Saket Pateriya\textsuperscript{1}, Vineet Ghule\textsuperscript{1}, Divyang S Jadav\textsuperscript{1}, Deekshant Kumar\textsuperscript{1}, Karthick V.\textsuperscript{1}, and Vikram M Gadre\textsuperscript{1}

\textsuperscript{1}Affiliation not available

May 25, 2024
SSNet: Novel Approach for Fingerprint Recognition in Data-Scarce Scenarios

Saket Pateriya, Vineet Ghule, Divyang S. Jadav, Deekshant Kumar, Karthick V., and Vikram M. Gadre

Abstract—State-of-the-art (SOTA) models typically use large datasets for pre-training and then fine-tune on smaller datasets for better performance. However, the high computational cost can be a barrier for many researchers. There is a need to focus on data size-independent models suited for data-scarce scenarios, which is essential for tasks like fingerprint recognition and could make research more accessible and generalizable in resource-limited environments. With this aim, this paper presents a novel approach to the difficulties in contactless fingerprint recognition, particularly with scarce and poor-quality challenging dataset images due to contactless acquisition. Our proposed system uses a 'Scattering using a Shearlet Network (SSNet)' to extract fingerprint features and a score-level fusion scheme to improve authentication accuracy. In contrast to the computationally expensive and mathematically less transparent dense deep learning networks such as vision transformers, attention networks, deep learning-based hybrid approaches, etc., SSNet is an economical framework with fixed filters. The SSNet is a replacement to the Scattering Wavelet Network (SWN) that utilizes a Complex Morlet Wavelet (CMW). Our model significantly improves verification and identification accuracy over SOTA approaches, particularly with scarce and poor-quality challenging datasets.

Index Terms—Contactless Fingerprint Recognition, Multiresolution Analysis, Scattering using Shearlet Network.

I. INTRODUCTION

Fingerprint recognition is a mechanism of verifying a user’s identity by assessing the distinctiveness of their fingerprints patterns. Recently, there has been a surge of interest among researchers in contactless fingerprint recognition, mainly after the COVID-19 pandemic, to prevent the spread of viruses and diseases [1], [8]. However, contactless fingerprint acquisition poses several challenges, including variable lighting conditions, nonuniform backgrounds, camera focus adjustment, proper distance maintenance, scaling, rotations, contactless fingerprint data scarcity, and various class variabilities [1], [19] (see Fig.1, and Fig.2).

The fingerprint minutiae extraction model designed by MENet and FingerNet better performed on standard contact-based fingerprint dataset [11], [13], indicating that fingerprint recognition using traditional minutiae points has poor performance. In [24], authors proposed a multi-Siamese network, which led to a breakthrough step enabling correct matching between contactless and traditional contact-based fingerprint images while using handcrafted features with deep learning insights. In [12], the pre-trained ResNet models are fine-tuned on the PolyU fingerprint dataset to provide high recognition rates based on the effectiveness of deep transfer learning. On the other hand, PoreNet [10] presents a new CNN-based pore descriptor for high-resolution fingerprint images to precisely identify the pores in fingerprints. Recently attention-based Vision Transformers (ViTs) combined with CNN pave the way for new headways in biometric recognition [21]–[23], [29], showing the model’s suitability in catching complex
patterns that underlie biometric data. Contactless biometric recognition using CNNs overlooks various factors, including the best number of convolutional layers, the stride values, and the selection of a non-linearity function. We can say that the choice of optimization and loss functions for training CNNs are mathematically non-correlated. It is worth noting that CNNs exclude crucial local time-frequency information required in making critical decisions and computations. The dense and data-intensive structures of CNNs mandate a significant training time requirement [3], [5], [8], [24].

Recent advancements in wavelet theory underscore the significant leap from one-dimensional to two-dimensional applications, particularly in enhancing image analysis techniques. The crux of this evolution lies in extending Multiresolution Analysis (MRA) to two-dimensional spaces, which amplifies wavelet transformations utility for complex image data, including fingerprints. It focuses on multiscale geometric analysis with wavelets, ridgelets, curvelets, and shearlets, significantly improving feature extraction and classification accuracy suited for biometric applications, including fingerprint recognition [2], [18], [28]. The drawback of MRA such as wavelets, ridgelets, curvelets is that they use rotations, which distort the integer grid in the discrete domain and do not have a single generating function [16], [18], [25], [26]. On the other hand, shearlets have emerged as a potent alternative to overcome these limitations since they scale with anisotropicity, detect wavefront sets, and converges quickly [16], [18], [28]. Their localization properties allow shearlets to provide exact control of scale, direction, and position in a manner quite suitable for applications in biometric fields, such as fingerprint recognition.

Morlet-type wavelet filter banks are mostly applied in the SOTA Two Dimensional (2D) Scattering Wavelet Network (SWN) under the limited-data setting, as discussed in [3], [17], [27]. The rotations in SWN that use Morlet wavelet tend to distort the discrete spatial grid. However, shearlets preserve the discrete spatial grid since they apply shearing instead of rotation. Without the discrete domain grid distortion, the shearlet transform is more straightforward to implement from a single generator function, offering better multiresolution, direction selectivity, and position control than a filter bank with Morlet wavelets [18], [27], [30]. Our model proposes new methodologies that leverage techniques to make fingerprint recognition more efficient and computationally economical by automating some decisions using multi-scale, direction selectivity on anisotropic scaling, affine, and stable to small deformations, which constitutes a Parseval frame and offer optimally sparse features representation to overcome the limitations of deep learning networks and SWN [1], [2], [8], [18], [28].

To test our hypothesis, we are evaluating a solution presented to us by the Indian Army to address the challenge of securing food consignments during testing and dispatch. The process involves bulk procurement of food rations, followed by inspection and quality control by the Army. Random samples are drawn, and the remaining stock from which the samples are taken is sealed. The samples are then transported back to the Army lab for further testing using various unsecured means if transport. Once the samples pass inspection, the sealed stock is shipped to its destination across India. While the samples are sealed and signed to ensure their sanctity during transit, the current system presents a vulnerability. Malicious actors could potentially tamper with or replace the samples with inferior quality products, compromising the quality of the entire stock.

To address this security gap, the Indian Army proposed using biometric validation as an additional security measure. This approach involves linking unique biometric identifiers, such as fingerprints, to the samples, sealed containers, and remaining stock during transport. Since biometric data is unique to each individual and difficult to replicate, this method significantly reduces the risk of tampering. Our proposed implementation leverages contactless fingerprint capture. A blue-ink fingerprint is captured on white paper and securely attached to the consignment. Fingerprint verification is achieved by matching a photograph of the captured fingerprint with a pre-stored database containing the authorized individual’s fingerprint.

The prime contributions of this work are detailed as follows:

- We have developed a dataset based on fingerprints and thumbprints captured using blue ink on paper from both hands of 15 subjects.
- We have developed a novel architecture, SSNet, scattering wavelet network using fixed weights nonseparable shearlet filters for better directional selectivity and efficient feature extraction from fingerprint images. To the best of our knowledge, this is the first application of SSNet for fingerprint feature extraction.
- We have developed a multiple unit and multiple instance fingerprint recognition system by employing a score-level fusion technique to identify and verify a person using multiple fingerprint images, which presents authentication improvement over single unit fingerprint systems.

The rest of the paper is organized as follows. The description of proposed system architecture is detailed in section II. The experimental results and their analysis is discussed in section III. The paper is concluded with future direction in section IV.

II. PROPOSED METHODOLOGY

Scaling, translation, and rotation are rigid transformations and may introduce inter-class and intra-class variability in the images [3], [5], [8]. For example, in digit classification, a small deformation may lead to a change of class. SWN is
invariant to rigid transformation, additive noise perturbation, and small deformations, preserving higher-order moments for inter-class discrimination. We refer the reader to reference [3] to understand SWN in detail. In the following sections we have explained various sections of fingerprint recognition system outlined in Fig. 3.

A. Cone Adaptive Shearlet Filters

Inspired by the seminal works, we have borrowed the notations from [16], [18] to describe the details of the shearlet system that we are using in this work.

This framework allows us to analyze images using filters with different scales and shear levels, capturing features across various scales and orientations with high precision. We write our shearlet for a particular (relative) scale level \( j \) and shear level \( k \) as

\[
\hat{\psi}_{j,k}(\xi_1, \xi_2) = 2^{-\frac{3j}{2}} \hat{\psi}_1(4^{-j} \xi_1) \hat{\psi}_2 \left( 2^j \frac{\xi_2}{\xi_1} + k \right)
\]  

where \( j = 0, \ldots, J - 1 \), \( |k| \leq 2^j \), \( k \in \mathbb{Z} \). \( J \) denotes the total number of scale levels which is given as input while constructing the shearlet system. The absolute scale of the shearlets is chosen depending upon the image grid size and \( J \), such that shearlets of scale \( j \) can be downsampled (in space) by a factor of \( 4^{j-1-j} \) and the lowpass filter \( \phi \) can be downsampled (in space) by a factor of \( 4^j \) without any spectral aliasing.

For symmetrical treatment of the horizontal and vertical cones of the frequency plane, it is tiled into three sets of shearlets \( \hat{\Psi}_h, \hat{\Psi}_v, \hat{\Psi}_x \), defined as follows and visualized in frequency and spatial domains in Fig. 4:

\[
\hat{\Psi}_h_j := \{ \hat{\psi}_{j,k}(\xi_1, \xi_2) : |k| \leq 2^j - 1 \} \quad (2)
\]

\[
\hat{\Psi}_v_j := \{ \hat{\psi}_{j,k}(\xi_2, \xi_1) : |k| \leq 2^j - 1 \} \quad (3)
\]

\[
\hat{\Psi}_{x,j}(\xi_1, \xi_2) := \begin{cases} 
\hat{\psi}_{j,k}(\xi_1, \xi_2) & \text{if } \xi_1 \geq \xi_2 \\
\hat{\psi}_{j,k}(\xi_2, \xi_1) & \text{if } \xi_1 < \xi_2 
\end{cases} \quad (4)
\]

\[
\hat{\Psi}_{x} := \{ \hat{\psi}_{x,j}(\xi_1, \xi_2) : |k| = 2^j \} \quad (5)
\]
We also define the combined shearlet system as follows

\[ \hat{\Psi}_j := \hat{\Psi}_{h_j} \cup \hat{\Psi}_{v_j} \cup \hat{\Psi}_{x_j} \]  

(6)

Over a two-dimensional grid of size 512 × 512, for each \( j = 0, 1, \ldots, J - 1 \) (where \( 4^J \leq 512 \)) there are \( |\hat{\Psi}_j| = 2^j + 2 \) shearlet filters, which are along different directions. In total, including the lowpass filter \( \phi \), there are \( 1 + \sum_{j=0}^{J-1} 2^j + 2 \) filters comprising the shearlet system. This shearlet system comprehensively covers the entire frequency plane, which we confirmed by summing the spectral energies \( = \sum_{j,l} |\hat{\psi}_{j,l}|^2 = 1 \), as shown in the right top corner in Fig. 4.

**Fig. 4:** Shearlet filters for scale \((J=2)\) across frequency and spatial domains and also in the right top corner shows total spectral energy.

### B. Fundamental Theory of SSNet Feature Extractor

The method exploits the shearlet filters and scattering transform to extract anisotropic features in images in a robust manner and thus furnishes a light yet powerful alternative to conventional deep learning models with no need for training.

We adopted the Kymatio SWN structure [3], [17], which is then tailored for a shearlet system. Unlike the Kymatio implementation, where the number of directions is fixed, the shears or directions vary with scale \( j \) in our model since we use a shearlet system. The SSNet features maps are obtained as outputs of different rooted paths in the SSNet tree. The feature maps of layers 0, 1 and 2 are the outputs of 0, 1 and 2 length paths respectively. The set of feature maps of layers 0, 1 and 2 (along paths specified by the sequence of scales of filters applied along them) are defined as follows by equations (7), (8), and (9) respectively:

\[
C_J() := \{ x(u) * \phi(u) \downarrow 4^J \}
\]

(7)

\[
C_J(j_1) := \{ |x(u) * \psi_{j_1}(u) \downarrow 4^{J-1-j_1}| * 4^{J-1-j_1} \phi (4^{J-1-j_1} u) \downarrow 4^{j_1+1} : \hat{\psi}_{j_1} \in \hat{\Psi}_{j_1} \}
\]

(8)

\[
C_J(j_1,j_2) := \left\{ \left( |x(u) * \psi_{j_1}(u) \downarrow 4^{J-1-j_1}| * 4^{J-1-j_1} \phi (4^{J-1-j_1} u) \downarrow 4^{j_1+1} : \hat{\psi}_{j_1} \in \hat{\Psi}_{j_1} \right) \right\}
\]

(9)

The resultant downsampling along every path stays consistent and seen by the following relation:

\[
J = (J-1 - j_1) + (j_1 - j_2) + \ldots + (j_m - 1 - j_m) + (j_m + 1)
\]

It follows from the constraint \( J - 1 \geq j_1 \geq \ldots \geq j_m \).

1) **First Order SSNet \((S_{sn1})\):** In the first order of the SSNet, filters within the shearlet system are utilized to capture different bands of the frequency region. In the spatial domain, these filters yield high correlation coefficients in areas where the fingerprint’s contours align with the filter’s orientation. The process is then made robust to small local deformations through the use of a modulus operation and averaging function, \( \phi \), thereby providing a degree of invariance to local translation and local rotation. The output is a set of spatially downsampled (by a factor of \( 4^J \)) feature maps with one feature map per shearlet. The first-order SSNet can capture information about the local orientation and spacing between the ridges on a fingerprint, as demonstrated in equation (10) and Fig. 3.

\[
S_{sn1} = C_J() \cup \left( \bigcup_{j = 0}^{J-1} C_J(j_1) \right)
\]

(10)

2) **Second Order SSNet \((S_{sn2})\):** The \( S_{sn2} \) operation is a complex multi-layer analysis of images by their decomposition into components belonging to different scales to extract image details at various resolutions. In simple terms, the modulus of the first layer carries amplitudes of the signal filtered (using a particular wavelet). In the second layer, the wavelet would correlate with patterns in this amplitude signal, thereby bearing similarity to that wavelet in the spatial domain. Using some wavelets that vary smoothly in space (i.e., in the low-frequency area) will respond to capturing substantial information about the local curvature of the ridges. The second-order equation (11) and Fig. 3 show this change concerning space in the orientation and spacing of ridges.

\[
S_{sn2} = C_J() \cup \left( \bigcup_{j = 0}^{J-1} C_J(j_1) \cup \left( \bigcup_{j_2 = 0}^{j_1} C_J(j_1,j_2) \right) \right)
\]

(11)

The equations for \( S_{sn2} \) and \( S_{sn1} \) describe the complete first order and second order sets of SSNet feature maps. To reduce the feature dimensionality and computational complexity, we have used a subset of the set of these feature maps in some of the model approaches described later.
III. Experimental Results and Discussion

A. Datasets

We have validated our model using the Indian Institute of Technology Bombay Contactless Fingerprint Dataset (IITB-CFD) [7], a standard public dataset [4], [7], [15], [19], [29], and the Indian Army Inkprint Dataset. Fig. 1 and Fig. 2 are the sample images from IITB-CFD and the Indian Army Inkprint datasets, respectively. The following subsections describe the organization of datasets.

1) IITB-CFD [7]: The IITB-CFD comprises fingerprints of 200 individuals captured indoors using specialized non-contact imaging technology. Each subject provided 4 samples of a single finger, totaling 800 images. These images often exhibit slight blurriness, making the fingerprint ridges not always distinctly visible. For our experiments, this dataset was divided into 600 training images and 200 testing images, maintaining a 3:1 ratio of training to test images per individual.

2) Indian Army Inkprint Dataset: The Indian Army Inkprint dataset contains ink-based fingerprints collected voluntarily from the Indian Army personnel in Mumbai, India, in September 2022. Each participant provided four samples on white paper from each finger and thumb of both hands. A total of 15 subjects participated in dataset creation, resulting in 600 images (15 Persons × 5 fingers × 4 images of each finger × 2 hands). The physical prints were digitized at 600 dpi, and each image was formatted to 750 x 900 pixels resolution. The dataset is relatively small and has many challenges in processing as most images are partially smudged and have random horizontal and vertical cuts, distortion, ink spread, and background variations. We have captured these inkprints using an ordinary Android mobile phone. Hence, touchless acquisition variations are also added to the dataset images. For our experiment, we are using only left- and right-hand thumbprints from this dataset (i.e., 120 images). We have used three thumbprint images, each from the right and left hand of a subject, for training and one for testing.

B. Preprocessing

1) IITB-CFD: We have applied three preprocessing steps: histogram equalization, adaptive thresholding, and finally, fingerprint enhancement using Gabor filters, as shown in Fig. 3. Histogram equalization can be omitted to reduce execution time since it does not cause a significant difference in the preprocessed output. Adaptive thresholding locally binarizes the ridges and valleys. The binarization of the images may add spurious noise and minutiae points. Hence, the final fingerprint enhancement is done using oriented Gabor filters to extend the ridges [1].

2) Indian Army Inkprint Dataset: The preprocessing pipeline for the Indian Army inkprint dataset initiates with image normalization and identifying the region of interest. In this step, the image is normalized to ensure that the ridge regions have zero mean and unit standard deviation. The image is then divided into blocks, and the standard deviation within each block is calculated to determine the fingerprint region of interest (ROI). Following this, the orientation of the ridges in the fingerprint is estimated using image gradients and a block processing approach. This results in an orientation image that indicates the direction of the ridges. Subsequently, the ridge frequency is estimated. This process involves analyzing blocks of the image to determine the ridge count in each block, utilizing the previously identified ridge regions and orientation data. This step provides an image indicating the estimated ridge frequencies and a median frequency value for the valid regions of the image. The next vital step is enhancing the fingerprint image using oriented Gabor filters [1], [9].

C. Approaches Followed

The filters are compactly supported in frequency but not space. In the spatial domain, these filters have an overlap with the repeated signal (image) at the boundaries. To reduce this overlap, the preprocessed images are zero-padded to 512 x 512 images. The padded images are used to extract features at various layers and scales. In this approach, effectiveness can be gauged by taking two arrays of features and measuring the distance between them using the Euclidean metric. The performance has been visually interpreted by plotting the verification performance; the Receiver Operating Characteristics (ROC) along with Equal Error Rate (EER) and Precision-Recall (PR) score curves for pairwise image features, providing insights into the performance of the proposed methods. The Cumulative Match Characteristic (CMC) score is calculated by minimizing and ranking the Euclidean distance metrics across all classes or individuals, followed by averaging all test images. The Cumulative Match Characteristics (CMC) curve evaluates the identification performance. [14], [15].

A parent-to-child branch in the SSNet tree consists of filtering by a specific filter, followed by down-sampling (based on filter scale) and then taking modulus operation in the spatial domain. At the first layer, the network extracts information related to edge directionality. At the second layer, which takes the children corresponding to the first layer branches as inputs, the filters (of relatively lower scale and having support in lower frequency bands) capture information about the change in the ridge direction with respect to the image space.

1) Baseline SWN Model (Order, O = 2, Scale, J = 5): We first used the Kymatio library’s scattering network with scaled and rotated complex Morlet wavelets. We used scales, J = 5, for the experimentation, giving five filter sets corresponding to the scales, j = 0, 1, 2, 3, 4. Each such set contains 8 filters oriented in one of the 8 directions on the frequency plane. The first layer of the network uses all these filters in a quantity of 8 × 5 = 40. The second layer of the network uses filters at scales lower than their corresponding parent filters. For example, the directed outputs from a filter at scale, j = 3, would come from filtering with scales, 0, 1, and 2. The total number of 2D feature maps developed through this process is 681. The input is averaged with the Gaussian scaling function in the zeroth layer, providing one feature map. The padded image is converted into a feature map array with dimensions 681 × 16 × 16, the latter size attributed to the division of 512 by 2^5 = 16. Hence, the total dimensionality of the feature array is 174336. The scales and orientations well covered in this comprehensive approach make it a strong feature extraction method for any complex image datasets.
2) SSNet (Order, O = 1, Scale, J = 1): A notable aspect of this approach is the absence of a secondary filtering layer. Among the array of models we have investigated, this one stands out for its simplicity and reduced computational complexity. It incorporates just one layer, containing eight filters, which facilitates a swift processing and conversion time. Specifically, this model transforms an image array of $512 \times 512$ pixels into a feature array with the dimensions of $8 \times 32 \times 32$, amounting to a total feature dimensionality of 8192. We use only layer 1 coefficients in this approach.

$$S_{sn1} = C_J(J - 1)$$  \hspace{1cm} (12)

3) SSNet (Order, O = 2, Scale, J = 1): This SSNet gives us 104 image (each of size $32 \times 32$) feature maps ($8 + 8 \times (4+8)$), making the features in this model a comprehensive superset of those from the first model. We use layer 1 and layer 2 coefficients for this approach.

$$S_{sn2} = C_J(J - 1) \cup \left( \bigcup_{j_2=0}^{J-1} C_J(J - 1, j_2) \right)$$  \hspace{1cm} (13)

4) SSNet (Order, O = 2, Scale, J = 3): The model generates a total of 589 2D feature maps. Here, we also account for including the zeroth layer, achieved by averaging the input with the shearlet averaging function. This approach applies a transformation given by equation (11) with $J = 3$.

This approach effectively transforms a $512 \times 512$ padded image array into a $589 \times 8 \times 8$ feature array. The reduced size, $8 \times 8$, results from dividing 512 by $4^3$ (the scale factor is 4 as successive scales increase by a factor of 4). The total dimensionality of this feature array is 37696. We used layers 0, 1, and 2 coefficients in this approach.

5) SSNet (Order, O = 2, Scale, J = 4): This approach effectively transforms a $512 \times 512$ padded image array into a $2541 \times 2 \times 2$ feature array. The reduced size, $2 \times 2$, results from dividing 512 by $4^4$ (the scale factor is 4 as successive scales increase by a factor of 4). The total dimensionality of this feature array is 10,164. We use layer 0, layer 1 and layer 2 coefficients in this approach. This approach applies a transformation given by equation (11) with $J = 4$.

D. Score Level Fusion

In biometrics [6], [14], [20], score-level fusion is a valuable technique for enhancing system performance. It involves the combination of scores from fingerprint recognition systems dedicated to the right and left thumbs, a new distance-based score to leverage the complementary information. Each system generates a score based on its feature vector, explicitly trained on images from the respective thumb. These scores are fused through a convex combination, assigning equal weights to both the right and left thumb scores. We have applied the score level fusion technique only for the Indian Army Inkprint datasets as IITB-CFD does not provide multiple fingerprint information. The mathematical representation of the fused score, $S_{fused}$, is given by:

$$S_{fused} = 0.5 \cdot S_R + 0.5 \cdot S_L$$  \hspace{1cm} (14)

where $S_R$ and $S_L$ represent the scores from the right and left thumb biometric systems, respectively. This method leverages each thumb’s complementary information to improve the accuracy and reliability of biometric verification and identification processes.

The results are reported in the form of Table and graphical representation as follows:

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Verification</th>
<th>Identification</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSNet- (O=1, J=1)</td>
<td>0.96</td>
<td>10.33</td>
<td>0.62</td>
</tr>
<tr>
<td>SSNet- (O=2, J=1)</td>
<td>0.97</td>
<td>9.88</td>
<td>0.66</td>
</tr>
<tr>
<td>SSNet- (O=2, J=3)</td>
<td>0.99</td>
<td>5.13</td>
<td>0.79</td>
</tr>
<tr>
<td>SSNet- (O=2, J=4)</td>
<td>0.97</td>
<td>8.50</td>
<td>0.63</td>
</tr>
</tbody>
</table>

The mathematical representation of the fused score, $S_{fused}$, is given by:

$$S_{fused} = 0.5 \cdot S_R + 0.5 \cdot S_L$$  \hspace{1cm} (14)
TABLE II: Performance Metrics for Left Thumbs (LT) Images on Indian Army Data

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Verification</th>
<th>Identification</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC</td>
<td>EER%</td>
<td>AUPR</td>
</tr>
<tr>
<td>SWN- (O=2, J=5)</td>
<td>0.80</td>
<td>28.87</td>
<td>0.53</td>
</tr>
<tr>
<td>SSNet- (O=2, J=3)</td>
<td>0.84</td>
<td>23.33</td>
<td>0.61</td>
</tr>
<tr>
<td>SSNet- (O=2, J=4)</td>
<td>0.91</td>
<td>17.32</td>
<td>0.71</td>
</tr>
</tbody>
</table>

TABLE III: Performance Metrics for Right Thumbs (RT) Images on Indian Army Data

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Verification</th>
<th>Identification</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC</td>
<td>EER%</td>
<td>AUPR</td>
</tr>
<tr>
<td>SWN- (O=2, J=5)</td>
<td>0.84</td>
<td>25.60</td>
<td>0.56</td>
</tr>
<tr>
<td>SSNet- (O=2, J=3)</td>
<td>0.84</td>
<td>23.33</td>
<td>0.61</td>
</tr>
<tr>
<td>SSNet- (O=2, J=4)</td>
<td>0.84</td>
<td>17.74</td>
<td>0.64</td>
</tr>
</tbody>
</table>

TABLE IV: Performance Metrics with Score Level Fusion on Indian Army Data

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Verification</th>
<th>Identification</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUROC</td>
<td>EER%</td>
<td>AUPR</td>
</tr>
<tr>
<td>SSNet-(O=2, J=4)-LT</td>
<td>0.91</td>
<td>17.32</td>
<td>0.71</td>
</tr>
<tr>
<td>SSNet-(O=2, J=4)-RT</td>
<td>0.91</td>
<td>17.32</td>
<td>0.71</td>
</tr>
<tr>
<td>SSNet-(O=2, J=4)-Fusion</td>
<td>0.965</td>
<td>11.11</td>
<td>0.84</td>
</tr>
</tbody>
</table>

IV. Conclusion

The work presented in this paper investigates the verification and identification performance of various SWN models. The main goal is to observe the performance difference after integrating non-separable shearlet filters instead of CMW in SWN. Increasing the scales and layers makes the model complex due to the increase in feature vector size but improves the verification and identification performance, as evident from Tables I, II, III, and IV and Figures 5, 6, 7a, and 7b.

Our model with the order, O = 2, and scale, J = 3 and J = 4, outperformed the Kymatio library’s SWN and other SOTA approaches on both the datasets, as evident from Tables I, II, III, and IV and Figures 5, 6, 7a, and 7b. Evaluating our model on IITB-CFD provided the highest performance parameter metrics values with Receiver Under the Area Operating Characteristics (AUROC) of 0.99, Area Under the Precision Recall curve (AUPR) of 0.81, EER of 5.13%, 97% Rank-1 accuracy, and 98% Rank-5 accuracy. Similarly, we have evaluated our model on the Indian Army Inkprint dataset using the same performance metrics and observed improved performance with the score level fusion technique. We have achieved performance metrics values of an AUROC of 0.965, AUPR of 0.84, EER of 11.11%, 93.3% Rank-1 accuracy, and 100% Rank-5 accuracy on the Indian Army Inkprint datasets. The improved accuracy using the score level fusion technique accounts for greater and complementary information being available from a single subject for verification and identification compared to unimodal systems.

Unlike SOTA deep learning models, our model employs fixed-weight shearlet filters that can adapt to the new dataset without training. Thus, our model’s performance is independent of the dataset size and can be easily scaled as per the dataset size. We have also presented that our model’s choice of order and scale is mathematically supported, thus paving a path for explainable deep learning frameworks. In future, we will explore the hybrid application of machine learning and deep learning models, such as a Hybrid SSNet-DL (HSSNet) and a Learnable SSNet (LSSNet).

ACKNOWLEDGEMENT

We are thankful to the National Center of Excellence in Technology for Internal Security, Indian Institute of Technology Bombay (NCEITIS-IITB), and the Indian Army for supporting this work. We also thank Dr. Azhar Yousuf, who is currently working as a Postdoc fellow at the Department of Computer Science and Engineering, IIT Bombay, for our deliberations regarding this research work, which greatly helped in improving the article to the present stage.

REFERENCES


Saket Pateriya received the M.Tech. degree in instrumentation and control engineering with process control specialization from Netaji Subhash Institute of Technology (NSIT- University of Delhi), Delhi, India, in 2013. He was an Assistant Professor in the Discipline of Electrical and Electronics Engineering at the SRM University Chennai NCR campus in Delhi, India, from September 2013 to December 2018. He is currently pursuing a Ph.D. degree in the Discipline of Electrical Engineering at the Indian Institute of Technology (IIT) Bombay, Mumbai, India since January 2019. His current research interests lie in signal processing, focusing on multiresolution and machine learning approaches, particularly in biometrics.

Vineet Ghuule is currently pursuing the B.Tech. + M.Tech. (Dual) degree in Electrical Engineering with Communication and Signal Processing specialization, a minor degree in Computer Science and Engineering and a minor degree in Artificial Intelligence and Data Science at the Indian Institute of Technology (IIT) Bombay, India. His current research interests include signal processing, focusing on multiresolution approaches.

Divyang S. Jadav has a M.Tech. degree in Electrical Engineering with Microelectronics specialization from Indian Institute of Technology Bombay (IITB), India. He is currently working as an Assistant Professor at the School of Technology Management & Engineering at Narsee Monjee Institute of Management Studies, Navi-Mumbai campus. He is also pursuing a PhD degree in Electrical Engineering at the IITB. His research interest is in the signal and image processing field.

Deekshant Kumar is the Director of Software Engineering at HERE Technologies. He has been recently recognized as 40 under 40 by AI Magazine-2023. He currently works at HERE Technologies and leads the Technology Innovation Lab team globally and the tooling team that makes maps for OEMs such as BMW, Audi etc., and has 14 patents to his credit. He is also pursuing a PhD degree from IIT Bombay. His current research interests lie in signal processing, focusing on multiresolution and machine learning approaches.

Karthick V. was commissioned into the Indian Army from Officers Training Academy, Chennai, into the 10th Battalion of The Dogra Regiment and had extensive experience in the Kashmir Valley in Counter-Terrorist/Counter-Insurgency operations along the Line of Control. Major Karthick is currently serving with the Army Service Corps, the Logistics Arm of the Indian Army, which looks after the Army Fleet of vehicles and the supply of food rations to the troops. He is now involved in the quality control of rations supplied by vendors across India. He has collaborated with a research team from IIT Bombay to develop a touchless biometrics validation system to enhance the security of stock tendered while ensuring the sanctity of the supply chain.

Vikram M. Gadre received the B.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology (IIT), New Delhi, India, in 1989 and 1994, respectively. He is currently a Professor with the Department of Electrical Engineering, IIT Bombay, Mumbai, India. His research interests broadly include communication and signal processing, with emphasis on multiresolution approaches. Dr. Gadre was the recipient of the President of India Gold Medal from IIT Delhi in 1989 and the Award for Excellence in Teaching from IIT Bombay in 1999 and 2004. He was also the recipient of the SVC Aiya Memorial Award and the Prof. K. Sreenivasan Medal from the Institute of Electronics and Telecommunication Engineers for contribution to education in Electronics and Telecommunication.