

Deep Feature Collaboration for Challenging 3D Finger Knuckle Identification

Kevin H. M. Cheng, Ajay Kumar

Abstract—Contactless 3D finger knuckle pattern is a new biometric identifier which offers highly discriminative features for the finger knuckle based personal identification. State-of-the-art methods for object recognition, a more generic problem, employ deep neural network based approaches and demonstrate superior effectiveness. However, any direct applications from those methods do not outperform specialized hand-crafted feature description approaches for the problem addressed in this paper. In addition, such deep neural network based methods have to address challenges associated with emerging biometrics, e.g. availability of very limited training data, large intra-class or train-test sample variations as observed for the real applications, etc. This paper attempts to address the above challenges and introduces a new deep neural network based approach for the contactless 3D finger knuckle identification. Our approach simultaneously encodes and incorporates deep features from multiple scales to form a more robust deep feature representation. Such collaborative feature representations are robustly matched using an efficient alignment scheme with a fully convolutional architecture to accommodate involuntary finger variations during the contactless imaging. Comparative experimental results in the two-session 3D finger knuckle images database, acquired from over 200 subjects and is publicly introduced from this paper, illustrate superior performance over the state-of-the-art methods, e.g. offering ~22% GAR improvement at extremely low FAR under challenging comparison scenarios. Additional experiments in other publicly available databases including 3D palmprint, 3D fingerprint, and 2D finger knuckle further validate the effectiveness and demonstrate the generalizability of the proposed approach.

Index Terms—3D hand biometrics, finger knuckle recognition, convolutional neural network

I. INTRODUCTION

BIOMETRIC recognition offers convenient and reliable solutions for many emerging applications such as unlocking smartphones, authentication for crossing borderlines, and E-business. Among various biometric identifiers, face [1, 2], fingerprint [3, 4], and iris [5, 6] are the most popular choices of biometrics. However, every biometrics has its own applications and limitations. For examples, a NIST report [7] submitted for the US Congress stated that about 2% of the population does not have usable fingerprints; iris recognition requires high-quality images; face recognition is vulnerable to presentation attacks with sophisticated make-up. In such scenarios, finger knuckle recognition provides an alternative

TABLE I: COMPARATIVE SUMMARY OF OUR METHOD WITH STATE-OF-THE-ART APPROACHES

	Conventional Learning Approaches	Hand-crafted Approaches	This Paper
Recognition Mechanism	Prediction from Deeply Learnt Abstract Features	Comparison with Spatial Feature Templates with Alignment	Prediction from Deeply Learnt Multi-Scale Features and with Alignment
Key Limitations	Challenging when Train-Test Sample Variation is Large	Difficult to Design Complex Feature Descriptors	Requires More Computational Time for Alignment

with several advantages: finger knuckle biometrics itself provides invariant and discriminative information for reliable recognition; combining the complementary information in addition to other biometric identifiers enables higher recognition performance; the acquisition of finger knuckle images is convenient and can be simultaneously with fingerprint. Therefore, finger knuckle recognition [8-11, 21] has recently become an active research frontier.

The recent advancement of finger knuckle recognition utilized 3D information for more accurate recognition [21]. The current bottleneck of such 3D finger knuckle recognition lies in the limited capability from the hand-crafted approach. Meanwhile, the effectiveness of neural network technologies has been well demonstrated in various computer vision tasks such as object segmentation [12-15] and object recognition [16, 17]. The applications of such approach have also been investigated for several specific biometric problems such as fingerprint recognition [18], iris recognition [19], and face recognition [20]. Therefore, it is highly motivated to investigate a customized neural network to advance the recent 3D finger knuckle recognition framework.

A. Limitations and Challenges

There are several challenges associated with neural network based approach for addressing the 3D finger knuckle recognition problem. Firstly, state-of-the-art methods introduced for various applications often adopt deep architectures with many layers, which require a large number of training samples. However, in real biometric or forensic scenarios, only one or a few samples are available. As for research investigation, there is only one publicly available 3D

finger knuckle database [21] to-date where the images are acquired from 130 subjects, each with six forefinger and middle finger images, while two-session images are available from 105 subjects. Therefore, effectively training a neural network with a few samples is very difficult. Secondly, it is further challenging for neural networks to perform well given the different statistical distributions of the training data and test data. In real biometric application scenarios, finger knuckle images during operation phase can be acquired from different cameras or lenses, which can result in large intensity, deformation, and geometric changes, as compared to respective samples acquired during the user registration. When the acquisition of gallery and probe images is from different cameras, resulting image variations can contribute to noticeable differences between the training and test sets. The 3D finger knuckle database provides such scenarios and has been acquired under two-session imaging using different camera lenses. Therefore, the performance degradation, using state of art methods, is noticeable under such cross comparison scenarios. Thirdly, the development of neural network based approach in this new 3D finger knuckle recognition problem requires empirical cycles for customizing network architecture and tuning hyperparameters. Lack of any such research efforts to this new biometric identifier makes it difficult to develop such deep neural networks from scratch.

B. Our Work and Key Contributions

Our work includes the development of a specialized neural network which collaborates multi-scales feature information and enables the efficient alignment of biometric images which are usually associated with translational variations, for contactless 3D finger knuckle recognition. The key contributions of this paper can be summarized as follows.

(i) This paper introduces a new neural network approach based on collaborating features for more accurate contactless 3D finger knuckle identification. Our approach incorporates deep features from multiple scales to form a more reliable deep feature representation. Such collaborative feature representations are robustly matched using an efficient alignment scheme with a fully convolutional architecture to accommodate involuntary finger variations during the contactless imaging. Challenges posed due to the availability of very limited training samples are mitigated by incorporating transfer learning of generic image features from bottom layers of *ResNet* [16], which has been well trained on very large datasets for object recognition. Besides, the complexity of our biometric problems is less when comparing to generic object recognition problems, therefore a neural network with less layers is expected to produce adequate performance, while excessive layers may induce more serious overfitting issues. Our framework therefore incorporates a compact network architecture to address the 3D finger knuckle identification problem considered in this paper.

Contactless 3D finger knuckle patterns illustrate irregular ridges and valleys with varying sizes. Since convolution neural network based approach usually pools information from bottom layers to top layers and therefore the higher or top layers can

illustrate the feature maps at lower scales which can also be more informative in discriminating 3D knuckle patterns at coarse level. Therefore, jointly utilizing the features from such multiple scales or layers can offer more discriminant feature representation, for more accurate finger knuckle recognition, and is the key motivation of incorporating collaborative feature representation in our framework. Another difficulty for comparing finger knuckle images is to accommodate the frequently observed translational variations, which is more significant than those in other similar hand biometric identifiers such as palmprint and fingerprint because the region of interest in finger knuckle patterns lacks a well-defined boundary. Therefore, an efficient alignment scheme for the frequently observed translational changes is incorporated in our framework to generate more accurate matches between the registered images and the test images with significant/higher intra-class variations. Such arguments for the development of our framework are systematically and empirically validated during the ablation studies and can be observed along with the comparative experimental results in Section IV.B.

(ii) Although the focus of our work is on contactless 3D finger knuckle identification, we also demonstrate the generalizability of the proposed approach in other similar contactless hand-biometric identification problems and present comparative experimental results using publicly available contactless 3D palmprint, 3D fingerprint and 2D finger knuckle databases. These outperforming experimental results, presented in Section IV, are highly promising and validate the effectiveness of the proposed approach.

(iii) This paper also develops and introduces a significantly larger contactless 3D finger knuckle database in public domain to advance further research in this area. This database is acquired from 228 different subjects (with two-session from 190 different subjects) and is extended from the earlier version acquired from 130 different subjects (with two-session images from 105 subjects). Six 3D forefinger images and six 3D middle finger images are acquired and available for each subject per session. This new 3D finger knuckle database contains challenging images that can represent real-world scenarios where the second session images were acquired under different imaging environment or with different imaging lens and illumination.

Table I summarizes the difference between our proposed method, conventional learning approaches and hand-crafted approaches while Table II summarizes the comparative performance using Genuine Acceptance Rate (GAR) for a high security condition where False Acceptance Rate (FAR) is at 10^{-4} . These results are selected from the best performing

TABLE II: COMPARATIVE RECOGNITION PERFORMANCE FOR HIGH SECURITY

GAR at FAR= 10^{-4}	Conventional Learning Approaches	Hand-crafted Approaches	This Paper
3D Finger Knuckle	39.4%	77.5%	94.7%
2D Finger Knuckle	75.9%	75.8%	86.6%
3D Palmprint	96.7%	95.6%	98.0%
3D Fingerprint	95.8%	92.5%	94.2%

methods of individual approaches, while the comprehensive experimental results will be presented in Section IV. In order to ensure the reproducibility, the newly collected 3D finger knuckle images dataset, the network model and implementation codes for the experiments described in this paper can be downloaded via the weblink [34].

Rest of this paper is as follows: Section II discuss the related work; Section III presents the technical details of the proposed approach; Section IV presents the comparative experimental results; Section V concludes this paper and present the further research direction.

II. RELATED WORK

Contactless 3D finger knuckle recognition is a new area for automatic personal recognition. The use of 3D finger knuckle patterns for biometric recognition was first studied together with the use of finger dorsal surfaces [23]. Unfortunately, the attention on 3D finger knuckle recognition has not been aroused probably because the effectiveness of using 3D finger knuckle information was constrained by the low resolution of 3D finger dorsal images being studied and the ineffective feature descriptor employed for extracting discriminative features, which is a generic surface shape descriptor, the Shape Index [22, 26]. The attentions on 3D finger knuckle recognition has been raised by a recent research [21] which specifically investigates the use of 3D finger knuckle patterns for biometric recognition. This study investigated various aspects of 3D finger knuckle recognition, such as the feature description, the possibility of presentation attacks towards a finger knuckle recognition system, the individuality of finger knuckle, the comparisons between 2D and 3D finger knuckle recognition, and provides a benchmark database for further research and investigation. The effectiveness and potentials of using 3D finger knuckle images for biometric recognition has been validated. However, there are other open questions on using this new biometric trait such as whether using neural network can do a better job.

As the studies on 3D finger knuckle recognition is new, they are supported by the literatures from finger knuckle recognition [8-11, 25], palmprint recognition [24, 27, 32-33, 43] and palm vein recognition [40, 42]. A survey paper on finger knuckle recognition [10] summarized existing work in 2D finger knuckle recognition. However, it has been shown in [21] that a state-of-the-art method for 2D palmprint recognition, *Difference of Normal (DoN)* [27], outperforms other competing methods including *Fast-RLOC* and *Fast-ComCode* [33] for the 2D finger knuckle recognition evaluation. Therefore, this method is considered as a baseline for the evaluation of 2D

finger knuckle recognition. On the other hand, *Surface Code* [32] and *Binary Shape* [33] are the two competing methods for the performance comparison of 3D finger knuckle recognition in [21]. Both methods are originally developed for 3D palmprint recognition, which also compute binary biometric templates from 3D depth images. Unlike these two methods, the surface gradient derivative (*SGD*) feature descriptor [21] computes binary biometric templates from 3D surface normal images. It is currently the state-of-the-art method for 3D finger knuckle recognition.

Beside the hand crafted feature approaches, neural network based methods have been actively investigated for many applications while convolutional neural network (CNN) based approach is one of the leading state-of-the-art in computer vision related tasks such as object recognition [16-17, 28], instance segmentation [12-13] and biometric recognition [19, 41]. *AlexNet* [28] is one of the representatives from CNN approach. *VGGNet* [17] investigated the use of a small kernel size (i.e. 3×3) for extracting robust features. *ResNet* [16] is another more recent CNN based method, which has shown to offer state-of-the-art performance for object recognition. In addition, collaborating multiple deep features can offer promising performance for object recognition (e.g. *DenseNet* [46], *SE-ResNet* [47]) and semantic segmentation [44, 45]. Besides, learning from one or a few images is also an important problem for real world applications. The one-shot learning problem was first addressed using a Bayesian approach [37]. Recent attentions attempt to address this problem with specially designed networks such as Memory-Augmented Neural Networks [30] and Matching Networks [31]. Although neural network models are expected to be quite generalizable, its application on specific biometric problems requires customized development.

Currently, there are several challenges when employing neural network approach for 3D finger knuckle recognition. For examples, many existing neural network models requires a lot of training data, but it is not practical to acquire many training data for each human subject, especially for research investigation with limited resources. Besides, neural network approach can perform well on simple recognition tasks if the distribution of the training and testing samples are similar. However, it is more difficult to perform well if the distribution of the training and testing samples are different. Last but not least, the development of neural network approach for 3D finger knuckle recognition requires empirical efforts from researchers, which is a promising area yet to be developed.

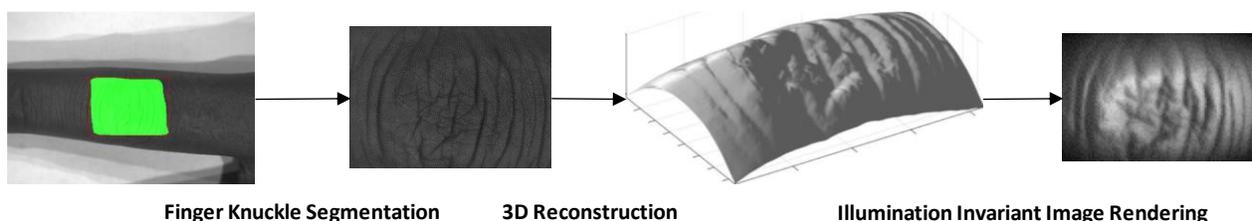


Fig. 1. Illustration of key steps for finger knuckle image pre-processing.

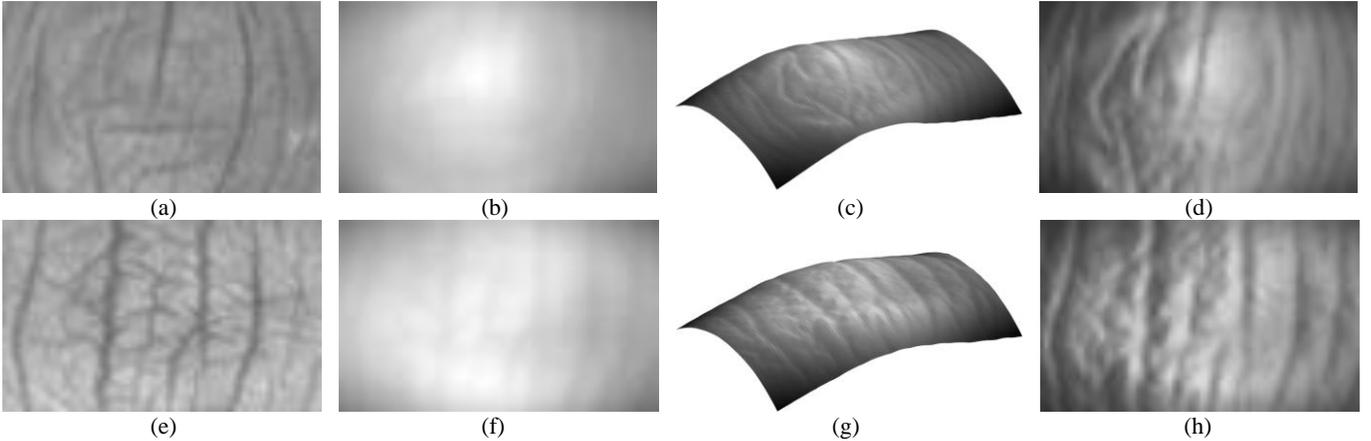


Fig. 2. Sample Images for: (a)/(e) Intensity Image (2D); (b)/(f) Depth Image (3D); (c)/(g) Rendered Image (3D View); (d)/(h) Rendered Image (Top View).

III. 3D FINGER KNUCKLE RECOGNITION

This section presents the technical details of the proposed 3D finger knuckle recognition framework.

A. Preprocessing of 3D Finger Knuckle Images

Figure 1 illustrates the key steps for preprocessing finger knuckle images. We first segment the finger knuckle area of interest (also known as finger knuckle detection), followed by 3D reconstruction and image rendering.

One of the reliable finger knuckle segmentation method described in [21] employed a simple edge pixel counting mechanism. Despite this method produce acceptable segmentation performance, it fails for some challenging samples which limits the finger knuckle recognition performance. Therefore, we attempted to incorporate a popular neural network method, *Mask R-CNN* [12] for finger knuckle segmentation. We first prepare a training set while the ground truth masks are prepared by applying the finger knuckle segmentation method detailed in [21]. This method first computes the edge image from the original finger knuckle image, followed by counting the number of edge pixels within a fixed size sliding window. The sliding window is shifted vertically and horizontally along the image. The location where the maximum number of edge pixels within the sliding window is considered as the area of interest. Since the original segmentation method may produce incorrectly segmented images, we manually remove those unsuccessful samples and utilize the remaining samples for generating the ground truth for the training of *Mask R-CNN* with fine tuning from the COCO dataset [29]. The effectiveness of improved segmentation for more accurate finger knuckle recognition will be verified by the ablation study presented in Section IV.B.

Emerging research on the use of 3D information for finger knuckle recognition suggested that 3D information provides more and stable information due to its property of illumination invariant. Therefore, rather than using the 2D intensity images, we attempt to utilize the 3D information from finger knuckle patterns for biometric recognition. It is not the same as to normalize 2D intensity images, which does not incorporate 3D

information. We found that using the normalized depth images as the network input degrades the recognition performance when compared to using the conventional 2D intensity images, probably because of the low contrast in the depth images. Therefore, we attempt to render artificial images from 3D finger knuckle images which is expected to be illumination invariant. We design a single directional light source at infinity and disable the specular reflections from the finger knuckle surface. Such rendered images can inherit the embedded 3D information as well as providing good contrast for the networks. Using the rendered images which embeds 3D information helps finger knuckle recognition despite the existence of the reconstruction errors, which will be verified by the ablation study presented in Section IV.B. Figure 2 shows the sample images for visualization. Those rendered images are used as inputs for training and testing the finger knuckle networks presented in Section III.B.

B. Finger Knuckle Networks (FKNet)

In order to address the challenges of employing neural network approach for 3D finger knuckle recognition, we develop a new neural network from a popular network architecture, *ResNet* [16], which has been shown to offer excellent performance on a wide range of applications including object recognition and instance segmentation. The problem of lacking training samples can be mitigated from the transfer learning of generic image features from bottom layers of *ResNet*, which has been well trained on large object recognition datasets. Using such weights also enable a promising initialization of the training process. However, if *ResNet* is directly applied to the 3D finger knuckle recognition problem, its performance is poor due to the challenges described in Section I, while comparative experimental results will also be shown in Section IV. Therefore, we attempted to investigate the useful component of *ResNet* and customize the networks for our application.

Similar to the original paper [16], a building block of the residual structure is defined as

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}, b) + \mathbf{x} \quad (1)$$

where x and y is the input and output of the layer respectively,

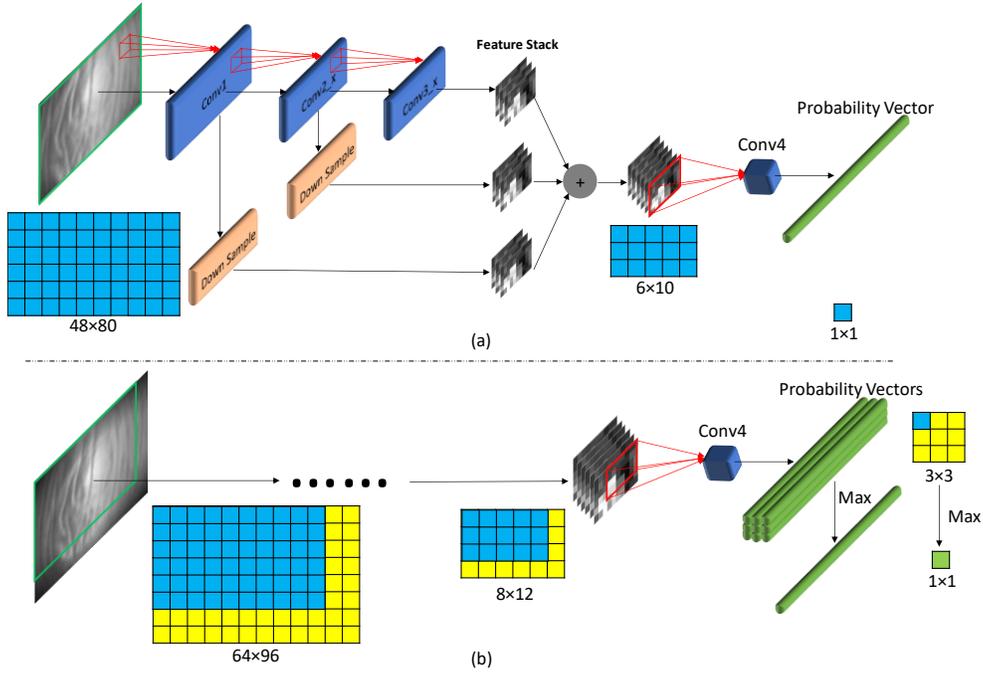


Fig. 3. Simplified network architectures for *FKNet*: (a) Network Training; (b) Identification using Trained Network

w is the weights, b is the bias and f is a generic function. With our investigation, it is observed that an average pooling layer is located before the last classification layer. This layer is expected to serve the purpose of integrating high level spatial information to be an abstract feature vector. However, the spatial relationship of finger knuckle structural patterns can be helpful for the identification and therefore the operation of average pooling may cause the loss of such spatial relationship information and may not be necessary.

Besides, a branch of deep learning methods incorporates convolution layers in a series manner and make use of the last layer to proceed for recognition tasks because the information in the last layer is expected to contain high level and abstract information from previous layers. However, it is also possible that fine feature information from bottom layers has not been incorporated to the upper coarse feature maps. Moreover, a series of convolution layers usually contains down sampling operations implemented by pooling layers. Discriminative finger knuckle features, e.g. irregular ridges and valleys patterns with different size and scales, may not be fully utilized. Therefore, we also attempted to utilize that information from different layers corresponding to multi-scales feature maps. Collaborating that feature level information provides a more complete utilization of spatial information with different scales and such collaboration can offer significant performance improvement. The respective experiments were performed to show such improvement and those results are presented in Section IV.B.

Besides, the number of convolution layers required depends on the complexity of the problem being addressed, while excessive layers may induce more serious overfitting issues. Unlike object recognition problems requires the learning of high-level abstract features, finger knuckle recognition requires the learning of the invariant structural patterns. Existing

TABLE III: LAYER CONFIGURATIONS FOR THE CONTACTLESS 3D FINGER KNUCKLE DATASETS DURING TRAINING

Layer	Output Size	Kernel Size
<i>Conv1</i>	$24 \times 40 \times 64$	$7 \times 7, 64$, stride 2
<i>Conv2_x</i>	$12 \times 20 \times 256$	3×3 max pool, stride 2
<i>Conv3_x</i>	$6 \times 10 \times 512$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
<i>Conv4</i>	$1 \times 1 \times 190$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$ $6 \times 10, 190$, <i>softmax</i>

successful finger knuckle recognition methods utilizing hand-crafted features of low complexity, which may indicate that the discriminative information embedded in finger knuckles do not have a high degree of complexity. Therefore, we attempted to simplify the network by reducing the number of top convolution layers. We found that using the convolution layers of *ResNet-50* up to *Conv3_x* is adequate for the contactless 3D finger knuckle recognition and such experimental results of this ablation study are presented in Section IV.B.

With the above theoretical arguments, we defined the *FKNet* as follows:

$$\mathbf{p} = f_4 [f_3 (f_2 (f_1(I))) + g_2 (f_2 (f_1(I))) + g_2 (f_1(I))] \quad (2)$$

where I is the input image, f_1, f_2, f_3, f_4 are the functions corresponding to the convolution layers, g_1, g_2 are the functions corresponding to the down sampling layers and \mathbf{p} is the output of the network with a dimension of $1 \times N$ and N is the number of classes. When comparing with *ResNet-50*, our network enables less parameters and thus higher efficiency. Lastly, the output \mathbf{p} further processed with a *softmax* function:

$$p'_i = \frac{e^{p_i}}{\sum_{i=1}^N e^{p_i}}, \quad i \in [1, N] \quad (3)$$

The output vector \mathbf{p}' represents the probabilities of belonging

to a class. During training, the cross-entropy loss is adopted:

$$L = -\sum_{i=1}^N t_i \log(p'_i) \quad , \quad i \in [1, N] \quad (4)$$

where \mathbf{t} is the ground truth label of the training samples. During testing, the probability vector \mathbf{p}' is considered as the similarity scores for a test image. When comparing to the hand crated approach, this network style enables quite efficient identification tasks because a probability vector for all classes can be generated at once.

In summary, we develop a customized *FKNet* for more accurate comparison of 3D finger knuckle images. Figure 3 shows the simplified network architecture of the proposed *FKNet* while Table III shows the layer configurations with an input image of size 48×80 and a total number of 190 classes. Down sampling is performed by *Conv3_1* with a stride of 2. The configurations and building blocks of *Conv1*, *Conv2_x* and *Conv3_x* also adopt a residual structure as in *ResNet*.

C. Finger Knuckle Alignment and Comparison

From the knowledge of comparing finger knuckle images using hand-crafted features, the alignment of feature templates by translations is a crucial step to ascertain the recognition performance. Although CNN based approach has some capability of tolerating translational shifting because the upper layers can aggregate information from lower layers with pooling layers, such approach only accommodates a limited degree of translational shifting. Since the area of interest of finger knuckle is loosely defined, a sharp boundary does not exist along the finger knuckle regions. Despite the accurate finger knuckle segmentation by *Mask R-CNN*, those segmented images from the same subject can still contain quite a large degree of translational variation of the finger knuckle patterns, which is challenging to CNN based approach.

Therefore, we attempted to introduce a new training and testing strategy incorporating an alignment scheme with neural network approach. During the training stage, we provide the rotational shifted versions of the sample images as the only data augmentation. Those images are also cropped with a fixed smaller window located at the center of each image, which enables translational shifting operations in the testing stage. During the testing stage, we can present multiple translationally shifted versions of testing samples to the trained networks. This simple technique is denoted as the *simple alignment scheme*. For each presented test image, a vector representing the probability of the test image belonging to each class will be the output. While presenting the images of shifted versions to the classifier, a matrix containing the probability vectors will be the output of the network. We obtain the entire probability vector containing the maximum probability value among the matrix as the final output. Those values of the final probability vector are considered as the comparison scores between the test image and the enrolled classes.

The major limitation of the *simple alignment scheme* lies in its bulky and inefficient computation of inputting many shifted versions of images. However, this issue can be addressed using a convolutional implementation of the sliding window approach, which was first suggested in the object detection

literature [39]. The inputting of shifted versions of images can

Algorithm 1:	Comparing Contactless 3D Finger Knuckle Test Images using <i>FKNet</i>
Input:	I : unknown image with dimension $m \times n$ (i.e. 64×96); s : down sampling factor of the network (i.e. 8); d : depth of the feature stack (i.e. 832); u, v : kernel size of <i>Conv4</i> (i.e. 6×10) N : number of classes;
Output:	\mathbf{p}' : final comparison scores
1:	Convolve I with the trained network to generate a feature stack of dimension $\frac{m}{s} \times \frac{n}{s} \times d$;
2:	Compute the probability vectors of dimension $\left(\frac{m}{s} - u + 1\right) \times \left(\frac{n}{s} - v + 1\right) \times N$;
3:	Compute the maximum probability among all vectors;
4:	Compute the final comparison scores \mathbf{p}' using the probability vector containing the maximum probability.

be considered as a typical sliding window approach, which can also be implemented using a fully convolutional approach. Figure 3 shows a schematic diagram of this alignment scheme. During the training stage, the dimension of training images is 48×80 and those images are cropped from the images with original sizes of 64×96 . The dimension of the input feature map before *Conv4* is 6×10 , which is reduced by 8 times. The *Conv4* layer, together with *softmax* function, generates a probability vector of size $1 \times 1 \times 190$. During the testing stage, instead of presenting multiple translationally shifted versions of the test samples to the trained network, we can present the testing images with the original size (64×96) directly to the networks because our networks are fully convolutional. Such configuration results in a size of 8×12 for the input feature map before *Conv4*. In this case, the *Conv4* layer, together with *softmax* function, generates 9 probability vectors of size $3 \times 3 \times 190$. Similar to the *simple alignment scheme*, we obtain the entire probability vector of 190 classes containing the maximum probability value among the matrix as the final output. Furthermore, in order to precisely implement the sliding window approach containing all possible translationally shifted versions, we also present a combination of eight horizontally shifted and eight vertically shifted versions of the test samples to the trained network because the network will reduce the size of the input images by eight times. Again, we obtain the entire probability vector of 190 classes containing the maximum probability value among the matrix as the final output. Those values of the final probability vector are considered as the final comparison scores between the test image and the enrolled classes. Algorithm 1 summarizes the procedures to match the test images using *FKNet*.

IV. EXPERIMENTS AND RESULTS

In order to examine the performance of the proposed 3D finger knuckle recognition framework, we present comparative experimental results using a publicly available 3D finger knuckle database. In addition, the effectiveness of our approach is also validated using two others larger 2D finger knuckle

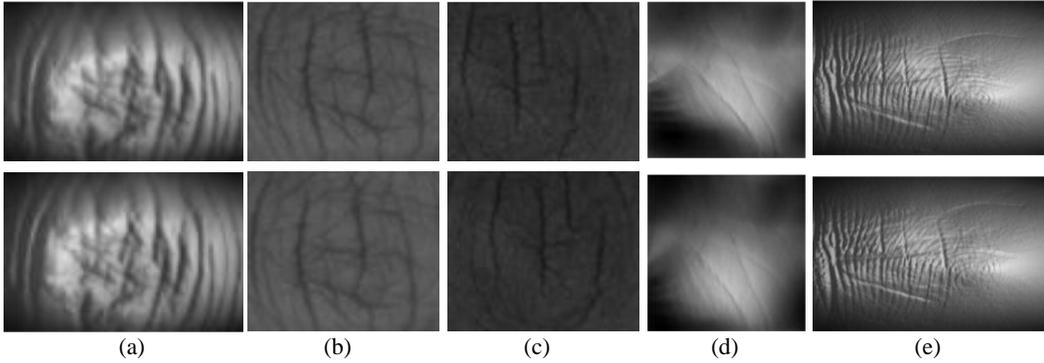


Fig. 4. Sample images from five employed databases: (a) HKPolyU 3D Finger Knuckle Images Database; (b) HKPolyU Contactless Finger Knuckle Images Database; (c) HKPolyU Contactless Hand Dorsal Images Database; (d) HKPolyU Contact-free 3D/2D Hand Images Database; (e) HKPolyU 3D Fingerprint Images Database.

database, a 3D palmprint database and a 3D fingerprint database. Figure 4 shows some sample images of these datasets. We present the experimental results using receiver operating characteristics (ROC) curves for verification scenarios and cumulative match characteristics (CMC) curve for identification scenarios. State-of-the-art methods are selected as baselines for the performance comparisons.

A. Contactless 3D Finger Knuckle Database

Our database has been acquired using a photometric stereo approach for the 3D imaging. The biometric imaging device consists of a camera, seven evenly distributed illuminations, a control circuit, and a computer as in [4]. The acquired finger knuckle images are preprocessed using the methods described in Section III.A.

The HKPolyU 3D Finger Knuckle Images Database [21] is a recently available benchmark two-session database with 2D and 3D finger knuckle images. It contains 1410 forefinger images and 1410 middle finger images from 130 subjects, among them 105 subjects contain two-session images. Since the database is quite small, we attempted to expand the database by acquiring more images with 1098 forefinger images and 1098 middle finger images from 98 subjects. The new dataset contains 2508 forefinger images and 2508 middle finger images from 228 subjects, among them 190 subjects contain two-session images. Six forefinger images and six middle finger images are available for each subject per session.

The use of different lenses can enable the imaging under a less constrained environment and provides larger image variations similar to those acquired in real deployment scenarios. For examples during the border-crossing inspections, it is not uncommon to observe imaging systems at different checkpoints from the different vendors or a part of imaging devices can be different from the remaining due to system maintenances and upgrades. In more challenging scenarios like forensics, images of suspects are to be compared with those already stored in the government databases, where the imaging devices are likely to be different. The images acquired in our case using two lenses, therefore attempts to present similar challenges: one lens enables a smaller field of view while another lens enables a broader field of view.

Among the 190 subjects, the first session images of 96 subjects and 94 subjects are acquired from lens-A and lens-B

respectively, while the second session images of 7 subjects and 183 subjects are acquired from lens-A and lens-B respectively. This imaging scenario enables that images from 89 subjects are acquired from different lens, which contain large illumination and deformation variations. Such variations create huge challenges for accurately extracting the irregular ridge and valley features for finger knuckle recognition. Moreover, when adopting a standard two-session biometric protocol, i.e. comparing second session images to the first session images, many imposter images are acquired from the same lens while genuine images are acquired from different lens. This configuration creates special challenges to deep learning methods because the statistical distribution of training and testing data are very different. Therefore, the new database includes scenarios that considers possible situations in the real world and is challenging for validating the performance of the proposed 3D finger knuckle recognition framework.

B. Evaluation with 3D Finger Knuckle Recognition

Ablation Studies

To begin with, we present the experimental results of ablation studies to justify our design of *FKNet* using a subset of the 3D finger knuckle database. We utilized 94 subjects of which both sessions' images are acquired using lens-B. By excluding the cross-lens challenges, we can fairly compare the performance of our network with *ResNet-50*. Mask R-CNN are used for segmenting the finger knuckle images for our experimentation in this sub-dataset. We adopt a standard two-session evaluation protocol, which consider the first session images as the training set and the second session images as the testing set. This evaluation protocol generates 52452 ($94 \times 93 \times 6$) imposter comparison scores and 564 (94×6) genuine comparison scores.

The first experiment aims at comparing the recognition performance between with and without aggregating the spatial relationship information using an average pooling layer before the last classification layer and the results are shown in Figure 5(a). It can be observed that retaining the feature of size 6×10 (denoted as ResNet without avg. pool) produces a much better performance than averaging the spatial relationship information to be size 1×1 (denoted as ResNet (Original)). These finding verify our theoretical arguments presented in Section III.B.

The second ablation experiment aims at verifying the

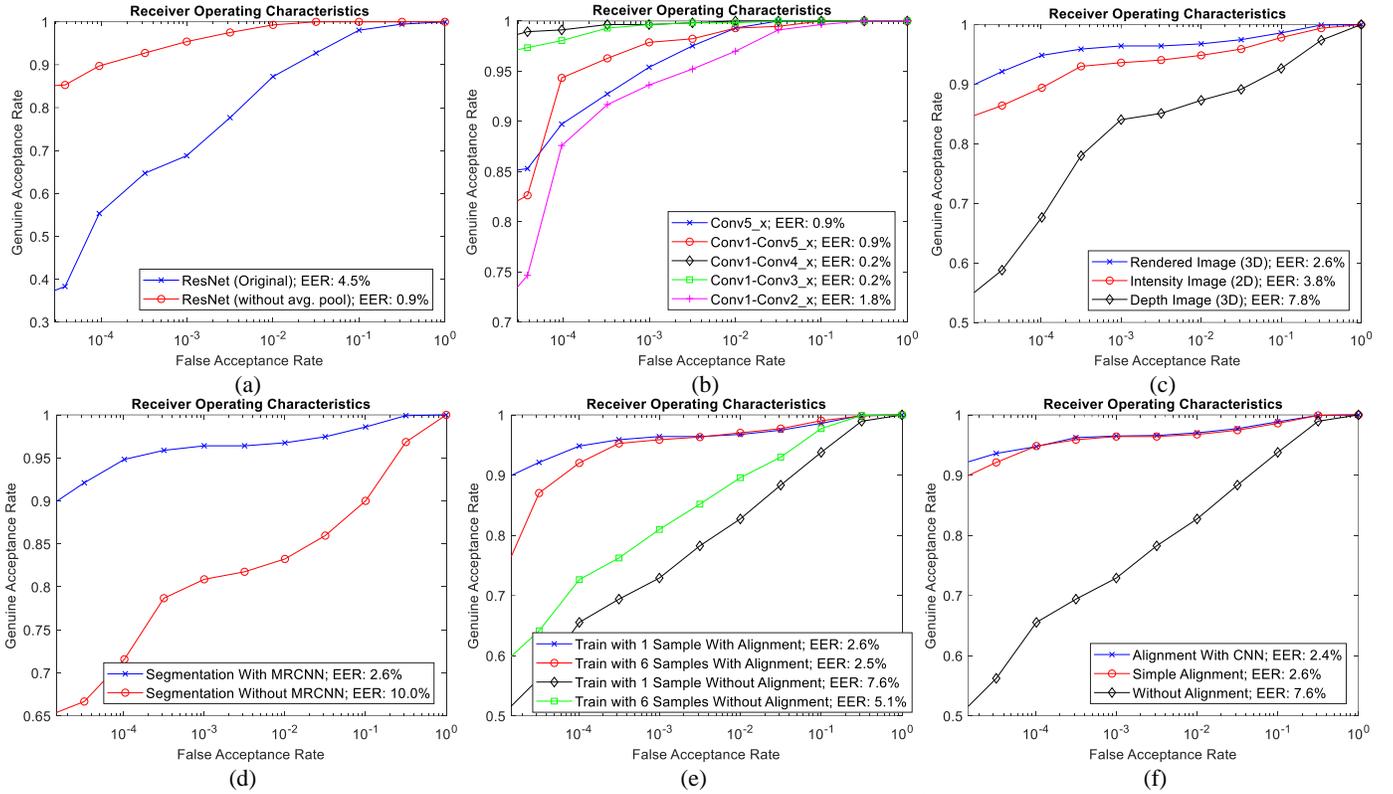


Fig. 5. Experimental Results of Ablation Studies in the Contactless 3D Finger Knuckle Datasets for Comparing the effects of: (a) the last pooling layer; (b) collaborating multi-scales information; (c) using different images; (d) improved segmentation using Mask R-CNN; (e) using one/six samples for the training (f) alignment.

usefulness of collaborating the multi-scales feature information from the previous layers and the results are shown in Figure 5(b). We first compare the recognition performance between using only the information from the last convolution layer without the collaboration (denoted as Conv5_x) and also using the multi-scales information from previous layers (denoted as Conv1-Conv5_x). Our experimental results indicate that using such collaboration enables outperforming recognition accuracy than the original approach without the collaboration, which verify our theoretical arguments described in Section III.B. After that, we further compare the recognition performance with simplifying the network by reducing the number of convolution layers, i.e. using the information from the first convolution layer up to the fourth (denoted as Conv1-Conv4_x), the third (denoted as Conv1-Conv3_x) and the second (denoted as Conv1-Conv2_x) convolution layer blocks respectively. The experimental results indicate that both Conv1-Conv4_x and Conv1-Conv3_x can produce superior performance while Conv1-Conv2_x significantly degrades the performance. Since Conv1-Conv3_x can produce comparable results with Conv1-Conv4_x, we suggest using Conv1-Conv3_x due to its simplicity.

Our next ablation studies attempt to justify our final configurations which will be adopted in the comparative experiments with the state-of-art-methods. For these studies, we performed verification experiments using all 190 subjects with two-sessions images, where first session images are used for the training and second session images are used for the testing. This evaluation protocol generates 215460 ($190 \times 189 \times 6$) imposter

comparison scores and 1140 (190×6) genuine comparison scores. When training our network, we rotate the images from minus ten to ten degrees with a step of one degree for the data augmentation. This data augmentation technique allows more adaptive ability for the trained network. We initialize our network with the weights provided along with *ResNet-50* [16].

Figure 5(c) compare the effect of using various sources of information for inputting to our network. We use the same configurations for other steps including segmentation and alignment and compare the performance between using the rendering approached described in Section III.A, using 3D depth images and using 2D intensity images. Although using 3D images are expected to outperform using 2D images due to its illumination invariant property, however when the normalized depth images are used, the performance is quite poor probably because the contrast of such images are low and therefore those finger knuckle patterns are blurred. Instead, when we render the illumination invariant images from 3D images using a fixed illumination, the performance is better than only using 2D images. The experimental results presented in Figure 5(c) justify the need for rendering the illumination invariant images.

Figure 5(d) compare the performance improvement by incorporating *Mask R-CNN* for more accurately segmenting finger knuckle images. We use the same configurations for other steps including alignment and rendered images are used. For training the *Mask R-CNN*, we employed the one-session images from the remaining 38 subjects to ensure no overlapping for the training and testing sets for the *Mask R-CNN*. The

experimental results presented in Figure 5(d) show that improving the segmentation using *Mask R-CNN* enable significant performance improvement.

Figure 5(e) compare the performance between using only one raw image versus six raw images for training the network. We use the same configurations for other steps including segmentation and rendered images are used. It is a common believe that more training images is always preferable. In fact, this argument is valid for a general approach when the alignment technique is not incorporated and can be observed from comparing the last two curves (black and green). This finding concurs with other CNN based approach where more training images enable the network with better capabilities. However, when we attempt to address the translational variation problem with the proposed alignment technique, such more training images may not be necessary. In fact, when we use more training images for each subject, those images contain translational variation may help the network to enhance tolerance to genuine samples but may also induces higher tolerance for imposter samples. The experimental results presented in Figure 5(e), first two curves (blue and red), show that our proposed network architecture along with the alignment technique can perform well in the challenging scenario that only one raw image per subject is used for the training.

Figure 5(f) compare the performance improvement by the proposed alignment technique for enhancing the finger knuckle verification. We use the same configurations for other steps including segmentation and rendered images are used. We argue that the pooling layers are expected to be the key elements from the CNN based approach for accommodating the translational variations. It can only accommodate a small translational shifting within the pooling kernel. However,

unlike other recognition tasks such as iris recognition, the area of finger knuckle pattern does not have a clear edge or boundaries, so that the segmented images would still contains large translational variations. When the variation exceeds the limitations of the network (e.g. kernel of pooling layers), the recognition performance is expected to be degraded. With our suggested alignment technique, this issue can be well addressed. The experimental results presented in Figure 5(f) show that our alignment technique is very effective for improving the recognition performance. While Alignment using the fully convolutional approach described in Section III.C. enables further improved results. Same trends can also be observed from the experimental results shown in Figure 5(e).

In summary, the experimental results of ablation studies have justified the theoretical arguments presented in Section III. Therefore, in the following comparisons with state-of-the-art methods, we employed the *Mask R-CNN* for segmenting finger knuckle images, and only one raw image per subject is used for the training, for all the evaluated methods.

Comparison with State-of-the-Art Methods

We further compare the performance of our proposed method with the state-of-the-art methods in the challenging scenario where only one raw image per subject is used for the training, while the images are segmented using *Mask R-CNN*. This evaluation protocol generates 215460 ($190 \times 189 \times 6$) imposter comparison scores and 1140 (190×6) genuine comparison scores. The *Surface Gradient Derivatives (SGD)* [21] is currently the state-of-the-art method for 3D finger knuckle recognition. This method is also reported to outperform other state-of-the-art 3D finger knuckle and palmprint recognition methods including *Surface Code* [32] and *Binary Shape* [33]. Since this method essentially extract the discriminative features from 3D surface normal images, those images are used for the

TABLE IV: COMPARATIVE PERFORMANCE IN HKPOLYU 3D FINGER KNUCKLE IMAGES DATABASE FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-3}$	$= 10^{-4}$	$= 10^{-5}$
Ours	96.5%	94.7%	91.4%
<i>SGD (TPAMI20)</i>	85.4%	77.5%	68.7%
<i>CR_L1_DALM (TPAMI15)</i>	46.3%	37.9%	32.8%
<i>CR_L2 (TPAMI15)</i>	44.8%	36.1%	29.2%
<i>ResNet (CVPR16)</i>	28.7%	13.6%	5.5%
<i>VGG (ICLR15)</i>	22.5%	9.4%	2.7%

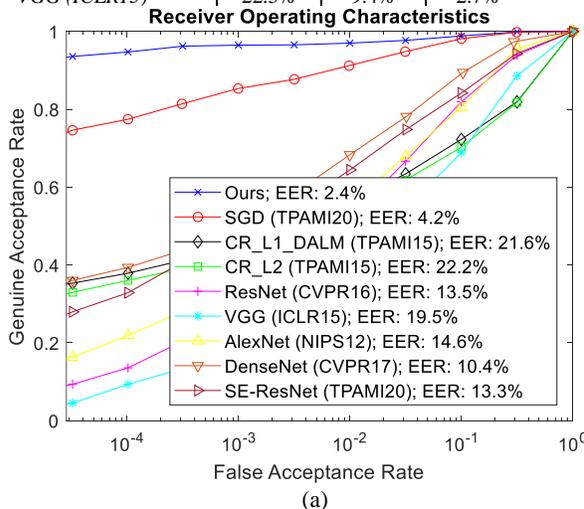


TABLE V: COMPARATIVE COMPUTATIONAL TIME (IN MILLISECONDS)

Method	Computational Time (in ms)
<i>CR_L2 (TPAMI15)</i>	6.1
<i>SGD (TPAMI20)</i>	24.0
Ours	36.6
<i>AlexNet (NIPS12)</i>	36.6
<i>CR_L1_DALM (TPAMI15)</i>	89.4
<i>DenseNet (CVPR17)</i>	89.9

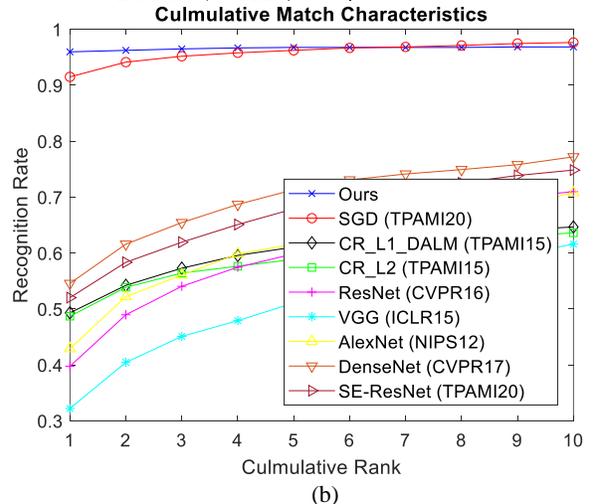


Fig. 6. Comparative Experimental Results in Contactless 3D Finger Knuckle Datasets: (a) ROC; (b) CMC.

evaluation.

Beside the most competent method for 3D finger knuckle recognition, we also attempted to compare the performance with another state-of-the-art 3D palmprint recognition method, using collaborative representation based framework with L1-norm (*CR_L1_DALM*) [38] and L2-norm regularizations (*CR_L2*) [38], which has reported superior performance. In order to fairly compare with this method, we first investigate the variations between their reported database and our database. The reported database contain square size images with 128 pixels, it is reasonable to resize and crop our images, without changing the aspect ratio, to be the same size as the images in the reported database, and employ the same parameters provided along with *CR_L1_DALM* and *CR_L2*, which is already optimized for their reported database. Since this method essentially extract the discriminative features from 3D depth images, those images are used for the evaluation.

Furthermore, we also compare the performance with various popular deep learning methods, including *ResNet-50* [16], *VGG-16* [17], *AlexNet* [28], *DenseNet-121* [46] and *SE-ResNet-50* [47] to verify our arguments on the challenges of employing neural network approach for 3D finger knuckle recognition. While training these networks, we modify the last classification layer to an output of 190 classes to fit the experimental condition. Except this layer, we initialized the weights from the respective trained model available along with the respective methods for *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. The inputs to all the networks are the same, i.e. rendered images, to ensure the fairness in comparisons.

Figure 6 presents such comparative experimental results using the baseline methods selected in this paper and Table IV highlights the comparative performance for high security conditions. It can be observed that our proposed method significantly outperforms all the methods. However, unlike 3D palmprint recognition, *CR_L1_DALM* and *CR_L2*, does not produce comparable performance, probably due to the challenging nature of our database that, some images are acquired using different camera lens so that the statistical distribution between the gallery and probe samples are significantly different. It is also worth mentioning that CNN based methods suffer heavily from this experimental condition, probably due to the following reasons.

The major reason can be attributed to the cross-lens situations which is quite challenging for the CNN based approach because the images acquired from lens-A and lens-B are indeed very different, which results is large variations between the statistical distribution of training and testing samples. The translational

variations in finger knuckle images are also challenging because of the limited capability of CNN based approach for accommodating large translational variations, which is expected to rely on the pooling layers. Another reason can be attributed to the challenging scenario while only one raw image sample are presented. Despite these challenging conditions, our proposed approach addresses these issues by learning discriminative information from one raw image and match the test images using the alignment technique. These reproducible experimental results [34] validate the effectiveness of our proposed method for 3D finger knuckle recognition.

Evaluation of Computational Time

We also comparatively evaluated the computational time for our proposed approach with the best performing hand-crafted approach, *SGD* [21] and other learning based methods including *CR_L1_DALM*, *CR_L2*, *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. We evaluate the computational time required for generating a score vector of 190 classes when a preprocessed image of dimension (64×96) is presented to respective approaches, while image alignment is not considered for all the methods in this comparison. The experiments were performed on a machine with Intel Core i7-6700 (3.40GHz) using MATLAB 2015b, CentOS 7. All methods were executed using CPU only to ensure fairness in the comparisons. Table V presents the computational time per sample for each of the considered approach.

Among the compared methods, *CR_L2* is the fastest method with its low computational complexity, however, this method does not deliver promising accurate performance in the challenging dataset evaluated in this paper. Besides, *SGD* is the second fastest method due to its simple feature extraction and comparison approach while also achieving accurate recognition performance, which shows that it is a promising baseline using hand-crafted approach with both accuracy and efficiency. Our approach ranked the third together with *AlexNet*, due to the simpler network architectures when comparing to *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. The comparative computational time presented in this section indicates that although our method does not outperform all methods with efficiency, but the computational time required is quite comparable with the best methods. Meanwhile, it is expected that a slight increase of computational time is required if the alignment scheme using the fully convolutional approach

TABLE VI: COMPARATIVE PERFORMANCE HKPOLYU CONTACTLESS FINGER KNUCKLE IMAGES DATABASE FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-3}$	$= 10^{-4}$	$= 10^{-5}$
Ours	87.9%	86.6%	81.3%
<i>DoN (TPAMI16)</i>	83.2%	75.8%	47.3%
<i>ResNet (CVPR16)</i>	85.4%	75.9%	55.9%
<i>VGG (ICLR15)</i>	81.3%	67.9%	47.5%
<i>AlexNet (NIPS12)</i>	85.5%	71.6%	60.0%
<i>DenseNet (CVPR17)</i>	78.4%	64.4%	46.9%
<i>SE-ResNet (TPAMI20)</i>	70.9%	55.2%	36.3%

TABLE VII: COMPARATIVE PERFORMANCE IN HKPOLYU CONTACTLESS HAND DORSAL IMAGES DATABASE FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-3}$	$= 10^{-4}$	$= 10^{-5}$
Ours	78.2%	74.5%	66.8%
<i>DoN (TPAMI16)</i>	62.8%	46.7%	18.4%
<i>ResNet (CVPR16)</i>	63.6%	42.9%	21.4%
<i>VGG (ICLR15)</i>	56.7%	34.8%	12.6%
<i>AlexNet (NIPS12)</i>	66.0%	47.1%	28.8%
<i>DenseNet (CVPR17)</i>	39.0%	20.7%	8.2%
<i>SE-ResNet (TPAMI20)</i>	37.3%	21.0%	8.9%

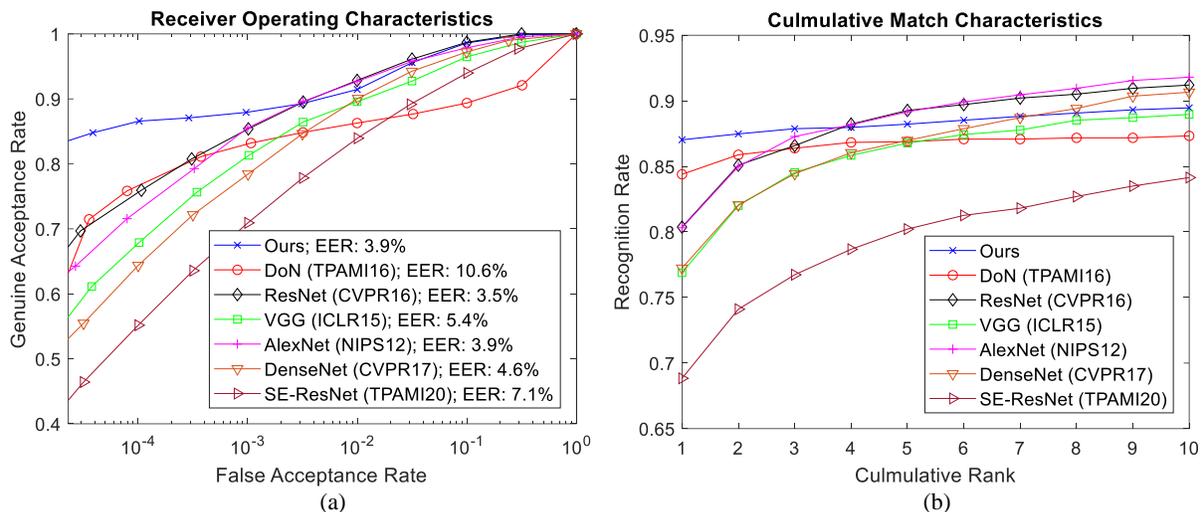


Fig. 7. Comparative Experimental Results in HKPolyU Contactless Finger Knuckle Images Database: (a) ROC; (b) CMC.

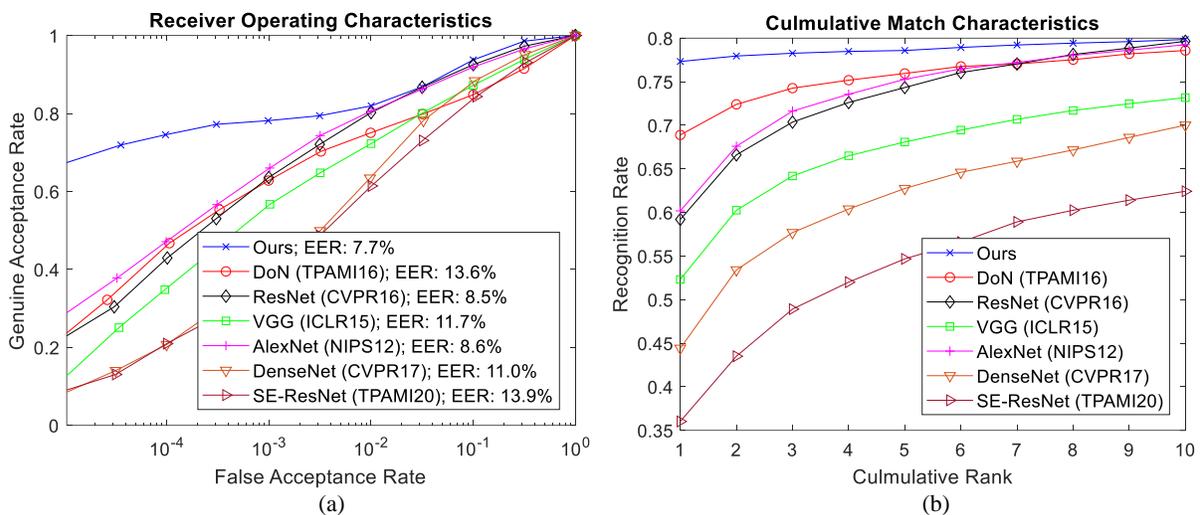


Fig. 8. Comparative Experimental Results in Contactless Hand Dorsal Images Database: (a) ROC; (b) CMC.

is enabled. However, the effectiveness of using such alignment scheme can be justified by its remarkable recognition performances.

C. Evaluation with 2D Finger Knuckle Recognition

Despite our framework is developed for 3D finger knuckle recognition, it is expected to be applicable for 2D finger knuckle recognition. Therefore, we also provide supportive experimental results using a large 2D finger knuckle images database for further validating the effectiveness of our proposed method. The HKPolyU Contactless Finger Knuckle Images Database (Version 1.0) [35] contains 2515 images from 503 subjects, each with five images. This database provides loosely segmented finger knuckle images with low imaging qualities. We also adopt a challenging scenario using the major finger knuckle images where only the first image of each subject is used for the training and the remaining four images of each subject are used for the performance evaluation. This evaluation protocol generates 1,010,024 ($503 \times 502 \times 4$) imposter comparison scores and 2012 (503×4) genuine comparison scores. We compare our method with a state-of-the-art 2D palmprint recognition method, *DoN* [27], which is reported in

also offer state-of-the-art finger knuckle recognition method in [21], and other popular CNN based methods including *ResNet-50* [16], *VGG-16* [17], *AlexNet* [28], *DenseNet-121* [46] and *SE-ResNet-50* [47]. While training these networks, we modify the last classification layer to an output of 503 classes to fit the experimental condition. Except this layer, we initialized the weights from the respective trained model available along with respective methods for *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. Since the *SGD* [21], *CR_L1_DALM* [38] and *CR_L2* [38] require 3D images, we are unable to compare with this method in this database.

Figure 7 presents the comparative experimental results with the baseline methods paper and Table VI highlights the comparative performance for high security conditions. It can be observed that our method outperforms other methods especially in straight security conditions where the false acceptance rates must be very low. It is worth mentioning that unlike the experimental results presented in Section IV.B, CNN based methods achieved comparable performance with the state-of-the-art method using hand-crafted features, probably because the challenges of cross-lens situations do not exist. Since there some segmented finger knuckle images which are completely

out of the finger knuckle region, the performance improvement of our method is less obvious when comparing with the datasets described in Section IV.A.

Besides, we also provide supportive experimental results using a very large 2D finger knuckle images database. The HKPolyU Contactless Hand Dorsal Images Database [36] contains 3560 images from 712 subjects, each with five images. This database also provides loosely segmented finger knuckle images with low imaging qualities. Again, we also adopt a challenging scenario using the major index finger knuckle images where only the first image of each subject is used for the

training and the remaining four images of each subject are used for the performance evaluation. This evaluation protocol generates 2,024,928 ($712 \times 711 \times 4$) imposter comparison scores and 2848 (712×4) genuine comparison scores. While training the networks, we modify the last classification layer to an output of 712 classes to fit the experimental condition.

Figure 8 presents the comparative experimental results with the baseline methods paper and Table VII highlights the comparative performance for high security conditions. It can be observed that our method significantly outperforms all other methods in both verification and identification scenarios, which

TABLE VIII: COMPARATIVE PERFORMANCE IN HKPOLYU CONTACT-FREE 3D/2D HAND IMAGES DATABASE FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-3}$	$= 10^{-4}$	$= 10^{-5}$
Ours	98.4%	98.0%	96.5%
SGD (TPAMI20)	96.8%	95.6%	94.8%
CR_L1_DALM (TPAMI15)	97.6%	96.7%	95.9%
CR_L2 (TPAMI15)	96.7%	95.9%	94.1%
ResNet (CVPR16)	82.3%	68.0%	53.2%
VGG (ICLR15)	74.0%	59.0%	42.8%
AlexNet (NIPS12)	86.2%	75.3%	59.2%
DenseNet (CVPR17)	78.2%	67.9%	52.1%
SE-ResNet (TPAMI20)	53.6%	39.4%	30.3%

TABLE IX: COMPARATIVE PERFORMANCE IN HKPOLYU 3D FINGERPRINT IMAGES DATABASE FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-2}$	$= 10^{-3}$	$= 10^{-4}$
Ours	97.9%	95.8%	94.2%
SGD (TPAMI20)	97.5%	94.1%	92.5%
CR_L1_DALM (TPAMI15)	92.5%	86.3%	83.3%
CR_L2 (TPAMI15)	87.5%	83.8%	75.8%
ResNet (CVPR16)	99.6%	98.3%	94.2%
VGG (ICLR15)	99.2%	96.3%	92.9%
AlexNet (NIPS12)	99.2%	98.8%	95.8%
DenseNet (CVPR17)	99.2%	92.1%	77.1%
SE-ResNet (TPAMI20)	93.8%	82.5%	68.8%

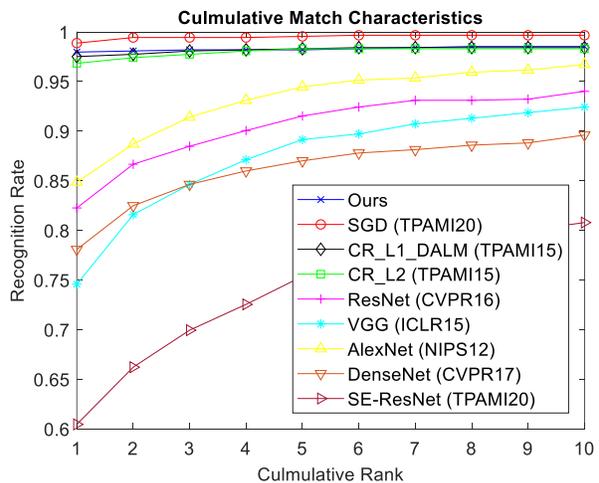
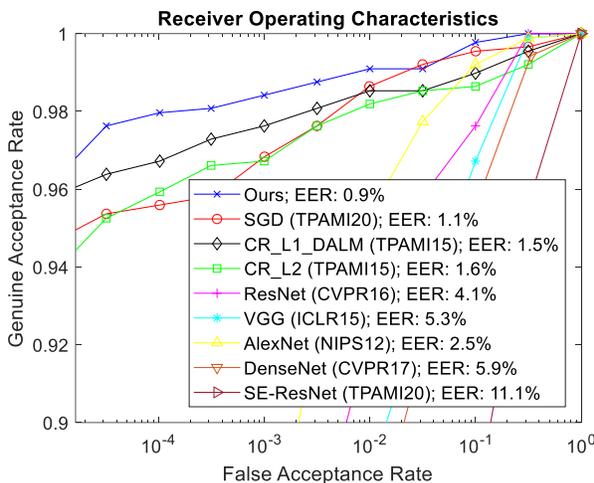


Fig. 9. Comparative Experimental Results in the HKPolyU Contact-free 3D/2D Hand Images Database: (a) ROC; (b) CMC.

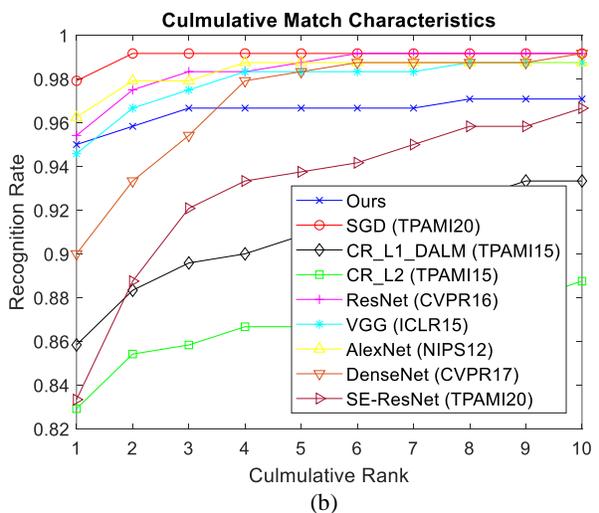
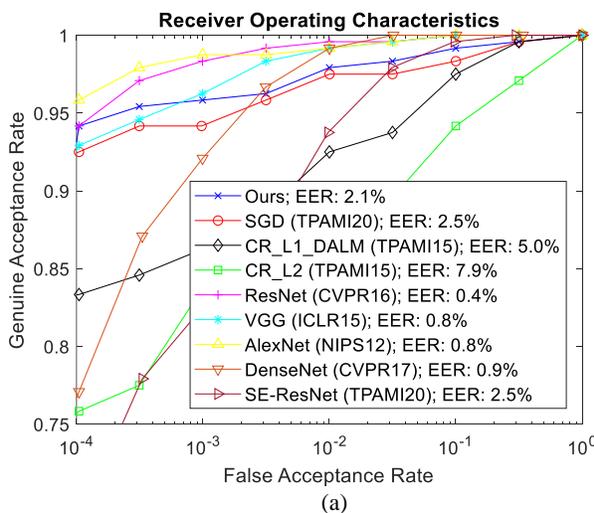


Fig. 10. Comparative Experimental Results in the HKPolyU 3D Fingerprint Images Database: (a) ROC; (b) CMC.

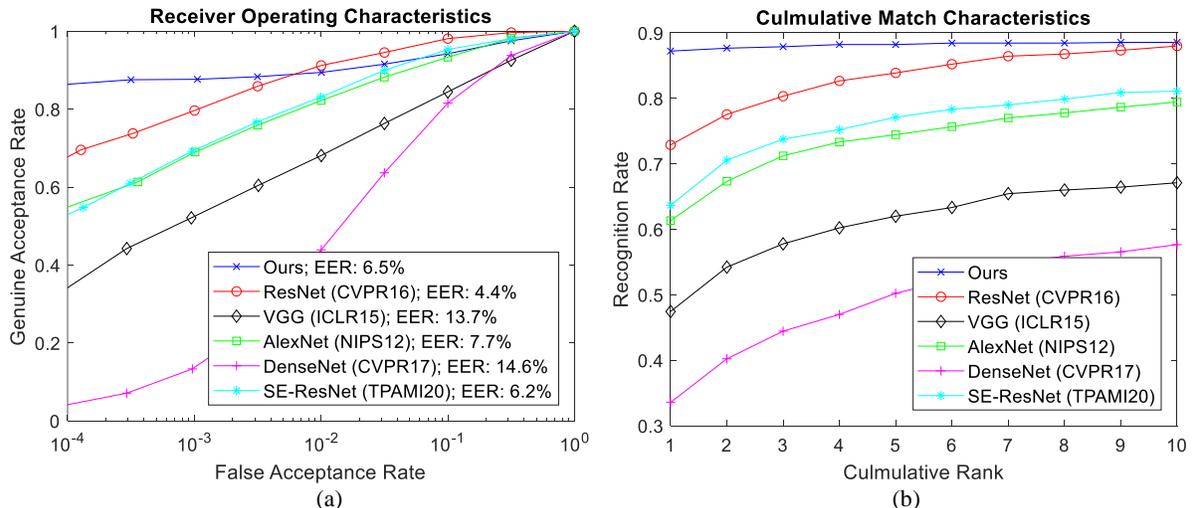


Fig. 11. Comparative Experimental Results in the 3D Fingerprint Database [49]: (a) ROC; (b) CMC.

has validated the effectiveness of our proposed method.

D. Evaluation with 3D Palmprint Recognition

Although the scope of this paper is on 3D finger knuckle recognition, we also attempt to evaluate the performance of our approach using other 3D biometrics datasets for investigating the generalizability of the proposed approach. Similar to [21], we employed the same 3D palmprint dataset, the HKPolyU Contact-free 3D/2D Hand Images Database Version 1.0 [32], containing two-sessions images from 177 subjects (each with five images per session), for the performance evaluation. In order to obtain consistent experimental results as reported in earlier references, we adopt the same evaluation protocol using all first session images for the training and the second session images for the testing, which generates 885 (177×5) genuine and 155760 ($177 \times 176 \times 5$) imposter comparison scores.

Similar to Section IV.B, 3D surface normal images are used for *SGD* [21], 3D depth images are used for *CR_L1_DALM* and *CR_L2* [38], while the artificially rendered images are used for our method and other neural network based methods including *ResNet-50* [16], *VGG-16* [17], *AlexNet* [28], *DenseNet-121* [46] and *SE-ResNet-50* [47]. We also modify the last classification layer to an output of 177 classes and initialized the weights from the respective trained model. Except this layer, we initialized the weights from the respective trained model available along with respective methods for *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. Since rotational alignment was not considered in this dataset by neither the recent baseline method *SGD* nor another former baseline method *Binary Shape* [33], we disabled our data augmentation with rotations to ensure fairness in the comparisons.

Figure 9 presents such comparative experimental results using the baseline methods and Table VIII highlights the comparative performance for high security conditions. Our approach significantly outperforms other methods for the verification scenario. For the identification scenario, *SGD* is still the best performing method which slightly outperforms ours. However, it can be observed that our approach

TABLE X: COMPARATIVE PERFORMANCE IN 3D FINGERPRINT DATABASE [49] FOR HIGH SECURITY CONDITIONS

GAR at FAR	$= 10^{-2}$	$= 10^{-3}$	$= 10^{-4}$
Ours	89.4%	87.7%	86.2%
<i>ResNet (CVPR16)</i>	91.2%	79.7%	69.6%
<i>VGG (ICLR15)</i>	68.1%	52.1%	30.9%
<i>AlexNet (NIPS12)</i>	82.2%	69.0%	53.4%
<i>DenseNet (CVPR17)</i>	43.9%	13.3%	3.7%
<i>SE-ResNet (TPAMI20)</i>	83.1%	69.2%	54.8%

outperforms other popular neural network based methods.

E. Evaluation with 3D Fingerprint Recognition

Besides 3D palmprint recognition, we also evaluate the performance of our proposed approach for 3D fingerprint recognition. Similar to [21], we also employed the 3D fingerprint dataset, the HKPolyU 3D Fingerprint Images Database [4], containing one-session images from 240 subjects (each with six images), for the performance evaluation. The original experimental protocol employed an all-to-all evaluation protocol, which is not applicable for CNN based methods requiring the splitting of data into training and testing sets. Therefore, we select a relatively less challenging protocol to avoid performance degradations of the baseline methods. This protocol uses the first five images for the training and the remaining last image for the testing. This protocol generates 240 (240×1) genuine and 57360 ($240 \times 239 \times 1$) imposter comparison scores. Similar to Section IV.B and IV.D, 3D surface normal images are used for *SGD* [21], 3D depth images are used for *CR_L1_DALM* and *CR_L2* [38], while the artificially rendered images are used for our method and other evaluated methods including *ResNet-50* [16], *VGG-16* [17], *AlexNet* [28], *DenseNet-121* [46] and *SE-ResNet-50* [47]. We also modify the last classification layer to an output of 240 classes and initialized the weights from the respective trained model. Except this layer, we initialized the weights from the respective trained model available along with respective methods for *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. Similar to Section IV.D, rotational alignment was not considered in this dataset by neither the recent baseline method *SGD* nor another former baseline method *Finger*

Surface Code [4], we disabled our data augmentation with rotations to ensure fairness in the comparisons.

Figure 10 presents such comparative experimental results using the baseline methods and Table IX highlights the comparative performance for high security conditions. Despite most evaluated method achieves comparable performance due to the less challenging protocol used for this dataset, we show that our approach can also produce comparable performance on 3D fingerprint recognition without further customizing our network.

Besides, we also provide supportive experimental results using another 3D fingerprint database. This 3D fingerprint database [49] contains 3000 images from 150 subjects, each with ten fingers, two images per finger. We follow the original reference that those images from six fingers (right thumb, right index, right middle finger, left thumb, left index and left middle finger) are used for the experimentation. While the images provided in this database are not truly 3D, other large-scale 3D databases can help to advance further research. More details on this database can be referred to [49]. The original experimental protocol is not suitable for CNN based methods which generally requires the splitting of data into training and testing sets. Therefore, similar to our other experiments, we adopt a two-session like evaluation protocol that the first image from each finger is used for the training and the second image from each finger is used for the performance evaluation. This evaluation protocol generates 809,100 ($900 \times 899 \times 1$) imposter comparison scores and 900 (900×1) genuine comparison scores. Since this database only provides rendered images from 3D fingerprint, we are unable to compare with *SGD* [21], *CR_L1_DALM* [38] and *CR_L2* [38]. Instead, we compare our methods with other state-of-the-art CNN based methods including *ResNet-50* [16], *VGG-16* [17], *AlexNet* [28], *DenseNet-121* [46] and *SE-ResNet-50* [47]. We also modify the last classification layer to an output of 900 classes and initialized the weights from the respective trained model. Except this layer, we initialized the weights from the respective trained model available along with respective methods for *ResNet-50*, *VGG-16*, *AlexNet*, *DenseNet-121* and *SE-ResNet-50*. Since there is only one raw image per class, we enabled our data augmentation with rotations for all methods to enhance the recognition performance and to ensure fairness in the comparisons.

Figure 11 presents such comparative experimental results using the baseline methods and Table X highlights the comparative performance for high security conditions. It can be observed that our method generally outperforms all other methods in both verification and identification scenarios, which has again validated the effectiveness of our proposed method.

In summary, we demonstrate the effectiveness of our proposed recognition framework on 3D finger knuckle, 3D palmprint and 3D fingerprint, which indicates the good generalizability on various 3D biometrics problem.

V. CONCLUSIONS AND FUTURE WORK

Despite deep learning approaches have been widely developed in object recognition, the direct applications from such approaches do not outperform specialized hand-crafted

feature description approaches for the problem addressed in this paper. Moreover, such approaches have to address challenges associated with biometrics, e.g. availability of very limited training data, large intra-class or train-test sample variations as observed for the real applications. This paper develops a new neural network based approach for the contactless 3D finger knuckle identification. Our network is compact to avoid unnecessary overfitting, simultaneously collaborates multi-scales feature which are informative for personal identification and incorporates an efficient alignment scheme with a fully convolutional architecture to accommodate involuntary finger variations during the contactless imaging. Our reproducible [34] experimental results using publicly available databases including contactless 3D finger knuckle, 3D palmprint, 3D fingerprint and 2D finger knuckle, not only demonstrate the effectiveness but also the generalizability of our proposed approach.

Since it is the first work to develop a neural network approach for 3D finger knuckle recognition, several limitations with this paper has to be addressed in the future extension of this work. The key limitation of the proposed approach lies in its relatively more computational time when compared to other state-of-the-arts methods because of the additional alignment scheme introduced. However, the need of such alignment can be justified by its effectiveness to significantly improve the recognition performance. Secondly, our network is fine-tuned from *ResNet*, mainly because of the lack of large amount of training data. The need for pre-trained models can be mitigated by acquiring more diversified 3D finger knuckle images in the future. Meanwhile, our trained model is also provided along with this paper [34] for further research development. Thirdly, similar to object recognition frameworks, e.g. *ResNet* and *VGG*, and a periocular recognition research [48], the last classification layer limits the application on open-set identification problems. The current closed-set solutions enable the embedding of registered subjects' information into model parameters and therefore provide efficient closed-set identification. Our proposed approach can be deployed in a small-scale environment like offices, houses, or personal devices, where the training is very easy. Open-set solutions will be considered in the future work with the investigation of one-shot learning using Siamese or triplet architectures. Fourthly, the current approach considered segmentation and recognition as two separate problems. It is also interesting to simplify the networks, e.g. reusing some deep features from segmentation, or to develop an end-to-end network architecture for real world applications in the future work. Despite the above challenges, this paper provides an important foundation for further research and investigation to advance such technologies.

REFERENCES

- [1] Y. Shi, X. Yu, K. Sohn, M. Chandraker and A. K. Jain. Towards Universal Representation Learning for Deep Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [2] S. Z. Gilani and A. Mian. Learning from Millions of 3D Scans for Large-Scale 3D Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

- [3] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. Handbook of Fingerprint Recognition. *Springer Science & Business Media*, 2009.
- [4] A. Kumar and C. Kwong. Towards Contactless, Low-Cost and Accurate 3D Fingerprint Identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [5] J. Daugman. How Iris Recognition Works. *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21-30, 2004.
- [6] H. Proenca and J. C. Neves. IRINA: Iris Recognition (Even) in Inaccurately Segmented Data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [7] Summary of NIST Standards for Biometric Accuracy, Temper Resistance and Interoperability. *NIST Report*, 2002.
- [8] L. Zhang, L. Zhang, D. Zhang and H. Zhu, "Online Finger-Knuckle-Print Verification for Personal Authentication," *Pattern Recognition*, 43(7), pp.2560-2571, 2010.
- [9] A. Kumar and Ch. Ravikanth. Personal Authentication using Finger Knuckle Surface. *IEEE Transactions on Information Forensics and Security*, 4(1), pp.98-110, 2009.
- [10] G. Jaswal, A. Kaul, and R. Nath. Knuckle Print Biometrics and Fusion Schemes—Overview, Challenges, and Solutions. *ACM Computing Surveys (CSUR)*, 49(2), pp.1-46, 2016.
- [11] J. Kim, K. Oh, B. S. Oh, Z. Lin and K. A. Toh. A Line Feature Extraction Method for Finger-Knuckle-Print Verification. *Cognitive Computation*, 11(1), pp.50-70, 2018.
- [12] K. He, G. Gkioxari, P. Dollár and R. Girshick. Mask R-CNN. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [13] R. Girshick. Fast R-CNN. In *IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [14] S. Ren, K. He, R. Girshick and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2015.
- [15] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [16] K. He, X. Zhang, S. Ren and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [17] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations (ICLR)*, 2015.
- [18] C. Lin and A. Kumar. Contactless and Partial 3D Fingerprint Recognition using Multi-View Deep Representation. *Pattern Recognition*, 83, pp.314-327, 2018.
- [19] Z. Zhao and A. Kumar. Towards More Accurate Iris Recognition Using Deeply Learned Spatially Corresponding Features. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [20] O. M. Parkhi, A. Vedaldi and A. Zisserman. Deep Face Recognition. In *British Machine Vision Conference (BMVC)*, 2015.
- [21] K. H.M. Cheng and A. Kumar. Contactless Biometric Identification using 3D Finger Knuckle Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), pp.1868-1883, 2020.
- [22] J. J. Koenderink and A. J. Vandoorn. Surface Shape and Curvature Scales. *Image and Vision Computing*, 10(8), pp.557-564, 1992.
- [23] D. L. Woodard and P. J. Flynn. Finger Surface as a Biometric Identifier. *Computer Vision and Image Understanding*, 100(3), pp.357-384, 2005.
- [24] K. Ito, T. Aoki, H. Nakajima, K. Kobayashi, T. Higuchi. A Palm-Print Recognition Algorithm using Phase-Only Correlation. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 91(4), pp.1023-1030, 2008.
- [25] A. Shoichiro, K. Ito, and T. Aoki. A Multi-Finger Knuckle Recognition System for Door Handle. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013.
- [26] C. Dorai and A. K. Jain. COSMOS - A Representation Scheme for 3D Free-form Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10), pp.1115-1130, 1997.
- [27] Q. Zheng, A. Kumar and G. Pan. A 3D Feature Descriptor Recovered from a Single 2D Palmprint Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6), pp.1272-1279, 2016.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [29] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C.L. Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)*, 2014.
- [30] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap. One-shot Learning with Memory-Augmented Neural Networks. arXiv preprint arXiv:1605.06065, 2016.
- [31] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra. Matching Networks for One Shot Learning, In *Advances in Neural Information Processing Systems (NIPS)*, 2016
- [32] V. Kanhangad, A. Kumar, and D. Zhang. A Unified Framework for Contactless Hand Verification. *IEEE Transactions on Information Forensics and Security*, 6(3), pp.1014-1027, 2011.
- [33] Q. Zheng, A. Kumar, and G. Pan. Suspecting Less and Doing Better: New Insights on Palmprint Identification for Faster and More Accurate Matching. *IEEE Transactions on Information Forensics and Security*, 11(3), pp.633-641, 2016.
- [34] Weblink for downloading the 3D finger knuckle images datasets, the network model and implementation codes: <http://www4.comp.polyu.edu.hk/~csajaykr/fknet.zip>
- [35] The HK PolyU Contactless Finger Knuckle Images Database (V-1.0): <http://www4.comp.polyu.edu.hk/~csajaykr/fn1.htm>
- [36] The HK PolyU Contactless Hand Dorsal Images Database: <http://www4.comp.polyu.edu.hk/~csajaykr/knuckleV2.htm>
- [37] F. Li, R. Fergus, and P. Perona. One-shot Learning of Object Categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4), pp.594-611, 2006.
- [38] L. Zhang, Y. Shen, H. Li, and J. Lu. 3D Palmprint Identification Using Block Wise Features and Collaborative Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8), pp.1730-1736, 2015.
- [39] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. In *International Conference on Learning Representations (ICLR)*, 2014.
- [40] F. Ahmad, L. Cheng and A. Khan. Lightweight and Privacy-Preserving Template Generation for Palm-Vein-Based Human Recognition. *IEEE Transactions on Information Forensics and Security*, 15, pp.184-194, 2019.
- [41] Y. Liu, A. Kumar, "Contactless palmprint identification using deeply learned residual features," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 2, pp. 172–181, Apr. 2020.
- [42] W. Kang and Q. Wu. Contactless Palm Vein Recognition Using a Mutual Foreground-Based Local Binary Pattern. *IEEE Transactions on Information Forensics and Security*, 9(11), pp.1974-1985, 2014.
- [43] F. Yue, B. Li, M. Yu and J. Wang. Hashing Based Fast Palmprint Identification for Large-Scale Databases. *IEEE Transactions on Information Forensics and Security*, 8(5), pp.769-778, 2013.
- [44] S.J. Park, K.S. Hong and S. Lee. Rdfnet: RGB-D Multi-Level Residual Feature Fusion for Indoor Semantic Segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [45] Z. Zhang, X. Zhang, C. Peng, X. Xue and J. Sun. Exfuse: Enhancing Feature Fusion for Semantic Segmentation. In *European Conference on Computer Vision (ECCV)*, 2018.
- [46] G. Huang, Z. Liu, L. Van Der Maaten and K.Q. Weinberger. Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [47] J. Hu, L. Shen, S. Albanie, G. Sun and E. Wu. Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), pp.2011-2023, 2020.
- [48] H. Proença and J. C. Neves. Deep-prwis: Periocular Recognition Without the Iris and Sclera using Deep Learning Frameworks. *IEEE Transactions on Information Forensics and Security*, 13(4), pp.888-896, 2018.
- [49] W. Zhou, J. Hu, I. Petersen, S. Wang and M. Bennamoun A Benchmark 3D Fingerprint Database. In *IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2014.