, g	100	100
RLT [25]	Number of iterations, N	3000	3000
	Cross-sectional width, W	11	11
	Radius, r	1	1
NMRT [20]	Lattice size, L	16	16
Hessian phase [20]	Sigma, σ	0.5 0.8, 1.2 1.8 2.5	0.5 0.8, 1.2 1.8 2.5

compute the matching scores between the biometric feature templates extracted using MPC, WLD and RLT techniques, we adopted the maximal matched pixel ratio (MPR) as a measure of similarity. In order to perform translation and rotation invariant matching, feature templates F_g and F_p extracted from the gallery and probe images, respectively, were compared. In particular, F_g was compared with F_p and its translated and rotated versions which constitute F_p' with the translation offset of 10 pixels and rotation offset of 4°, both with increments of 2. For each pair of F_g and a template in F_p', ratio of the number of matched pixels to the total number of pixels in the feature templates was computed and the maximum value corresponds to MPR.

$$MPR(I_g, I_p) = \max_{F \in F_{p'}} \frac{\sum_{x,y} 2 * F_g(x, y) * F(x, y)}{\sum_{x,y} (F_g(x, y) + F(x, y))} \quad (3)$$

In case of NMRT- and Hessian phase-based approaches, we adopted the patch-based hamming distance for matching as proposed by the authors in [20]. Overlapping patches of size 25×25 pixels with an overlap of 3 pixels between the patches were considered, and the hamming distances for the corresponding patches were summed up to form the final dissimilarity score. The results of the experiments in verification and identification scenarios are presented in the following sections.

4.2.1 Verification experiments

In order to perform experiments, we considered the left- and right- hand samples from a subject as belonging to separate individuals, effectively resulting in 200 and 370 subjects in CASIA and IITI datasets, respectively. Thus, there are six hand images per subject in both IITI and CASIA datasets. Each of these images yields a palm-vein and four finger-vein images. For each biometric modality, we considered the first three samples of each subject as gallery and the remaining as probe. Matching scores between the gallery and probe
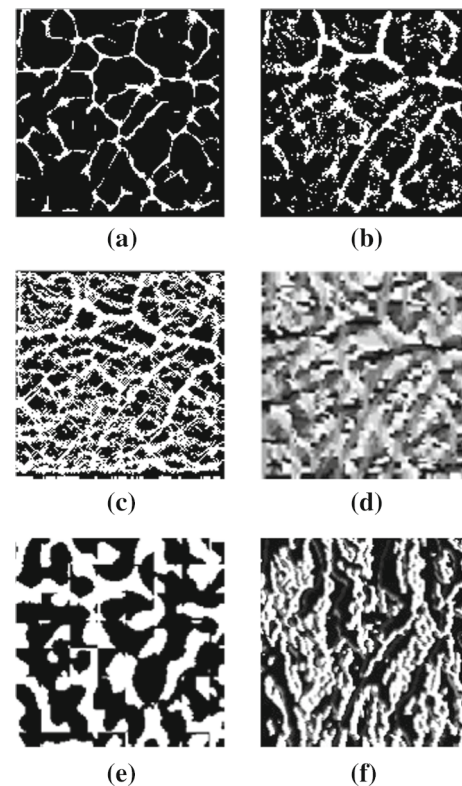
Fig. 13 Feature templates extracted from the palm-vein ROIs using **a** MPC, **b** WLD, **c** RLT, **d** NMRT, **e** Hessian phase and **f** CS-LBP
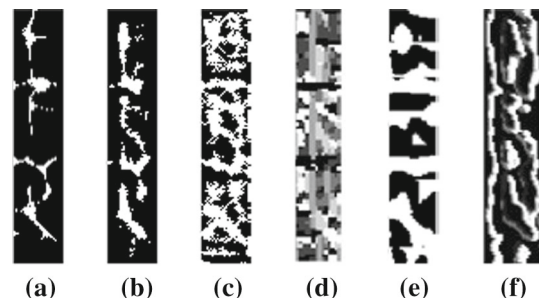
Fig. 14 Feature templates extracted from the finger-vein ROIs using **a** MPC, **b** WLD, **c** RLT, **d** NMRT, **e** Hessian phase and **f** CS-LBP

Table 2 Performance of the individual modalities in the verification framework

| Approach | EER (%) | | | | | | | | | |
| | IITI | | | | | CASIA | | | | |
	Palm	Index	Ring	Middle	Little	Palm	Index	Ring	Middle	Little
WLD [26]	3.25	6.01	6.51	5.10	10.19	6.08	14.08	12.98	13.99	19.14
MPC [44]	2.47	4.69	5.22	4.19	7.29	4.54	12.06	12.59	11.55	15.96
RLT [25]	8.96	8.16	12.37	7.64	20.14	17.17	17.17	28.79	28.92	30.87
NMRT [20]	6.18	6.76	6.78	6.38	9.44	9.57	19.17	17.72	19.33	20.68
Hessian phase [20]	7.16	9.16	9.73	8.98	13.49	9.39	22.66	21.26	23.23	22.91
Proposed	**0.40**	**2.54**	**3.12**	**2.98**	**4.12**	**2.61**	**6.24**	**7.23**	**6.23**	**10.97**

Bold indicates the best performance

Table 3 Performance of the individual modalities in the identification framework

| Approach | Rank-1 identification rate (%) | | | | | | | | | |
| | IITI | | | | | CASIA | | | | |
	Palm	Index	Ring	Middle	Little	Palm	Index	Ring	Middle	Little
WLD [26]	98.83	97.39	97.20	97.12	94.50	97.50	90.50	88.83	91.33	84.17
MPC [44]	99.19	97.38	97.03	97.03	94.50	97.17	89.00	89.83	90.50	81.67
RLT [25]	93.51	95.76	87.39	95.86	73.60	83.00	76.17	59.50	61.33	51.54
NMRT [20]	94.86	94.68	94.32	94.50	91.08	93.50	74.83	73.33	74.67	63.33
Hessian phase [20]	93.42	90.18	90.45	91.26	86.39	92.33	65.50	64.83	64.17	57.00
Proposed	**100**	**98.28**	**98.19**	**98.29**	**97.20**	**98.00**	**91.33**	**94.50**	**94.00**	**85.30**

Bold indicates the best performance

samples were computed, resulting in 1800 and 3330 genuine scores as well as 358,200 and 1,228,770 impostor scores for CASIA and IITI datasets, respectively. Based on these scores, two error rates, false match rate (FMR) and false non-match rate (FNMR), were computed. The decision threshold was determined to be the operating point at which FMR equaled FNMR. Finally, the equal error rates (EERs) for palm-vein as well as four finger-vein modalities are reported in Table 2. The EER values corresponding to the best performance for each modality have been highlighted in boldface.

Two observations can be made from the results presented in Table 2. Firstly, the proposed technique clearly outperforms the other techniques with minimum EER values for palm-vein and finger-vein modalities on both the datasets. EER values of 0.40 and 2.61% were obtained for palm-vein matching on IITI and CASIA datasets, respectively. These values are lower than the EER values achieved by the other techniques. Secondly, the performance of palm-vein modality was better than that of finger-vein modalities. This may be attributed to the fact that the features extracted from the palm-vein are more discriminatory compared to the features extracted from the finger-vein modalities owing to the latter's sparser vascular networks. Among the finger-vein modalities, the best EER values of 2.54 and 6.23% were achieved by the index and middle fingers of IITI and CASIA datasets, respec-

tively. Both of these values are higher than the EER values obtained for palm-vein. The little finger performs poorly, as it contains the least amount of vascular information. It may also be noted that the performance of all the approaches is better on IITI dataset as compared to the CASIA dataset. This could be due to the higher degree of hand-pose variations in terms of rotation and scale in CASIA hand images. Nevertheless, the merit of the proposed technique is evident since the results obtained are consistently better than the other techniques for each modality on both datasets.

4.2.2 Identification experiments

The rank-1 identification rates obtained for each of the modalities are reported in Table 3. The cumulative match characteristic (CMC) curves on the IITI and CASIA datasets are presented in Figs. 15 and 16, respectively. The results presented in the identification scenario are consistent with those achieved in the verification scenario, in which it was found that the proposed technique consistently outperforms the other techniques considered in this study. Rank-1 identification rates greater than 97% were achieved for all the modalities on IITI dataset with the best performance of 100% for palm-vein and the worst performance of 97.20% for little finger-vein. On the CASIA dataset, palm-vein achieves
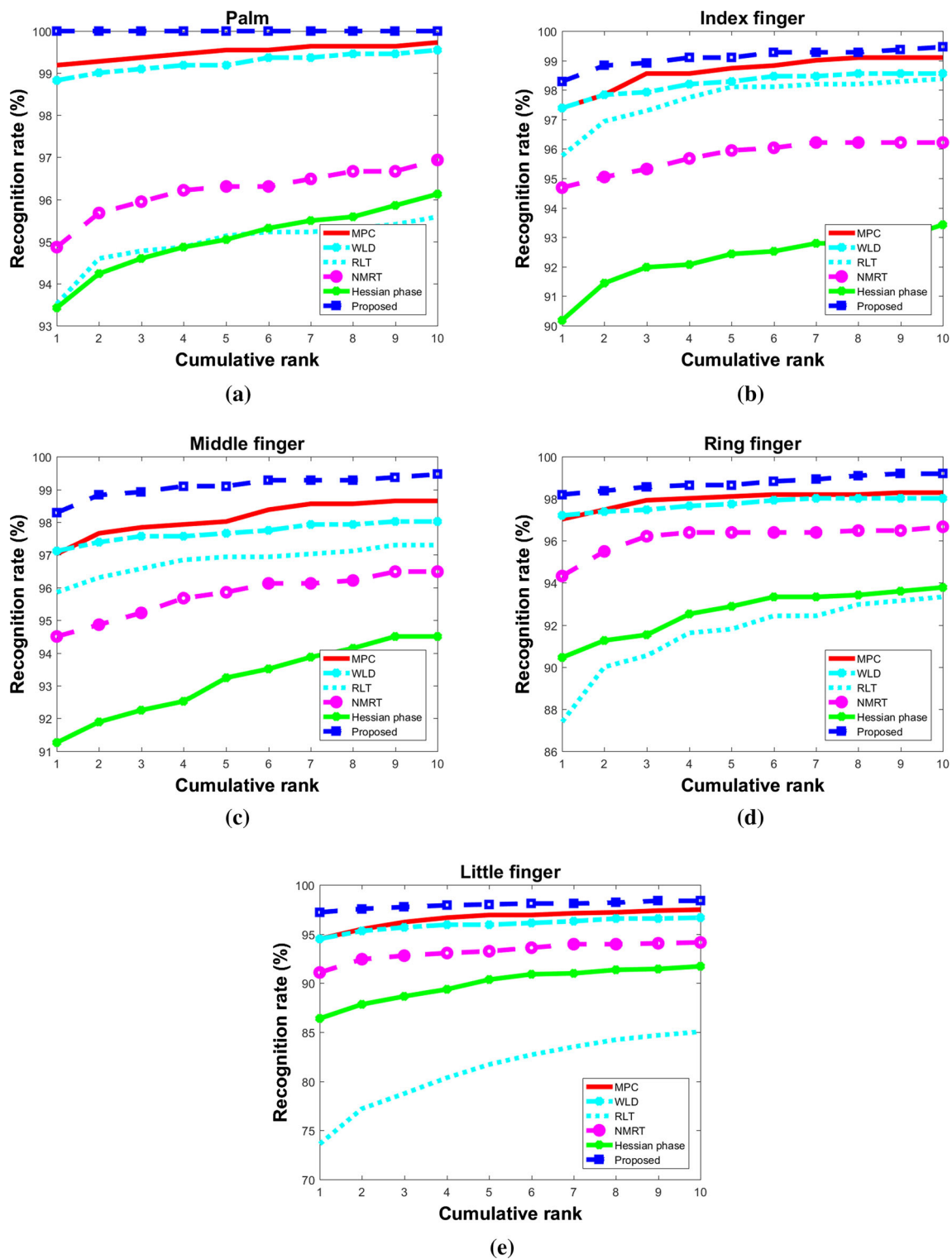
Fig. 15 Performance of the individual modalities in the identification framework on IITI hand-vein dataset

an accuracy of 98%, while the best performance among the finger-veins was achieved by the middle finger with a recognition rate of 94.50%. Next, we present the results in the multimodal framework for biometric recognition in which palm-vein and finger-vein modalities are combined.
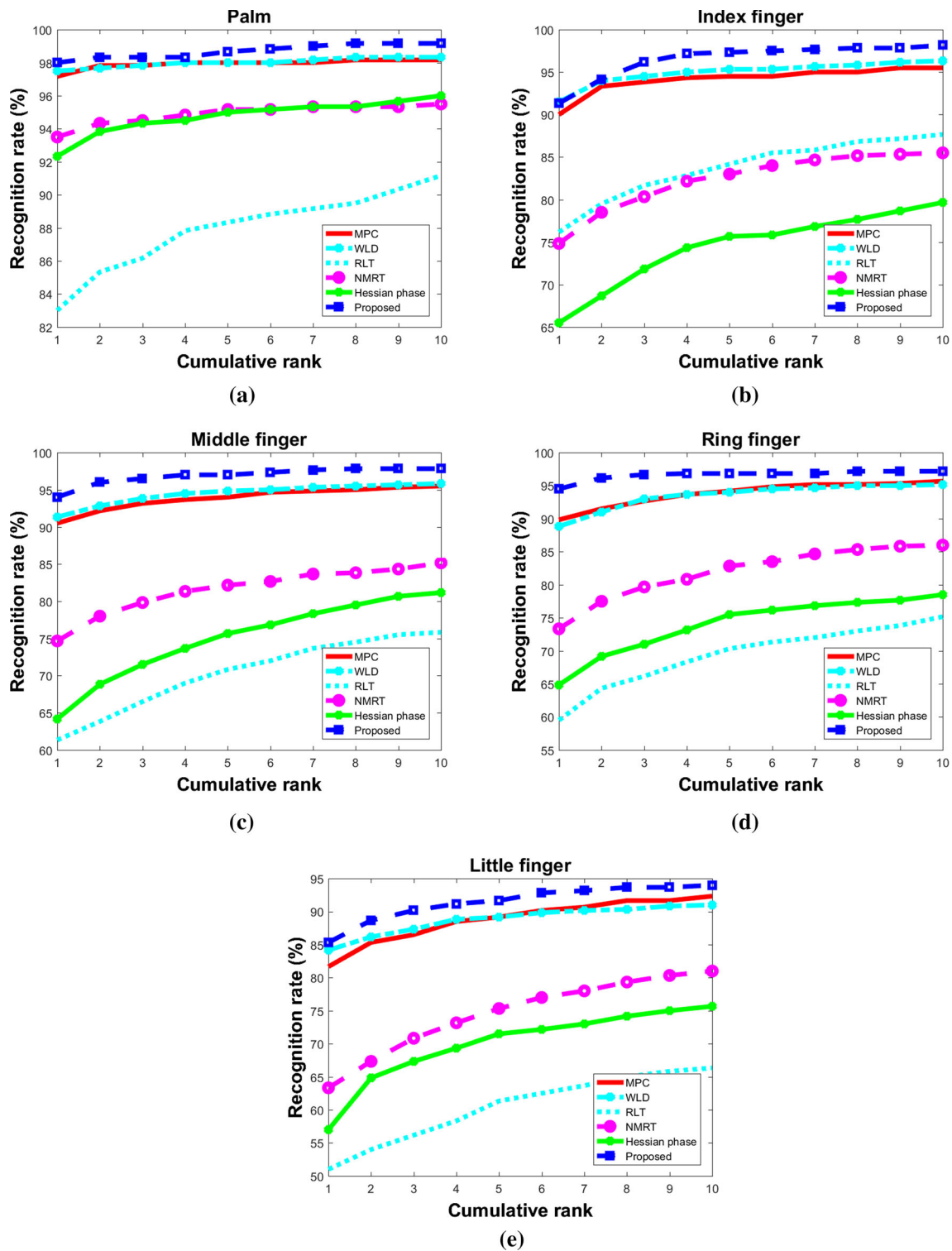
Fig. 16 Performance of the individual modalities in the identification framework on CASIA dataset

4.3 Performance evaluation in the multimodal framework

The second set of experiments were performed in order to analyze the performance upon fusion of the palm-vein and finger-vein modalities. In particular, we assessed the effect of augmenting the vascular information present in the palm-vein with that in the finger-veins at the matching score level for all the techniques considered in the previous section. All possible combinations of finger-veins were considered for

Table 4 Recognition performance on IITI and CASIA datasets upon matching score-level fusion in the verification scenario

Dataset	EER (%)									
	IITI					CASIA				
Approach	P	P+1F	P+2F	P+3F	P+4F	P	P+1F	P+2F	P+3F	P+4F
WLD [26]	3.25	2.27	2.37	**2.25**	2.34	6.08	6.11	5.43	6.15	**5.23**
MPC [44]	2.47	2.52	1.19	0.90	**0.73**	4.54	4.92	3.59	**3.26**	3.32
RLT [25]	8.81	8.79	6.79	**6.48**	8.79	17.17	18.56	16.23	14.44	**13.98**
NMRT [20]	6.18	4.12	3.09	2.61	**2.25**	9.58	7.88	6.58	6.08	**6.06**
Hessian phase [20]	7.16	6.10	4.97	4.17	**3.91**	9.39	9.09	8.27	7.60	**7.24**
Proposed	0.40	0.39	0.30	0.17	**0.13**	2.61	1.98	1.68	1.37	**1.21**

Bold indicates the best performance

Table 5 Rank-1 identification rates for IITI and CASIA datasets upon matching score-level fusion in the identification scenario

Dataset	Rank-1 identification rate (%)									
	IITI					CASIA				
Approach	P	P+1F	P+2F	P+3F	P+4F	P	P+1F	P+2F	P+3F	P+4F
WLD [26]	98.83	99.00	99.17	99.28	**99.50**	97.5	97.5	97.64	98.12	**98.37**
MPC [44]	99.19	99.09	99.32	99.55	**99.72**	97.17	98.33	98.08	98.25	**98.50**
RLT [25]	93.51	96.13	97.46	98.04	**98.38**	83.85	74.45	78.5	81.91	**84.17**
NMRT [20]	94.86	97.00	98.26	98.69	**99.00**	93.50	94.29	95.86	96.46	**96.83**
Hessian phase [20]	93.42	94.59	95.84	96.50	**97.39**	92.33	93.13	94.03	94.67	**95.67**
Proposed	100	100	100	100	**100**	98.00	98.83	99.47	99.83	**100**

Bold indicates the best performance

fusion with the palm-vein. P+nF denotes the combination of the palm-vein with n finger-veins where $n \in [1, 2, 3, 4]$. The score-level fusion of the normalized scores was performed using the weighted sum rule. We present the performance evaluation in the verification and identification scenarios in the following sections.

4.3.1 Verification experiments

Table 4 shows the average EER values for each of the P+nF cases on IITI and CASIA datasets. It can be observed that for all the techniques, including the proposed technique, the multimodal case achieves better performance than the unimodal (palm-vein only) case. This shows that the information from finger-vein images supplements the palm-vein images and enhances the biometric performance. As seen in the previous experiment, the finger-vein modality by itself is not a very rich source of vein patterns. Nevertheless, fusing the finger-vein information with the palm-vein leads to better overall performance. It can also be observed that among all the techniques listed in the table, the proposed technique is the best and achieves an EER of 0.13% when the scores from all the finger-veins were fused with that of the palm-vein. The best fusion performance among all the combinations (P+nF), for every technique, is highlighted in boldface. The results suggest that the best values were obtained for a higher value of n, i.e., when either three or four finger-veins were fused with the palm-vein. The proposed approach achieves a per-

formance improvement of 67.5%, from 0.40% when only palm-vein is employed to 0.13% on fusion of all finger-veins with the palm-vein (P+4F). Similarly, the average improvement of 45.62% and 30.82% was obtained on IITI and CASIA datasets, respectively.

4.3.2 Identification experiments

Table 5 shows the average rank-1 identification rates achieved on fusion of the finger-veins with the palm-vein at the matching score level. It can be seen from the table that there is definite performance improvement when all finger-veins were used for fusion, for every technique considered in this study. Figures 17 and 18 exhibit an ascending trend in recognition rate as n increases. In particular, the rank-1 and rank-10 identification rates are higher for $n = 3$ or 4 as compared to the palm-vein-only case (P) on IITI and CASIA datasets. It may also be noted that the identification rate is 100% on both IITI and CASIA datasets upon score-level fusion for the proposed method. Thus, it can be concluded from the results presented in this section that the fusion of information leads to improved performance (Fig. 19).

4.4 Comparison with existing approach

In the previous experiment, we observed that supplementing the palm-vein with the finger-veins leads to increased performance. Another way of utilizing the vascular infor-
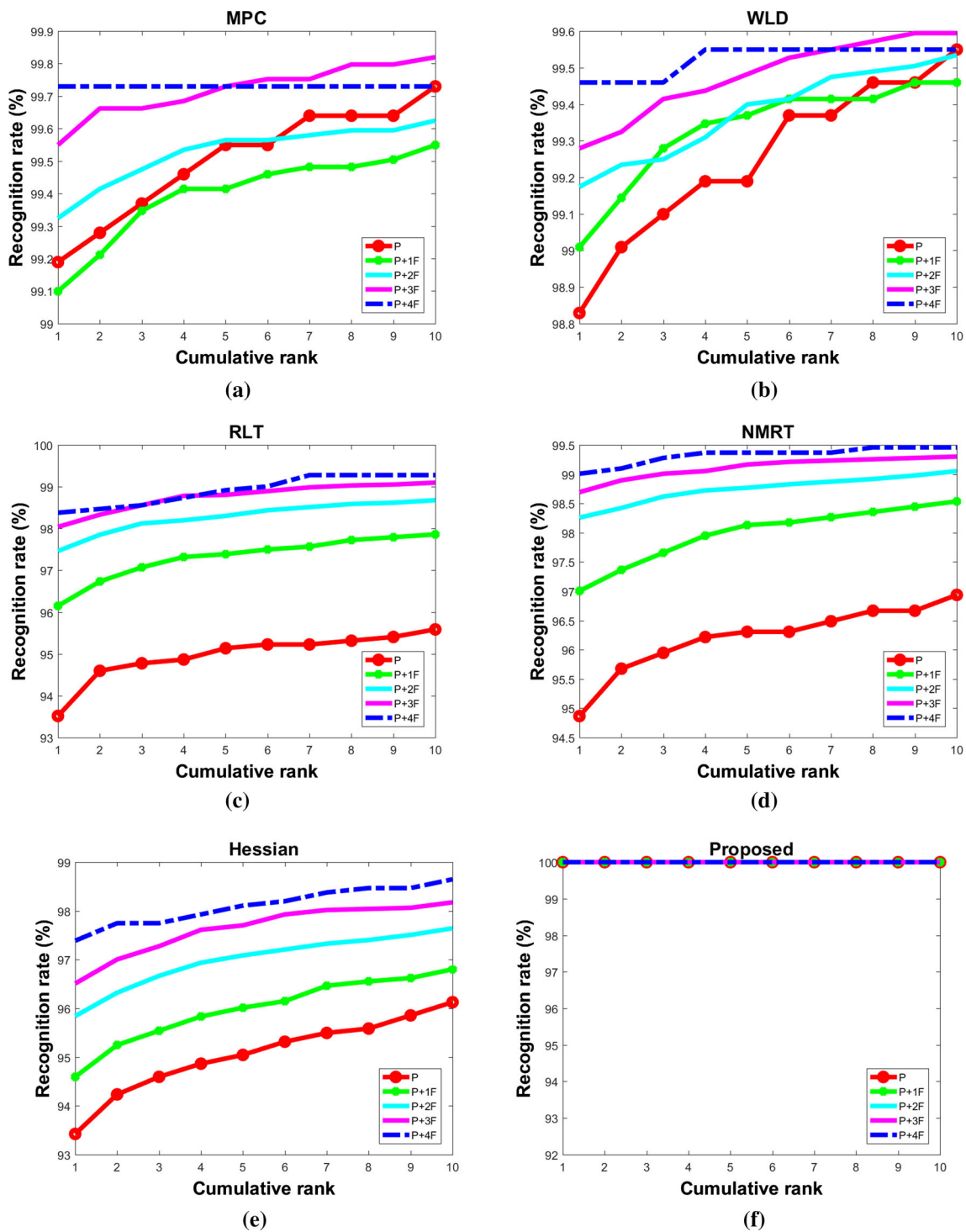
Fig. 17 Cumulative match characteristics showing the improvement in identification performance on IITI dataset when fusion is performed with finger-vein modalities using **a** MPC, **b** WLD, **c** RLT, **d** NMRT, **e** Hessian phase and **f** proposed approach

mation present in the finger-vein regions is by considering the whole hand-vein image at once instead of separating them into palm and finger regions, as was done in [40]. In this section, we compare the performance of the proposed approach with the multimodal framework presented in [40].

The hand-vein images were enhanced using the difference of Gaussian technique followed by histogram equalization. On the enhanced hand-vein images, RootSIFT algorithm [44] was applied for feature extraction and matching. In order to increase the accuracy of keypoint matching, neighbor-
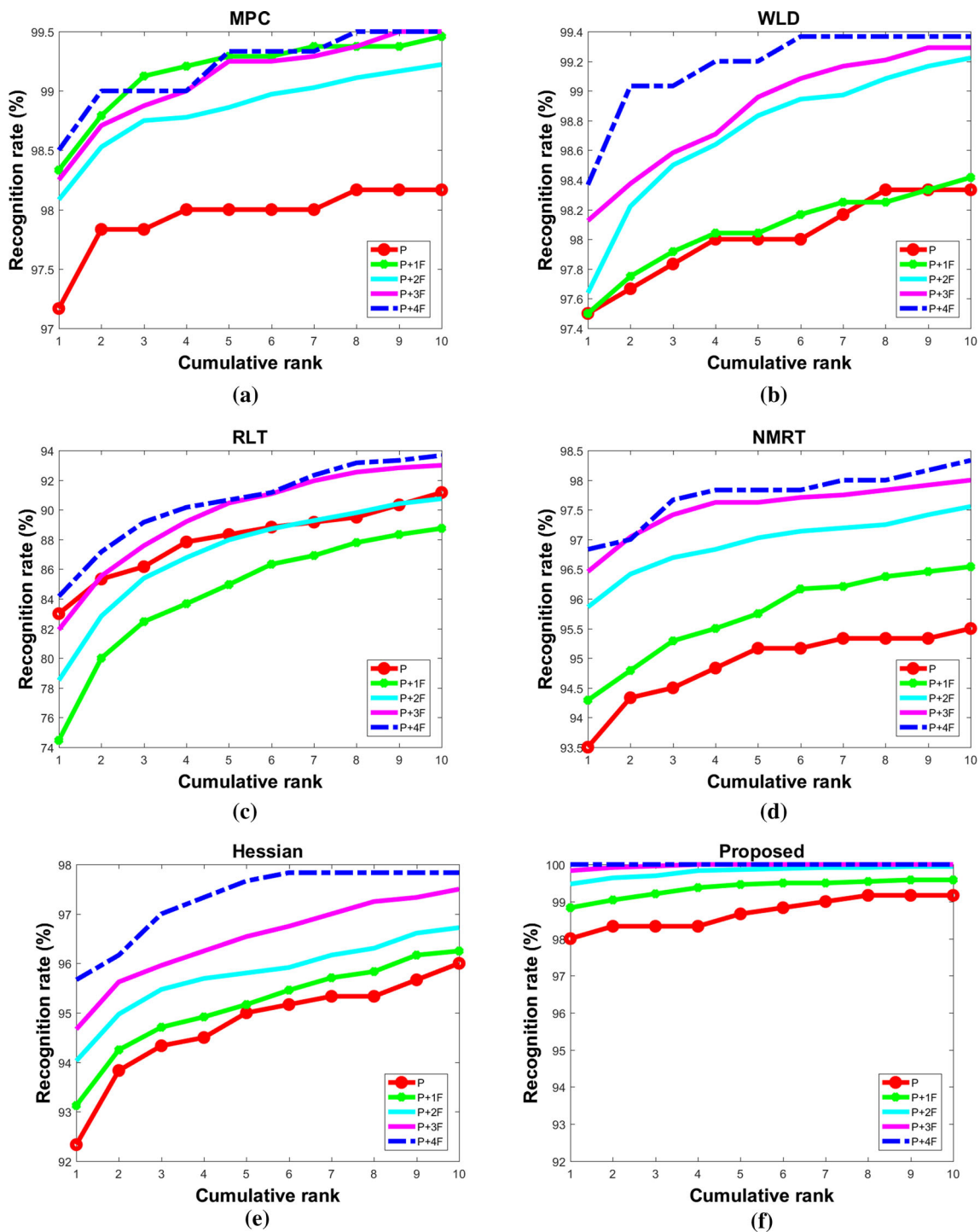
Fig. 18 Cumulative match characteristics showing the improvement in identification performance on CASIA dataset when fusion is performed with finger-vein modalities using **a** MPC, **b** WLD, **c** RLT, **d** NMRT, **e** Hessian phase and **f** proposed approach

hood and LBP-based mismatch removal techniques [44] were employed. Finally, the number of matches was considered to be the similarity score. Table 6 shows the EER and rank-1 identification rates of this set of experiments. The results suggest that the proposed approach achieves better performance than the approach in [40].
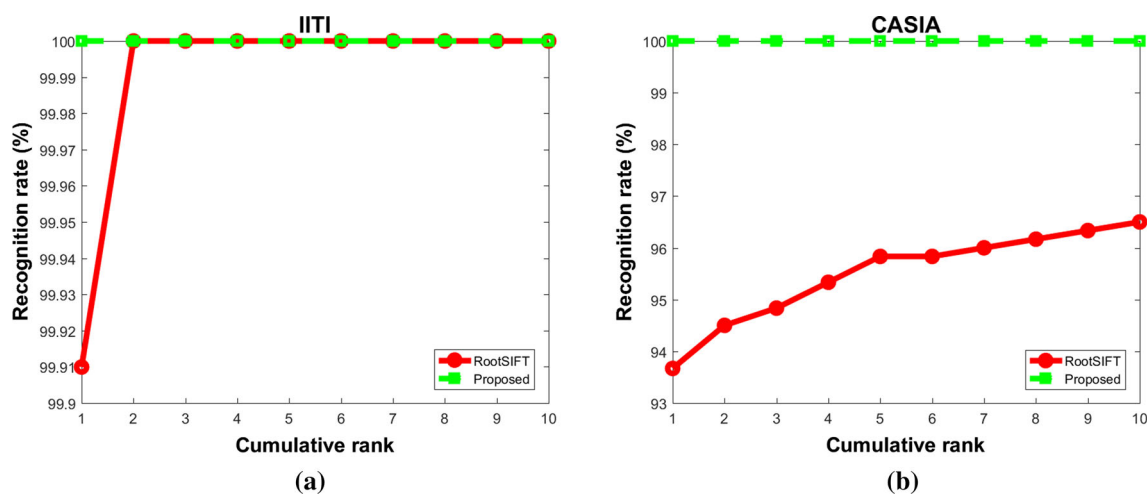
Fig. 19 Cumulative match characteristics showing the recognition performance of the proposed approach and the approach in [40] on **a** IITI and **b** CASIA datasets

Table 6 Comparison with an existing approach

Approach	EER (%)		Rank-1 identification rate	
	IITI	CASIA	IITI	CASIA
RootSIFT [39]	0.67	2.29	99.90	93.67
Proposed	**0.13**	**1.21**	**100**	**100**

5 Conclusion

Multimodal biometric systems have been shown to provide reliable biometric performance. In this paper, we have presented a single-sensor multimodal approach for hand-vein-based biometric recognition. From the acquired hand-vein image, the constituent palm-vein and four finger-vein ROIs were extracted. After the preprocessing center-symmetric local binary pattern (CS-LBP) technique was employed to transform the extracted ROIs into CS-LBP image representations. Subsequently, a hierarchical non-rigid matching technique inspired by the architecture of deep convolutional neural networks was employed for matching the CS-LBP images. Finally, in order to utilize the information present in the finger-vein regions, weighted sum rule-based matching score-level fusion was performed to combine the palm-vein and the four finger-veins. An in-house database was collected from the left and right hands of 185 subjects. In order to evaluate the proposed approach, three sets of experiments were performed on the in-house and the publicly available CASIA datasets. The results of the first experiment suggest that the proposed approach consistently achieves better performance than the existing techniques considered in this study. In particular, the proposed approach achieves equal error rates of 0.13 and 1.21%, and rank-1 identification rates of 100 and 100% on the in-house and CASIA datasets, respectively.

Results of the second set of experiments suggest that there is considerable performance improvement when the palm-vein is supplemented by finger-veins for all the feature extraction techniques considered in this work. Further, we compared our technique with an existing technique that extracts features from the entire hand-vein image. The experimental results suggest that considering the constituent palm-vein and finger-vein images instead of the entire hand-vein image achieves better recognition performance. As part of the future work, we plan to utilize hand geometry features to further improve the recognition performance of the proposed system. The in-house database and the extracted palm-vein and finger-vein ROIs will be made publicly available on http://www.iiti.ac.in/people/~kvivek/index.html to further the research in this area.

References

1. Brunelli, R., Falavigna, D.: Person identification using multiple cues. IEEE Trans. Pattern Anal. Mach. Intell. **17**(10), 955–966 (1995)
2. Prakash, S., Gupta, P.: Fusion of ear with other traits. In: Ear Biometrics in 2D and 3D 107-116 (2015)
3. Shekhar, S., Patel, V.M., Nasrabadi, N.M., Chellappa, R.: Joint sparse representation for robust multimodal biometrics recognition. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 113–126 (2014)
4. Yang, J., Zhang, X.: Feature-level fusion of fingerprint and finger-vein for personal identification. Pattern Recogn. Lett. **33**(5), 623–628 (2012)
5. Wang, J.G., Yau, W.Y., Suwandy, A., Sung, E.: Person recognition by fusing palmprint and palm vein images based on laplacianpalm representation. Pattern Recogn. **41**(5), 1514–1527 (2008)

6. Yang, W., Huang, X., Zhou, F., Liao, Q.: Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion. Inf. Sci. **268**, 20–32 (2014)

7. Ramalho, S.M., Correia, P.L., Soares, L.D., et al.: Biometric identification through palm and dorsal hand vein patterns. In: EUROCON International Conference on Computer as a Tool, pp. 1–4 (2011)

8. Heenaye, M., Khan, M.: A multimodal hand vein biometric based on score level fusion. Proc. Eng. **41**, 897–903 (2012)

9. Veluchamy, S., Karlmarx, L.R.: System for multimodal biometric recognition based on finger knuckle and finger vein using feature-level fusion and k-support vector machine classifier. IET Biom. **6**(3), 232–242 (2017)

10. Uhl, A., Wild, P.: Personal recognition using single-sensor multimodal hand biometrics. Image and Signal Process. **5099**, 396–404 (2008)

11. Kumar, A., Zhang, D.: Combining fingerprint, palmprint and hand-shape for user authentication. In: 18th International Conference on Pattern Recognition (ICPR), vol. 4, pp. 549–552 (2006)

12. Sanches, T., Antunes, J., Correia, P.L.: A single sensor hand biometric multimodal system. In: 15th European Signal Processing Conference, pp. 30–34 (2007)

13. Kimura, T., Makihara, Y., Muramatsu, D., Yagi, Y.: Single sensor-based multi-quality multi-modal biometric score database and its performance evaluation. In: International Conference on Biometrics (ICB), pp. 519–526 (2015)

14. Bhilare, S., Kanhangad, V., Chaudhari, N.: A study on vulnerability and presentation attack detection in palmprint verification system. Appl, Pattern Anal (2017). https://doi.org/10.1007/s10044-017-0606-y

15. Gottschlich, C., Marasco, E., Yang, A.Y., Cukic, B.: Fingerprint liveness detection based on histograms of invariant gradients. In: IEEE International Joint Conference on Biometrics (IJCB), pp. 1–7 (2014)

16. Chen, H., Valizadegan, H., Jackson, C., Soltysiak, S., Jain, A. K.: Fake hands: spoofing hand geometry systems. In: Biom. Consort (2005)

17. Patil, I., Bhilare, S., Kanhangad, V.: Assessing vulnerability of dorsal hand-vein veryfication system to spoofing attacks using smartphone camera, In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1–6 (2016)

18. Mirmohamadsadeghi, L., Drygajlo, A.: Palm vein recognition with local binary patterns and local derivative patterns. In: International Joint Conference on Biometrics (IJCB), pp. 1–6 (2011)

19. Zhou, Y., Kumar, A.: Contactless palm vein identfication using multiple representations. In: Biometrics: Theory Applications and Systems (BTAS), pp. 1–6 (2010)

20. Zhou, Y., Kumar, A.: Human identification using palm-vein images. IEEE Trans. Inf. Forensics Secur. **6**(4), 1259–1274 (2011)

21. Yan, X., Kang, W., Deng, F., Wu, Q.: Palm vein recognition based on multi-sampling and feature-level fusion. Neurocomputing **151**, 798–807 (2015)

22. Kang, W., Wu, Q.: Contactless palm vein recognition using a mutual foreground-based local binary pattern. IEEE Trans. Inf. Forensics Secur. **9**(11), 1974–1985 (2014)

23. Liu, Z., Song, S.: An embedded real-time finger-vein recognition system for mobile devices. IEEE Trans. Consum. Electron. **58**(2), 522–527 (2012)

24. Yang, G., Xi, X., Yin, Y.: Finger vein recognition based on a personalized best bit map. Sensors **12**(2), 1738–1757 (2012)

25. Miura, N., Nagasaka, A., Miyatake, T.: Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. Mach. Vis. Appl. **15**(4), 194–203 (2004)

26. Huang, B., Dai, Y., Li, R., Tang, D., Li, W.: Finger-vein authentication based on wide line detector and pattern normalization. In: 20th International Conference on Pattern Recognition (ICPR), pp. 1269–1272 (2010)

27. Song, W., Kim, T., Kim, H.C., Choi, J.H., Kong, H.J., Lee, S.R.: A finger-vein verification system using mean curvature. Pattern Recogn. Lett. **32**(11), 1541–1547 (2011)

28. Wang, Y., Xie, W., Yu, X., Shark, L.K.: An automatic physical access control system based on hand vein biometric identification. IEEE Trans. Consum. Electron. **61**(3), 320–327 (2015)

29. Huang, D., Tang, Y., Wang, Y., Chen, L., Wang, Y.: Hand vein recognition based on oriented gradient maps and local feature matching. In: Asian Conference on Computer Vision (ACCV), pp. 430–444 (2012)

30. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Computer Vision and Pattern Recognition, pp. 1701–1708 (2014)

31. Sun, Y., Wang, X., Tang, X.: Deeply learned face representations are sparse, selective, and robust. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2892–2900 (2015)

32. Qin, H., El-Yacoubi, M.A.: Deep representation-based feature extraction and recovering for finger-vein verification. IEEE Trans. Inf. Forensics Secur. **12**(8), 1816–1829 (2017)

33. Jaswal, G., Nigam, A., Nath, R.: DeepKnuckle: revealing the human identity. Multimed. Tools Appl. **76**(18), 18955–18984 (2017)

34. Heikkila, M., Pietikainen, M., Schmid, C.: Description of interest regions with center-symmetric local binary patterns. In: ICVGIP, pp. 58–69 (2006)

35. Revaud, J., Weinzaepfel, P., Harchaoui, Z., Schmid, C.: Deep-matching: hierarchical deformable dense matching. Int. J. Comput. Vis. **120**(3), 300–323 (2016)

36. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)

37. Kanhangad, V., Kumar, A., Zhang, D.: A unified framework for contactless hand verification. IEEE Trans. Inf. Forensics Secur. **6**(3), 1014–1027 (2011)

38. Heikkila, M., Pietikainen, M.: A texture-based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)

39. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. Pattern Recogn. **38**(12), 2270–2285 (2005)

40. Kang, W., Liu, Y., Wu, Q., Yue, X.: Contact-free palm-vein recognition based on local invariant features. PLoS ONE **9**(5), e97548 (2014)

41. Zhang, D., Guo, Z., Lu, G., Zhang, L., Zuo, W.: An online system of multispectral palmprint verification. IEEE Trans. Instrum. Meas. **59**(2), 480–490 (2010)

42. Tome, P., Marcel, S.: On the vulnerability of palm vein recognition to spoofing attacks. In: The 8th IAPR International Conference on Biometrics (ICB) (2015)

43. CASIA, M.: Palmprint v1 database [online]. http://www.cbsr.ia.ac.cn/mspalmprint. Accessed Mar 2017

44. Choi, J.H., Song, W., Kim, T., Lee, S.R., Kim, H.C.: Finger vein extraction using gradient normalization and principal curvature. In: SPIE-IS&T Electronic Imaging (2009)

45. Arandjelovific, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: Computer Vision and Pattern Recognition (CVPR), pp. 2911–2918 (2012)

Shruti Bhilare received B.E. degree in computer engineering from Institute of Engineering and Technology, Devi Ahilya university Indore, India, in 2009, and M.E. degree in computer engineering with specialization in Software Engineering from Institute of Engineering and Technology, Devi Ahilya university Indore, India, in 2011. She is currently pursuing Ph.D. in computer science and engineering with

the Indian Institute of Technology Indore, Indore, India. Her research interest includes pattern recognition and biometrics.

Gaurav Jaswal received his M.Tech. and Ph.D. degrees (not awarded) in electrical engineering from National Institute of Technology, Hamirpur, India in 2010 and 2018 respectively. He is currently a research associate in the School of Computing and electrical engineering at Indian Institute of Technology Mandi, India. His research interest includes Multimodal Biometrics and Deep Learning.

Vivek Kanhangad received the Ph.D. degree from the Hong Kong Polytechnic University in 2010. Prior to joining Hong Kong PolyU, he received M.Tech. degree in electrical engineering from Indian Institute of Technology Delhi and worked for Motorola India Electronics Ltd, Bangalore for a while. From June 2010 to December 2012, he was a faculty member at International Institute of Information Technology Bangalore. He is currently an associate professor in the discipline of electrical engineering at Indian Institute of Technology Indore, India. He regularly serves on the technical program committees of the reputed international conferences on biometrics. Currently, he is an editorial board member of IEEE Biometrics Compendium Journal. His research interests are in the overlapping areas of digital signal and image processing, pattern recognition with focus on biometrics and biomedical applications.

Aditya Nigam received M.Tech. and Ph.D. degrees in computer science and engineering from Indian Institute of Technology Kanpur, India, in 2009 and 2015 respectively. He is currently working as an Assistant Professor in the School of computing and electrical engineering at Indian Institute of Technology Mandi, India. His research interests include Multimodal Biometrics, Computer Vision and Deep Learning.