Person Identification from Fingernails and Knuckles Images using Deep Learning Features and the Bray-Curtis Similarity Measure

Mona Alghamdi[1], Plamen Angelov, Alvaro Lopez Pellicer

Computing and communication, Lancaster university, UK

Abstract

In this paper, an approach that makes use of knuckle creases and fingernails for person identification is presented. It introduces a framework for automatic person identification that includes localisation of the region of interest (ROI) of many components within hand images, recognition and segmentation of the detected components using bounding boxes, and similarity matching between two different sets of segmented images. The following hand components are considered: i) the metacarpophalangeal (MCP) joint, commonly known as the base knuckle; ii) the proximal interphalangeal (PIP) joint, commonly known as the major knuckle; iii) the distal interphalangeal (DIP) joint, commonly known as the minor knuckle; iv) the interphalangeal (IP) joint, commonly known as the thumb knuckle, and v) the fingernails. Crucial elements of the proposed framework are the feature extraction and similarity matching. This paper exploits different deep learning neural net-

Email addresses: m.alghamdi@lancaster.ac.uk (Mona Alghamdi), p.angelov@lancaster.ac.uk (Plamen Angelov), a.lopezpellicer@lancaster.ac.uk (Alvaro Lopez Pellicer)

[1]Computer Science Department, College of Science and Humanities, Imam Abdulrahman Bin Faisal University, Jubail, Kingdom of Saudi Arabia

works (DLNNs), which are essential in extracting discriminative high-level abstract features. We further use various similarity measures for the matching process. We validate the proposed approach on well-known benchmarks, including the 11k Hands dataset and the Hong Kong Polytechnic University Contactless Hand Dorsal Images known as PolyU. The results indicate that knuckle patterns and fingernails play a significant role in the person identification framework. The 11K Hands dataset results indicate that the left-hand results are better than the right-hand results and the fingernails produce consistently higher identification results than other hand components, with a rank-1 score of 100%. In addition, the PolyU dataset attains 100% in the fingernail of the thumb finger.

Keywords: Biometric, Hand, Segmentation, Feature extraction, Deep learning, Fine-tuning, Similarity matching, Person Identification

1. Introduction

The interest and demand for automated person identification have significantly increased and led to growing development in computer vision and machine learning. Identification and verification are types of personal authentication that allow access to buildings, cars, computers, and mobile devices and are vital in various other applications. There are traditional ways to verify user authentication, such as user credentials and one-time passwords (OTPs), to protect data from unauthorised access. However, these conventional methods have significant limitations; they might be lost, hacked, or stolen. Biometric identification is one of the most reliable solutions that overcome the earlier drawbacks and authorise safe access. Person identification

is based on detecting, segmenting and extracting pattern information to find valid individuals. A recognition system is a field of study that combines advanced machine learning, including DLNN, and biometrics from individuals based on their physical and behavioural traits [1, 2]. These traits can be the ear [3], face [4], hand gesture [5], finger-vein [6], or fusion of finger vein and knuckle pattern [2].

More attention has been paid to the characteristics of the hand. The hand biometric is critical in the human identification system because many distinctive features are available in hands, which can distinguish and identify the individual [7]. The approaches of extracting features can be divided into two main categories. The first category includes traditional techniques that deal with many visual characteristics. However, the traditional feature descriptor method is problem-specific and usually requires intervention to choose the descriptor [8, 9]. The second and most recent technique is the DLNN, which outperforms the traditional methods in many applications [2].

DLNN algorithms improve the recognition performance in different domains such as object localisation, pattern recognition, and image segmentation [10]. Furthermore, transfer learning of a DLNN that has been trained in a specific domain can be employed for another task and can be used in biometric systems. For example, these systems can use a fusion of the biometrics as mentioned above of finger veins, and finger knuckles [2], or different knuckle patterns for person identification [11]. In this survey, many studies used DLNN in the biometric systems for person identification [12]. However, the survey did not include knuckle patterns or fingernails for biometric systems. DLNN in the form of convolutional neural networks (CNNs) and

transfer learning have been employed in various biometric systems, which used knuckle patterns and fingernails [10, 11]. These systems can be based on primary biometrics, such as face [13], iris [14], fingerprint [15], and finger-vein [6].

The current research introduces a unique framework for automatic person identification using hand components such as knuckles and fingernails. The detection, recognition, segmentation and matching of distinct regions are part of the proposed system. We extended the previously introduced method known as the ***p***erson ***i***dentification based on ***f***ingernails and dorsal ***k***nuckle patterns (PIFK) [10]. In this paper, we experiment with fine-tuning DenseNet201 for feature extraction and utilise Bray-Curtis (BC) distance metric for the features' comparisons. The reason for selecting the DenseNet201 and Bray-Curtis combination for extracting features and matching is that among many base CNN models and similarity metrics these combination achieved good performance.

1.1. Feature extraction of finger knuckles patterns and fingernails

In pattern recognition problems, feature extraction is very critical. Features are used to map the characteristics of an object and are essential to support the matching process [16]. Coding methods[17, 18]; subspace approaches [19]; [20]; and texture analysis methods [21] can be categorised as the common algorithms for knuckle crease recognition [16]. This work explores the initial study for person identification by focusing on the skin folds, and crease patterns on the major knuckle [22]. This technique in the following study [23] essentially aligned the hand based on its shape before extracting the features, which increases the system's performance.

Sid et al. [24] offered a biometric system based on the finger knuckle pattern (FKP) that employs a DLNN technique called discrete cosine transform network (DCTNet) and a support vector machine (SVM) classifier. The proposed DCTNet design consists of a convolution layer, a binary hashing, and a block-wise histogram extraction. Gao et al. [25] executed a reconstruction in the Gabor filter response and applied a score level adaptive binary fusion algorithm to fuse the matching distances before and after reconstruction adaptively. Zhai et al. offered a batch-normalised CNN with a DLNN approach [26]. More distinguishing characteristics may be retrieved using the suggested CNN, satisfying recognition performance.

DLNN was used by Chlaoua et al. [27] to construct a multimodal biometric system based on images of FKP modalities. The principal component analysis network (PCANet) extracted FKP characteristics. Joshi et al. [28] trained a Siamese CNN for FKP detection using images collected from a preprocessed knuckle ROI. The Euclidean distance is used in network training to maximise the dissimilarity between images from different subjects or fingers and reduce the distance between images from the same finger.

Choudhury et al. [29] offered a solution for personal identification based on DLNN and combining two biometric traits, fingernail plates and finger knuckles. Kim et al. [30] propose a biometric method based on a finger-wrinkle image captured by a smartphone camera. A deep residual network is used to improve identification performance that has been impacted by misalignment and lighting variance during the image capture stage. Thapar et al. [31] proposed matching the whole dorsal finger rather than just the major/minor ROIs. FKIMNet, a Siamese-based CNN matching framework

that generates 128-D image features, is employed by including a dynamically adaptable margin, hard negative mining, and data augmentation. Usha et al. [32], Jaswal et al. [33], and Sadik et al. [34] provided an overview of the suggested FKP-based authentication techniques.

The fingernail biometric trait is another physiological characteristic that can play a significant role in the person identification and authentication system [35]. However, the requirements of a particular biometric trait to fully satisfy security applications have not yet been met [36]. The work reported in this paper [36] is the first attempt at developing a biometric framework based on the usability of such characteristics obtained from low-resolution nail-plate images. According to nail anatomy research [37], only the nail plate regrows when new cells are formed. Therefore, throughout a person's life, the spacing between the grooves of the nail bed remains relatively constant. Thus, unlike other traits that vary with age, these nail surface characteristics may be precious for identification over an individual's whole lifespan [36].

The fingernail ridge patterns that form on the outer nail surface are very distinguishable, even in the case of identical twins [38] or between various fingernails of an individual. The rigid nail structure resists any breakdown or environmental influences, except for abnormalities produced by nail illnesses and disorders, making the nail surface ridge pattern preferable to other biometric characteristics for identification [37]. The fingernail could expose to many diseases that lead to nail deformation. Examples of these diseases are Onychomychosis, Psoriasis and Beau's line [39] [36].

In this work [40], a deep learning approach for human authentication based on the fingernail and finger knuckle print was proposed, and AlexNet

is used as a pre-trained model. Normalisation and fusion techniques were combined at different levels [40]. In this study [10], end-to-end deep CNNs were used to extract discriminative high-level abstract features from knuckle patterns and fingernails. The matching procedure was based on the Bray-Curtis similarity. The index, middle, and ring fingernail plates were fused at the rank level utilising the Ant Colony Optimization approach in this study [41]. Three unique pre-trained models, AlexNet, ResNet-18, and DenseNet-20 were used to extract deep learning feature sets of the three nail plates [41].

1.2. Contributions and outline of this paper

The paper contributions can be summarised as follows:

1. A new framework for person identification is presented known as the *p*erson *i*dentification based on *f*ingernails and dorsal *k*nuckles (PIFK$^+$), contains a combination of dorsal hand components detection, segmentation, feature extraction and matching.
2. The paper investigates many base models of CNN for the best performance based on feature extraction and similarity metrics. Then, fine-tuning using the best performing model, the DenseNet201, a subset of the 11kHands [42] and PolyU [43] datasets based on each hand component to improve the retrieval of features.
3. Utilising the best matching metric, the Bray-Curtis, for estimating the similarity per hand's component.

   The improvement in this work is in the feature extraction stage, in which we fine-tune the DenseNet201 on segments of hands from the

well-known datasets 11kHands [42], and PolyU [43]. The remainder of this paper is structured as follows: section 2 explores the proposed framework (PIFK$^+$), section 3 shows the results and discussion, and section 4 contains the conclusion of this paper.

2. The person identification based on fingernails and knuckles (PIFK$^+$)

An overview architectural diagram of the developed PIFK$^+$ framework is presented in figure 1. Figure 2, 3, and 4 display the stages of the framework in more details. The PIFK$^+$ is expanded from the recently published method in [10].

These stages will be explained in the following sub-sections:

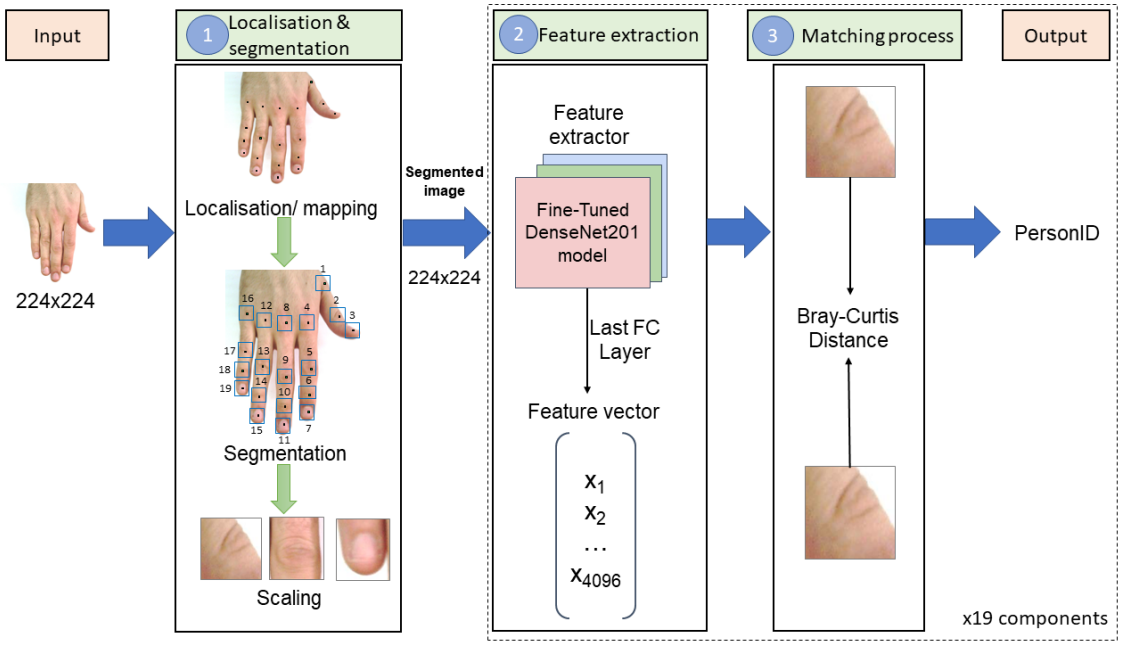

Figure 1: An overview schematic diagram of the developed framework for person identification based on dorsal fingernails and knuckle patterns (PIFK$^+$)

2.1. First phase: Knuckle and fingernail regions detection and segmentation from the hand image

This subsection explores the first phase of the developed framework (PIFK^{+}). This phase is essential before the feature extraction phase. In this stage, the detection, localisation, projection, segmentation, resizing, and re-scaling were conducted from the original hand image as displayed in figure 2. The outcomes from this step are 19 parts for each hand. These parts are various hand components: the base knuckles or MCP, major knuckles or PIP, minor knuckles or DIP, major knuckles of the thumb or IP, and fingernails of both right and left hands.

To detect the location of a hand region like a knuckle and fingernail, the method devoted to this purpose is multi-view bootstrapping for hand pose estimation [44]. This method detects the keypoints of the hand's major components. The original image of the hand was resized to 224×224 to get the best result of the keypoint localisation. The detected keypoints from the resized image were then mapped to the original high-definition image to get a bigger size segmented image. The keypoints obtained were then utilised to segment each hand image. After segmentation, the sub-images were re-scaled (normalised), which allows the original pixel values from the sub-images to range from 0 to 1. The approach of multi-view bootstrapping for hand pose estimation uses a keypoint detector $d(.)$. The keypoint detector maps an input image of the hand $I \in \mathbb{R}^{w \times h \times 3}$, where w indicates the vertical and h indicates the horizontal dimension of the image and three corresponds to the RGB channels, to P components located in x positions as follows:

$$d(I) \mapsto \{(x_p, c_p) \text{ for } p \in [1...P]\}, \tag{1}$$

each confidence score c_p is associated with a location x_p that corresponds to one component p, such as PIP or DIP. A hand has a total of 19 components.

We followed the detection and segmentation method to find a bounding box per hand's major component, as introduced in [10]. Figure 2 shows the stages of the hand sub-images segmentation.

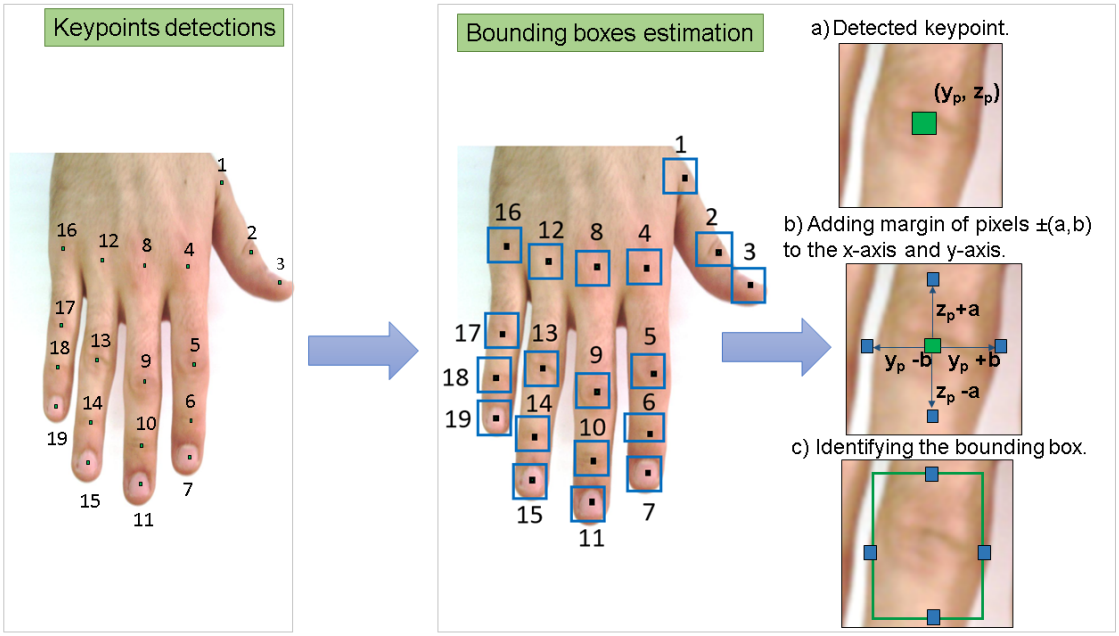

Figure 2: Sample of the localisation including keypoints/landmarks detection and bounding boxes estimation for fingernail and knuckle crease regions.

The detection approach from [44] is extended to segment each hand component. The segmentation was done by estimating the size of each region's bounding box based on the identified centre point. As a result, the required area may be divided into two stages. In the first stage, using the multi-view bootstrapping approach to determine the centre of the segmented section, we can re-define the keypoint as follows:

$$y_p, z_p \leftarrow x_p, \tag{2}$$

where y_p and z_p are the location of the point in the x-axis and y-axis, respectively.

The second stage involves defining the bounding box by specifying a height and width per type of component to the centre point. Next, we estimated a margin of pixels (a) and (b) to both axes around a detected keypoint. A rectangular box was then defined as illustrated in figure 2 to produce a segment (S) using the following equation:

$$S = \begin{cases} \mathrm{y}_p + b \\ \mathrm{y}_p - b \\ \mathrm{z}_p + a \\ \mathrm{z}_p - a \end{cases} \tag{3}$$

2.2. Second phase: feature extraction

Traditional image processing methods have been popular in many computer vision applications for decades. However, these methods often require extensive human intervention [45] in the design of the feature extractor. This paper uses these abstract high-level features extracted from each segmented part of the hand. In the beginning, we tried many base models of CNN to evaluate the best-performing model in the '11k Hands' dataset. We found that the base model of DenseNet201 and the Bray-Curtis for feature extraction and similarity measuring achieved excellent results in different parts of the hands. Figure 5 presents the most base (pretrained) models with

high performance in different sub-images from the right and left hands in the '11k Hands' dataset. We then attempted to fine-tune parts of the segmented images per ROI from the '11k Hands' and 'PolyU' datasets using the DenseNet201 model as feature extractors.

We chose to fine-tune the DenseNet201 model among various DLNNs, such as VGG16, ResNet50V2, MobileNet, and others. Because this model showed high performance for extracting high-level and abstract features [10, 11]. Regarding the fine-tuning, the first 700 layers were frozen from the original model of DenseNet201. We then added 2D global average pooling, batch normalisation with 0.90 momentum, dropouts equal to 0.5, dense with 4096 vectors, a Relu activation function, 0.6 dropouts, batch normalisation with 0.9 momentum, and a classification layer with 170 dimensions and softmax activation function. The network was trained with 150 epochs. A stochastic gradient descent was used as the learning optimisation, the learning rate was 0.001, with a Nesterov momentum of 0.9, and the loss function was categorical cross-entropy as illustrated in figure 3. Finally, the same network was retrained with all its layers and settings with 100 epochs.

Several types of augmentations were applied to the training samples before fine-tuning to reduce the problem of small training data. The augmentation includes rescaling the pixels from 0 to 1, horizontal flipping, a randomised rotation range of 30, a randomised width and height shift range value of 30, and a randomised zoom in the range of 0.9 to 1.1. The training data was enriched with these augmentations, which enhanced model learning.

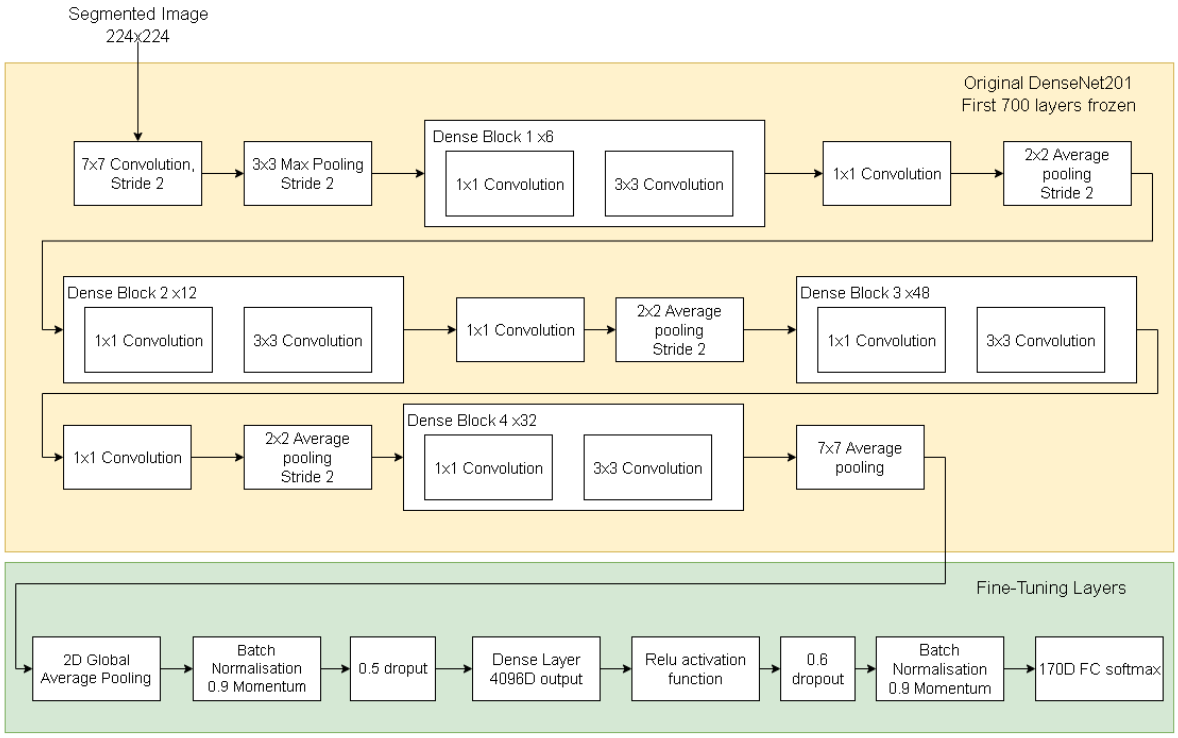

Figure 3: A demonstrative diagram of the fine-tuning of the DenseNet201 layers for feature extractor.

2.3. Third phase: Similarity estimation and matching

The third phase is the similarity evaluation and decision (matching). This paper assessed many metrics, such as the Bray-Curtis, Cosine, and Cityblock, with the feature extractor of the base models mentioned above. The best recognition results are then used for the next fine-tuning stage to improve the performance further. The similarity metric is the inverse of the distance metric and measures the similarity between two segments of a person's hand. These hand segments were first mapped to feature spaces using the feature extractors. Each feature vector length is 4096. Secondly, the similarity met-

rics were used to match individuals per hand component. Figure 4 presents a sample diagram of this stage. A brief definition of the similarity metrics is in the following sub-sections:

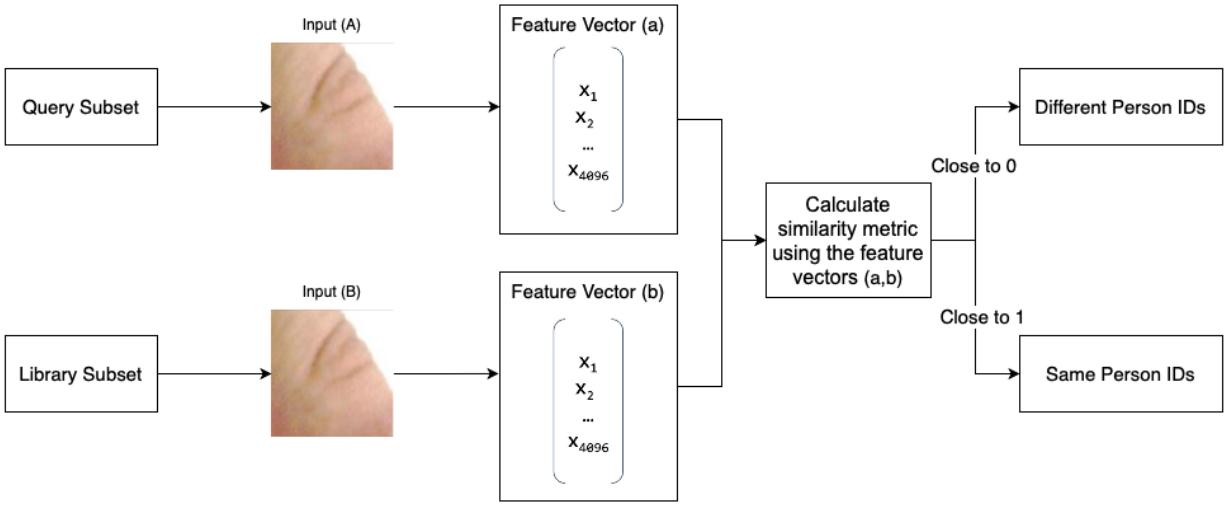

Figure 4: An illustrative diagram of the similarity measuring using the Bray-Curtis metric.

Suppose we have two segments, which are S_a and S_b. Their vectors can be represented as $S_a = (a_1, a_2, ..., a_p)$ and $S_b = (b_1, b_2, ..., b_p)$. We can define the various distance metric as the following:

The Bray-Curtis (BC) metric. is a distance of two vectors for two segments that can be indicated as $d_{BC}(S_a, S_b)$ and defined as follows:

$$d_{BC}(S_a, S_b) = \frac{\sum_{i=1}^{P} |a_i - b_i|}{\sum_{i=1}^{P} |a_i + b_i|} \tag{4}$$

The Cosine (C) metric. can measure the distance between two vectors as follows:

$$C(S_a,\ S_b) = \frac{S_a.S_b}{\|S_a\|\|S_b\|} \tag{5}$$

The Cityblock (CB) metric. may calculate the distance between two vectors as shown in 6. Suppose the vectors are composed of two dimensions (x, y), the CB can be defined as the following:

$$CB(S_a,\ S_b) = \|x_{S_a} - x_{S_b}\| + \|y_{S_a} - y_{S_b}\| \tag{6}$$

The degree of similarity (Sim) between two vectors is the inverse of the distance (Dis). The shorter the distance between two vectors, the greater the degree of similarity, and vice versa, and may be described as $Sim(S_a, S_b) = 1 - Dis(S_a, S_b)$. As follows, the closest match found using a similarity metric corresponds to the ID as follows [10]:

$$\hat{ID} = argmin\{Dis\} = argmax\{Sim\} \tag{7}$$

The proposed framework of PIFK$^+$ is outlined by Algorithm 1 in the form of a pseudo-code.

Algorithm 1: The proposed framework of PIFK^{+}

Data: Dataset from 11kHands and PolyU

while *Hand image (I) is available* **do**

  Detect the keypoints of the hand $d(I)$ using (1);

  Re-define the center point using (2);

  Define the bounding box of the segmented component using (3);

  Discard any non-well-segmented sub-image;

end

Collect sub-images of each component as a dataset;

for $p \leftarrow 1$ **to** 19 **do**

  Split the dataset to training, validation, and testing;

  Set the network configuration; Generate images augmentation; Train with defined epochs; Produce the network weight (W) of the best validation accuracy; Extract features from the pairs S_a and S_b using W;

  Calculate the similarity using (4);

  if

$$d_{BC}(S_a, S_b) \simeq 0$$

  then

$$S_a, S_b \in ID$$

  end

  if

$$d_{BC}(S_a, S_b) \simeq 1$$

  then

$$S_a, S_b \notin ID$$

  end

end

3. Result and discussion

The experiment we performed aimed to identify a person from a group of individuals based on the input of hand images. This section will evaluate the proposed framework on the 11k Hands and PolyU datasets.

3.1. Datasets description

There are many benchmarks for hand images, such as CASIA [46], Bosphorus [47], IITD [47], GPDS150hand [48], however they consist of only the palm side of the hand. The 11k Hands benchmark contains RGB images of the dorsal and palm surface of 190 subjects' right and left hands. The images have a resolution of 1600 x 1200 pixels. In this paper, we are interested in the dorsal only for the person identification. The number of the right-hand segments is $54,404$, and of the left hand is $52,369$. We used 47,719 and 46,890 segmented images from the left and right 11k Hands dataset, respectively, for the training and validation of the fine-tuned model. We also divided the dataset into 77% and 33% in training and validation. The testing part included different sets of sub-images. There were 380 sub-images in the query and 2,517 in the gallery for the left hands. They were 380 sub-images in the query, and 2,517 sub-images in the gallery for the left-hands. Also, 380 sub-images in the query and 4,528 sub-images in the gallery for the right-hands.

The Hong Kong Polytechnic University Contactless Hand Dorsal Images Database [43] was also considered, which consisting of 4650 surfaces of the right-hand dorsal images in a flat position from 501 subjects. These images with resolution (1600×1200 pixels) were captured using mobile and handheld cameras. However, this dataset has only the right hands, which limits its use

of the dataset. The training and validation total sub-images were 75,312, and we divided the data into 77% and 33% in training and validation, respectively. The testing included 916 and 6,426 segmented images in the query and gallery sets, respectively.

3.2. Experimental protocol

The identification framework in our experiments matches two different sub-images taken from the same person and vice versa. We evaluated our framework using a *rank*-1 recognition rate, cumulative matching characteristic (CMC) and standard deviation (SD). The rank-1 of recognition can be defined as follows:

$$rank\text{–}1 = \frac{N_i}{N} \times 100, \tag{8}$$

where N_i is the number of samples that were correctly assigned to the correct person, and N is the total number of samples that were attempted to be recognised.

The CMC, which indicates the accuracy performance of $rank - n$ [49], was also utilised to demonstrate the overall performance. The datasets used to evaluate the suggested strategy, and the outcomes will be described in the next section.

The SD is a metric that reflects how widely the data is spread out in a given space. The samples are highly close to the mean when the standard deviation is extremely low. On the other hand, a high standard deviation indicates that the samples are spread out over a wide range of values [50]. The function of SD can be defined as follows:

$$SD = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}, \tag{9}$$

where x_i is the sample's value, either matched indicated by the value 1 or not-matched indicated by the value 0, and the value $\bar{x}$ is the mean of the matched samples.

3.3. Pre-processing phase

As shown in figure 1, the pre-processing phase was explained in sub-section 2. Therefore, the details of PIFK$^+$ about this phase will be covered in this sub-section.

Step 1: *Detection and recognition of the knuckle creases and fingernails:* The detection of the knuckle creases and fingernails is described in sub-section 2.1. First, the image is resized to 224×224, and fingernails and knuckle creases were identified and localised using the Multi-view Bootstrapping method [44]. Then, automatic indexing of each segmented ROI was utilised. The total number of indexes was 19 distributed in different hand regions. Finally, the keypoints locations were mapped to the original high-definition size of the image. The dimensions of the bounding boxes were then estimated in terms of width and height.

Step 2: *Resizing and re-scaling:* The segmented images were resized into 224×224 pixels and re-scaled. The resizing was important to be compatible with the feature extractor model.

Step 3: *The query and library data sets:* The data was structured into the query and library sets in this step. We considered an evaluation of the method Leave-One-Out Cross-Validation (LOOCV) to evaluate the proposed

algorithm's performance. The LOOCV had only one fold. Furthermore, because the samples were independent, each prediction in the identification problem was independent of the others [51]. As a result, we considered only one sub-image of a subject per ROI in the query. As a result, we had 20 sub-images of 20 different subjects per component in the query set. The rest sub-images were 239 of the same subjects per component in the library. We also had 19 ROI per individual of the left and right hand, respectively.

We disregarded 87 images of left hands and 6 right hands in the 11k Hands dataset due to inaccurate keypoint detection. Following segmentation due to erroneous detection, 428 sub-images of left-hand segments and 573 of right-hand segments were eliminated. The total number of eliminated sub-images in the PolyU collection was 5,733 out of 88,392. Disregarded images and sub-images were not included in the framework evaluation, which has been considered a limitation of the study.

3.4. Deep learning for feature extractor

For visual object recognition, deep learning has become the dominant machine learning approach [52]. In this paper, we evaluate the recognition performance using two different steps. In the first phase, we investigated many base models of CNN pretrained on the popular database' ImageNet' to extract the features from sub-images. The architecture is varied among these pretrained models. However, they shared a significant and complicated structure that significantly impacted the model performance [53]. Studying many models is to identify the most sophisticated ones in extracting discriminative features. This step is crucial and leads to better individual identification. Interestingly, the best performing base models in terms of

315 rank-1 recognition accuracy, from highest to lowest, are DenseNet201 [53], MobileNet [54], MobileNetV2 [24], DenseNet169, and ResNet50V2 [55]. Figure 5 and table 1 display the best base models in rank-1 recognition rate from the right and left hands in the '11k Hands' dataset.

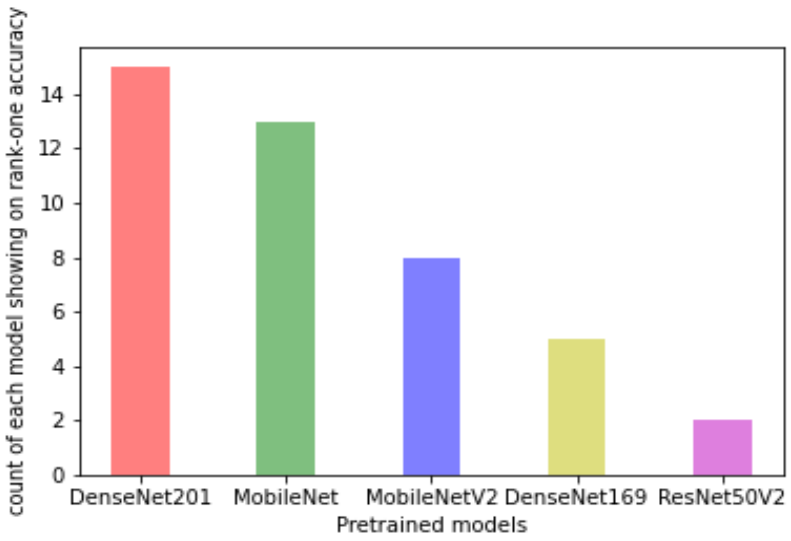

Figure 5: Count of each model showing on rank-1 accuracy for fingernails and knuckles-print regions the '11k Hands' dataset.

Table 1: The rank-1 recognition accuracy of 11k-hands database (shown in %) using different pretrained models and similarity distances.

Side	Region	Finger	Rank-1	Distance	model
Left	Base Knuckle	Thumb	87.83	Bray-Curtis	DenseNet201
		Index	82.01	Bray-Curtis	ResNet50V2
		Middle	80.95	Cityblock	DenseNet201

		Ring	98.94	Cosine	DenseNet201
		Little	99.47	Cosine	MobileNetV2
	Major Knuckle	Thumb	86.77	Cosine	MobileNet
		Index	85.63	Cosine	MobileNet
		Middle	86.70	Bray-Curtis	MobileNetV2
		Ring	99.47	Cosine	MobileNet
		Little	100	Bray-Curtis	DenseNet169
	Minor Knuckle	Index	84.12	Cosine	MobileNet
		Middle	86.77	Bray-Curtis	DenseNet201
		Ring	99.47	Bray-Curtis	DenseNet201
		Little	99.47	Cityblock	DenseNet201
	Fingernail	Thumb	87.83	Bray-Curtis	DenseNet169
		Index	91.00	Cosine	MobileNet
		Middle	100	Bray-Curtis	DenseNet201
		Ring	99.47	Bray-Curtis	DenseNet201
		Little	89.00	Cosine	MobileNet
Right	Base Knuckle	Thumb	85.26	Brycurtis	DenseNet201
		Index	81.57	Brycurtis	DenseNet201
		Middle	80.00	Brycurtis	MobileNetV2
		Ring	80.00	Bray-Curtis	MobileNetV2
		Little	83.15	Bray-Curtis	DenseNet169
	Major Knuckle	Thumb	84.21	Brycurtis	DenseNet201
		Index	83.15	Brycurtis	DenseNet201
		Middle	85.18	Cosine	MobileNet

	Ring	83.68	Bray-Curtis	DenseNet201
	Little	80.00	Bray-Curtis	MobileNetV2
Minor Knuckle	Index	80.00	Brycurtis	MobileNetV2
	Middle	85.18	Brycurtis	ResNet50V2
	Ring	83.15	Brycurtis	DenseNet169
	Little	80.00	Bray-Curtis	MobileNetV2
Fingernail	Thumb	85.78	Cosine	MobileNet
	Index	90.00	Bray-Curtis	MobileNet
	Middle	91.05	Bray-Curtis	MobileNet
	Ring	93.68	Bray-Curtis	MobileNet
	Little	84.73	Cosine	MobileNet

This paper evaluated the results of the first step in the experiment using the rank-1 recognition accuracy as displayed in the table 1 w.r.t. the '11k-hands' benchmark. This study considered several distance metrics and pretrained models to identify the best-performing ones. Table 1 demonstrates the best distances and models performing on the base, major, minor knuckles, and fingernails of five fingers on the left and right hands of the '11k-hands' database. Surprisingly, the results, as shown in table 1, indicate that the left components of the hand are more identifiable than the right ones. We also observed that the identifiability of fingernail patches from both hand sides is relatively higher than different knuckle patches. This result may refer to the nail being consistent, having a unique shape, and the proposed approach performing exceptionally well.

Figure 5 refers to the best models that appear on rank-1 accuracy among all regions of both hands. From the highest to the lowest, we can observe that the best models are DenseNet201, MobileNet, MobileNetV2, DenseNet169, and ResNet50V2. In addition, many distance metrics are studied to examine their impact on the matching result. Indeed, many distances appeared on rank-1 accuracy; however, we showed the most repeated distance between different regions. We can notice that Brycurits, cosine, and city-block are the most achieved distances, from the highest to the lowest on rank-1.

In the second phase, fine-tuning the DenseNet201 model was considered for extracting features. Figure 3 displays the structure of the retrained DenseNet201 model. This model was considered because it showed the best performance, as demonstrated by the phase one results. The base model of DenseNet201 [52] was originally pretrained on the well-known ImageNet dataset. This model has many advantages: DenseNets201 naturally scales to hundreds of layers, providing no optimization challenges; DenseNets201 requires fewer parameters and processing time.

We achieved high performances in training and validation using the fine-tuned DenseNet201 in subsets of each ROI from 11k Hands and PolyU. The last fully connected (FC) layer was utilized from the fine-tuned model to extract the features per hand component. Finally, pairs of the query and library feature vectors were generated from the last FC layer to calculate the similarity using Bray-Curtis.

3.5. Similarity estimation and matching

Estimating the similarity between two sub-images of the same hand is a part of the matching process. In the matching process, one segmented

image from the query can match with one or more corresponding segmented image(s) in the library. We estimated the similarity based on the feature vectors gained from the last FC layers of the fine-tuned DenseNet201 model as described in section 2.2. We then utilised the Bray-Curtis similarity metric for the matching process. As a result, the proposed PIFK^{+} framework gained a high performance compared to the state-of-the-art methods. Table 2 shows our proposed framework's rank-1 recognition rate and standard deviation (SD) and a comparison of our method and other state-of-the-art methods.

Table 2: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k Hands-L	11k Hands-R	PolyU
Fingernail	Thumb	Our (Rank-1)	100	95.00	100
		Our (SD)	0	0.22	0
		PIFK [10]	87.83	84.21	93.81
		[11]	-	-	-
	Index	Our (Rank-1)	100	100	95.83
		Our (SD)	0	0	0.2
		PIFK [10]	89.42	88.95	90.40
		[11]	-	-	-
	Middle	Our (Rank-1)	100	100	91.83
		Our (SD)	0	0	0.27
		PIFK [10]	90.48	89.47	87.65
		[11]	-	-	-
	Ring	Our (Rank-1)	100	100	93.88

Table 2: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k Hands-L	11k Hands-R	PolyU
		Our (SD)	0	0	0.24
		PIFK [10]	93.65	91.58	85.10
		[11]	-	-	-
	Little	Our (Rank-1)	100	100	95.83
		Our (SD)	0	0	0.2
		PIFK [10]	84.13	80.00	87.30
		[11]	-	-	-
Minor Knuckle	Index	Our (Rank-1)	100	95.00	91.67
DIP		Our (SD)	0	0.22	0.27
		PIFK [10]	84.66	76.84	72.47
		[11]	86.35	89.56	74.81
	Middle	Our (Rank-1)	100	95.00	85.71
		Our (SD)	0	0.22	0.35
		PIFK [10]	85.19	82.11	68.92
		[11]	93.69	93.17	84.57
	Ring	Our (Rank-1)	100	95.00	91.84
		Our (SD)	0	0.22	0.27
		PIFK [10]	84.13	77.89	71.46
		[11]	91.45	89.56	80.80
	Little	Our (Rank-1)	100	95.00	91.67
		Our (SD)	0	0	0.2

Table 2: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k Hands-L	11k Hands-R	PolyU
		PIFK [10]	81.48	76.84	78.43
		[11]	83.91	80.83	73.25
Major Knuckle	Thumb	Our (Rank-1)	95.00	95.00	91.11
PIP		Our (SD)	0.22	0.22	0.28
		PIFK [10]	85.71	84.21	76.83
		[11]	-	-	-
	Index	Our (Rank-1)	100	95.00	97.87
		Our (SD)	0	0.22	0.14
		PIFK [10]	82.45	83.16	79.71
		[11]	93.28	93.93	91.23
	Middle	Our (Rank-1)	100	95.00	93.75
		Our (SD)	0	0.22	0.24
		PIFK [10]	85.11	82.54	76.64
		[11]	94.70	95.26	88.90
	Ring	Our (Rank-1)	90.00	95.00	91.49
		Our (SD)	0.30	0.22	0.28
		PIFK [10]	80.95	83.68	83.23
		[11]	94.30	92.03	90.34
	Little	Our (Rank-1)	95.00	100	97.87
		Our (SD)	0.22	0	0.14
		PIFK [10]	83.07	75.79	82.06

Table 2: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k Hands-L	11k Hands-R	PolyU
		[11]	88.59	86.34	88.57
Base Knuckle	Thumb	Our (Rank-1)	100	100	79.59
MCP		Our (SD)	0	0	0.40
		PIFK [10]	87.30	85.26	58.35
		[11]	-	-	-
	Index	Our (Rank-1)	100	90.00	95.92
		Our (SD)	0	0.3	0.19
		PIFK [10]	78.84	81.58	64.34
		[11]	78.82	84.82	62.38
	Middle	Our (Rank-1)	100	95.00	97.96
		Our (SD)	0	0.22	0.14
		PIFK [10]	80.95	77.37	67.27
		[11]	84.32	85.96	61.49
	Ring	Our (Rank-1)	100	80.00	97.96
		Our (SD)	0	0.22	0.14
		PIFK [10]	80.42	77.37	66.47
		[11]	76.99	73.62	57.16
	Little	Our (Rank-1)	100	100	97.96
		Our (SD)	0	0	0.14
		PIFK [10]	84.13	80.53	67.94
		[11]	83.71	81.78	59.16

Table 2: The rank-1 recognition rate (shown in %) and SD for the 11k Hands and PolyU datasets.

Region	Finger	Method	11k Hands-L	11k Hands-R	PolyU

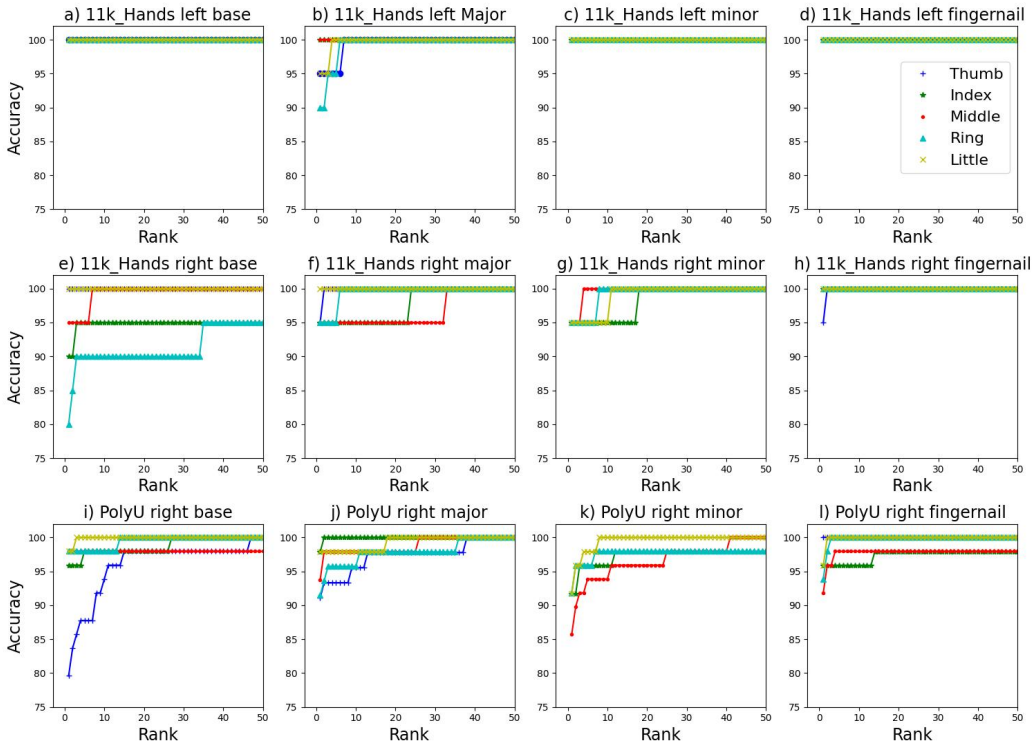

Figure 6: The CMC of the proposed PIFK^{+}; the a) base; b) major; c) minor knuckles; d) fingernails of the left hand; e) base; f) major; g) minor h) fingernails of the right hands from the 11k Hands dataset; i) base; j) major; k) minor l) fingernails of the right hands from the PolyU dataset.

The CMC of the knuckles and fingernails of the left (chart a-d) and right (chart e-h) hands in the 11k Hands are illustrated in figure 6. The

left hand—including the fingernails and knuckles from 11K Hands—is more recognisable than the right hand for most of the fingers, as shown by the results (see table 2 and figure 6). This result also complies with the study shown in [10]. We also found that fingernail sub-images from both left and right hands from the 11k Hands and PolyU datasets achieved better results than the recognition using all knuckle sub-images. We can observe that the SD is close to 0 when the rank-1 is the highest value and is more than 0 when the rank-1 is lower than 100. These results are shown in table 2 and the CMC chart 6.

In general, the proposed framework PIFK$^+$ outperformed the state-of-the-art methods [10, 11, 31]. We had the best results for both hands in almost all fingernails. The fingernails rank-1 accuracies were 100% in the 11k Hands dataset and 100% on the thumb and ring finger in the PolyU dataset. Overall, we can observe that the DIP/minor knuckles got minimal performance compared to the other hand components in all datasets. However, the PIP components' performance was slightly better than the DIP components in both datasets. The MCP components were also higher than the PIP and DIP in the two datasets. The PIP obtained a rank-1 accuracy of 100% in the index and middle fingers in the left hands and right little finger of the 11k Hands dataset and 97.87% in the index and little fingers in the PolyU dataset, compared to the middle finger that achieved 94.70% and 85.11% in the left-hands of the 11k Hands dataset in [11], and [10], respectively.

The DIP component in each finger of the left 11k Hands dataset achieved a rank-1 accuracy of 100%, whereas the one in the ring finger obtained 91.84% in the PolyU dataset. This observation is also valid among the right-hand

parts of the PolyU dataset. The PIP outperforms the DIP component, with rank-1 accuracies of 97.87% for the index and little fingers. For the MCP/base knuckle component, we gained 100% in most fingers in the 11k Hands dataset and 97.96% in the middle, ring, and little fingers from the PolyU datasets in comparison to [10], and [11] that got lower results using the same component as displayed in table 2.

Although there is a lack of studies similar to our approach, we compared our study with the previous most similar ones like [10, 31, 11]. In addition, we analysed the fingernails of our result with this method [31]. We found that the proposed framework PIFK$^+$ achieved excellent performance for most fingers, with a rank-1 accuracy of 100%. However, this method [31], which scored 94.83% in the fingernail identification, did not specify which hand (right or left) or finger the component locate. The best performance was 98.60% using the fusion of the right hands in the major, minor, fingernail, and finger components in the [31] method using the PolyU dataset. However, in our work, using only one component, such as the fingernail of the thumb finger, we achieved a higher result with a rank-1 score of 100%. The study in [11] gained 100% in the fusion of all components and fingers of a hand. However, we achieved a 100% in most components of the fingernail in the index, middle, ring, and little finger of the 11kHands dataset and the thumb finger of the left hands in the 11k Hands and PolyU datasets.

4. Conclusion

PIFK$^+$ is a comprehensive framework for automated person identification for forensic and security applications presented in this study, where

the hand image plays a significant role in identifying a person. The framework used the dorsal surface of the five human hand components (fingernails, MCP, PIP, DIP, and IP joints of the five fingers). The proposed approach PIFK^{+} utilises all 19 hand components from two popular datasets: the 11k Hands and PolyU datasets. The approach fine-tunes segmented subsets per component using the DenseNet201 model and applies the Bray-Curtis similarity metric to extract abstract features and find the corresponding pairs' similarities. In addition, the framework automatically labels and generates bounding boxes around the detected hand components. We evaluated the proposed approach on these two most common datasets, and our results outperform the state-of-the-art methods. Furthermore, we discovered intriguing facts about the hands and their components. Such as the continuous difference in outcomes between the left and right hands, favouring the left. In addition, the fingernails were found to be the best-performing components in both datasets. Developing the framework using other fine-tuned CNN models for feature extraction may be considered for future studies. Furthermore, evaluating both the quality-segmented images and real-world poor-quality segmented images due to false ROI localisation in the recognition system may be relevant to future studies.

References

[1] M. Anbari, A. M. Fotouhi, Finger knuckle print recognition for personal authentication based on relaxed local ternary pattern in an effective learning framework 32 (2021) 55. doi:10.1007/s00138-021-01178-6. URL https://doi.org/10.1007/s00138-021-01178-6

[2] S. Daas, A. Yahi, T. Bakir, M. Sedhane, M. Boughazi, E. B. Bourennane, Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion, IET Image Processing 14 (2020) 3859–3868. doi:10.1049/iet-ipr.2020.0491.

[3] A. Kamboj, R. Rani, A. Nigam, R. Ranjan Jha, CED-Net: context-aware ear detection network for unconstrained images, Pattern Analysis and Applications 24 (2021) 779–800. doi:10.1007/s10044-020-00914-4.
URL https://doi.org/10.1007/s10044-020-00914-4

[4] F. Juefei-Xu, E. Verma, M. Savvides, Deepgender2: A generative approach toward occlusion and low-resolution robust facial gender classification via progressively trained attention shift convolutional neural networks (ptas-cnn) and deep convolutional generative adversarial networks (dcgan), in: Deep Learning for Biometrics, Springer, 2017, pp. 183–218.

[5] J. Wu, P. Ishwar, J. Konrad, Two-stream cnns for gesture-based verification and identification: Learning user style, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 42–50.

[6] R. S. Kuzu, E. Piciucco, E. Maiorana, P. Campisi, On-the-fly finger-vein-based biometric recognition using deep neural networks, IEEE transactions on information forensics and security 15 (2020) 2641–2654. doi:10.1109/TIFS.2020.2971144.

[7] A. Kumar, Can we use minor finger knuckle images to identify humans?, in: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), IEEE, 2012, pp. 55–60. doi:10.1109/BTAS.2012.6374558.

[8] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern recognition 71 (2017) 158–172. doi:10.1016/j.patcog.2017.05.025.

[9] L. Yann, B. Yoshua, H. Geoffrey, Deep learning, Nature 521 (2015) 436. doi:10.1038/nature14539.

[10] M. Alghamdi, P. Angelov, B. Williams, Automated person identification framework based on fingernails and dorsal knuckle patterns, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2021, pp. 01–08.

[11] R. Vyas, H. Rahmani, R. Boswell-Challand, P. Angelov, S. Black, B. M. Williams, Robust end-to-end hand identification via holistic multi-unit knuckle recognition, in: 2021 IEEE International Joint Conference on Biometrics (IJCB), IEEE, 2021, pp. 1–8.

[12] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, D. Zhang, Biometrics recognition using deep learning: A survey, arXiv preprint arXiv:1912.00271.

[13] Y. V. Vizilter, V. S. Gorbatsevich, A. V. Vorotnikov, N. A. Kostromov, Real-time face identification via cnn and boosted hash-

ing forest, Computer Optics 41 (2017) 254–265. doi:10.18287/2412-6179-2017-41-2-254-265.

[14] R. R. Jha, G. Jaswal, D. Gupta, S. Saini, A. Nigam, Pixisegnet: pixel-level iris segmentation network using convolutional encoder–decoder with stacked hourglass bottleneck, IET biometrics 9 (2020) 11–24. doi:10.1049/iet-bmt.2019.0025.

[15] H.-R. Su, K.-Y. Chen, W. J. Wong, S.-H. Lai, A deep learning approach towards pore extraction for high-resolution fingerprint recognition, IEEE, 2017, pp. 2057–2061. doi:10.1109/ICASSP.2017.7952518.

[16] G. Jaswal, A. Kaul, R. Nath, Knuckle print biometrics and fusion schemes - Overview, challenges, and solutions, ACM Computing Surveys 49 (2). doi:10.1145/2938727.

[17] L. Zhang, L. Zhang, D. Zhang, Finger-knuckle-print verification based on band-limited phase-only correlation (2009).

[18] Z. Le-Qing, Z. San-Yuan, Multimodal biometric identification system based on finger geometry, knuckle print and palm printdoi:10.1016/j.patrec.2010.05.010.

[19] D. S. Guru, K. B. Nagasundara, S. Manjunath, Feature level fusion of multi-instance finger knuckle print for person identification, ACM International Conference Proceeding Series (2010) 186–190doi:10.1145/1963564.1963595.

[20] Z. S. Shariatmadar, K. Faez, A novel approach for Finger-Knuckle-Print

recognition based on Gabor feature fusion, Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011 3 (2011) 1480–1484. doi:10.1109/CISP.2011.6100450.

[21] L. Zhang, L. Zhang, D. Zhang, H. Zhu, Ensemble of local and global information for fingerknuckle-print recognition, Pattern Recognition 44 (9) (2011) 1990–1998. doi:10.1016/j.patcog.2010.06.007.
URL http://dx.doi.org/10.1016/j.patcog.2010.06.007

[22] D. L. Woodard, P. J. Flynn, Finger surface as a biometric identifier, Computer vision and image understanding 100 (2005) 357–384. doi:10.1016/j.cviu.2005.06.003.

[23] A. K. Jain, N. Duta, Deformable matching of hand shapes for user verification, Vol. 2, 1999. doi:10.1109/icip.1999.823019.

[24] K. Sid, F. Laallam, D. Samai, A. Tidjani, Finger knuckle print features extraction using simple deep learning method, Int. J. Comput. Sci. Commun. Inf. Technol. (CSCIT) 5 (2017) 12–18.

[25] G. Gao, P. Huang, S. Wu, H. Gao, D. Yue, Reconstruction in gabor response domain for efficient finger-knuckle-print verification, in: 2018 Australian & New Zealand Control Conference (ANZCC), IEEE, 2018, pp. 110–114.

[26] Y. Zhai, H. Cao, L. Cao, H. Ma, J. Gan, J. Zeng, V. Piuri, F. Scotti, W. Deng, Y. Zhi, et al., A novel finger-knuckle-print recognition based on batch-normalized cnn, in: Chinese conference on biometric recognition, Springer, 2018, pp. 11–21.

[27] R. Chlaoua, A. Meraoumia, K. E. Aiadi, M. Korichi, Deep learning for finger-knuckle-print identification system based on pcanet and svm classifier, Evolving Systems 10 (2) (2019) 261–272.

[28] J. Joshi, S. Nangia, K. Tiwari, K. Gupta, Finger knuckleprint based personal authentication using siamese network, in: 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2019, pp. 282–286.

[29] S. H. Choudhury, A. Kumar, S. H. Laskar, Biometric authentication through unification of finger dorsal biometric traits, Information Sciences 497 (2019) 202–218.

[30] C. S. Kim, N. S. Cho, K. R. Park, Deep residual network-based recognition of finger wrinkles using smartphone camera, IEEE Access 7 (2019) 71270–71285.

[31] D. Thapar, G. Jaswal, A. Nigam, Fkimnet: a finger dorsal image matching network comparing component (major, minor and nail) matching with holistic (finger dorsal) matching, in: 2019 international joint conference on neural networks (IJCNN), IEEE, 2019, pp. 1–8.

[32] K. Usha, M. Ezhilarasan, Finger knuckle biometrics–a review, Computers & Electrical Engineering 45 (2015) 249–259.

[33] G. Jaswal, A. Kaul, R. Nath, Knuckle print biometrics and fusion schemes–overview, challenges, and solutions, ACM Computing Surveys (CSUR) 49 (2) (2016) 1–46.

[34] M. A. Sadik, M. N. Al-Berry, M. Roushdy, A survey on the finger knuckle prints biometrie, in: 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), IEEE, 2017, pp. 197–204.

[35] S. Z. Li, A. Jain, Encyclopedia of biometrics, Springer Publishing Company, Incorporated, 2015.

[36] A. Kumar, S. Garg, M. Hanmandlu, Biometric authentication using finger nail plates, Expert systems with applications 41 (2) (2014) 373–386.

[37] A. A. Diaz, A. F. Boehm, W. Rowe, Comparison of fingernail ridge patterns of monozygotic twins, Journal of Forensic Science 35 (1) (1990) 97–102.

[38] C. R. Daniel III, B. M. Piraccini, A. Tosti, The nail and hair in forensic science, Journal of the American Academy of Dermatology 50 (2) (2004) 258–261.

[39] W. Taylor, D. Roberts, J. Boyle, Guidelines for treatment of onychomychosis, British Journal of Dermatology 148 (2002) 402–410.

[40] H. Heidari, A. Chalechale, Biometric authentication using a deep learning approach based on different level fusion of finger knuckle print and fingernail, Expert Systems with Applications 191 (2022) 116278.

[41] S. H. Choudhury, A. Kumar, S. H. Laskar, Adaptive management of multimodal biometrics—a deep learning and metaheuristic approach, Applied Soft Computing 106 (2021) 107344.

[42] M. Afifi, 11K Hands: Gender recognition and biometric identification using a large dataset of hand images, Multimedia Tools and Applications 78 (2019) 20835–20854. doi:10.1007/s11042-019-7424-8.
URL https://doi.org/10.1007/s11042-019-7424-8

[43] A. Kumar, X. Zhihuan, Personal identification using minor knuckle patterns from palm dorsal surface, IEEE Transactions on Information Forensics and Security 11 (2016) 2338–2348. doi:10.1109/TIFS.2016.2574309.

[44] T. Simon, H. Joo, I. Matthews, Y. Sheikh, Hand Keypoint Detection in Single Images using Multiview Bootstrapping, Tech. rep. arXiv:1704.07809v1.

[45] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recognition 71 (2017) 158–172.

[46] Z. Sun, T. Tan, Y. Wang, S. Z. Li, Ordinal palmprint represention for personal identification [represention read representation], in: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Vol. 1, IEEE, 2005, pp. 279–284.

[47] E. Yoruk, E. Konukoglu, B. Sankur, J. Darbon, Shape-based hand recognition, IEEE transactions on image processing 15 (7) (2006) 1803–1815.

[48] M. A. Ferrer, A. Morales, C. M. Travieso, J. B. Alonso, Low cost multimodal biometric identification system based on hand geometry, palm

and finger print texture, in: 2007 41st Annual IEEE International Carnahan Conference on Security Technology, IEEE, 2007, pp. 52–58.

[49] A. K. Jain, P. J. Flynn, A. A. Ross, ProQuest, Handbook of biometrics, Springer, New York, 2008.

[50] J. Han, M. Kamber, J. Pei, Data mining concepts and techniques, 3rd Edition, Morgan Kaufmann series in data management systems, Elsevier, Burlington, Mass., 2012.

[51] L. Zhang, H. Li, Encoding local image patterns using riesz transforms: With applications to palmprint and finger-knuckle-print recognition, Image and Vision Computing 30 (2012) 1043–1051. doi:10.1016/J.IMAVIS.2012.09.003.

[52] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[53] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Janua (2017) 2261–2269. arXiv:1608.06993, doi:10.1109/CVPR.2017.243.

[54] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, arXivarXiv:1704.04861.

[55] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.