Automated Person Identification Framework Based on Fingernails and Dorsal Knuckle Patterns

Mona Alghamdi
School of Computing and Communications
Lancaster University, UK
Department of Computer Science, Imam Abdulrahman Bin Faisal University
Al Jubail, Saudi Arabia
m.alghamdi@lancaster.ac.uk; msmalghamdi@iau.edu.sa

Plamen Angelov
School of Computing and Communications
Lancaster University
Lancaster, UK
p.angelov@lancaster.ac.uk

Bryan Williams
School of Computing and Communications
Lancaster University
Lancaster, UK
b.williams6@lancaster.ac.uk

Abstract*—Hand images are of paramount importance within critical domains like security and criminal investigation. They can sometimes be the only available evidence of an offender's identity at a crime scene. Approaches to person identification that consider the human hand as a complex object composed of many components are rare. The approach proposed in this paper fills this gap, making use of knuckle creases and fingernail information. It introduces a framework for automatic person identification that includes localisation of the regions of interest within hand images, recognition of the detected components, segmentation of the region of interest using bounding boxes, and similarity matching between a query image and a library of available images. The following hand components are considered: i) the metacarpohalangeal, commonly known as base knuckle; ii) the proximal interphalangeal joint commonly known as major knuckle; iii) distal interphalangeal joint, commonly known as minor knuckle; iv) the interphalangeal joint, commonly known as thumb's knuckle, and v) the fingernails. A key element of the proposed framework is the similarity matching and an important role for it is played by the feature extraction. In this paper, we exploit end-to-end deep convolutional neural networks to extract discriminative high-level abstract features. We further use Bray-Curtis (BC) similarity for the matching process. We validated the proposed approach on well-known benchmarks, the '11k Hands' dataset and the Hong Kong Polytechnic University Contactless Hand Dorsal Images known as 'PolyU HD'. We found that the results indicate that the knuckle patterns and fingernails play a significant role in the person identification. The results from the 11K dataset indicate that the results for the left hand are better than the results for the right hand. In both datasets, the fingernails produced consistently higher identification results than other hand components, with a rank-1 score of 93.65% on the ring finger of the left hand for the '11k Hands' dataset and rank-1 score of 93.81% for the thumb from the 'PolyU HD' dataset.

I. Introduction

In recent years, considerable attention has been paid to biometric recognition and its suitability to play a major role in person identification. Biometrics augmented by advanced machine learning techniques is a field of study aiming to recognize and identify subjects based on their characteristics. Examples of biometric identifiers are the face [1], iris [2], ear [3], hand [4], and footprint [5]. Each type has its strengths and weaknesses [6].

More recently, there has been growing attention to the investigation of the hand as a biometric identifier. The hand plays an important role in our daily activities, including the way we interact with the world [7]. The hand has specific characteristics which can identify an individual [8] and has many advantages as a biometric reservoir. The process of acquiring hand data is user-friendly, simple, cost-efficient, and time-efficient [9], [4]. The hands have many characteristic components such as knuckle creases (including the metacarpophalangeal (MCP) joints known as the base knuckle, proximal interphalageal (PIP) joint known as the major knuckle, the distal interphalangeal (DIP) joint known as the minor knuckle, and interphalangeal (IP) joint known as the major knuckle of the thumb), fingernails, hand shape, fingerprints, hand veins and palm creases [4].

The finger knuckle has a specific pattern, which includes creases, curves, lines, and textures. A knuckle pattern is a discriminating feature that has the potential for identification. Utilising knuckle creases as a biometric trait could be of value for human identification in different scenarios, such as criminal investigations [9].

The present paper presents a novel framework for automatic person identification based on different hand components including the knuckles and fingernails of the hands. This framework includes detection, and segmentation of the various regions. It involves feature extraction using the pretrained CNN model known as the DenseNet201, dissimilarity measuring using the Bray-Curtis distance metric to find the best match between individuals for person identification in a forensic application.

The biometric framework for person identification typically consist of several stages including image acquisition, image pre-processing, image segmentation, feature extraction, simi-

larity evaluation, and ranking the results. In the next two subsections, we will present these stages in more detail.

A. Knuckle creases and fingernail detection

Image pre-processing may include many techniques such as image enhancement, filtering, normalization, and illumination correction. [6]. In addition, image segmentation and localisation of regions-of-interests (ROIs) may play an important role in the efficiency of the overall framework. In general, detection and region localisation methods of the hand elements can be divided into two main categories. In the first category, several algorithms have been designed to extract low-level features, such as edges, texture, convexity and colour [10], [11], enhancing the colour and contrast information [12]. In the second category, some research papers have been published which propose that convolutional neural networks (CNN) models are trained on manually annotated data of knuckles, and use part of the trained CNN as abstract features. An example of this approach is the method proposed in [13] to train a state-of-the-art region-based CNN using different bounding boxes as ground truths for segmenting the PIP (major knuckles). The same method was extended to segment the PIP and DIP (minor knuckles), and fingernails [14].

Many papers have focused on the PIP [13], [14]. However, in real-world scenarios, knuckles must be extracted from the hand image with various rotations or quality variations [15]. A framework to localise and recognise different regions of the hand for person identification was presented in [15] in which a faster region-based convolutional neural network (R-CNN) and DenseNet201 models were trained on annotated data of DIP, PIP, and MCP (base knuckles) from hand images from the '11k Hands' and 'PolyU HD' datasets. The trained R-CNN model was used to localise and recognise different hand regions. However, the paper did not consider the fingernails or the thumb. The study [14] reports the use of the DIP, PIP and fingernails based on the Hong Kong Polytechnic University finger knuckle image (FKI) database [16], which consists of raw images of middle fingers only. This data is then passed through a Siamese network to identify individuals.

Most methods for deep-learning-based segmentation require extensive labelling of data for training the network, which is often hard to fulfil for sizeable datasets [17]. It has to be stressed that to date, the study of fingernails from hand images has not attracted attention within biometric systems research. One reason may be that fingernails are prone to change.

B. Feature extraction and matching

Feature extraction and selection play a fundamental role in pattern recognition problems. Features define the data space, and together with the distance metric used, are vital for the effectiveness of further analysis, including the matching stage. The typical algorithms for knuckle crease recognition can be categorised into coding methods [18], [19], subspace methods [20], [21] and texture analysis methods [22]. These are analysed and examples explored in this survey [6].

To the best of our knowledge, the first study to focus on the PIP for person identification problems is [23]. The author applied a 2D Gabor filter to extract orientation information around the finger knuckle pattern (FKP) to represent the features. To the best of our knowledge, the first study suggesting that minor finger knuckle patterns are instrumental for human identification is [8]. The author in this study compared different methods, including LBP, ILBP and 1D log Gabor filter. An attempt to investigate the holistic information on the dorsal surface of the hand was presented in [24]. An investigation conducted for the recovery and matching task of endpoint and bifurcation minutiae patterns of the finger knuckle from 120 different subjects was reported in [25]. [15] investigated the base, major, and minor knuckles of the four fingers to identify individuals. The fine-tuning of DenseNet201 deep learning was been utilised for feature extraction and the cosine distance used for matching.

On a brief review of the feature extraction methods of the hand, the studies did not consider feature extraction via deep learning without the need of retraining. Therefore, this paper focuses on utilising a pre-trained deep learning model as a feature extractor.

C. Contributions and outline of this paper

The contributions of this paper can be summarised as follows:

1) A novel framework for person identification is introduced using deep CNN to extract high level discriminative features. The proposed framework consists of detecting key hand components, recognition by automatic labelling, and similarity matching per component.
2) A segmentation method is proposed based on an extension of the state-of-the-art multi-view bootstrapping [7], which was extended to extract regions of the human hand's keypoints and made use of the automatic labelling. The hand's detected locations include the MCP, PIP, DIP, and the fingernails of the five fingers of the left and right dorsal surface of the hands.
3) The developed framework uses the four components, which are the MCP, PIP, DIP, and fingernails of all fingers from both right and left hands. To the best of our knowledge there is no publication that uses the thumb knuckles and fingernails.
4) We employ Bray-Curtis similarity to compare the extracted features. To the best of our knowledge, this is the first work to use this metric for feature comparison in a biometric application.

The rest of this paper is organized as follows: Section II introduces the design of the proposed framework for ***p***erson ***i***dentification based on ***f***ingernails and ***k***nuckle patterns localisation from hand images (PIFK). Section III describes the experimental results and evaluation, and section IV presents the conclusion and outlines future research.

II. The Proposed PIFK Framework

Figure 1 depicts the schematic diagram of the proposed PIFK framework. The proposed system consists of two phases: i) acquisition, and ii) identification. In the acquisition phase, a person presents the dorsal hand to capture images using the acquiring device. The captured images are then passed to pre-processing and generate cropped sub-images of the hand's keypoints as shown in figure 1. Each hand comprises 19 different components, namely:

1) 5 base knuckles or MCP,
2) 4 major knuckles or PIP,
3) 4 minor knuckles or DIP,
4) 1 major knuckle of the thumb or IP, as well as
5) 5 fingernails.

These are then stored in the relevant data-store d_n ($n \in \{1, 19\}$) as explained in sub-section III-B. A pre-trained deep neural network is then used to extract abstract high-level features from each finger's component, as described in more detail in sub-section II-B. These extracted features are sophisticated and instrumental to discriminate each sub-image.

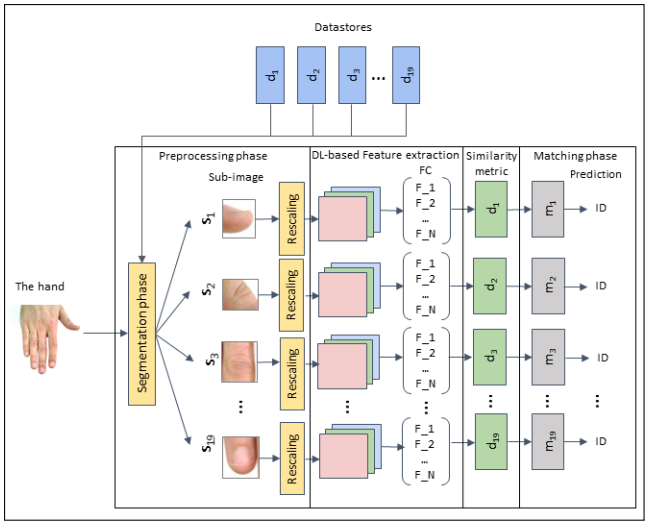

Fig. 1: The schematic diagram of the proposed framework for ***p***erson ***i***dentification based on ***f***ingernails and dorsal ***k***nuckle patterns (PIFK) from the hand image.

The proposed PIFK framework consists of the following phases (see figure 1):

- pre-processing: the knuckles and fingernails segmentation,
- feature extraction,
- (dis-) similarity evaluation and decision (matching).

These stages are detailed in the following subsections.

A. Detection and segmentation of knuckle and fingernail keypoints

In the pre-processing phase, firstly the original image of the hand was resized to 224 by 224 pixels. Each image was then segmented based on the detected keypoints. Then rescaling (normalization) of these sub-images was conducted to speed up the process of extracting features. The original pixel values of the sub-images range from 0 to 255. By rescaling, the pixel values are transformed to the interval from 0 to 1. In this paper, unlike other published approaches, segmentation of five types of region of interest was considered, namely: MCP, PIP, DIP, IP as well as the fingernail from both the left and right hands. The keypoints of these elements were obtained using the multi-view bootstrapping method for hand pose estimation [7].

The method reported in [7] was extensively trained on large datasets of annotated hand keypoints including challenge occlusion using a multi-camera setup to capture a multi-view of the hand. These datasets are the MPII human pose dataset [26], and images from the New Zealand sign language (NZSL) exercises of the Victoria university of Wellington [27]. Therefore, the method in this framework has not been trained on new data because it is already pre-trained for the hand regions localisation using the mentioned databases.

The basic concept of this method [7] is to use multiple views of multiple frames of the hand. The method can detect and project the position of the 3D keypoints. The method utilises a keypoint detector $d(.)$, which converts a cropped input image patch of a hand $I \in \mathbb{R}^{w \times h \times 3}$ to P component x locations as follows:

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

where confidence c_p is associated with each detection; a location point p matches one component, such as PIP or DIP; w denotes vertical and h denotes horizontal dimension of the image and 3 corresponds to the RGB channels.

Each location was detected using a pre-trained deep neural network. The detector was pre-trained on the previously mentioned databases with corresponding keypoint annotations, $(I^f, \{y_p^f\}$, where f indicates an image frame, and the annotated keypoints for the image I are in the set $\{y_p \in \mathbb{R}^2\}$.

The purpose of utilising this method was to estimate the approximate centre of the hand components. Each detected keypoint has a coordinate location x_p, a confidence score c_p, and an associated index. In this paper, the associated index $p \in [1...P]$ was used to support the labelling of each of the 19 ROI automatically with the confidence measure $c_p \in [0, 1]$ ranging from the lowest to highest score.

In this paper, we used the image of the hand represented by one frame and one view of the hand: $f = 1$, and $v = 1$, respectively. The confidence score c for the view v and keypoint p is considered as the degree of certainty, where combined with the training of the keypoints.

We extend the detection method to segment each component of the hand. The segmentation was conducted by determining the dimensions of a bounding box of each region empirically based on the detected centre-point. Therefore, the segmentation of the required region can be conducted in two stages: firstly, identifying the centre of the segmented part using the mult-iview bootstrapping method; and secondly, defining the bounding box by specifying a fixed height and width per type of component in regards to the centre-point. The dimensions

of the five types of ROI (MCP, PIP, DIP, IP and fingernails) were defined. Figure 2 displays the stages of the segmentation.

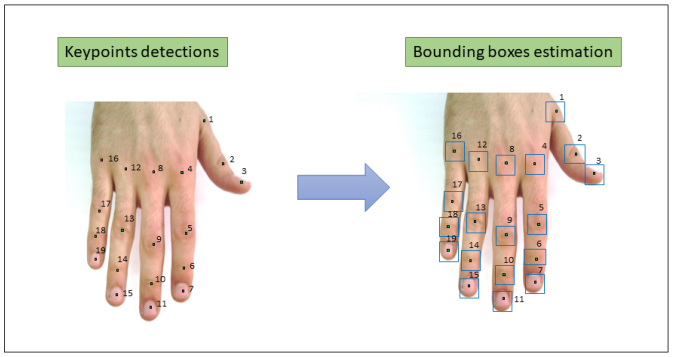

Fig. 2: Example of the localisation including keypoints/landmarks detection and bounding boxes for all fingernail and knuckle crease regions.

The following steps demonstrate the detection of the ROI and related bounding box estimation for image segmentation. Because the detection method requires a fixed size of the input image with dimensions 224 by 224 pixels, each detected keypoint assigned as (x', y') should be mapped to the same keypoint position denoted as (x, y) on the original high dimensional image. The image was resized and the coordinates transformed accordingly.

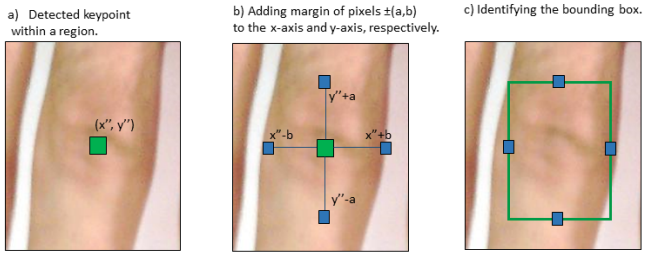

Fig. 3: An illustration of forming the bounding box around an identified keypoint.

Thirdly, we identify the bounding box to generate a sub-image S_n of the region:

Suppose a centre location of a keypoint is (x'', y''), to identify a bounding box, an estimated margin of pixels defined by the user assigned as a and b is added to both positive and negative sides of the y-, and x-axis, respectively. Figure 3 illustrates this concept using a patch of the knuckle area.

B. Feature extraction phase

Many image processing techniques consider low-level characteristics, such as edges, shapes, geometries, and lighting changes. However, these techniques often require extensive human intervention in the designing stages. By contrast, deep learning and specifically CNNs, started to be widely used as a common mechanism for extracting features and learning the pattern representation from high dimensional data such as images [28]. These are usually pre-trained on large image sets, such as ImageNet [29].

A deep neural network model pre-trained on one domain can transit the representation of learned data to another domain or task using a technique called transfer learning [30]. It offers an architecture that enables extracting abstract and high-level features [29], which usually represent the outputs of the final fully connected (FC) layer in a vector form, the dimensions of which vary depending on the pre-trained model. In this paper, we did not do fine-tuning for the original model of DenseNet201. Instead, we used the original DenseNet201 model [31] that was trained on the popular 'ImageNet' dataset [29] as a feature extractor. The model has a more complex architecture which makes it a more powerful feature extractor. The architecture of the model [31] considers specific layer connectivity for data flow. In traditional CNNs, each layer has l inputs, which are feature-maps from all previous convolutional blocks. Instead, in the DenseNet201 network, each layer l, receives data inputs from all previous layers, maps them into features, and forwards them to all succeeding layers. The output layer is denoted as x_l. A layer achieves a concatenation of features by applying $\frac{l(l+1)}{2}$ connections, compared to a typical CNN architecture that only uses a l layer network. The DenseNet201 network requires fewer parameters compared to other CNNs, enhancing data flow and gradients through the model. Figure 4 shows the network layout with three dense blocks.

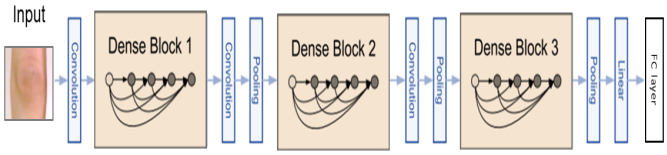

Fig. 4: DenseNet201 architecture with three dense blocks.

The feature-maps from all preceding layers are received as input to the l^{th} layer as follows:

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]), \tag{2}$$

where $[x_0, x_1, ..., x_{l-1}]$ denotes feature-map concatenation output in layers $0, ..., l-1$, and $H_l(.)$ refers to compound operation that comprises of batch normalization, followed by a rectified linear unit and a 3×3 convolution [31]. The segmented sub-images were resized to 224 by 224 pixels to be compatible with this network for feature extraction. The sub-images were rescaled to a new value domain ranging from 0 to 1. In this way, all features are given an equal weight, which is useful for feature extraction and the similarity metric [32].

C. Similarity estimation and matching

Similarity metrics can play an essential role in biometric systems. In this paper, the Bray-Curtis distance metric [33] was used to identify the dissimilarity or distance

among different ROI as part of the decision phase of the person identification task. After exploring the matching results among other distances like cosine and euclidean, the Bray-Curtis performed better. BC distance between vectors $a = (a_1, a_2, ..., a_p)$ and $b = (b_1, b_2, ..., b_p)$ is denoted as $d_{BC}(a, b)$ and can be defined as follows:

$$d_{BC}(a,b) = \frac{\sum_{i=1}^{P} |a_i - b_i|}{\sum_{i=1}^{P} |a_i + b_i|}, \tag{3}$$

The similarity between two vectors is the opposite to the distance (the lower the distance between two vectors the higher the degree of similarity and vice versa) and can be defined as:

$$Sim_{BC}(a,b) = 1 - d_{BC}(a,b) \tag{4}$$

The ID is assigned to the closest match determined through the Bray-Curtis similarity as follows:

$$\hat{ID} = argmin\{d_{BC}\} = argmax\{Sim_{BC}\} \tag{5}$$

D. Experimental protocol

In our experiments we consider a valid recognition the one that matches two different sub-images taken from the same person and vice versa. We used $rank$-1 recognition rate which is computed as follows:

$$rank\text{–}1 = \frac{N_i}{N} \times 100, \tag{6}$$

where N_i denotes the number of images correctly assigned to the right individual, and N indicates the overall number of images attempted to be identified.

In order to demonstrate the overall performance we also used the cumulative matching characteristic (CMC) which shows the accuracy performance in terms of $rank - n$ [34]. The description of the datasets used to evaluate the proposed method, and the results, will be demonstrated in the next section.

III. Experimental results and evaluation

This section reports the results of the evaluation of the proposed PIFK on the '11k Hands' dataset [35].

A. Datasets description

First, we used the '11-k Hands' dataset [35], which consists of RGB images of the dorsal and palm surface of the right and left hands of 190 subjects (in this study, we focused on the dorsal only). Each image has a resolution 1600×1200 pixels.

Next, we consider the Hong Kong Polytechnic University Contactless Hand Dorsal Images Database [24], which consists of 4650 surfaces of right-hand dorsal images in a flat position from 501 subjects. These images with the same resolution (1600×1200 pixels) are captured using mobile and handheld cameras.

B. Pre-processing phase

The pre-processing phase (see figure 1) was described in sub-section II. In this sub-section, the specifics of PIFK regarding this phase will be detailed.

Step 1: Fingernail and knuckle crease detection:

This was described in sub-section II and illustrated in figures 1 and 3.

Step 2: Recognition of knuckle creases and fingernails:

PIFK uses an automatic indexing of each ROI; for example, index 1 and index 2 support labelling of the MCP and IP of the thumb, respectively (see figure 2).

Step 3: Bounding box estimation and segmentation:

The estimated dimension of the bounding box in terms of (width, height) was predefined in both datasets as follows: (150, 168) pixels for the MCP, (150, 160) for the PIP and for the IP, (150,140) for the DIP, and (160, 184) for the fingernails. These values were found through experimentation.

Step 4: Splitting the data into two data sets:

In this phase, we split the sub-images into two data sets:

- a query data set, and
- a library data set.

In this study, we used the Leave-One-Out Cross-Validation (LOOCV) evaluation method since the number of images is not very large. The method assessed the performance of the proposed algorithm. In LOOCV, each fold has only one sample, and random partitioning of the data into training and testing does not exist. Also, every prediction in the identification problem is independent of each other because the samples are independent [36].

Therefore, we consider a single individual subject in the query at a time (and vary this averaging the results at the end) and the remaining individuals in the library set (also replacing one at a time). In addition, we have for each individual images of both, left and right hand and 19 ROI per image. In the '11k Hands' dataset, due to false detection of keypoints, we ignored 87 images of left hands and 6 images of right hands. After segmentation due to the erroneous detection, 428 sub-images of left hands' components were removed and 573 of right hands' components. As a result, the total number of sub-images in the query set for all components of the left hands were 3,589 and 3,609 for the right hands. Finally, the total number of images in the library set for all components of the left and right hands were 48,780 and 50,795, respectively. In the 'PolyU HD' dataset, the total number of removed sub-images are $5,733$ out of $88,392$. Therefore, the total sub-images in the query set are $9,411$, whereas the library set is $73,248$.

C. Pre-trained deep learning based feature extractor

In regards to the feature extraction, in this study we utilised a DenseNet201 neural network pre-trained on the ImageNet dataset. We considered the outputs of the last fully connected (FC) layer which has dimension (1×1920), namely FC2 as a vector that represents the abstract high-level features.

The evaluation of the proposed algorithm is a subject-independent manner. The data used to train the original

DenseNet201 model is the 'ImageNet' dataset. It is different from the data used to evaluate the proposed method. We used '11khands' and 'PolyU' datasets.

D. Similarity estimation and matching

The matching process includes estimating the similarity between two sub-images of the same hand element (one from the query and one from the library). BC distance was used to estimate the similarity between pairs of sub-images as detailed in equations (3), (4) and (5) from sub-section II-C.

TABLE I: The rank-1 recognition rate (shown in %) for the '11k Hands' and 'PolyU HD' datasets

Region	Finger	11k Hands-L	11k Hands-R	PolyU-R
Fingernail	Thumb	87.83	84.21	93.81
	Index	89.42	88.95	90.40
	Middle	90.48	89.47	87.65
	Ring	93.65	91.58	85.10
	Little	84.13	80.00	87.30
Minor Knuckle	Index	84.66	76.84	72.47
DIP	Middle	85.19	82.11	68.92
	Ring	84.13	77.89	71.46
	Little	81.48	76.84	78.43
Major Knuckle	Thumb	85.71	84.21	76.83
PIP	Index	82.45	83.16	79.71
	Middle	85.11	82.54	76.64
	Ring	80.95	83.68	83.23
	Little	83.07	75.79	82.06
Base Knuckle	Thumb	87.30	85.26	58.35
MCP	Index	78.84	81.58	64.34
	Middle	80.95	77.37	67.27
	Ring	80.42	77.37	66.47
	Little	84.13	80.53	67.94

In figure 5, the CMC of the major knuckles of the left (chart a)) and right (chart b)) hands and fingernails of the left (chart c)) and right (chart d)) hands from the '11k Hands' dataset are shown, respectively.

Interestingly, as indicated by the results (see table I and figure 5), the left hand—including the fingernails and knuckles from '11K Hands'- is more identifiable than the right hand for the majority of the fingers.

We also observed that the recognition using fingernail sub-images from both hands and both datasets is higher than that of all knuckle sub-images, which can clearly be seen in table I for both data sets and visualized in figure 7 for the '11K Hands' dataset. In particular, we achieved the best performance results on the left fingernails, with rank-1 accuracies of 93.65% on the ring finger, 90.48% on the middle, and 89.42% on the index from '11K Hands'. Among the right fingernails, the highest performance -as demonstrated by table I- was achieved by using the ring, middle, and index, with accuracies of 91.58%, 89.47%, and 88.95%, respectively. The best results for the

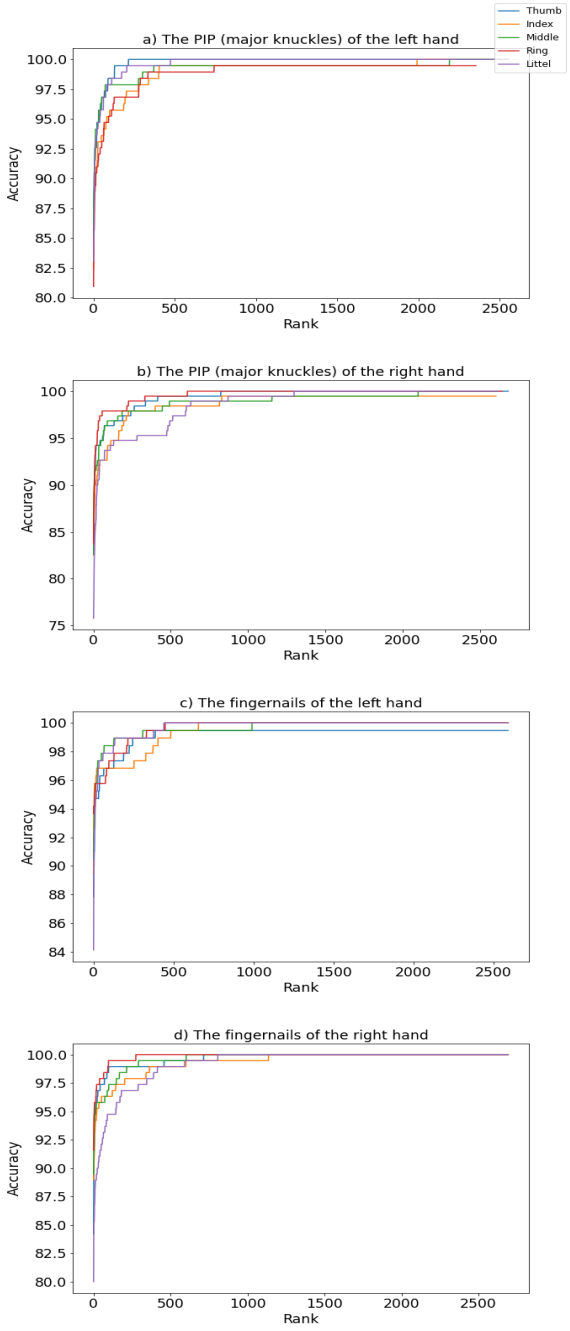

Fig. 5: The CMC of the proposed PIFK; a) major knuckles of the left hands; b) major knuckles of the right hands; c) fingernails of the left hands; d) fingernails of the right hands from the '11k Hands' dataset.

'PolyU HD' dataset were also achieved for the fingernails with rank-1 accuracies 93.81% and 90.40% for the thumb and the index finger, respectively. These results may be linked to the fact that the fingernails have a more rigid and defined shape [14] than the knuckle creases.

In addition, we can observe that the MCP are relatively insignificant compared to the other components of the hand.

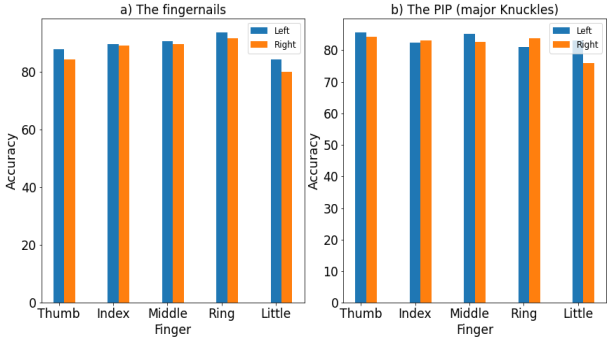

Fig. 6: The recognition accuracy for the left and right hands from '11K Hands' dataset based on fingernails (left) and the major knuckles, PIP (right).

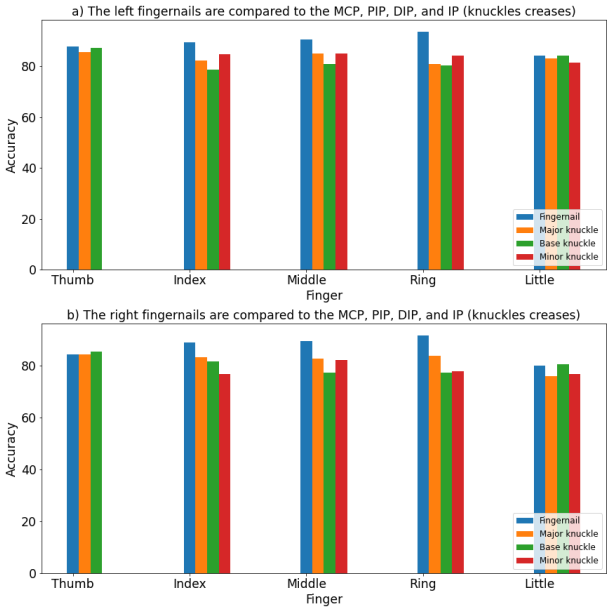

Fig. 7: The performance of fingernail regions is compared to the knuckle regions for: a) the left, and b) the right hands in the '11k Hands'.

The PIP perform slightly better than the DIP, particularly on the thumb and middle fingers of the left hands, with rank-1 accuracies 85.71% and 85.11%, respectively for the '11K hands' dataset. Among the right-hand parts, the IP and the PIP obtain better results, with rank-1 accuracies 84.21% and 83.68% respectively for the same dataset. This observation is also valid for the 'PolyU HD' dataset with rank-1 accuracies 83.23% and 82.06% for the ring and little fingers, respectively. This may refer to the fact that the PIP patterns contain more discriminative creases, which might increase the identification performance over the DIP patterns.

The results also revealed that the IP and MCP of both thumbs achieved higher results than the other fingers from '11k Hands' dataset; the IP (major knuckles) of the left and right thumb obtained rank-1 accuracies 85.71% and 84.21%, respectively, while the MCP of the left and right thumb obtained rank-1 accuracies 87.30% and 85.26%, respectively.

In this study, for the first time, we study 19 components of human hands at the same time and the results are shown in table I. The comparison of the results with existing state-of-the-art approaches is complicated by the fact that there are no publications that use the full variety of components. However, we validated our approach on two different datasets. Furthermore, we made a general comparison with another state-of-the-art approach, namely [14] which only used the PolyU FKI [16] and PolyU FKP [4] databases. In this study the PolyU FKI [16] database was used to segment 15 components, namely the fingernails, PIP (major knuckles), and DIP (minor knuckles) from finger images. However, the study did not mention on which finger these components where located. Furthermore, the '11k Hands' dataset contains the hand images of both hands and is larger than the PolyU FKI database. The PolyU FKP [4] dataset only contains the right and left PIP and it is not clear from which finger or hand these images were captured.

The performance using the FKIMNet method [14] achieved a rank-1 score 94.83% for the fingernail, 90.52% for the PIP and 88.73% for the DIP using the PolyU FKI dataset. These results are similar in magnitude to the results we report. However, our results are on a larger and more complex dataset and clearly indicate where the keypoints are located.

Furthermore, we report interesting observations confirmed on two different datasets that fingernails provide better results than other hand components. In addition, we also observed that the left hands achieved higher results than the right hands.

IV. Conclusion

This paper introduces a framework for automated human identification for a forensic application, where the hand image might be the only existing evidence to identify offenders at a crime scene. The framework used the dorsal surface of the five components of the hand (fingernails, MCP, PIP, DIP, and IP of the five digits). The proposed approach is the first of its kind to utilise all 19 hand components for identifying individual subjects. The framework begins with the localisation and identifying of these 19 components from the input hand image. It further automatically labels them and builds bounding boxes around identified keypoints. In this study, we evaluated the proposed method using two popular datasets ('11k Hands' and 'PolyU HD') and compared our results with another study [14]. We observed interesting findings that concern fingernails and the consistent difference of results between the left and the right hand in favour of the left hand. The best-performing elements were shown to be the fingernails for both datasets. In addition, the framework extracted abstract and discriminating features using a pre-trained CNN model DenseNet201. Further, we employed Bray-Curtis distance metric for the matching process. The proposed method achieved high-performance results. We will improve the proposed framework in our future work by investigating different similarity metrics and CNN models for feature extraction.

References

[1] H. M. Sim, H. Asmuni, R. Hassan, and R. M. Othman, "Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face

images," *Expert systems with applications*, vol. 41, no. 11, pp. 5390–5404, 2014.
[2] R. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.
[3] A. Kamboj, R. Rani, A. Nigam, and R. Ranjan Jha, "CED-Net: context-aware ear detection network for unconstrained images," *Pattern Analysis and Applications*, vol. 24, pp. 779–800, 2021.
[4] A. Kumar and C. Ravikanth, "Personal Authentication using Finger Knuckle Surface," Tech. Rep. 1, 2009.
[5] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Computer Vision and Image Understanding*, vol. 167, pp. 1–27, feb 2018.
[6] G. Jaswal, A. Kaul, and R. Nath, "Knuckle print biometrics and fusion schemes - Overview, challenges, and solutions," *ACM Computing Surveys*, vol. 49, no. 2, 2016.
[7] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," tech. rep.
[8] 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)*, pp. 55–60, IEEE, 2012.
[9] G. Jaswal, A. Nigam, and R. Nath, "DeepKnuckle: revealing the human identity," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18955–18984, 2017.
[10] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, "Online finger-knuckle-print verification for personal authentication," *Pattern Recognition*, vol. 43, pp. 2560–2571, jul 2010.
[11] A. Nigam, K. Tiwari, and P. Gupta, "Multiple texture information fusion for finger-knuckle-print authentication system," *Neurocomputing*, vol. 188, pp. 190–205, may 2016.
[12] G. Schaefer, M. I. Rajab, M. E. Celebi, and H. Iyatomi *Computerized Medical Imaging and Graphics*, pp. 99–104.
[13] R. Ranjan Jha, D. Thapar, A. Nigam, and S. M. Patil, "UBSegNet: Unified Biometric Region of Interest Segmentation Network," tech. rep.
[14] D. Thapar, G. Jaswal, and A. Nigam, "FKIMNet: A Finger Dorsal Image Matching Network Comparing Component (Major, Minor and Nail) Matching with Holistic (Finger Dorsal) Matching," 2019.
[15] R. Vyas, H. Rahmani, R. Boswell-Challand, P. Angelov, S. Black, and B. Williams, "Robust End-to-End Hand Identification via Holistic Multi-Unit Knuckle Recognition," *IJCB 2021, to appear*, 2021.
[16] A. Kumar, "Importance of Being Unique From Finger Dorsal Patterns: Exploring Minor Finger Knuckle Patterns in Verifying Human Identities," *IEEE transactions on information forensics and security*, vol. 9, no. 8, pp. 1288–1298, 2014.
[17] N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, "Hand segmentation with structured convolutional learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9005, pp. 687–702, 2015.
[18] L. Zhang, L. Zhang, and D. Zhang, "Finger-Knuckle-Print Verification Based on Band-Limited Phase-Only Correlation," tech. rep., 2009.
[19] Z. Le-Qing and Z. San-Yuan, "Multimodal biometric identification system based on finger geometry, knuckle print and palm print," 2010.
[20] D. S. Guru, K. B. Nagasundara, and S. Manjunath, "Feature level fusion of multi-instance finger knuckle print for person identification," *ACM International Conference Proceeding Series*, pp. 186–190, 2010.
[21] Z. S. Shariatmadar and 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*, vol. 3, pp. 1480–1484, 2011.
[22] L. Zhang, L. Zhang, D. Zhang, and H. Zhu, "Ensemble of local and global information for fingerknuckle-print recognition," *Pattern Recognition*, vol. 44, no. 9, pp. 1990–1998, 2011.
[23] Lin Zhang, Lei Zhang, and D. Zhang, "Finger-knuckle-print: A new biometric identifier," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 1981–1984, IEEE, 2009.
[24] A. Kumar and X. Zhihuan, "Personal Identification Using Minor Knuckle Patterns From Palm Dorsal Surface," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2338–2348, 2016.
[25] A. Kumar and B. Wang, "Recovering and matching minutiae patterns from finger knuckle images," *Pattern Recognition Letters*, vol. 68, pp. 361–367, 2015.
[26] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," tech. rep.
[27] M. D. McKee, Rachel and D. Alexander, "Nz sign language exercises," 2010.
[28] S. Khan, *A Guide to Convolutional Neural Networks for Computer Vision*. 2018.
[29] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE.
[30] L. Yann, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017.
[32] J. Han, M. Kamber, and J. Pei, *Data mining concepts and techniques*. Morgan Kaufmann series in data management systems, Burlington, Mass.: Elsevier, 3rd ed. ed., 2012.
[33] R. Shyam and Y. N. Singh, "Face recognition using augmented local binary pattern and Bray Curtis dissimilarity metric," *2nd International Conference on Signal Processing and Integrated Networks, SPIN 2015*, pp. 779–784, 2015.
[34] A. K. Jain, P. J. Flynn, A. A. Ross, and ProQuest, *Handbook of biometrics*. New York: Springer, 2008.
[35] M. Afifi, "11K Hands: Gender recognition and biometric identification using a large dataset of hand images," *Multimedia Tools and Applications*, vol. 78, pp. 20835–20854, 2019.
[36] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015.