On Matching Finger-selfies using Deep Scattering Networks

Aakarsh Malhotra*, *Student Member, IEEE,* Anush Sankaran*, *Member, IEEE,* Mayank Vatsa, *Senior Member, IEEE,* and Richa Singh, *Senior Member, IEEE*

Abstract—With the advancements in technology, smartphones' capabilities have increased immensely. For instance, the smartphone cameras are being used for face and ocular biometric-based authentication. This research proposes *finger-selfie* based authentication mechanism, which uses a smartphone camera to acquire a selfie of a finger. In addition to personal device-level authentication, finger-selfies may also be matched with livescan fingerprints present in the legacy/national ID databases for remote or touchless authentication. We propose an algorithm which comprises of segmentation, enhancement, Deep Scattering Network based feature extraction, and Random Decision Forest to authenticate finger-selfies. This paper also presents one of the largest finger-selfie database with over $19,400$ images. The images in the IIIT-D Smartphone Finger-selfie Database v2 are captured using multiple smartphones and include variations due to background, illumination, resolution, and sensors. Results and comparison with existing algorithms show the efficacy of the proposed algorithm which yields equal error rates in the range of $2.1 - 5.2\%$ for different experimental protocols.

Index Terms—Finger-selfie, touchless fingerprint recognition, verification, fingerprint database, deep scattering networks

1 Introduction

Attributed to the technological advancements, the capabilities of smartphones have increased immensely. The inclusion of cameras in smartphones has increased user convenience and has even aided the biometric-based authentication process. The smartphone camera-based biometric authentication using face [35], ocular [34], and iris [32] has been shown to be secure, flexible, and easy to use. As illustrated in Fig. 1, the idea can also be further extended to finger-selfie[1] based authentication [27]. Finger-selfie employs the in-built camera for acquiring an image of the finger(s) which can then be used for authentication.

In the current COVID19 pandemic situation and the post COVID19 world, touchless biometric modalities would be of significant interest to the community. Finger-selfies can provide a good alternative to touch-based fingerprint recognition. The usage of finger-selfies for authentication can be two-fold: (i) both the query and probe samples are finger-selfie images and (ii) in the scenarios of remote authentication, finger-selfie may be matched with livescan fingerprints present in the legacy/national ID databases. The cross-domain matching of finger-selfies as a query against livescan fingerprints as a template holds a pivotal role in forensic investigations [38]. For instance, the South Wales Police recently received a finger-selfie of a drug dealer holding drugs on his finger [49]. The drug dealer was circulating an image of drugs placed over his fingers via Whatsapp. The scientific support unit was able to utilize the finger-selfie to find the offender

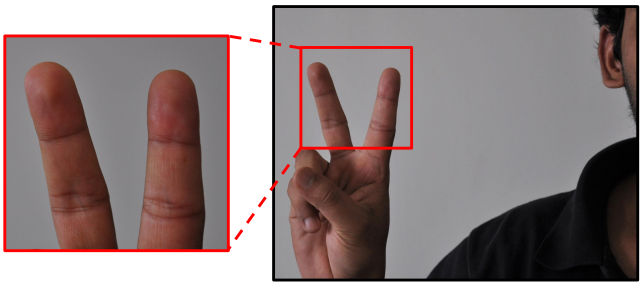


Fig. 1. A high resolution finger-selfie can provide distinctive ridge information for authentication purposes.

by comparing the finger-selfie with fingerprints of past offenders. Similarly, a hacker was able to generate fingerprints of a German minister from an image of the minister's fingers taken from a DSLR camera from 3 meters away [12]. Additionally, the police may establish the identity of deceased or unconscious individual [33] by transferring finger-selfies of the unknown subject to a remote server. At the remote server, the finger-selfie may be matched with the live-scan fingerprint databases to reveal the potential identity and contact the concerned people. Such applications showcase the need to develop an automated algorithm for finger-selfie recognition, that can operate simultaneously for device access control and forensic/law enforcement applications.

This research aims to design an efficient algorithm for automated finger-selfie recognition. This includes both finger-selfie-to-finger-selfie and finger-selfie-to-livescan fingerprint matching. As shown in Fig. 2, a finger-selfie can be captured under any environment with any kind of smartphone camera, thereby introducing many challenges in recognizing the image. The major challenges [28] that affect a finger-selfie matching algorithm are:

- Finger-selfie can be obtained in a controlled indoor illumina-

- *A. Malhotra is with IIIT-Delhi, New Delhi, India 110020 (email: aakarshm@iiitd.ac.in)*
- *M. Vatsa, and R. Singh are with IIT Jodhpur, India, 342037 (email: {mvatsa, richa}@iitj.ac.in)*
- *A. Sankaran is with Deeplite, Montreal, Canada (e-mail: anushs@iiitd.ac.in)*

* *A. Malhotra and A. Sankaran contributed equally.*

1. Finger-selfies are hand finger ridge impressions captured directly using a camera in a touchless method. Fingerprint images are ridge impressions captured using a touch-based live-scan sensor.

TABLE 1
A literature review of algorithms for pre-processing and comparison of finger-selfies obtained using mobile phones.

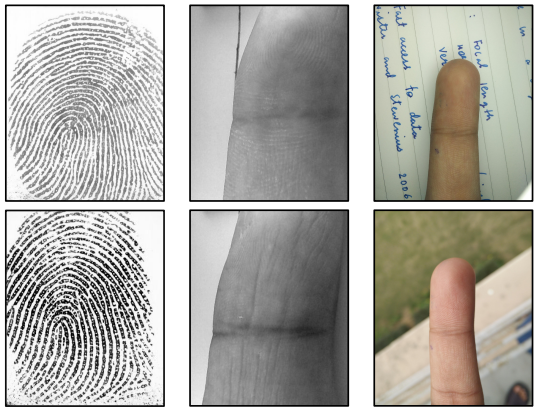
Research	Database	Challenges					Algorithm	Results
		Illumination	Background	Resolution	Position	Video		
Lee et al., 2005 [18]	840 images from 168 fingers						Segmentation using color model, ridge gradient extraction	GAR: 75.00 at 0.1% FAR
Lee et al., 2008 [19]	120 images from 120 fingers, 1200 fingerprints				✔		Pose, quality estimation using gradient coherence and symmetry	Rejection rate: 5.67%, EER: 3.02%
Stein et al., 2012 [42]	492 images from 82 fingers				✔		Quality estimation using edge density	EER: 19.10%
Derawi et al., 2012 [9]	1320 images from 220 fingers						Matching using VeriFinger SDK	EER: 4.50%
Li et al., 2012 [21]	2100 images from 100 fingers	✔	✔				Matching using VeriFinger SDK and NFIS	EER: 24.80% - 49.60%
Li et al., 2013 [20]	2100 images from 100 fingers	✔	✔				Quality estimation using 12 features and SVM	Spearman correlation = 0.53
Stein et al., 2013 [41]	990 images from 74 fingers, 66 fingervideos					✔	Reflection based spoofing detection	EER in the range: 1.20% - 3.00%
Tiwari and Gupta,2015 [44]	156 images from 50 fingers				✔		Matching using scale invariant features	EER: 3.33%
Sankaran et al., 2015 [37]	5100 fingerphotos from 128 fingers	✔	✔				Segmentation, enhancement, matching using ScatNet and RDF	EER in the range: 5.1% - 6%
Minaee and Wang, 2015 [29]	1480 fingerprint images				✔		Matching using ScatNet, PCA, and SVM	EER: 8.10%
Lin and Kumar, 2018 [24]	1800 3D contactless fingerprints from 300 fingers				✔		Deformation correction model, with minutiae and ridge matching	EER: 4.46%
Deb et al., 2018 [8]	2472 fingerphotos from 1236 fingers			✔			Matching using COTS fingerprint matcher	TAR: 92.4 − 98.6% at 0.1% FAR
Chopra et al., 2018 [7]	3450 fingerphotos from 230 fingers	✔	✔	✔	✔		Segmentation using VGG SegNet, recognition using ResNet50	EER: 35.48%
Wasnik et al., 2018 [47]	720 fingerphotos from 48 fingers						Verification using LBP, HOG, BSIF, and VeriFinger	EER in the range: 6.05 − 12.84%
Lin and Kumar, 2018 [23]	3920 3D contactless fingerprint from 300 fingers				✔		Siamese networks to learn multi-view representation	EER in the range: 0.81 − 8.35%
Wild et al., 2019 [48]	1728 fingerphoto from 108 fingers, 2582 fingerprints			✔			Skin-color segmentation, QA and matching by VeriFinger & NFIQ	GAR: 98.70 at 0.1% FAR
Proposed	**19456 images (17024 finger-selfies and 2432 live scan fingerprints) from 304 fingers**	✔	✔	✔			**Saliency based segmentation, enhancement, DSN + RDF matching**	**EER in the range: 2.11% - 5.23%**

tion or in an uncontrolled outdoor illumination setting. During outdoor capture, ambient lighting during day and night time varies a lot which affects the quality of images. Further, usage of camera *flash* makes the preprocessing difficult.
- Real-time capture of images with varying background. The distance of the closest background object may vary, making foreground segmentation an arduous task.
- Cameras in various smartphones have features such as resolution, auto-focus, and flash LEDs. These factors can alter the quality of the obtained image.
- Cross domain matching between finger-selfie and traditional livescan fingerprints.
- Varying orientation and distance of the finger from the camera results in affine and projective transformations.

1.1 Related Work

In the literature, researchers have proposed algorithms focusing on matching finger images obtained from smartphone camera (or web-camera). Table 1 summarizes the existing studies and their respective performances. In 2005, Lee et al. [18] proposed a ridge segmentation algorithm using a color model for the foreground skin region. They enhanced the ridge information by computing the ridge orientation using the gradient for a database of 400 images. In 2008, Lee et al. [19] performed finger quality estimation using gradient information coherence in the local region, under varying poses. They collected a private dataset with four subsets, with 120 fingerphoto sequences and 1200 fingerprint images.

In 2012, Stein et al. [42] emphasized the need of a quality estimation algorithm and proposed an algorithm using the ridge edge density in a local region for a dataset of 41 subjects from two mobile devices. In 2012, Li et al. [21] studied the performance of ten-print matchers such as Verifinger by Neurotechnology and NFIS from NIST. Their dataset constituted of 2100 fingerphotos captured using three different mobile phones with varying background and illumination. Based on this study, Li et al. [20] observed that minutiae extraction using existing commercial



(a) Live-scan fingerprints (b) Webcam finger images (c) Smartphone finger-selfie

Fig. 2. Sample images illustrating variations due to different acquisition methods: (a) live-scan fingerprints, (b) contactless fingerprints using webcam [17], and (c) contactless finger-selfies using smartphone cameras.

matchers such as Verifinger is extremely noisy and produces lots of spurious minutiae. They proposed a learning based quality estimation algorithm using fingerprint specific features along with SVM classifier. Stein et al. [41] and Taneja et al. [43] presented a study using a sequence of fingerphotos to avoid spoofing. The data was collected in a controlled environment. Recently, Malhotra et al. [26] showed the need for privacy preservation of finger-selfies on social media. They learn perturbation using an adversarial model to misclassify finger-selfies against genuine fingerprints.

Minaee and Wang [29] proposed ScatNet features followed by PCA and SVM for matching touchless fingerprints. The images were acquired in a controlled environment. No segmentation or enhancement were considered, and experiments were shown on PolyU HRF dataset [45]. As seen in Fig 2, the smartphone captured finger-selfies are challenging and visually different from webcam captured touchless fingerprints. Lin and Kumar [24] and Deb et al. [8] also proposed contactless fingerprint databases. The former had 1800 contactless fingerprints from 300 classes while the latter had 2472 fingerphotos from 1236 classes. Lin and Kumar reported an of EER of 4.46% when Robust Thin-Plate Spline Model and Deformation Correction Model are used together to perform contactless to livescan fingerprint matching. Deb et al. used COTS to perform fingerphoto to livescan fingerprint matching. Under different combinations of fingerphotos, the reported TAR varied between 92.39% to 98.55%.

1.2 Research Contributions

The use case scenario of most of the existing studies restrict to matching finger-selfie to either finger-selfie or livescan fingerprints. These studies primarily focus on designing algorithms for one or two steps of the recognition pipeline. However, as highlighted by [14], matching contactless fingerprints between various contactless devices provide very poor results. Hence, additional research is necessary to propose a robust algorithm for finger-selfie recognition. Moreover, there is no publicly available database with a large number of finger-selfies and fingerprint images.

In this research, we propose an end to end pipeline with segmentation, ridge structure enhancement, feature representation, and verification. This is an extension of our research [37], in which we proposed a ScatNet based matching pipeline for fingerphoto images[2] captured from smartphones. We created the IIIT-D Smartphone FingerPhoto Database v1 (ISFPDv1), focusing on background and illumination variations for finger-selfie matching with 5100 fingerphotos.The key contributions of this research are:

- A novel combination of saliency and skin color based finger-selfie region of interest segmentation algorithm to remove the highly varying noisy background. The preliminary version of the algorithm [37] used only a skin color-based segmentation.
- Study the effect of camera resolution and environmental variations in the acquired finger-selfie, using Deep Scattering Network representation [40] and matching algorithm.
- Creating and publicly releasing IIIT-D SmartPhone Finger-selfie Database v2 (ISPFDv2). The database consists of 17024 finger-selfies from 304 unique fingers, acquired using smartphones OnePlus One and MicroMax Canvas Knight. Additionally, 2432 livescan fingerprints are also collected using Secugen Hamster IV sensor.
- Extensive experimental analysis on the proposed dataset to study the effect of segmentation and enhancement approaches. Further, we study the impact of (i) noisy background, (ii) environmental illumination, (iii) camera resolution, and (iv) cross-domain scenario, on the performance of finger-selfie authentication.

2 Proposed Finger-selfie Recognition Algorithm

The proposed finger-selfie matching pipeline has four steps: (i) finger-selfie segmentation, (ii) ridge pattern enhancement, (iii) Deep Scattering Network (DSN) representation extraction, and (iv) matching. The individual steps are illustrated in Fig. 3 and explained in the sections below.

2.1 Finger-selfie Segmentation

The process of segmentation involves finding a binary mask in a captured finger-selfie, that represents the *distal phalanges* of the finger. As shown in Fig. 4, even though a finger-selfie can be captured with different kinds of background objects, it is safe to assume that the finger is the closest object to the smartphone camera. Under this assumption, we observe two distinguishing features that separate the finger-selfie from the rest of the background: (i) the skin color of the finger region, and (ii) the salient nature [2] of the finger region in the image. Thus, the proposed algorithm combines the region covariance based saliency [10] along with skin color measurements for effective finger-selfie segmentation. The steps involved in the proposed segmentation algorithm are:

Step 1: For each pixel, seven visual features (represented further as d-dimensional) are extracted: (i) intensity in $L*a*b$ color space, (ii) the edge orientation along x and y directions, and (iii) x and y locations of the pixel. The $L*a*b$ color space is chosen due to its closeness with human perception[3].

Step 2: For a region R_i in an image, a 7×7 covariance matrix C_i is constructed using the 7-dimensional feature vector (extracted

2. Fingerphoto and finger-selfie in our context are same as they are captured by the smartphone camera by the user themselves.

3. $L*a*b$ color space has perceptual uniformity [50] [52]. Further, the a and b components resemble human chromatic system, whereas the L component approximates to the human perception of lightness [3]

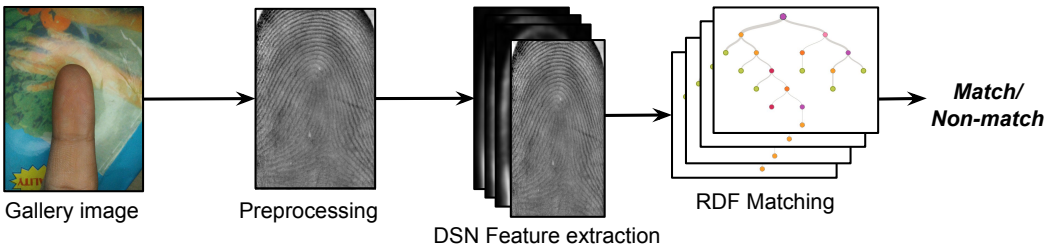


Fig. 3. Illustrating the procedure involved in the proposed Deep Scattering Network (DSN) based finger-selfie verification pipeline.

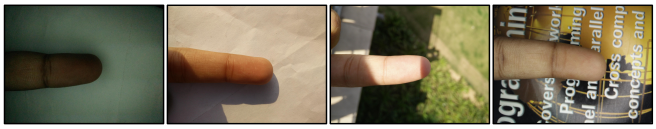


Fig. 4. Challenging cases for segmentation due to natural background and indoor/outdoor illumination variations.

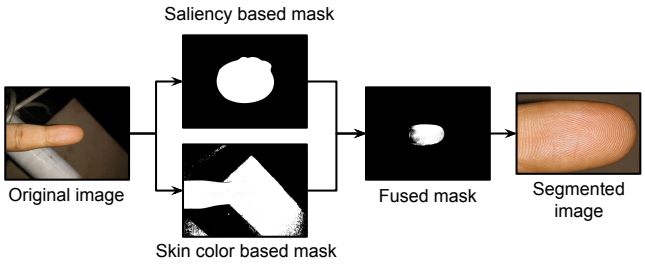


Fig. 5. Illustration of the segmentation algorithm combining the saliency based and skin color based segmentation maps.

in Step 1). The first order statistics of C_i are computed and the extracted statistics are non-linearly aggregated, as follows,

$$\xi(C_i) = (\mu, s_1, \ldots, s_d, s_{d+1}, \ldots, s_{2d})^T \quad (1)$$

where, μ and s are the mean and standard deviation, respectively. The size of the region R_i is chosen based on the image resolution.

Step 3: The dissimilarity between any two regions, $D(R_i, R_j)$ is computed as a normalized difference between the representation obtained by covariance matrices, $\xi(C_i), \xi(C_j)$ as follows,

$$D(R_i, R_j) = \frac{||\xi(C_i) - \xi(C_j)||}{1 + ||x_i - x_j||} \quad (2)$$

where, x_i and x_j denote the pixel location of center of the image regions R_i and R_j, respectively.

Step 4: The covariance based saliency map [10] of any region, $map_{sal}(R_i)$, is computed by a non-linear aggregation of the dissimilarities of the current region R_i with the surrounding m most similar regions of the image, as follows,

$$map_{sal}(R_i) = \frac{1}{m} \sum_{j=1}^{m} D(R_i, R_j) \quad (3)$$

Step 5: To compute the skin color based segmentation, the RGB image is converted into CMYK color space. The normalized magenta channel is used as the skin color map, map_{skin}, as magenta channel retains ample amount of skin color [39].

Step 6: The overall segmentation map is obtained by combining the saliency and the skin color based maps using a weighted sum fusion. The segmentation map map_{seg} is obtained by,

$$map_{seg} = w_1 \times map_{sal} + w_2 \times map_{skin} \quad (4)$$

where, w_1 and w_2 represent the weights associated to the normalized saliency map and normalized skin map, respectively.

Step 7: Otsu's thresholding method [31] is applied on the obtained map, map_{seg}, to obtain the binary segmented mask.

Fig. 5 shows a visual illustration of the fusion of skin color based segmentation and saliency based segmentation map to obtain the final segmented result.

2.2 Finger-selfie Enhancement

The aim of enhancement is to improve the image quality so that the finger ridge patterns can be efficiently extracted from the captured finger-selfie. The primary challenge for ridge extraction is the noise induced by the surrounding illumination variation which affects the contrast between valleys and ridges. The segmented image of the previous step is first converted to gray scale. Speckle noise is removed using median filtering and histogram equalization is applied to mitigate the effect of illumination variation. Next, the image is sharpened to improve the contrast between valleys and ridges. In the resultant image, the ridge information constitutes the high-frequency components while the valley and noise components constitute the low-frequency information. To enhance the difference between the ridge and valley information, we sharpen the image by subtracting the Gaussian blurred image ($\sigma = 2$) from the previous image. Fig. 6 shows sample output provided by the enhancement algorithm.

2.3 Deep ScatNet (DSN) based Feature Representation

Finger-selfie images captured at different instances observe variations in rotation, translation, and scaling along with variations due to the environment. While researchers have explored the use of minutiae, low resolution and varying ridge-valley contrast in finger-selfies makes it difficult to accurately detect minutiae [20]. We propose feature representation for finger-selfies using Deep Scattering Networks [4], [5]. DSN is a local descriptor which incorporates multi-directional and multi-scale information. DSN is calculated with a set of wavelet decompositions and complex modulus. It has been shown that DSN based features effectively encode texture patterns in images in presence of rotation, scaling, and deformations [40]. As seen in Fig. 6, the enhanced finger-selfie has good ridge-valley patterns and we assert that DSN features can encode these patterns.

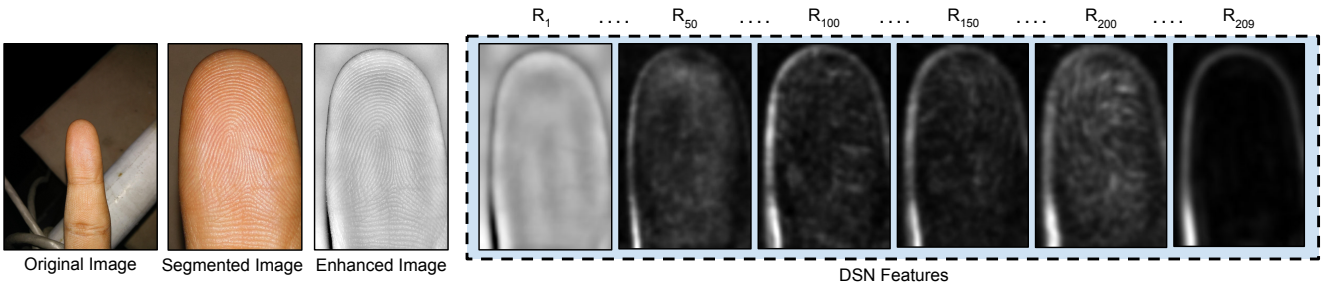


Fig. 6. DSN representation of the enhanced image where R_1 to R_{209} represent the responses from different depth of DSN.

Let the signal $x(u) \in \mathbb{R}^2$ (image in this case) represent the input. DSN learns the features S_n at each level $n \in J$, where $0 \leq n \leq J$, J represents the network depth. The output is a representation obtained by concatenation of DSN features from each level S_n.

Let $\phi_J(u) = 2^{-2J}\phi(2^{-J}u)$ be a low-pass averaging filter and u be the set of parameters corresponding to the locations x and y in the image space. Similarly, let $\psi_{\theta,j}(u) = 2^{-2j}\psi(2^j\theta^{-1}u)$ be the high-frequency, quadrature phase, complex wavelet filter bank. This high frequency wavelet bank ψ is built by changing the rotation θ and the scale (dilation) 2^j parameters. The first 2D wavelet transform for the signal x is denoted by W_1. It consists of convolutions with both low pass and high pass filters. Convolution with the low pass filter computes a locally affine invariant representation S_0 as follows:

$$S_0x(u) = x * \phi_J(u) \tag{5}$$

These coefficients are the 0^{th}-order DSN coefficients. The representation obtained in the above equation is translation invariant up to 2^J pixels. Since the representation lacks the high frequency information, it is recovered by convolving x with $\psi_{\theta,j}(u)$. To make the term $x * \psi_J(u)$ invariant to local translation and diminish deviation amongst the coefficients, the complex phase is eliminated using modulus operator. Hence, the wavelet-modulus transform operator is given as follows:

$$|W_1|x = (x * \phi(u), |x * \psi_{\lambda_1}(u)|) \tag{6}$$

where, λ_1 corresponds to the set of first level filtering parameters (θ, j). The high frequency filters supplement information to the representation procured in Equation 5. Thus, we perform the second wavelet transform W_2 on the second term of the above equation. This again gives two outputs, one from low pass filter $\phi(u)$, and other from high pass filter $\psi(u)$. The output from low pass filter gives affine invariant representation of the high frequency part, as described below:

$$S_1x(u, \lambda_1) = |x * \psi_{\lambda_1}(u)| * \phi_J(u) \tag{7}$$

$S_1x(u, \lambda_1)$ are the first-order DSN coefficients. $S_1x(u, \lambda_1)$ has all the filter responses of the wavelet bank $\psi_{\lambda_1}(u)$ (each successively applied with a low pass filter). For the high pass output, the complex phase is removed by modulus operator. Higher-order coefficients can be extracted by recursively building deeper wavelet filter banks:

$$|W_2||x * \psi_{\lambda_1}(u)| = (|x * \psi_{\lambda_1}(u)| * \phi, ||x * \psi_{\lambda_1}(u)| * \psi_{\lambda_1}|) \tag{8}$$

These deeper DSN coefficients gives a stable translation and rotation invariant representation for the finger-selfies, as illustrated in Fig. 6. Since DSN filters are pre-designed or handcrafted, obtaining a DSN representation is convolving these filters with the finger-selfie and there is no learning involved. Hence, it can be achieved with the computational capabilities of smartphones. The effective representation for a finger-selfie is the concatenation of all n-order outputs as: $\{S_0, S_1, \ldots, S_n\}$.

In this research, we experimentally observe an optimal depth of the DSN as $J = 2$, i.e., computing the second order DSN coefficients for all the finger-selfies. Let the enhanced finger-selfie, I_{enh} be of size $w \times h$. The concatenation of all responses up to the second order, $\{S_0, S_1, S_2\}$ contains a total of 209 filters, with each response of dimension $\frac{w}{8} \times \frac{h}{8}$. Thus, the size of DSN feature representation is $209 \times \frac{w}{8} \times \frac{h}{8}$.

2.4 Feature Matching

Let Q and T be the $1 \times N$ sized DSN representation vector of the query (probe) and the template (gallery) finger-selfies, respectively. A supervised binary classifier g:$(X \rightarrow Y)$ is learned to classify a DSN feature representation pair (Q, T) into either a genuine (match) or imposter (non-match) pair. The input feature X is the difference of query and template representations $(Q-T)$ and the classification labels Y are $\{$*genuine, imposter*$\}$. The classifier learns to predict if a representation pair is a genuine or an imposter pair. Two finger-selfies that belong to the same finger are defined as match pair, while images captured from different fingers or different subjects are considered as non-match pairs.

In this research, we use Random Decision Forest (RDF) [13] as the binary classifier for verifying the pair of finger-selfies. RDF is a non-linear ensemble based classifier with multiple decision trees [13]. It employs a repetitive random sub-sampling for bagging which assists in obtaining robust and faster results for correlated features. For a total of $\mathcal{D}$ finger-selfies in the training set, several bootstrap aggregates of size $r.\mathcal{D}$ are made with replacement, for a ratio r $(0.5 < r \leq 1)$. A forest having T trees is trained, where, each decision tree is trained with a single bootstrap of the data. Let M be the length of the vectorized DSN representation obtained for the finger-selfie. At each node of the tree, a feature sample m is chosen at random. This sample is utilized for taking the split decision to maximize information gain. Each tree is designed as a binary decision tree by assigning leaf nodes as $\{$*genuine, imposter*$\}$ corresponding to the training sample. Thus, each decision tree in the forest classifies the input pair of finger-selfies as a matching pair or a non-match pair. The final decision

TABLE 2
Summarizing the characteristics and the variations captured in the IIIT-D SmartPhone Finger-selfie Database v2.

Set	Name	Variations			Images
		Illumination	Background	Resolution	
I	WI	Controlled	White	13 MP	2432
	WO	Uncontrolled	White	13 MP	2432
II	NI	Controlled	Natural	13 MP	2432
	NO	Uncontrolled	Natural	13 MP	2432
III	RES5	Controlled	White	5 MP	2432
	RES8	Controlled	White	8 MP	2432
	RES16	Controlled	White	16 MP	2432
IV	LS	-	-	500 PPI	2432

is computed by taking a majority vote of all the decision trees in the ensemble.

2.5 Implementation Details

In a real-life scenario, the finger-selfie may be acquired at any PPI due to scale and resolution variations. Hence, we did not enforce any constraint on the PPI. Other researchers also downscale the acquired images to a fixed resolution to justify lower computation, model input dimensions, or making resolutions same as live-scan without taking PPI into consideration [7], [22], [23], [46]. Additionally, some algorithms operate directly on the resolution of the acquired finger-selfie without any scaling [25], [44]. The acquired image is of resolution 4208×3120 at approximately 1200 PPI. For different parameters, a grid search based approach is used to find the optimal values. For example, $\alpha = \sqrt{2}$ is used for region covariance based saliency map extraction. Similarly, optimal values for R_i is selected as 16×16 and r as 0.66. From all the images, the segmentation algorithm yields a fixed-size window of 1240×800 at approximately 1200 PPI. The segmented images are downscaled by half to 620×400 to reduce the computation of feature extraction. While MCC and NFIS based minutiae matching works at 500 PPI, we operate at a fixed resolution that approximates to 600 PPI at 13 MP.

The 2^{nd} order DSN representations are of length of 809,875 per sample. To reduce the dimensionality, we apply PCA [16] and preserve 99% Eigen energy. It yields a vector of length 95, which is provided as input to the classifier. Finally, T = 1000 (independent trees) is used in the RDF.

3 Database and Protocol

As a part of this research, we collect the IIIT-D SmartPhone Finger-selfie Database v2 (ISPFDv2). The details of the proposed database and its experimental protocol are elaborated below.

3.1 ISPFDv2: Database

The ISPFDv2 consists of more than 19400 images obtained from 304 unique fingers. In the proposed database, four instances from four fingers are collected for each of 76 subjects over two sessions. The finger-selfies are taken using smartphones OnePlus One and MicroMax Canvas Knight, while the corresponding livescan fingerprints are taken from Secugen Hamster IV. The indoor images in ISPFDv2 are captured in both constrained and unconstrained environments, while outdoor images are captured without flash during daylight and with flash during night. Auto-focus is always kept ON. The quality of these finger-selfies, as per NFIQ 2.0, is illustrated in Fig. 7. Based on the challenges, the database is divided into four subsets, as summarized in Table 2.

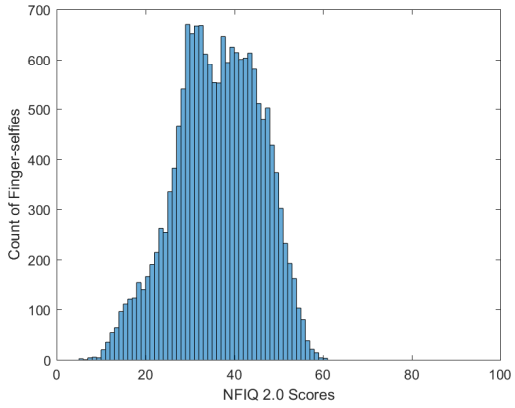


Fig. 7. NFIQ 2.0 scores for finger-selfies (quality: 35.83 $\pm$ 9.52).

Set I - White Background: With white background, finger-selfies are acquired indoors with controlled lighting (*WI*) and outdoors with varying illumination (*WO*). The subsets, illustrated in Fig. 8(a) and 8(b), capture the effect of illumination variations over a constrained white background. The images are taken using OnePlus One phone at 13MP resolution. Both *WI* and *WO* have 8 images each of right index, right middle, left index, and left middle fingers of 76 subjects, totalling 4864 images for Set I.

Set II - Natural Background: Finger-selfies are acquired indoors and outdoors, with unconstrained background. The subsets are shown in Fig. 8(c) and 8(d). The Natural Indoor (*NI*) subset highlights the outcome of background variations in controlled lighting, whereas, Natural Outdoor (*NO*) conveys the simultaneous effect of background and illumination variations. The images are captured using OnePlus One phone at 13MP resolution. Similar to Set I, Set II also has 4864 images.

Set III - Resolution: The set has finger-selfies captured in three different resolutions with controlled illumination and white background, as shown in Fig. 8(e). Two different smartphones, OnePlus One and MicroMax Canvas Knight, are used to capture the images at three different resolutions 5 MP, 8 MP, and 16 MP. Camera flash is turned OFF, whereas, auto-focus is kept ON. All the images are captured in an indoor lab environment, with uniform lighting and a blank white paper as the background. Under these settings, four instances of the index finger and middle finger of the right and left hand of 76 subjects are captured at all three resolutions, with a total of 7296 images.

Set IV - Livescan Fingerprints (LS): A subset of live-scan fingerprints are acquired using FBI certified Secugen Hamster IV fingerprint sensor, as shown in Fig. 8(f). These fingerprints are acquired at 500 PPI. This set comprises of 4 fingerprint images, each from the right index, right middle, left index, and left middle fingers from 76 subjects, acquired under 2 sessions (a week apart). Thus, the set has a total of 2432 images.

The IIIT-D SmartPhone Finger-selfie Database v2 is publicly available to the research community at http://iab-rubric.org/resources/spfd2.html.

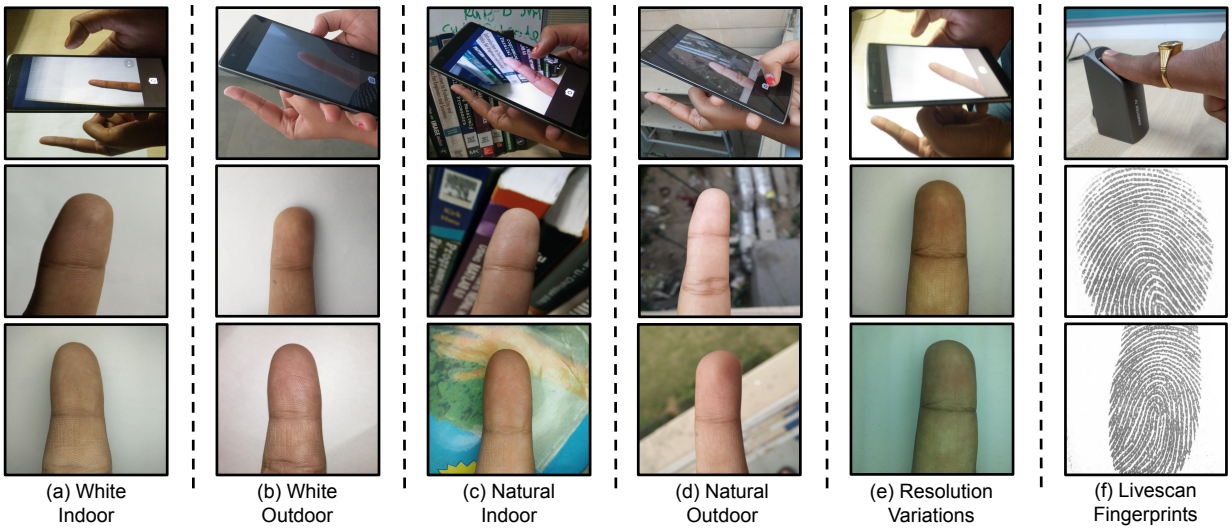


Fig. 8. Sample images from the IIIT-D SmartPhone Finger-selfie Database v2, illustrating the intra-class variations and noise present in the database.

3.2 Experimental Protocol

The experiments are performed with a subject disjoint 50%-50% train-test split with three times random cross-validation. From a total of 304 unique fingers (76 subjects×4 fingers), images corresponding to randomly selected 152 fingers (38 subjects×4 fingers) are used for training and the remaining 152 finger classes are used for testing. The effect of finger-selfie matching performance is studied using the following three experiments.

- **Background-Illumination (*Exp. 1*):** With the assumption that gallery images can be acquired in a controlled environment, White Indoor (*WI*) subset is the most constrained acquisition of finger-selfies. The *WI* subset is considered as gallery, whereas, {*WO*, *NI*, *NO*} subsets are taken as query, independently. *WI-WO* experiment shows the performance of finger-selfie matching under the influence of illumination, *WI-NI* highlights the influence of background variations, and *WI-NO* illustrates the simultaneous influence of illumination and background on the recognition performance.
- **Resolution (*Exp. 2*):** To simulate cross-resolution experiments, finger-selfies captured at 13MP in White Indoor (*WI*) environment are used as the gallery images. We use three different probe sets with capture resolution as 5MP, 8MP, and 16MP. Thus, two probe sets have lower resolution while one probe set has higher resolution than gallery.
- **Livescan-finger-selfie matching (*Exp. 3*):** To simulate cross-domain experiments, fingerprints (*LS*) are used as the gallery. We use seven different probe sets as WI, WO, NI, NO, 5MP, 8MP, and 16MP. Note that, since the livescan fingerprints are already tightly cropped, segmentation is not applied on fingerprint images. However, since since ridge flow patterns gets mirrored during acquisition by the fingerprint sensor, we perform a mirror operation on livescan fingerprint images before creating the pairs.

Thus, same gallery images are used for *Exp. 1* and *Exp. 2*, while *Exp. 3* uses livescan fingerprints as gallery. The probe sets comprise of finger-selfies, where subsets are varied to study the impact of capture variations on the matching performance.

4 Results and Analysis

The experiments are performed to evaluate the effectiveness of the proposed pipeline for finger-selfie recognition. This section is structured as follows: Section 4.1 elaborates on the results of *Exp. 1, 2,* and *3* in terms of Equal Error Rate (EER) and Receiver Operating Characteristic (ROC) curves. Section 4.2 provides an analysis of traditional minutia and non-minutia based methods in contrast to the proposed algorithm for finger-selfie recognition, followed by Section 4.3 highlighting the importance of enhancement. Lastly, Section 4.4 compares the proposed algorithm with recent algorithms for contactless fingerprint recognition.

4.1 Performance Analysis of Proposed Pipeline

We first perform a series of experiments to evaluate the effectiveness of the feature extractor and classifier in the proposed algorithm. Table 3 summarizes the algorithms used for segmentation, enhancement, feature extraction, and matching. The algorithms are validated with the proposed ISPFDv2 database and the results are shown in Table 4 and Fig. 11. The DSN+NN algorithm utilizes the same pipeline as DSN+RDF, except the fact that it utilizes fully connected layers to classify instead of RDF. The DSN+l_2 involves a training-free approach to match features using l_2 distance.

Analyzing the results in a column-wise fashion shows a comparison of the proposed algorithm against existing algorithms. It can be distinctly observed that the DSN features learnt with a supervised classifier provide better performance as compared to other algorithms, suggesting the importance of training a classifier for verification. DSN + RDF and DSN + NN provide EER in the range of $2.11 - 10.35\%$, while DSN + ℓ_2 distance matching provides EER in the range $15.72 - 20.82\%$. Further, RDF performs better than NN under all variations. We also observed that there is a negligible deviation in the performance of the proposed algorithm across the cross-validation experiments. With respect to existing feature descriptors, CompCode and MCC, DSN + RDF provides up to 20% improvement in EER. This can be attributed to the rich feature representation obtained using the high-frequency information in Deep Scattering Networks, and also its affine invariance property.

TABLE 3
The algorithmic pipeline of the different techniques used for comparison along with the proposed pipeline.

Type	Technique	Segmentation	Enhancement	Feature Extraction	Matching
Minutia based	NFIS	*nfseg*	✗	*mindtct*	*bozorth3*
	MCC [6]	Saliency + skin color	Median filter + sharpen	VeriFinger SDK	Minutiae Cylinder Code
	Wild et al. [48]	Skin color	✗	VeriFinger	VeriFinger
Non-Minutia	CompCode [53]	Saliency + skin color	Median filter + sharpen	Competitive Code	ℓ_2-distance
Deep Architectures	Chopra et al. [7]	VGG SegNet	✗	ResNet50 (pre-trained)	Cosine Similarity
	Lin and Kumar [23]	FCN	✗	6 Convolutional layers	Pairwise Siamese
DSN Based	DSN + ℓ_2	Saliency + skin color	Median filter + sharpen	DSN + PCA	ℓ_2-distance
	DSN + NN	Saliency + skin color	Median filter + sharpen	DSN + PCA	Neural network
	DSN + RDF	**(Saliency + skin color)**	**(Median filter + sharpen)**	**DSN + PCA**	**RDF**

TABLE 4
Performance of different feature extraction and matching algorithms in terms of EER (%). The preprocessing pipeline (segmentation + enhancement) is consistent for the first six columns. The last four columns use preprocessing as proposed in their respective papers.

	Gallery	Probe	DSN +l_2	DSN +NN	**DSN+RDF (Proposed)**	NFIS	MCC [6]	Comp Code [53]	Minaee & Wang [29]	Chopra et al. [7]	Lin & Kumar [23]	Wild et al. [48]
Exp. 1	White Indoor (WI)	WO	16.95	3.27	**3.00**	50.00	16.50	19.07	23.34	13.96	4.51	17.64
		NI	20.39	6.49	**3.21**	49.71	12.36	16.22	22.04	16.91	3.77	15.49
		NO	20.59	5.34	**2.11**	49.99	17.03	21.40	22.96	16.58	7.74	19.26
Exp. 2	Resolution (13 MP)	5 MP	15.72	7.53	**5.23**	49.96	10.35	14.32	13.65	15.20	9.79	8.44
		8 MP	17.93	5.42	**4.74**	49.88	10.01	13.03	16.21	14.41	5.09	8.97
		16 MP	17.07	3.73	**2.98**	50.00	48.48	10.84	21.38	11.78	4.69	44.50
Exp. 3	Livescan Fingerprints (LS)	WI	20.50	7.24	**3.41**	49.75	15.74	16.70	28.54	30.67	18.42	17.40
		WO	19.15	4.18	**3.29**	49.31	17.31	16.91	32.09	39.12	14.06	18.69
		NI	18.15	7.62	**2.79**	38.07	18.40	17.45	35.12	27.50	17.11	16.41
		NO	20.40	5.43	**2.49**	48.29	18.22	18.35	36.68	41.67	19.24	19.95
		5 MP	16.68	10.35	**5.05**	49.62	13.10	14.83	41.68	29.68	25.41	12.30
		8 MP	17.57	7.99	**3.88**	49.60	14.09	14.98	30.10	32.60	20.95	14.18
		16 MP	20.82	5.75	**2.91**	49.92	11.49	11.08	34.05	28.69	20.64	45.57

As explained in the protocol, three times random cross validation is performed and the standard deviation values reported in the supplementary file. It is observed that the standard deviation values under 0.4% for most cases of the proposed algorithm. Additionally, the ROC curve for finger-selfie to livescan matching is also included in Fig. 3 and Fig. 4 of the supplementary material.

We next evaluate the performance across different variations - *Exp. 1* and *Exp. 2*. From Table 4, it can be inferred that the proposed matching pipeline with RDF classifier provides the best performance across different variations. The consistently low error rates of $(2.1-3.2\%)$ for *Exp. 1* shows that the proposed algorithm is robust to the variations in background (*WO and NO*) and illumination (*NI and NO*). In the cross resolution matching experiments (*Exp. 2*), as summarized in the rows 4-6 of Table 4, we observe that matching high-resolution images (EER: 2.98%) yields slightly better results than matching low-resolution images (EER: 5.23%). It is important to note that different individuals have different kinds of phones and the camera resolution across different phones also varies. This result shows that using a phone camera with lower resolution leads to a small reduction in accuracy but not very significant - matching 8 MP with 13 MP finger-selfies gives 4.74% EER while matching 5 MP with 13 MP yields 5.23% EER.

To further study the impact of filter responses on overall performance, we perform response selection and response pruning experiments. In response selection, top-k responses are chosen based on Laplacian score for feature selection [11], Local Learning clustered feature selection [51], and ReliefF feature selection algorithm [36]. Experiments for response pruning are also performed where we select 50 random and 100 random responses/filters. We observe that in most of the cases, the proposed algorithm with all the filter responses outperforms selection and pruning approaches by at least 0.5%. It can be attributed to the fact that responses from different layers encode distinctive information, which is required for finger-selfie authentication. The results are presented in detail in Table 2 of the supplementary material.

The performance analysis for cross-domain experiment (*Exp. 3*) of matching finger-selfies with livescan fingerprints is shown in the last seven rows of Table 4. The robustness of the proposed algorithms can be inferred from consistently low error rates in (i) constrained setup (row 8: LS-WI), (ii) varying background-illumination (rows 9-11: LS-WO, LS-NI, and LS-NO), and (iii) resolution variations (row 12-14: LS-Res5, LS-Res8, and LS-Res16). On the contrary, the performance of other algorithms remains poor for all scenarios of *Exp. 3*. For instance, the performance of DSN+l_2 weakens when the resolution is increased to 16 MP, while, the error rate worsens for DSN+NN under the low-resolution scenario of 5 MP.

4.2 Minutia and Non-minutia Matching Pipeline

The acquired finger-selfie has a lot of background noise. To compare the performance of the proposed saliency-based segmentation algorithm, we segment finger-selfies with NFIS's *nfseg* module. *Nfseg* is a popular algorithm for the segmentation of fingerprints. The enhancement, feature extraction, and matching protocols remain the same during the comparison. The performance is summarized in Table 5 for both *Exp. 1* and *Exp. 2*. We observe that *nfseg* based segmentation has an overall EER in the range 11.4-15.9%, while for the proposed saliency-based

TABLE 5
Effect of pre-processing on finger-selfie matching using minutiae and the proposed matching framework. Subscript P denotes the proposed module. The results are reported in terms of EER (%).

Exp.	Gallery	Probe	Nfseg	Seg_P	Nfseg+ Enh_P	Seg_P+ Enh_P
			mindtct+bozorth		DSN+RDF	
1	WI	WO	50.00	48.68	14.78	3.00
		NI	49.71	49.09	15.61	3.21
		NO	49.99	48.44	15.91	2.11
2	13 MP	5 MP	49.96	49.26	11.38	5.23
		8 MP	49.88	49.55	13.45	4.74
		16 MP	50.00	49.75	14.73	2.98

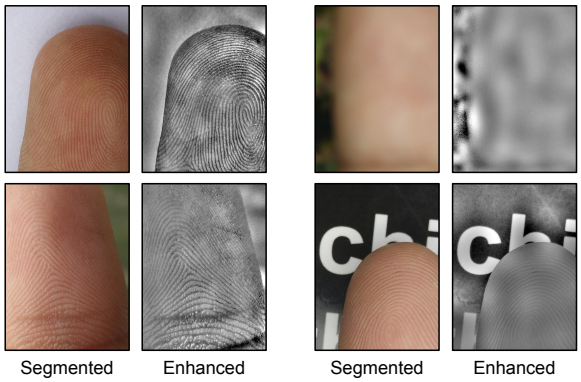
segmentation, the EER reduces to merely 2.1-3.2% for *Exp. 1* and 3.0-5.2% for *Exp. 2*. As seen in Fig. 10(c), the segmented image has a segregated ridge-valley region without noisy background, thereby, allowing better DSN features to be extracted. Similarly, we keep feature extraction and matching as mindtct+bozorth and vary segmentation algorithm to understand why NFIS algorithm performs poor. From Table 5, we observe that poor performance of NFIS algorithm can primarily be attributed to the spurious minutiae detection.

Analyzing the performance of MCC descriptor shows poor results for outdoor images in *Exp. 1*, as minutia extraction is highly spurious due to the capture variations. Further in *Exp. 2*, the performance of MCC descriptor drops suddenly when the resolution of probe images is higher than the gallery images. This can be attributed to the observation that MCC descriptor constructs fixed radius cylinders around each minutia to extract its descriptor. When the resolution of the probe image becomes higher than that of the gallery, no minutia is found within the constructed cylinder, and hence the matching performance drops. Dynamic prediction of the cylinder parameters can be performed for MCC descriptor, however, it is an independent research challenge.

The results of CompCode descriptor show that it is better at handling resolution variations as compared to handling environmental noise. However, DSN + RDF is more robust and is not much affected by the capture variations. Overall, we observe that partial finger-selfies and images with *out-of-focus* regions are better handled by DSN based matching algorithm. NFIS based pipeline also yields lower finger-selfie matching performance. Based on manual observation, we found that *nfseg* has a very high failure rate in segmenting the finger-selfie foreground region. It is primarily because *nfseg* is designed for fingerprints with a white background. Despite that, *nfseg* fails to segment finger-selfies with white backgrounds (WI and WO). Further, *mindtct* is not trained for extracting minutia from finger-selfies, thus, we observed more than 35% of the images to have zero minutia extracted. From these experiments, we infer that the public NFIS matcher from NIST cannot be used for matching smartphone captured finger-selfies.

4.3 Effectiveness of Enhancement

We next study the importance of the ridge-valley enhancement algorithm in the proposed finger-selfie verification pipeline. While segmentation is required to remove the background noise, enhancement improves the ridge-valley contrast and removes noise. The impact of enhancement on the matching performance is summarized in Table 6. The ROC curves are shown in Fig. 1 and Fig. 2 of the supplementary material. While the enhanced images in Fig. 9(a) may look similar to the grayscale version of



(a) Successful cases (b) Failure cases

Fig. 9. Successful and failure cases of enhancement on the ISPFDv2.

segmented images, the enhanced image has better ridge-valley contrast. For benchmarking, the matching results of only the segmented images (without any enhancement) are shown for both the experiments, *Exp. 1* and *Exp. 2*. These matching results for the segmented images are computed by converting the segmented image to grayscale, followed by feature extraction, and matching. The feature extraction and matching is performed by (i) MCC, (ii) CompCode, (iii) DSN+l_2, (iv) DSN+NN, and (v) DSN+RDF.

Without any enhancement, DSN+RDF produces EER in the range of $2.8 - 5.6\%$ for *Exp. 1* and in the range of $9.5 - 11.7\%$ for *Exp. 2*. With the proposed enhancement algorithm, the EER reduces to $2.1 - 3.2\%$ for *Exp. 1* and $3.0 - 5.2\%$ for *Exp. 2*. This highlights that the proposed ridge-valley enhancement in the pre-processing is essential to improve the performance of DSN+RDF, and is better than the performance of grayscaled segmented finger-selfie. The results also show that the enhancement algorithm has minimal influence on CompCode features. CompCode aims to extract the Gabor (variant) filter response of an image which allows the signals of a specific frequency band. During the enhancement phase, as image sharpening and smoothing are performed, certain frequency signals are removed that are essential for CompCode features. On the other hand, DSN extracts signals from all the frequency bands in a tree-like fashion and combines them. Similarly, as illustrated in *Exp. 2* in Table 6, enhancement algorithm shows limited improvement for MCC descriptor during cross-resolution matching. Further, DSN+NN demonstrate similar behavior. While there is no theoretical evidence to explain this pattern, it is hypothesized that these algorithms are robust against the proposed enhancement procedure.

As observed in Fig. 9, enhancement algorithm fails if there are no ridge-valley details visible due to a blurry acquisition. In a few scenarios where segmentation fails, the image tends to have more background regions than finger region. In such cases, the proposed algorithm might enhance some background detail instead of enhancing ridge-valley information. Hence, the success criteria of enhancement rely on the steps former to enhancement, namely, acquisition and segmentation.

4.4 Comparison with Recent Algorithms

In this subsection, we compare the performance of the proposed algorithm with recent algorithms for related tasks. The comparison

TABLE 6
Effect of enhancement on finger-selfie matching with background-illumination variations (*Exp. 1*) and cross-resolution variations (*Exp. 2*). The results are reported in terms of EER (%).

Algorithm	Template (Gallery)	Query (Probe) *Exp. 1*	Without Enhancement	With Enhancement	Query (Probe) *Exp. 2*	Without Enhancement	With Enhancement
MCC	WI (13 MP)	WO	19.05	16.50	5 MP	11.95	10.35
		NI	16.99	12.36	8 MP	12.61	10.02
		NO	19.87	17.03	16 MP	46.32	48.28
CompCode	WI (13 MP)	WO	18.26	19.07	5 MP	15.69	14.32
		NI	18.41	16.22	8 MP	15.14	13.03
		NO	21.40	21.40	16 MP	13.15	10.84
DSN + l_2	WI (13MP)	WO	22.20	16.95	5 MP	15.44	15.72
		NI	22.72	20.39	8 MP	19.48	17.93
		NO	18.81	20.59	16 MP	26.29	17.07
DSN+ NN	WI (13 MP)	WO	7.16	3.27	5 MP	7.41	7.53
		NI	6.83	6.49	8 MP	5.59	5.42
		NO	7.49	5.34	16 MP	4.22	3.73
DSN + RDF	WI (13MP)	WO	5.46	**3.00**	5 MP	9.67	**5.23**
		NI	5.58	**3.21**	8 MP	11.71	**4.74**
		NO	2.78	**2.11**	16 MP	9.48	**2.98**

is performed with different segmentation, feature extraction and matching algorithms.

4.4.1 Segmentation

Inspired from Chopra et al. [7], we utilized a supervised VGG SegNet [1] based framework for comparison. The VGG SegNet model is trained on the finger-selfie dataset [7] and tested on the ISPFDv2. However, the trained VGG SegNet model failed to segment the majority of fingers. Fig. 10 shows samples of finger-selfies and segmented images from the ISPFDv2.

It can be inferred that saliency-based finger-selfie segmentation outperforms segmentation using deep architectures. Additionally, we strongly believe that applications running on smartphones should use minimal resources (memory, time, and battery/power). The current deep learning algorithms require GPU resources, which is challenging in many smartphones. On the contrary, the proposed saliency-based segmentation has an added advantage of being computationally efficient.

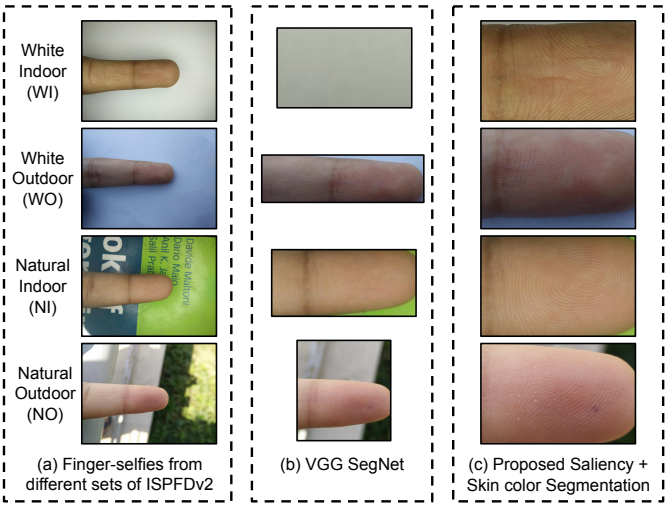


Fig. 10. Visual comparison between the segmented output of VGG SegNet and the proposed saliency based segmentation for finger-selfies.

4.4.2 Feature extraction and matching

A related work in the literature by Minaee and Wang [29] showed results on PolyU HRF dataset [45], which has high-resolution fingerprint images captured in a highly controlled environment. Their approach, when applied on a challenging task of matching smartphone captured finger-selfies on ISPFDv2 yields an EER in the range of $15 - 20\%$. They used ScatNet with a 2D filter bank, which can only incorporate spatial variance. Hence, their features are only robust towards translation invariance. On the contrary, the proposed method builds over 2D filter bank to incorporate orientation variations. The first step is to obtain a 3D signal by extracting rotation orbits over the 2D signal. The result is a 3-dimensional matrix, where the first two dimensions are the spatial position and the third dimension corresponds to orientation. Finally, a set of high pass and low pass Morlet filters are applied both along spatial and orientation variable to obtain a roto-translation convolution. Thus, the resultant feature representation is rotation and translation invariant. Additionally, our proposed algorithm is also aided with preprocessing stages of foreground segmentation and enhancement. We also compare results with a recent study by **Wild et al.** [48], which performs skin color segmentation, followed by feature extraction and matching using VeriFinger. As illustrated in Table 4, the proposed algorithm also outperforms the VeriFinger based recognition algorithm.

We next compare the feature extraction and matching with deep learning architectures. The results are compared with the following two algorithms:

(a) **Chopra et al. [7]:** Post semantic segmentation using VGG SegNet, we extract features from pre-trained ResNet50 (trained for ImageNet dataset classification) and matched using cosine similarity.
(b) **Lin and Kumar [23]:** The semantically segmented finger-selfies are used to train a 6-conv layer network using a Siamese configuration. During testing, the model is given finger-selfie pairs to evaluate if they belong to same finger or not. Note that, the original research used three siamese networks for 3D contactless fingerprint recognition for three different views of the finger (2 side views and one frontal). However, we used a single siamese network since finger-selfies in ISPFDv2 only have frontal views.

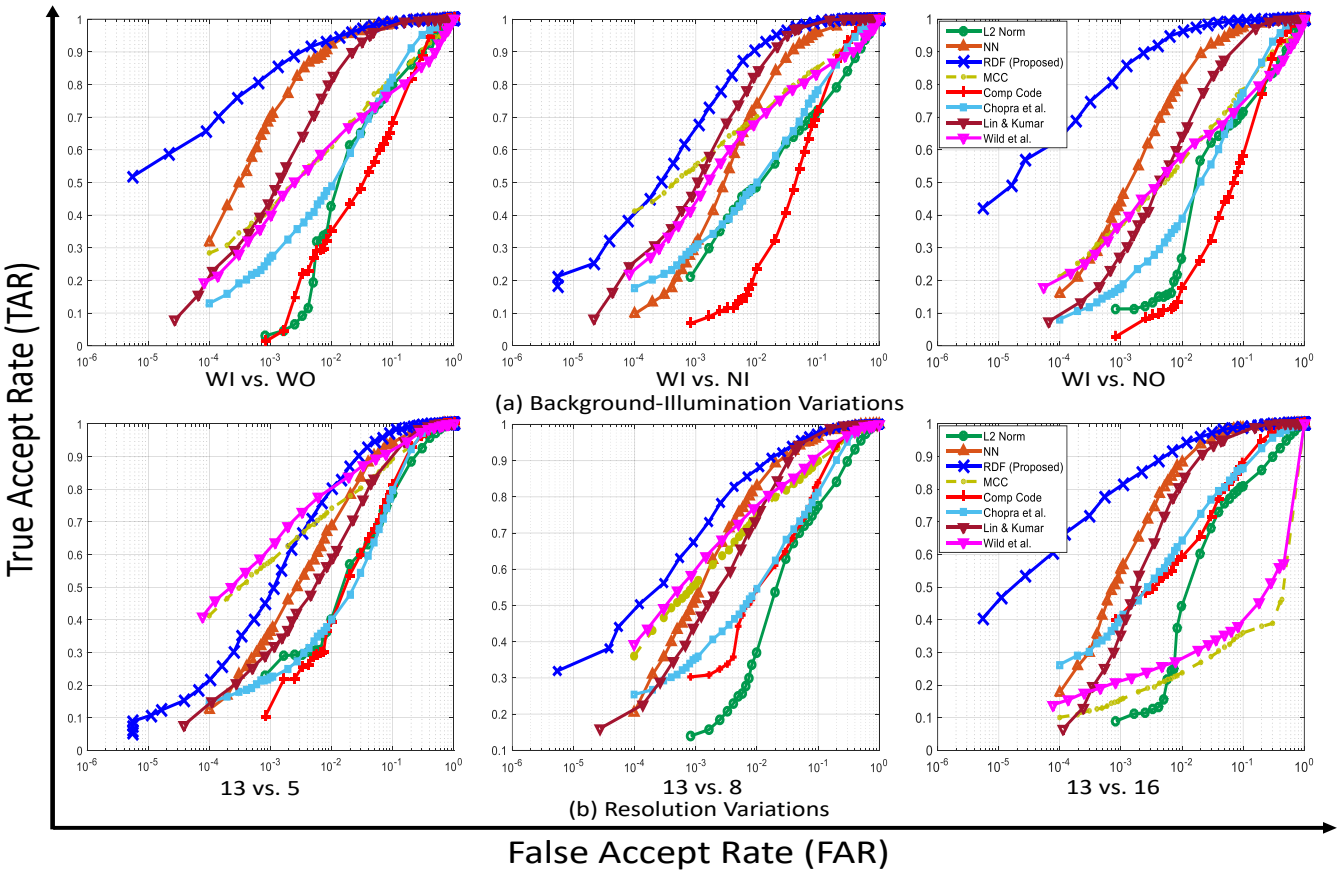


Fig. 11. ROC curves for different scenarios of feature extraction and matching algorithms. (Best viewed in color, x-axis in logscale).

Due to lack of source codes, the results on the ISPFDv2 database are reproduced by self-implementation of respective algorithms and summarized in Table 4 and Fig. 11. It can be observed that training deep networks significantly improves the recognition performance with EER in the range of 11.78% to 16.58% in Chopra et al. [7] and 3.77% to 9.79% in Lin and Kumar [23]). However, under different gallery-probe scenarios of ISPFDv2, the proposed algorithm has a lower EER by up to 5.63% in comparison to the state-of-the-art algorithm [23]. Also, compared to the baseline EER of 35.48% on the UNFIT database [7], the proposed algorithm outperforms with 31.74% EER.

Despite the effectiveness of deep learning architectures for different tasks of fingerprint matching such as pore extraction and liveness detection [15], [30], the research for contactless fingerprints is in nascent stages. *It is primarily because researchers have focused only on constrained scenarios of contactless fingerprint matching. In reality, variance introduced due to environmental and sensor variations can lead to an unconstrained acquisition*. Existing algorithms fail under these challenging cases. For instance, feature matching in Wild et al. [48] is performed using Verifinger. The scale variations due to 16 MP resolution results in a poor error rate of more than 44% in both *Exp. 2* and *Exp. 3*. Similarly, under highly unconstrained acquisition of natural outdoor (NO), all the four algorithms ([7], [23], [29], [48]) are found to be deficient in *Exp. 3*. For the same subset, the proposed DSN+RDF algorithm reports a competitive error rate of 2.49%. DSN, with filter banks varying in scale and orientation, encode ridge-valley contrast of finger-selfies. Hence, DSN+RDF based finger-selfie to finger-selfie and finger-selfie to livescan matching performance, aided by segmentation and enhancement framework, yields state-of-the-art performance for the ISPFDv2.

5 Conclusion

This paper highlights the challenges associated with the recognition of finger-selfies acquired by a smartphone camera. A Deep Scattering Network (DSN) based finger-selfie representation is proposed which is matched using an RDF classification. A finger-selfie segmentation and enhancement algorithm is also presented to assist the recognition process. To address the real-world challenges of: (i) background variation, (ii) environmental illumination, (iii) resolution of the camera, and (iv) matching legacy livescan fingerprints, IIIT-D SmartPhone Finger-selfie Database v2 (ISPFDv2) is created and made publicly available for the research community. The database consists of four sets with 19,456 images pertaining to 304 classes. The experimental results show a considerable performance improvement with the proposed algorithm over existing algorithms under both finger-selfie to finger-selfie matching and finger-selfie to livescan fingerprint matching.

6 Acknowledgment

The authors thank the reviewers for providing useful feedback on this research. The authors also acknowledge the participants for their cooperation during the data collection process. A. Malhotra

is partially supported through Visvesvaraya Ph.D. Scheme by MEITY, Govt. of India. A. Sankaran was partially supported by the TCS Ph.D. research fellowship. We extend our special thanks to A. Mittal for her support during the dataset collection.

REFERENCES

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (12):2481–2495, 2017.

[2] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12):5706–5722, 2015.

[3] A. Borji and L. Itti. Exploiting local and global patch rarities for saliency detection. In *IEEE Computer Vision and Pattern Recognition*, pages 478–485, 2012.

[4] J. Bruna and S. Mallat. Classification With Scattering Operators. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1561–1566, 2011.

[5] J. Bruna and S. Mallat. Invariant Scattering Convolution Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, 2013.

[6] R. Cappelli, M. Ferrara, and D. Maltoni. Minutia Cylinder-Code: A new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2128–2141, 2010.

[7] S. Chopra, A. Malhotra, M. Vatsa, and R. Singh. Unconstrained Fingerphoto Database. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 517–525, 2018.

[8] D. Deb, T. Chugh, J. Engelsma, K. Cao, N. Nain, J. Kendall, and A. K. Jain. Matching Fingerphotos to Slap Fingerprint Images. *arXiv preprint arXiv:1804.08122*, 2018.

[9] M. O. Derawi, B. Yang, and C. Busch. Fingerprint Recognition with Embedded cameras on Mobile Phones. In *Security and Privacy in Mobile Information and Communication Systems*, pages 136–147. Springer, 2012.

[10] E. Erdem and A. Erdem. Visual Saliency Estimation by Nonlinearly Integrating Features using Region Covariances. *Journal of Vision*, 13(4):1–20, 2013.

[11] X. He, D. Cai, and P. Niyogi. Laplacian Score for Feature Selection. In *Advances in Neural Information Processing Systems*, pages 507–514, 2006.

[12] A. Hern. Hacker fakes German minister's fingerprints using photos of her hands. https://tinyurl.com/y5qsrwlu. Accessed: 08-Nov-2019.

[13] T. K. Ho. Random Decision Forests. In *International Conference on Document Analysis and Recognition*, volume 1, pages 278–282, 1995.

[14] A. Inc. Evaluation of Contact versus Contactless Fingerprint Data (Final Report v2). https://www.ncjrs.gov/pdffiles1/nij/grants/245146.pdf, 2014. Accessed: 10-Feb-2020.

[15] H.-U. Jang, D. Kim, S.-M. Mun, S. Choi, and H.-K. Lee. DeepPore: Fingerprint Pore Extraction Using Deep Convolutional Neural Networks. *IEEE Signal Processing Letters*, 24(12):1808–1812, 2017.

[16] I. Jolliffe. *Principal Component Analysis*. Wiley Online Library, 2002.

[17] A. Kumar and Y. Zhou. Contactless Fingerprint Identification using Level Zero Features. In *IEEE Computer Vision and Pattern Recognition Workshops*, pages 114–119, 2011.

[18] C. Lee, S. Lee, J. Kim, and S.-J. Kim. Preprocessing of a fingerprint image captured with a mobile camera. In *Advances in Biometrics*, volume 3832, pages 348–355. 2005.

[19] D. Lee, K. Choi, H. Choi, and J. Kim. Recognizable-Image Selection for Fingerprint Recognition with a Mobile-Device Camera. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(1):233–243, 2008.

[20] G. Li, B. Yang, M. Olsen, and C. Busch. Quality Assessment for Fingerprints Collected by Smartphone Cameras. In *IEEE Computer Vision and Pattern Recognition Workshops*, pages 146–153, 2013.

[21] G. Li, B. Yang, R. Raghavendra, and C. Busch. Testing Mobile Phone Camera Based Fingerprint Recognition under Real-Life Scenarios. In *Norwegian Information Security Conference*, 2012.

[22] C. Lin and A. Kumar. Multi-Siamese Networks to Accurately Match Contactless to Contact-based Fingerprint Images. In *IEEE International Joint Conference on Biometrics*, pages 277–285, 2017.

[23] C. Lin and A. Kumar. Contactless and Partial 3D Fingerprint Recognition using Multi-view Deep Representation. *Pattern Recognition*, 83:314–327, 2018.

[24] C. Lin and A. Kumar. Matching Contactless and Contact-Based Conventional Fingerprint Images for Biometrics Identification. *IEEE Transactions on Image Processing*, 27(4):2008–2021, 2018.

[25] X. Liu, M. Pedersen, C. Charrier, F. A. Cheikh, and P. Bours. An Improved 3-step Contactless Fingerprint Image Enhancement Approach for Minutiae Detection. In *IEEE European Workshop on Visual Information Processing*, pages 1–6, 2016.

[26] A. Malhotra, S. Chhabra, M. Vatsa, and R. Singh. On Privacy Preserving Anonymization of Finger-selfies. In *IEEE Computer Vision and Pattern Recognition*, 2020.

[27] A. Malhotra, A. Sankaran, A. Mittal, M. Vatsa, and R. Singh. Fingerphoto Authentication Using Smartphone Camera Captured Under Varying Environmental Conditions. *Human Recognition in Unconstrained Environments*, pages 119–144, 2017.

[28] A. Malhotra, A. Sankaran, M. Vatsa, and R. Singh. Learning Representations for Unconstrained Fingerprint Recognition. *Deep Learning in Biometrics*, page 197, 2018.

[29] S. Minaee and Y. Wang. Fingerprint Recognition using Translation Invariant Scattering Network. In *IEEE Signal Processing in Medicine and Biology Symposium*, pages 1–6, 2015.

[30] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. Fingerprint Liveness Detection Using Convolutional Neural Networks. *IEEE Transactions Information Forensics and Security*, 11(6):1206–1213, 2016.

[31] N. Otsu. A Threshold Selection Method from Gray-level Histograms. *Automatica*, 11(285-296):23–27, 1975.

[32] K. B. Raja, R. Raghavendra, V. K. Vemuri, and C. Busch. Smartphone based Visible Iris Recognition using Deep Sparse Filtering. *Pattern Recognition Letters*, 57:33–42, 2015.

[33] R. Rajaram. Police grapple with establishing identity of the unknown dead. https://tinyurl.com/whxej9r. Accessed: 28-Nov-2019.

[34] A. Rattani and R. Derakhshani. On fine-tuning convolutional neural networks for smartphone based ocular recognition. In *IEEE International Joint Conference on Biometrics*, pages 762–767, 2017.

[35] D. J. Robertson, R. S. Kramer, and A. M. Burton. Face Averages Enhance User Recognition for Smartphone Security. *PloS one*, 10(3), 2015.

[36] M. Robnik-Šikonja and I. Kononenko. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning*, 53(1-2):23–69, 2003.

[37] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh. On Smartphone Camera based Fingerphoto Authentication. In *IEEE Biometrics Theory, Applications and Systems*, pages 1–7, 2015.

[38] A. Sankaran, M. Vatsa, and R. Singh. Latent fingerprint matching: A survey. *IEEE Access*, 2:982–1004, 2014.

[39] D. J. Sawicki and W. Miziolek. Human colour skin detection in cmyk colour space. *IET Image Processing*, 9(9):751–757, 2015.

[40] L. Sifre and S. Mallat. Rotation, Scaling and Deformation Invariant Scattering for Texture Discrimination. In *IEEE Computer Vision and Pattern Recognition*, pages 1233–1240, 2013.

[41] C. Stein, V. Bouatou, and C. Busch. Video-based Fingerphoto Recognition with Anti-spoofing Techniques with Smartphone Cameras. In *IEEE International Conference of the Biometrics Special Interest Group*, pages 1–12, 2013.

[42] C. Stein, C. Nickel, and C. Busch. Fingerphoto Recognition with Smartphone Cameras. In *IEEE International Conference of the Biometrics Special Interest Group*, pages 1–12, 2012.

[43] A. Taneja, A. Tayal, A. Malhorta, A. Sankaran, M. Vatsa, and R. Singh. Fingerphoto spoofing in mobile devices: a preliminary study. In *IEEE International Conference on Biometrics Theory, Applications and Systems*, pages 1–7, 2016.

[44] K. Tiwari and P. Gupta. A Touch-less Fingerphoto Recognition System for Mobile Hand-held Devices. In *IAPR International Conference on Biometrics*, pages 151–156, 2015.

[45] B. R. C. (UGC/CRC). The Hong Kong Polytechnic University (PolyU) High-Resolution-Fingerprint (HRF) Database. http://www4.comp.polyu.edu.hk/~biometrics/HRF/HRF_old.htm. Accessed: 08-Nov-2019.

[46] P. Wasnik, R. Ramachandra, M. Stokkenes, K. Raja, and C. Busch. Improved Fingerphoto Verification System Using Multi-scale Second Order Local Structures. In *IEEE International Conference of the Biometrics Special Interest Group*, pages 1–5, 2018.

[47] P. Wasnik, M. Stokkenes, M. Dunfjeld, R. Ramachandra, K. Raja, and C. Busch. Baseline Evaluation of Smartphone based Finger-photo Verification System: A Preliminary Study of Technology Readiness. In *Norwegian Information Security Conference*, 2018.

[48] P. Wild, F. Daubner, H. Penz, and G. F. Domínguez. Comparative Test of Smartphone Finger Photo vs. Touch-based Cross-sensor Fingerprint Recognition. In *IEEE International Workshop on Biometrics and Forensics*, pages 1–6, 2019.

[49] C. Wood. WhatsApp drug dealer caught by 'groundbreaking' work. http://www.bbc.com/news/uk-wales-43711477. Accessed: 11-Nov-2019.

[50] D. Xiang and B. Zhong. Scale-space saliency detection in combined color space. In *IEEE Chinese Automation Congress*, pages 726–731, 2015.

[51] H. Zeng and Y.-m. Cheung. Feature Selection and Kernel Learning for Local Learning-based Clustering. *IEEE Transactions on Pattern Analysis*

and Machine Intelligence, 33(8):1532–1547, 2010.

[52] J. Zhang and S. Sclaroff. Saliency Detection: A Boolean Map Approach. In *IEEE International Conference on Computer Vision*, pages 153–160, 2013.

[53] Q. Zheng, A. Kumar, and G. Pan. Suspecting Less and Doing Better: New Insights on Palmprint Identification for Faster and More Accurate Matching. *IEEE Transactions on Information Forensics and Security*, 11(3):633–641, 2016.

Aakarsh Malhotra received the B.Tech. degree in computer science from the IIIT-Delhi, India, in 2015. He is currently pursuing his doctoral degree at the IIIT-Delhi, India. He received Overseas Research Fellowship from IIIT-Delhi to visit West Virginia University as a visiting research scholar in 2016. His research interests include application of machine learning and image processing in biometrics. He is a recipient of the Visvesvaraya Ph.D. fellowship and is a student member of the IEEE, Computer Society.

Anush Sankaran received the B.Tech. degree in computer science from CIT, Coimbatore, India, in 2010. He completed his Ph.D. degree with the IIIT-Delhi, India. His research interests include deep learning, image processing, and their applications in biometrics. He was a recipient of the TCS Ph.D. Research Fellowship from 2010 to 2015, and the Best Poster Awards in the IEEE BTAS 2013 and the IEEE IJCB 2014. He worked as a Research Scientist at IBM Research, India and is currently working as Senior Research Scientist at Deeplite, Canada.

Mayank Vatsa received the M.S. and Ph.D. degrees in Computer Science from West Virginia University, USA, in 2005 and 2008, respectively. He is currently a Professor with IIT Jodhpur, India, and the Project Director of the Technology and Innovation Hub on Computer Vision and Augmented & Virtual Reality under the National Mission on Cyber Physical Systems by the Government of India. He is also an Adjunct Professor with IIIT-Delhi, India and West Virginia University, USA. His areas of interest are biometrics, image processing, machine learning, computer vision, and information fusion. He has co-edited books on Deep learning in Biometrics and Domain Adaptation for Visual Understanding. He is the recipient of the prestigious Swarnajayanti Fellowship from the Government of India, the A. R. Krishnaswamy Faculty Research Fellowship at the IIIT-Delhi, and several best paper and best poster awards at international conferences. He is an Area/Associate Editor of Information Fusion and Pattern Recognition, the General Co-Chair of IJCB 2020, and the PC Co-Chair of IEEE FG2021. He has also served as the Vice President (Publications) of the IEEE Biometrics Council where he started the IEEE Transactions on Biometrics, Behavior, And Identity Science.

Richa Singh received the Ph.D. degree in computer science from West Virginia University, Morgantown, USA, in 2008. She is currently a Professor at IIT Jodhpur, India, and an Adjunct Professor with IIIT-Delhi and West Virginia University, USA. She has co-edited book Deep Learning in Biometrics. Her areas of interest are pattern recognition, machine learning, and biometrics. She is a fellow of IAPR and a Senior Member of IEEE and ACM. She was a recipient of the Kusum and Mohandas Pai Faculty Research Fellowship at the IIIT-Delhi, the FAST Award by the Department of Science and Technology, India, and several best paper and best poster awards in international conferences. She is currently serving as a Program Co-Chair of IJCB 2020 and General Chair of FG 2021. She is also the Vice President (Publications) of the IEEE Biometrics Council and Associate Editor-in-Chief of Pattern Recognition.