

C2CL: Contact to Contactless Fingerprint Matching

Steven A. Grosz*, Joshua J. Engelsma, Eryun Liu, *Member, IEEE* and Anil K. Jain, *Life Fellow, IEEE*

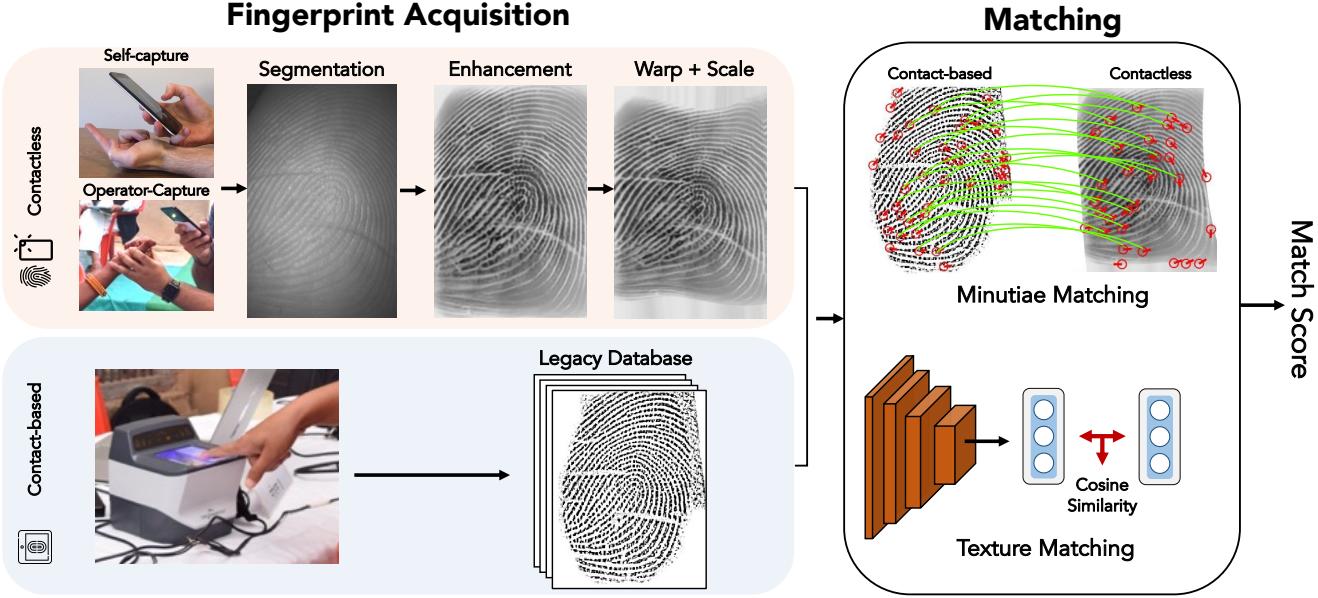


Fig. 1: Overview of matching contactless fingerprint images with a legacy database of contact-based fingerprint impressions. While only a specific scenario is shown here where contact-based images are obtained from optical FTIR readers (slap or single finger capture) and contactless images are captured by a smartphone camera, our approach can be applied to any heterogeneous fingerprint matching problem.

Abstract—Matching contactless fingerprints or finger photos to contact-based fingerprint impressions has received increased attention in the wake of COVID-19 due to the superior hygiene of the contactless acquisition and the widespread availability of low cost mobile phones capable of capturing photos of fingerprints with sufficient resolution for verification purposes. This paper presents an end-to-end automated system, called C2CL, comprised of a mobile finger photo capture app, preprocessing, and matching algorithms to handle the challenges inhibiting previous cross-matching methods; namely i) low ridge-valley contrast of contactless fingerprints, ii) varying roll, pitch, yaw, and distance of the finger to the camera, iii) non-linear distortion of contact-based fingerprints, and vi) different image qualities of smartphone cameras. Our preprocessing algorithm segments, enhances, scales, and unwarps contactless fingerprints, while our matching algorithm extracts both minutiae and texture representations. A sequestered dataset of 9,888 contactless 2D fingerprints and corresponding contact-based fingerprints from 206 subjects (2 thumbs and 2 index fingers for each subject) acquired using our mobile capture app is used to evaluate the cross-database performance of our proposed algorithm. Furthermore, additional experimental results on 3 publicly available datasets show substantial improvement in the state-of-the-art for contact to contactless fingerprint matching (TAR in the range of 96.67% to 98.30% at FAR=0.01%).

Index Terms—Fingerprint recognition, Sensor interoperability,

S.A. Grosz, J.J. Engelsma, and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824. E-mail: {groszste, engelsm7, jain}@cse.msu.edu
Eryun Liu is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. E-mail: eryunliu@zju.edu.cn
*Corresponding author.

Contact to contactless fingerprint matching

I. INTRODUCTION

Due to their presumed uniqueness and permanence, fingerprints are one of the most widely used biometric traits for secure authentication and search [1], [2]. Over the years many different types of fingerprint readers have been developed to obtain a digital image of a finger's friction ridge pattern. These readers vary in a number of different ways, including the underlying sensing technology (e.g., optical, capacitive, ultrasonic, *etc.*) or in the manner in which a user interacts with the reader (*i.e.* contactless, 4-4-2 slap, or single finger contact-based acquisition). Most prevailing fingerprint readers in use today necessitate physical contact of the user's finger with the imaging surface of the reader; however, this direct contact presents certain challenges in processing the acquired fingerprint images. Most notably, elastic human skin introduces a non-linear deformation upon contact with the imaging surface which has been shown to significantly degrade matching performance [3], [4], [5]. Furthermore, contact with the surface is likely to leave a latent impression on the imaging surface [6], which presents a security risk as an imposter could illegally gain access to the system through creation of a presentation (*i.e.*, spoof) attack.

In light of the ongoing Covid-19 pandemic, contactless fingerprint recognition has gained renewed interest as a hygienic alternative to contact-based fingerprint acquisition [7]. This is further supported by a recent survey that showed

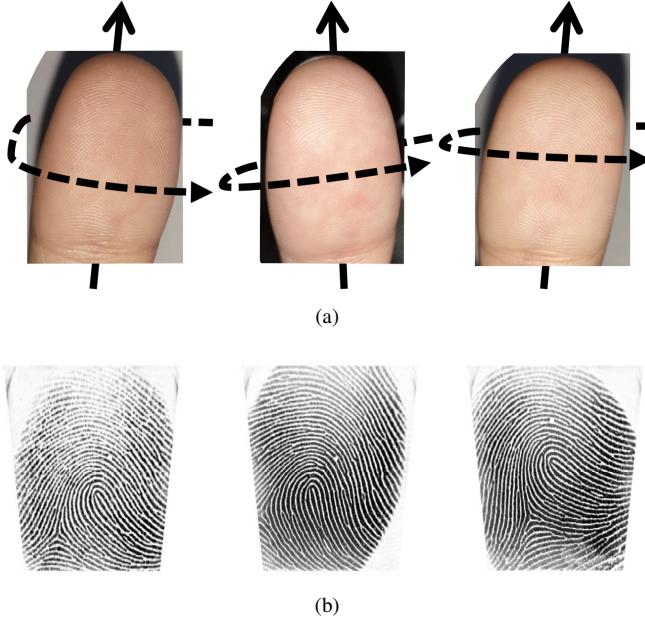


Fig. 2: Examples of contactless fingerprints (a) and their corresponding contact-based fingerprint images (b). Varying viewing angle, resolution, and illumination of contactless images and non-linear distortion of contact-based fingerprints contribute to the degradation of cross-matching performance. The contactless images shown are from the ZJU dataset.

that the majority of users prefer touchless capture methods in terms of usability and hygiene considerations [8]. Prior studies have explored the use of customized 2D or 3D sensing for contactless fingerprint acquisition [9], [10], [11], [12], [13], [14], while others have explored the low-cost alternative of using readily available smartphone cameras to capture “finger photos”¹ [15], [16], [17].

Despite the benefits of contactless fingerprint acquisition, imaging and subsequently matching a contactless fingerprint presents its own set of unique challenges. These include (i) low ridge-valley contrast, (ii) non-uniform illumination, (iii) varying roll, pitch, and yaw of the finger, (iv) varying background, (v) perspective distortions due to the varying distances of the finger from the camera, and (vi) lack of cross-compatibility with legacy databases of contact-based fingerprints (see Figure 2). For widespread adoption, contactless fingerprint recognition must overcome the aforementioned challenges and bridge the gap in accuracy compared to contact-contact fingerprint matching.

The most significant factor limiting the adoption of contactless fingerprint technology is cross-compatibility with legacy databases of contact-based fingerprints, which is particularly important for governmental agencies and large-scale national ID programs such as India’s Aadhaar National ID program which has already enrolled over 1 billion users based upon contact-based fingerprints. Several studies have aimed at improving the compatibility of matching legacy slap images to contactless fingerprint images [21], [20], [24], [25], [23], [19]; however, none have achieved the same levels of accuracy as

¹In general, contactless fingerprints refers to fingerprint images acquired by a contactless fingerprint sensor, whereas finger photo refers to fingerprint images acquired by a mobile phone. In this paper, we use the two terms interchangeably.

state-of-the-art (SOTA) contact-contact fingerprint matching (such as the results reported in FVC-ongoing [26] and NIST FpVTE [27]). Furthermore, all of these works focus on solving only a subset of the challenges in an effort to push the contact-contactless matching accuracy closer to the SOTA contact-contact matching systems. Indeed, to the best of our knowledge, this study presents the most comprehensive, end-to-end² solution in the open academic literature for contact-contactless fingerprint matching that addresses the challenges inherent to each step in the contact to contactless matching process (mobile capture, segmentation, enhancement, scaling, non-linear warping, representation extraction, and matching).

We show that our end-to-end matcher, called C2CL, is able to significantly improve contact-contactless matching performance over the prevailing SOTA methods through experimental results on a number of different datasets, collected by various research groups using their own app and fingerprint readers. We also demonstrate that our matcher generalizes well to datasets which were not included during training. This cross-database evaluation solves a shortcoming of many existing studies which train and evaluate algorithms on different training and test splits of the same contact-contactless dataset. Furthermore, despite multiple evaluation datasets, we train only a single model for our evaluations, rather than fine-tuning individual models to fit a specific dataset.

Concretely, the contributions of our work are stated as:

- 1) An end-to-end system, called C2CL, for contact-contactless fingerprint matching. C2CL is comprised of preprocessing (segmentation, enhancement, scaling, and deformation correction), feature extraction (minutiae and texture representations), and matching modules. Our preprocessing also benefits the Verifinger 12.0 commercial fingerprint SDK.
- 2) A fully automated, preprocessing pipeline to map contactless fingerprints into the domain of contact-based fingerprints and a contact-contactless adaptation of DeepPrint [28] for representation extraction. Our preprocessing and representation extraction is generalizable across multiple datasets and contactless capture devices.
- 3) SOTA cross-matching verification and large-scale identification accuracy using C2CL on both publicly available contact-contactless matching datasets as well as on a completely sequestered dataset collected at Zhejiang University, China. Our evaluation includes the most diverse set of contactless fingerprint acquisition devices, yet we employ just a single trained model for evaluation.
- 4) A smartphone contactless fingerprint capture app that was developed in-house for improved throughput and user-convenience. This app will be made available to the public to promote further research in this area³.
- 5) A new dataset of 9,888 2D contactless and corresponding contact-based fingerprint images from 206 subjects

²The Cambridge Dictionary defines end-to-end as “from the very beginning of a process to the very end”. Our method is end-to-end as it carries out the full process from data collection to recognition. Our use of “end-to-end” should not be confused with “end-to-end learning”.

³The project repository for the smartphone contactless fingerprint capture app is available at <https://github.com/ronny3050/FingerPhotos>.

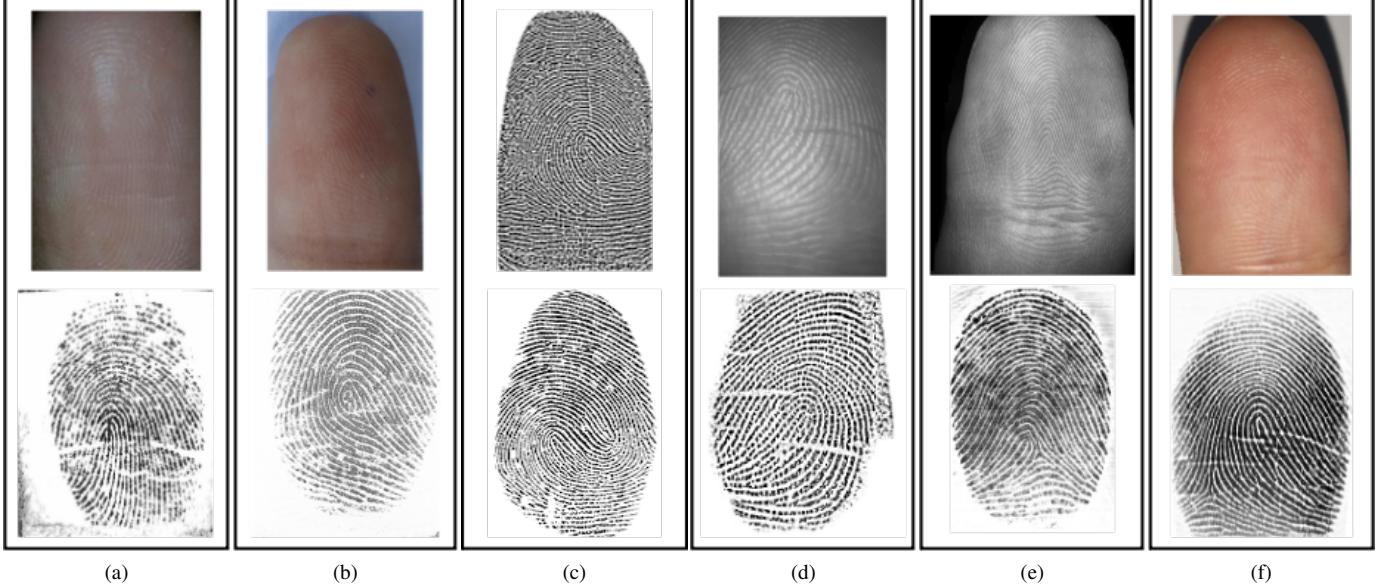


Fig. 3: Example contactless and contact-based fingerprint image pairs from databases which we have obtained from different research groups: (a) IIT Bombay [18], (b) ISPFDv2 [19], (c) MSU [20], (d) PolyU [21], (e) UWA [22], and (f) ZJU datasets. In general, contactless fingerprints suffer from low ridge-valley contrast, varying roll, pitch, and yaw, and perspective distortions, especially those captured by smartphone cameras (*e.g.*, (a), (b), (c) and (f)). We believe our study involves the largest collection of public domain databases of contactless and contact-based fingerprints.

(2 thumbs and 2 index fingers per subject), which will be made available in the public domain to advance much needed research in this area⁴.

II. PRIOR WORK

Prior studies on contact-contactless fingerprint matching primarily focus on only one of the sub-modules needed to obtain matching accuracy close to contact-contact based fingerprint matching systems (*e.g.*, segmentation, distortion correction, or feature extraction only). These studies are categorized and discussed below.

A. Segmentation

The first challenge in contact-contactless matching is segmenting the relevant fingerprint region from the captured contactless fingerprint images. Malhotra *et al.* [19] proposed a combination of a saliency map and a skin-color map to segment the distal phalange (*i.e.*, fingertip) of contactless fingerprint images in presence of varying background, illumination and resolution. Despite impressive results, the algorithm requires extensive hyperparameter tuning and still fails to accurately segment fingerprints in severe illumination conditions or noisy backgrounds. To alleviate these issues, we incorporate segmentation via an autoencoder trained to robustly segment the distal phalange of input contactless images.

B. Enhancement

One of the main challenges with contactless fingerprint images is the low ridge-valley contrast (Figure 3). The literature has addressed this in a number of different ways,

including adaptive histogram equalization, Gabor filtering, median filtering, and sharpening by subtraction of the Gaussian blurred image from the captured image ([25], [23], [19]). We also incorporate adaptive contrast enhancement in our work; however, one consideration that is lacking in existing approaches is the ridge inversion that occurs with Frustrated Total Internal Reflection (FTIR) optical imaging. In particular, the ridges and valleys of an FTIR fingerprint image will appear dark and light, respectively, while the opposite is true in contactless fingerprint images. Therefore, a binary inversion of the contactless fingerprint images is expected to improve the correspondence with their contact-based counterparts.

C. Scaling

After segmenting and enhancing a contactless fingerprint, the varying distances between fingers captured and the camera must be accounted for. In particular, since contact-based fingerprints are almost always captured at 500 pixels per inch (ppi), the contactless fingerprints need to be scaled to be as close to 500 ppi as possible. Previous studies have applied a fixed manual scaling, set for a specific dataset, or have employed contact-based fingerprint ridge frequency normalization algorithms that rely on accurate ridge extraction - which is often unreliable for contactless fingerprints. In contrast, we incorporate a spatial transformer network [30] which has been trained to automatically normalize the resolution of the contactless fingerprints to match that of the 500 ppi contact images. This scaling is performed dynamically, *i.e.* every input contactless fingerprint image is independently scaled.

D. Distortion Correction

A final preprocessing step for contact-contactless matching is non-linear distortion correction. To address this problem,

⁴This dataset will be available to interested readers after this paper has been accepted.

TABLE I: Summary of Published Cross-Matching Contact to Contactless Fingerprint Recognition Studies.

Study	Approach	Database	Accuracy [†]
Lin and Kumar, 2018 [21]	Robust TPS deformation correction model, minutiae and ridge matching	1,800 contactless and contact fingerprints from 300 fingers [21]. 2,000 contactless and 4,000 contact fingerprints from 1,000 fingers [22]	EER = 14.33% [21] EER = 19.81% [22]
Deb <i>et al.</i> , 2018 [20]	COTS matcher	2,472 contactless and contact fingerprints from 1,236 fingers [20]	TAR = 92.4% – 98.6% @ FAR = 0.1% [20]
Lin and Kumar, 2019 [23]	Fusion of three Siamese CNNs	960 contactless and contact fingerprints from 160 fingers [21]. 1,000 contactless and 2,000 contact fingerprints from 500 fingers [22]	EER = 7.93% [21] EER = 7.11% [22]
Wild <i>et al.</i> , 2019 [24]	Filtering based on NFIQ 2.0 quality measure, COTS matcher	1,728 contactless and 2,582 contact fingerprints from 108 fingers [24]	TAR = 95.5% – 98.6% @ FAR = 0.1% [24]
Dabouei <i>et al.</i> , 2019 [25]	TPS spatial transformer network for deformation correction and binary ridge-map extraction network, COTS matcher	2,000 contactless and 4,000 contact fingerprints from 1,000 fingers [22]	EER = 7.71% [22]
Malhotra <i>et al.</i> , 2020 [19]	Feature extraction with deep scattering network, random decision forest matcher	8,512 contactless and 1,216 contact fingerprints from 152 fingers [19]	EER = 2.11% – 5.23% [19]
Priesnitz <i>et al.</i> , 2021 [8]	Neural network-based minutiae feature extraction, open-source minutiae matcher	896 contactless from two different capture setups and 464 contact fingerprints from 232 fingers [8]	EER = 15.71% and 32.02% [8]
Proposed Approach	TPS spatial transformer for 500 ppi scaling and deformation correction of contactless fingerprints. Fusion of minutiae and CNN texture representations.	8,512 contactless and 1,216 contact fingerprints from 152 fingers [19]. 2,000 contactless and 4,000 contact fingerprints from 1,000 fingers [22]. 960 contactless and contact fingerprints from 160 fingers [21]. 9,888 contactless and 9,888 contact fingerprints from 824 fingers (ZJU Dataset)	EER = 1.20% [19] EER = 0.72% [22] EER = 0.30% [21] EER = 0.62% (ZJU Dataset)

[†] Some studies only report EER while other studies only report TAR @ FAR = 0.1%.

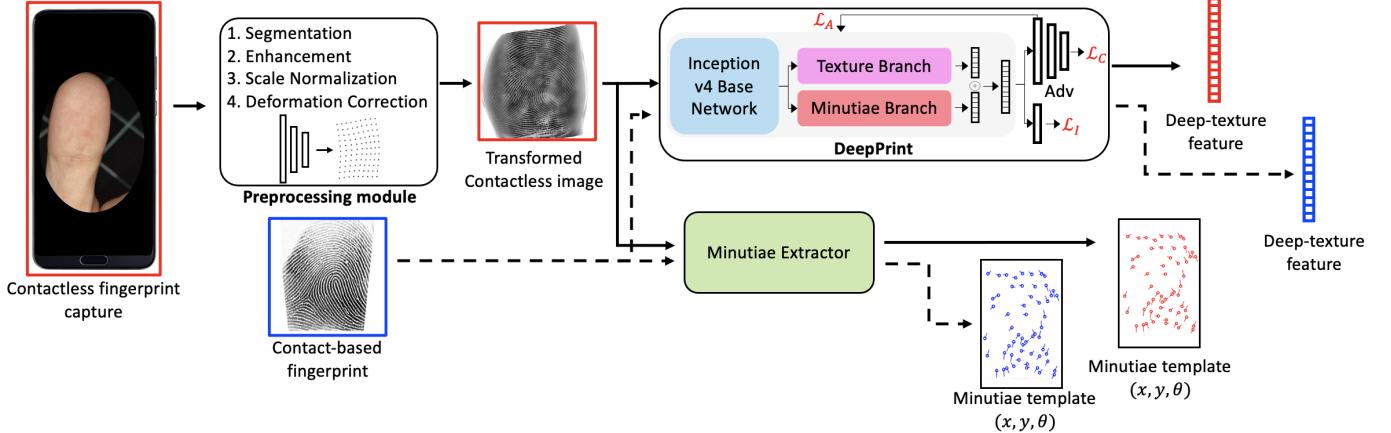


Fig. 4: System architecture of C2CL. (a) A contactless fingerprint is captured and used as input to the preprocessing module, consisting of segmentation, enhancement, 500 ppi ridge frequency scaling, and deformation correction; (b) the transformed image output by the preprocessing module is fed to DeepPrint [28], which extracts a texture representation (shown in red). Without performing any additional preprocessing, the corresponding contact-based fingerprint is again fed to DeepPrint to extract a texture representation (shown in blue). Simultaneously, a minutiae representation is extracted using the Verifinger 12.0 SDK from both the contactless and contact-based fingerprint images.

[21] used thin-plate-spline (TPS) deformation correction models (previously applied for contact-contact matching [3], [31], [32], [5], [33], [34]) using the alignment between minutiae annotations of corresponding contactless and contact fingerprints. A limitation is that the transformation is limited to one of six possible parameterizations. In a different study, Dabouei *et al.* [31] train a spatial transformer to learn the distortion correction that is dynamically computed for each input image. In [31], a contact-based image is used as the reference for learning the distortion correction for a contactless image. However, we argue that this is not a reliable ground truth since the deformation varies among different contact-

based fingerprint impressions. In our attempt to re-implement their algorithm, we found that this lack of a reliable and consistent ground truth makes training unstable, making it difficult to learn sound distortion parameters. In our work, rather than using the contact-based image as a reference, we use the match scores of our texture matcher as supervision for generating robust distortion correction. In other words, the distortion correction is optimized to maximize match scores between genuine contact-contactless fingerprint pairs.

TABLE II: Summary of contact to contactless fingerprint recognition datasets used in this study.

Dataset	# Subjects	# Unique Fingers	# Images (Contactless / Contact)	Contactless Capture Device	Contact Capture Device
UWA Benchmark 3D Finger-print Database, 2014 [22]	150	1,500	3,000 / 6,000	3D Scanner (TBS S120E)	CROSSMATCH Verifier 300 LC2.0
ManTech Phase2, 2015 [29]	496	4,960	N/A / N/A*	AOS ANDI On-The-Go (OTG), MorphoTrak Finger-On-The-Fly (FOTF), IDair innerID on iPhone 4.	Cross Match Guardian R2, Cross Match SEEK Avenger, MorphoTrak MorphoIDent, MorphoTrust TouchPrint 5300, Northrop Grumman BioSled
PolyU Contactless 2D to Contact-based 2D Images Database, 2018 [21]	N/A	336	2,976 / 2,976	Low-cost camera and lens (specific device not given)	URU 4000
MSU Finger Photo and Slap Fingerprint Database, 2018 [20]	309	1,236	2,472 / 2,472	Xiaomi Redmi Note 4 smartphone	CrossMatch Guardian 200, SilkID (SLK20R)
IIT Bombay Touchless and Touch-Based Fingerprint Database, 2019 [18]	N/A	200	800 / 800	Lenovo Vibe k5 smartphone	eNBioScan-C1 (HFUDU08)
ISPFDv2, 2020 [19]	76	304	17,024 / 2,432	OnePlus One (OPO) and Micromax Canvas Knight smartphones	Secugen Hamster IV
ZJU Finger Photo and Touch-based Fingerprint Database	206	824	9,888 / 9,888	Huawei P20, Samsung s9+, and OnePlus 8 smartphones	URU 4500

* The number of contact and contactless images acquired per finger varies for each device and the exact number is not provided.

E. Representation Extraction and Matching

After preprocessing a contactless fingerprint image to lie within the same domain as a contact-based fingerprint, a discriminative representation must be extracted for matching. In the prior literature there are two main approaches to feature representation: (i) minutiae representation ([31], [21]) and (ii) deep learning representation ([23], [19]). Minutiae-based approaches rely on clever preprocessing and other techniques to improve the compatibility of contactless fingerprint images for traditional contact-based minutiae extraction algorithms. On the other hand, deep learning approaches place less emphasis on preprocessing to manipulate the contactless fingerprints to improve correspondence with contact-based fingerprints, rather the responsibility is placed on the representation network to learn the correspondence despite the differences. For example, Lin and Kumar [35] and Dabouei *et al.* [31] both apply a deformation correction to the contactless image to improve the minutiae correspondence. In contrast, the deep learning approach taken in [23] applies very little preprocessing to the contactless image (just contrast enhancement and Gabor filtering) and leverages a Siamese CNN to extract features for matching. Similarly, Malhotra *et al.* [19] utilize a deep scattering network to extract multi-scale and multi-directional feature representations.

In contrast to prior studies, our approach utilizes both a texture representation and a minutiae representation. Given the lower contrast and quality of contactless fingerprints (causing missing or spurious minutiae) and the non-linear distortion and scaling discrepancies between contact and contactless fingerprints (negatively impacting minutiae graph matching algorithms) a global texture representation is useful to improve the contact-contactless matching accuracy. We demonstrate this hypothesis empirically in the experimental results.



Fig. 5: Example segmentation success (a) and failure (b) cases from images in the ISPFDv2 dataset using our segmentation algorithm. Sources of failure are presence of skin-like color tones in the background and varying skin complexion due to varying illumination.

III. METHODS

Our matcher, C2CL, aims to improve contact to contactless fingerprint recognition through a multi-stage preprocessing algorithm and matching algorithm comprised of both a minutiae representation and a texture feature representation. The preprocessing is employed to minimize the domain-gap between the contactless fingerprints residing in a domain \mathcal{D}_{cl} and contact-based fingerprints residing in another domain \mathcal{D}_c and consists of segmentation, enhancement, ridge frequency scaling to 500 ppi, and deformation correction through a learned spatial transformation network. After preprocessing, we extract deep-textural and minutiae representations (unordered, variable length sets $T = \{(x_1, y_1, \theta_1), \dots, (x_n, y_n, \theta_n)\}$) for matching. The final match scores are obtained via a score-level fusion between the texture and minutiae matching scores.

A. Preprocessing

Here we discuss the details of each stage of our preprocessing algorithm as illustrated in Figure 6.

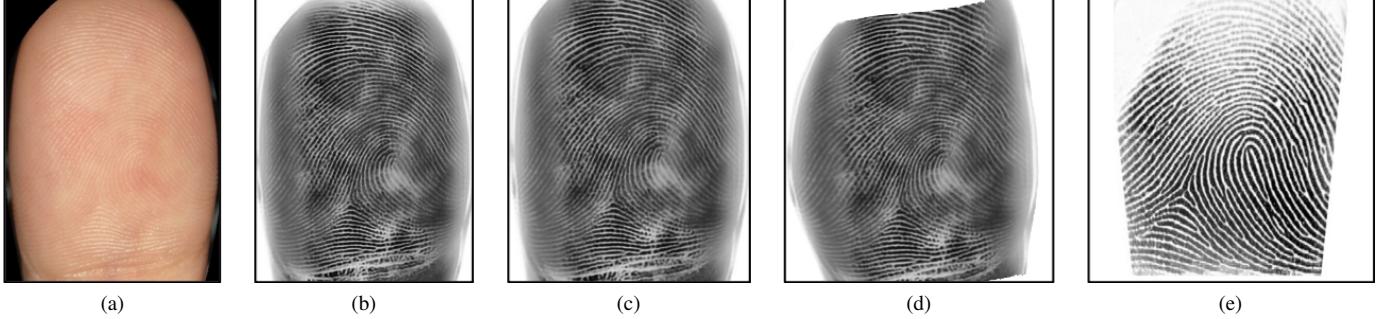


Fig. 6: Illustration of our preprocessing pipeline including (a) segmentation, (b) enhancement, (c) scaling, and (d) warping. For reference, a corresponding contact-based fingerprint is shown in (e).

1) Segmentation: Many contactless fingerprint datasets are unsegmented; for example, the ISPFDv2 dataset [19] contains unsegmented, $(4, 208 \times 3, 120)$ images with varying illumination, resolution, and background conditions. Thus, the first step in our preprocessing pipeline is to segment the distal phalange of the fingerprint using a U-net segmentation network [36]. Our segmentation algorithm is a network $S(\cdot)$ which takes as input the unsegmented contactless fingerprint I_{cl} of dimension $(m \times n)$ and outputs a segmentation mask $\hat{M} \in \{0, 1\}$ of dimension $(m \times n)$. The obtained segmentation mask, \hat{M} , is element-wise multiplied with I_{cl} to (i) crop out only the distal phalange of the contactless fingerprints and (ii) eliminate the remaining background to avoid detection of spurious minutiae in the later representation extraction stage. The segmented image I'_{cl} is then resized to 480×480 by maintaining the aspect ratio with appropriate padding for further processing.

For training $S(\cdot)$, we manually marked segmentation masks M of the distal phalange of 496 contactless fingerprints from the ISPFDv2 dataset⁵. Initially, 200 images were randomly selected to have varying resolutions of either 5MP, 8MP, or 16MP and another 200 were selected with varying backgrounds and illumination. An additional 96 images were specifically selected for their greater perceived difficulty, particularly images with skin tone backgrounds. The optimization function for training $S(\cdot)$ is a pixel-wise binary cross-entropy loss between \hat{M} and M (Eq. 1).

$$\mathcal{L}_{seg}(I_{cl}, I_c, M) = - \sum_{i,j} [M_{i,j} \log(\hat{M}_{i,j}|I_{cl}) + (1 - M_{i,j}) \log(1 - \hat{M}_{i,j}|I_{cl})] \quad (1)$$

2) Enhancement: Following segmentation, we apply a series of image enhancements $E(\cdot)$ to increase the contrast of the ridge-valley structure of the contactless images, including: (i) an adaptive histogram equalization to improve the ridge-valley contrast and (ii) pixel gray-level inversion to correct for the inversion of ridges between contact-based and contactless fingerprints. We also experimented with SOTA super resolution and de-blurring techniques, such as RDN [38], to further improve the contactless image quality, but found only minimal matching accuracy improvements at the expense of significant additional computational cost.

⁵We used the open source Labelme segmentation tool found on GitHub [37].

3) Distortion Correction and Scaling: After segmenting and enhancing the contactless fingerprints, the non-linear distortions that separate the domains of contactless and contact-based fingerprints must be removed. In particular, this includes both a perspective distortion (caused by the varying distance of a finger from the camera) and a non-linear distortion (caused when the elastic human skin flattens against a platen).

To correct for these discrepancies, we train a spatial transformer network (STN) [30] $T(\cdot)$ that takes as input a segmented, enhanced contactless image $I_{cl}^e = E(I'_{cl})$ and aligns the ridge structure to better match the corresponding contact-based image domain \mathcal{D}_c . The goal of the STN is two-fold: (i) an affine transformation $T_s(\cdot)$ to normalize the ridge frequency of the contactless images to match the 500 ppi ridge spacing of the contact-based impressions and (ii) a TPS deformation warping $T_d(\cdot)$ of the contactless images to match the deformation present in contact-based images due to the elasticity of the human skin.

Both $T_s(\cdot)$ and $T_d(\cdot)$ are comprised of a shared localization network $l(\cdot, w)$ and individual differentiable grid-samplers. Given an enhanced contactless fingerprint I_{cl}^e , $l(I_{cl}^e, w)$ outputs the scale (s), rotation (θ), and translation (t_x, t_y) of an affine transformation matrix A_s (Eq. 2) and a distortion field Θ which is characterized by a grid of $n \times n$ pixel displacements $\{(x_1, y_1) \dots (x_n, y_n)\}$. Subsequently, a scaled, warped image I_{cl}^w is obtained via Equation 3.

To learn the weights w of the localization network such that $T_s(\cdot)$ and $T_d(\cdot)$ correctly scale the contactless fingerprints to 500 ppi, and unroll them into a contact-based fingerprint, we minimize the distance between DeepPrint representations extracted from genuine pairs of scaled, warped contactless fingerprints (I_{cl}^w) and contact-based fingerprints (I_c). In particular, let $f(\cdot)$ be a frozen DeepPrint network pretrained on contact-based fingerprints. Then, we can obtain a pair of 192D DeepPrint identity representations R_{cl} and R_c via $R_{cl} = f(I_{cl}^w)$ and $R_c = f(I_c)$. Our loss can then be computed from Equation 4. By using the DeepPrint identity features extracted from contact-based fingerprint images to compute the loss, we are able to utilize the contact-based impressions as a ground truth of sorts. In particular, we are training our localization network to output better scalings and warpings such that the distortion and scale corrected contactless images have DeepPrint representations closer to their corresponding “ground truth” contact-based image.

We note that this approach has key differences to that which was proposed in [25] where the distortion corrected contactless image (scale was not learned in [25]) would be more directly compared to the ground-truth contact-based fingerprint via a cross-entropy loss between ‘‘binarized’’ versions of I_{cl}^w and I_c . We found that directly comparing the contactless and contact images via a cross entropy loss was quite difficult in practice since the ground truth contact image and the corresponding contactless image will have different rotations and translations separating them (even after scaling and distortion correction - resulting in a high loss value even if the scaling and distortion are correct). Furthermore, the contact-based image itself varies based upon the pressure applied during the acquisition, environmental conditions, sensor model, *etc.*, meaning that directly using the contact-based image as ground truth is unreliable. In contrast, since DeepPrint has been trained to be invariant to pressure, environmental conditions, and sensor model, our ground truth (DeepPrint representations from contact-based images) will remain stable across different contact-based impressions. In short, unlike [25], we learn both distortion correction **and** scaling correction simultaneously, and we use the DeepPrint identity loss to stabilize training of $T(\cdot)$ and to enable predictions of warpings and scalings which better improve matching accuracy.

$$A_s = \begin{bmatrix} s \cos(\theta) & -s \sin(\theta) & t_x \\ s \sin(\theta) & s \cos(\theta) & t_y \end{bmatrix} \quad (2)$$

$$I_{cl}^w = T(I_{cl}^e; A_s, \Theta) = T_d(T_s(I_{cl}^e, A_s), \Theta) \quad (3)$$

$$\mathcal{L}_{STN} = \|R_{cl} - R_c\|_2^2 \quad (4)$$

B. Representation Extraction

After performing all of the aforementioned preprocessing steps, we enter the second major stage of our contact-contactless matcher, namely the representation extraction stage. Our representation extraction algorithm extracts both a textural representation (using a CNN) and a minutiae-set. Scores are computed using both of these representations and then fused together using a sum score fusion.

1) *Texture Representation:* To extract our textural representation, we fine-tune the DeepPrint network proposed by Engelsma *et al.* in [28] on a training partition of the publicly available datasets which we aggregated (Table III). Unlike the deep networks used in [23] and [19] for extraction of textural representations, DeepPrint is a deep-network which has been specifically designed for fingerprint representation extraction via a built-in alignment module and minutiae domain knowledge. Therefore, in this work, we seek to adopt DeepPrint for contact-contactless fingerprint matching. As is common practice in the machine learning and computer vision communities, we are utilizing a pretrained DeepPrint network to warm start our model, which has been shown to improve over random initialization for many applications; for example, in fingerprint spoof detection [39].

Formally, DeepPrint is a network $f(\cdot)$ with parameters w that takes as input a fingerprint image I and outputs a fixed-length fingerprint representation R (which encodes the textural related features). During training, DeepPrint is guided to encode features related to fingerprint minutiae via a multi-task learning objective including: (i) a cross-entropy loss on both the minutiae branch identity classification probability \hat{y}_1 and texture branch identity classification probability \hat{y}_2 (Eq. 5), (ii) minimize the intra-class variance of class y via a center loss between the predicted minutiae feature vector R_1 and its mean feature vector R_1^y and the predicted texture feature vector R_2 and its mean feature vector R_2^y (where R_1 concatenated with R_2 form the full representation R), and (iii) a mean squared error loss on the predicted minutiae maps \hat{H} output by DeepPrint’s minutiae branch and ground truth minutiae maps H (Eq. 7). These losses are combined to form the DeepPrint identity loss, \mathcal{L}_{ID} (Eq. 8), where $\lambda_1 = 1$, $\lambda_2 = 0.00125$, $\lambda_3 = 0.095$ are set empirically.

$$\mathcal{L}_1(I, y) = -\log(\hat{y}_1^{j=y}|I, w) - \log(\hat{y}_2^{j=y}|I, w) \quad (5)$$

$$\mathcal{L}_2(I, y) = \|R_1 - \bar{R}_1^y\|_2^2 + \|R_2 - \bar{R}_2^y\|_2^2 \quad (6)$$

$$\mathcal{L}_3(I, H) = \sum_{j,k,l} (\hat{H}_{j,k,l} - H_{j,k,l})^2 \quad (7)$$

$$\mathcal{L}_{ID}(I, y, H) = \operatorname{argmin}_w \sum_{i=1}^N [\lambda_1 \mathcal{L}_1(I^i, y^i) + \lambda_2 \mathcal{L}_2(I^i, y^i) + \lambda_3 \mathcal{L}_3(I^i, H^i)] \quad (8)$$

Due to the differences in resolution, illumination, and backgrounds observed between different datasets of contactless fingerprint images, generalization to images captured on unseen cameras becomes critical. The problem of cross-sensor generalization in fingerprint biometrics (*e.g.*, optical reader to capacitive reader), of which contact to contactless matching is an extreme example, has been noted in the literature [40], [41], [42], [43], with many previous works aimed at improving the interoperability [44], [45], [46]. Motivated by the recent work employing adversarial learning to cross-sensor generalization of fingerprint spoof detection [47], we incorporate an adversarial loss to encourage robustness of DeepPrint to differences between acquisition devices. The adversarial loss \mathcal{L}_A is defined as the cross-entropy on the output of an adversary network $q(\cdot, \theta_A)$ across C classes of sensors, where the adversarial ground truth y' is assigned equal probabilities across these C classes (Eq. 9). The adversarial loss \mathcal{L}_A and identity loss \mathcal{L}_{ID} form the overall loss function \mathcal{L}_D used to train DeepPrint (Eq. 10), where $\lambda_4 = 0.1$ is empirically selected. The adversary network, $q(\cdot, \theta_A)$, is a two layer fully connected network, with weights θ_A , that predicts the probability of the class of input device used to capture each image, *i.e.*, minimizes the cross-entropy of the predicted device and the ground truth device label y (Eq. 11). Intuitively, if DeepPrint learns to fool the adversary, it has learned to encode identifying features which are independent of the acquisition device or camera.

TABLE III: Number of contactless and contact fingerprint images used in training each component of C2CL (# contactless / # contact).

Dataset	Segmentation	Deformation Correction & Scaling	DeepPrint
	$S(\cdot)$	$T(\cdot)$	$f(\cdot)$
UWA Benchmark 3D Fingerprint Database [22]	0/0	0/0	1,000/2,000
ManTech Phase2, 2015 [29]	0/0	0/0	21,352/28,574
PolyU Contactless 2D to Contact-based 2D Images Database [21]	0/0	1,920/1,920	1,920/1,920
MSU Finger Photo and Slap Fingerprint Database [20]	0/0	2,472/2,472	2,472/2,472
IIT Bombay Touchless and Touch-based Fingerprint Database [18]	0/0	800/800	800/800
ISPFDv2 [19]	496/0	8,400/1,200	8,400/1,200
ZJU Finger Photo and Touch-based Fingerprint Database	0/0	0/0	0/0
Total	496/0	13,592/6,392	35,944/36,966

$$\mathcal{L}_A(I, y') = - \sum_{c=1}^C y'_c \log q_A(y_c | f(I; w); \theta_A) \quad (9)$$

$$\begin{aligned} \mathcal{L}_D(I, y, H, y') = \operatorname{argmin}_w \sum_{i=1}^N & [\mathcal{L}_{ID}(I, y, H) \\ & + \lambda_4 \mathcal{L}_A(I^i, y'^i)] \end{aligned} \quad (10)$$

$$\mathcal{L}_C(I, y_c) = -y_c \log q_A(y_c | f(I; w); \theta_A) \quad (11)$$

In addition to the adversarial loss, we also increased the DeepPrint representation dimensionality from the original 192D to 512D and added perspective distortion and scaling augmentations during training. In an ablation study (Table VIII), we show how each of our DeepPrint modifications (fine-tuning, adversarial loss, perspective and scaling augmentations, and dimensionality change) improves the contact-contactless fingerprint matching performance.

2) *Minutiae Representation*: Finally, after extracting a textural representation with our modified DeepPrint network, we extract a minutiae-based representation from our preprocessed contactless fingerprints with the Verifinger 12.0 SDK.

C. Matching

Following feature extraction, from which we obtain texture representations (R_t^c, R_t^{cl}) and Verifinger minutiae representations (R_m^c, R_m^{cl}) for a given pair of contact and contactless fingerprint images (I_c, I_{cl}) , we compute a final match score as a weighted fusion of the individual scores computed between (R_t^c, R_t^{cl}) and (R_m^c, R_m^{cl}) . Concretely, let s_t denote the similarity score between (R_t^c, R_t^{cl}) and s_m denote the similarity score between (R_m^c, R_m^{cl}) , then the final similarity score is computed from a sum score fusion shown in Equation 12. For our implementation, $w_t = w_m = 0.5$ was selected empirically.

$$s = w_t s_t + w_m s_m \quad (12)$$

IV. EXPERIMENTS

In this section, we give details on various experimental evaluations to determine the effectiveness of C2CL for contact to contactless fingerprint matching. We employed various publicly available datasets for the evaluation of our algorithms, as well as a new database of contactless and corresponding contact-based fingerprints which was collected using our mobile-app in coordination with Zhejiang University (ZJU).

A. Datasets

Table II gives a detailed description of the publicly available datasets for contact to contactless matching used in this study and Figure 3 shows some example images from these datasets. For comparison with previous studies, we use the same train/test split of the PolyU dataset that was used in [23], which consists of 160 fingers for training with 12 impressions each and the remaining 160 fingers for testing with 6 impressions each. Similarly, we split the UWA Benchmark 3D dataset into 500 training fingers and 1,000 unique test fingers. Furthermore, following the protocol of Malhotra *et al.* [19], we split the ISPFDv2 dataset evenly into 50% train and 50% test subjects. Finally, we captured and sequestered a new dataset of contactless fingerprints and contact-based fingerprint images in coordination with ZJU for a cross-database evaluation (*e.g.*, not seen during training) to demonstrate generalizability of our algorithm. The cross-database evaluation is much more stringent than existing approaches which only train/test on different partitions of the same dataset. Indeed, the cross-database evaluation is a much better measure of how C2CL would perform in the real world.

The ZJU Finger Photo and Touch-based Fingerprint Database contains a total of 206 subjects, with 12 contactless images and 12 contact-based impressions per finger. The thumb and index fingers of both hands were collected for each subject, giving a total of 9,888 contactless and contact-based images each. The contactless images were captured using three commodity smartphones: HuaWei P20, Samsung s9+, and OnePlus 8, whereas the contact-based fingerprint impressions were captured on a URU 4500 optical-based scanner at 512 ppi. An Android fingerphoto capture app was developed to improve the ease and efficiency of the data collection. To initiate the capturing process, a user or operator enters the transaction ID for the user and uses an on screen viewing window to help guide and capture the fingerprint image. Furthermore, a counter displayed on the screen keeps track of subsequent captures to streamline the data collection process.

B. Implementation Details

All the deep learning components (segmentation network, deformation correction and scaling network, and DeepPrint) are implemented using the Tensorflow deep learning framework. Each network is trained independently and information regarding how many of the contactless and contact fingerprint

TABLE IV: Deformation Correction and Scaling Spatial Transformation Network Architecture, $T(\cdot)$.

Layer	#Filters, Filter Size, Stride	Output Dim.
0. Input	0, 0, 0	$480 \times 480 \times 1$
1. Convolution	$32, 3 \times 3, 2$	$240 \times 240 \times 32$
2. Convolution	$64, 3 \times 3, 2$	$120 \times 120 \times 64$
3. Convolution	$128, 3 \times 3, 2$	$60 \times 60 \times 128$
4. Convolution	$256, 3 \times 3, 2$	$30 \times 30 \times 256$
5. Max Pool	$256, 6 \times 4, 2$	$13 \times 14 \times 256$
6. Dense	—	1024
7. Dense	—	$2 \times n_0 + 4$

The final dense layer contains output neurons for a $2 \times n_0$ grid of $n_0 = n \times n$ pixel displacements and 4 neurons for the affine transformation matrix (s, θ, t_x and t_y). In our implementation, $n = 4$.

images from each of the datasets used in training each component of our algorithm is given in Table III.

1) *Segmentation Network*: A total of 496 contactless fingerprint images from the ISPFDv2 were manually labeled with segmentation masks outlining the distal phalange were used for training. Input images were down sampled to 256×256 during training to reduce the time to convergence, which occurred around 100,000 iterations using stochastic gradient descent (SGD) with a learning rate of $1e^{-3}$ and a batch size of 8 on a single NVIDIA GeForce RTX 2080 Ti GPU. During inference, the contactless fingerprint images are resized to 256×256 and resulting segmentation masks are upsampled back to the original resolution. Due to limited number of manually marked images, we employed random rotations, translations, and brightness augmentations to avoid over-fitting. Additionally, we incorporated random resizing of input training images within the range $[128 \times 128, 384 \times 384]$ to encourage robustness to varying resolution between capture devices.

2) *Deformation Correction and Scaling Network*: The pre-trained DeepPrint model in [28] was used to provide supervision of our spatial transformation network in line with Eq 4. The motivation for using a network pretrained on contact-based fingerprints, rather than our new finetuned model on contactless fingerprints, is that the goal of our transformation network is to transform the contactless fingerprint images to better resemble their contact-based counterparts. Thus, a supervisory network trained on solely contact-based fingerprint images is more suitable for this purpose. The architectural details of our STN localization network are given in Table IV. For our implementation, we set the number of sampling points for the distortion grid to $n = 4 \times 4$. Data augmentations of random rotations, translations, brightness adjustments, and perspective distortions were employed to avoid over-fitting. This network was trained for 25,000 iterations using an Adam optimizer with a learning rate of $1e^{-6}$ and a batch size of 16 on a single NVIDIA GeForce RTX 2080 Ti GPU.

3) *DeepPrint*: The DeepPrint network was trained on two NVIDIA GeForce RTX 2080 Ti GPUs with an RMSProp optimizer, learning rate of 0.01, and a batch size of 16. The added adversary network, which was trained in step with DeepPrint, also utilized an RMSProp optimizer with a learning rate of 0.01. A small validation set was partitioned from the DeepPrint

fine-tuning data outlined in Table III to stop the training (which occurred at 73,000 steps). Lastly, random rotation, translation, brightness, cropping, and perspective distortion augmentations were utilized during training.

C. Evaluation Protocol

To evaluate the cross-matching performance of our algorithms, we conduct both verification (1:1) and identification (1:N) experiments. For verification, we report the Receiver Operating Characteristic (ROC) curves at specific operating points and equal error rates (EER). Note that we report the True Acceptance Rate (TAR) at a False Acceptance Rate (FAR) of 0.01%, which is a stricter threshold than is currently reported in the literature, and which is also a threshold expected for field deployment. For the search experiments, the rank-one search accuracy is given against an augmented large scale gallery of 1.1 contact million fingerprints taken from an operational forensics database [2]. This is a much larger gallery than has previously been evaluated against in the literature and is again more indicative of what C2CL would face in the real world. Finally, we present ablation results on each significant component of our proposed system.

D. Verification Experiments

The verification experiments are conducted in a manner consistent with previous approaches to facilitate a fair comparison. In particular, (i) the PolyU testing dataset yields 5,760 ($160 \times 6 \times 6$) genuine scores and 915,840 ($160 \times 159 \times 6 \times 6$) imposter scores, (ii) the UWA Benchmark 3D dataset yields 8,000 ($1,000 \times 4 \times 2$) genuine and 7,992,000 ($1,000 \times 999 \times 4 \times 2$) imposter scores, (iii) the ISPFDv2 dataset (which is split into 7 different capture variations)⁶ yields 68,096 ($(152 \times 8 \times 8) \times 7$) genuine and 10,282,496 ($(152 \times 151 \times 8 \times 8) \times 7$) imposter scores, and (iv) the ZJU dataset yields 118,656 ($824 \times 12 \times 12$) genuine and 97,653,888 ($824 \times 823 \times 12 \times 12$) imposter scores. Due to the very high number of possible imposter scores for ZJU, we limit the number of imposter scores computed to only include the first impression of each imposter fingerprint. This process results in 678,152 imposter scores out of the possible 97,653,888 scores. It is assumed for all experiments that the contactless fingerprints and contact-based impressions are the probe and enrollment images, respectively.

Table V provides the Equal Error Rate (EER) and TAR @ FAR=0.01% of C2CL on the different datasets. For comparison with previous methods, rather than implement the relevant SOTA approaches that have been proposed and risk under representing those methods, we directly compare our approach to the results reported in each of the respective papers. In terms of EER, our method outperforms all the previous approaches in the verification setting. Not only does our individual performance of the minutiae and textural representations alone exceed that of the previous SOTA methods (in

⁶The 7 scenarios consist of different background, illumination, and resolution variations (e.g., white background & indoor lighting, white background & outdoor lighting, natural background & indoor lighting, natural background & outdoor lighting, 5MP resolution, 8MP resolution, and 16MP resolution. For our evaluation, we combine each of these into a single dataset.

TABLE V: Verification performance of C2CL.

Dataset	Verifinger 12.0		DeepPrint		DeepPrint + Verifinger 12.0		Previous SOTA
	EER (%)	TAR @ FAR=0.01%	EER (%)	TAR @ FAR=0.01%	EER (%)	TAR @ FAR=0.01%	EER (%)
PolyU	0.46	97.20	2.37	72.07	0.30	97.74	7.93 [23]
UWA	6.81	92.56	5.29	83.40	0.72	98.30	7.11 [23]
ISPFDv2	1.46	96.02	2.33	84.33	1.20	96.67	3.40 [‡] [19]
ZJU [†]	0.79	96.86	2.08	86.42	0.62	97.56	N/A

[‡] [19] reports results on the ISPFDv2 dataset per individual capture condition; 3.40 is the average EER across these data splits.

[†] Cross-database evaluation, *i.e.*, not seen during training.

TABLE VI: DeepPrint verification performance when finetuned on only PolyU.

Dataset	EER (%)	TAR (%) @ FAR=0.01%
PolyU	1.90	74.62
UWA	8.35	35.42
ISPFDv2	3.87	57.10
ZJU	2.99	68.99

particular, even if we remove Verifinger, we still beat SOTA in all cases), the fusion performance attains matching accuracy (EER = 0.30% – 1.20%), which is much closer to contact-contact fingerprint matching [26]. Even in the most challenging cross-database evaluation (ZJU), C2CL attains competitive performance with contact-contact matching - demonstrating the generalizability of C2CL to unseen datasets. Note that we report the TAR @ FAR=0.01% only for C2CL since most of the prior approaches only report EER and none report TAR @ FAR =0.01%.

Different from previous approaches, which train individual models on a train/test split for each evaluation dataset, we have trained just a single model for our evaluation across four different datasets. This protocol is actually more challenging than finetuning for each individual evaluation dataset. This is because despite having a smaller number of training samples, higher verification performance can more easily be achieved by individually trained models. To support this claim, we have finetuned an additional model on just the PolyU dataset using the same train/test split specified in [23] and recorded the verification performance in Table VI. We observe that our accuracy improves on PolyU from 2.37% EER for the model trained on our full combination of training datasets to 1.90% EER for the model trained on just PolyU; however, because of the lower performance on the other three datasets, we can see that this model is indeed over-fit to PolyU.

1) *Ablation study*: We present an ablation study (Table VII) to fully understand the contribution of the main components of our algorithm; namely, segmentation, enhancement, 500 ppi frequency scaling, and TPS deformation correction. From the ablation, we notice there is a substantial improvement in both EER and TAR @ FAR=0.01% just from incorporating proper enhancement of the contactless images. In most cases, there is almost a 50% reduction in EER from including both contrast enhancement and binary pixel inversion. For brevity, not shown in the table is the individual contribution of inverting the ridges of the contactless images aside from contrast enhancement. For reference, the EER of DeepPrint on ZJU warped images with only contrast enhancement is 2.49%.

This is in comparison to the EER of 2.08% on the warped images with both contrast enhancement and pixel inversion.

Furthermore, we observe that for the smartphone captured contactless fingerprints in the ISPFDv2 and ZJU datasets, there is a dramatic performance jump when incorporating our 500 ppi scaling network. Additionally, there is another noticeable improvement when incorporating the deformation correction branch of our STN, most notably for the ISPFDv2 dataset. Since the ZJU dataset contains equal numbers of thumb and index fingers, where the majority of our training datasets contain mostly non-thumb fingers, we observed that the deformation correction is less beneficial on average for the ZJU dataset compared to ISPFDv2. In fact, from Table IX, we see that the EER of just index fingers of ZJU is noticeably lower than the EER on thumbs.

To investigate whether the lower performance on thumbs is a limitation of the available training data or whether thumbs require a different distortion correction from non-thumbs, we retrained separate warping models on thumb data only and non-thumb data only. The test results for the ZJU dataset are in Table X. A couple of observations: i.) The performance (TAR @ FAR=0.01%) is highest for the model trained on both thumbs and non-thumbs, ii.) the model trained on non-thumbs performs slightly worse when applied to the test set of thumbs in the ZJU dataset, which indicates that the warping required for thumbs may be slightly different, and iii.) the performance of the thumb only model decreases on both thumbs and non-thumbs due to the limited number of thumb training examples.

2) *Multi-finger fusion verification*: The final set of verification experiments is to investigate the effects of finger position and multiple finger fusion in the verification accuracy for the ZJU dataset. Table IX, shows the individual performance per finger position and the fusion of multiple fingers; namely, thumb only, index only, fusion of right thumb and right index, fusion of left thumb and left index, and four finger fusion. The motivation for considering fusion of the thumb and index on each hand is that from a usability standpoint, a user may be able to use their dominant hand when capturing their own fingerprints. Notably, when fusing multiple fingers (*e.g.*, right index and left index), we obtain nearly perfect accuracy.

E. Search Experiments

For the identification (or search) experiments, we utilize the first impressions of both the contactless and contact-based fingerprints of the ZJU dataset. The contact-based fingerprints are placed in the gallery which is augmented with 1.1 million fingerprint images from an operational forensic database [2].

TABLE VII: Ablation study of C2CL using only Verifinger 12.0 for matching*. S = segmentation, E = enhancement, T_s = scaling, T_d = deformation correction.

Dataset	Modules				Overall (%)	
	S	E	T_s	T_d	EER	TAR @ FAR = 0.01%
PolyU	✓				0.86	93.19
	✓	✓			0.45	96.96
	✓	✓	✓		0.48	96.44
	✓	✓	✓	✓	0.46	97.20
UWA [‡]	✓				6.62	91.05
	✓	✓			6.81	92.56
ISPFDv2	✓				13.76	23.93
	✓	✓			7.83	38.53
	✓	✓	✓		2.02	93.3
	✓	✓	✓	✓	1.46	96.02
ZJU [†]	✓				3.35	82.8
	✓	✓			1.88	89.9
	✓	✓	✓		0.9	96.97
	✓	✓	✓	✓	0.79	96.86

* Ablation results for DeepPrint are not shown since only a single model was trained on the final $E+S+T_s+T_d$ images.

† We do not apply our STN here since these images are captured with a 3D scanner and are already unrolled and at a resolution of 500 ppi.

† Cross-database evaluation, *i.e.*, not seen during training.

TABLE VIII: DeepPrint Ablation Study

Method	ZJU EER (%)
DeepPrint [28]	4.07
+ finetune	2.68
+ 512D	2.64
+ Augmentations	2.35
+ Adversarial Loss	2.08

* Each row adds on to the previous row.

The contactless fingerprint images serve as the probes. We note that our 1.1 million augmented gallery is significantly larger than any of the existing galleries used to evaluate contact-contactless fingerprint search, and is more indicative of the real world use-case of cross fingerprint matching (*e.g.*, in a National ID system like Aadhaar where a large gallery of contact-based fingerprints is already enrolled and used for duplication).

We evaluate 3 different search algorithms on the ZJU augmented gallery: (i) Verifinger 1:N search, (ii) search via our DeepPrint texture matcher (scores s_t from Eq. 12 are computed between a given preprocessed, contactless probe, and all 1.1 million contact-based fingerprints in the gallery), and (iii) a two-stage search algorithm [28] where the DeepPrint texture scores are first used to retrieve the top-500 candidates, followed by a reordering using the 1:1 minutiae matching scores (s_m from Eq. 12) from Verifinger. The advantage of the two-stage search scheme is that it balances both speed and accuracy by utilizing the matching speed of DeepPrint to locate the first list of 500 candidates and the accuracy of Verifinger to further refine this list.

From Table XII, we observe that Verifinger outperforms

TABLE IX: Multi-finger fusion verification results of the proposed matcher on the ZJU dataset.

Finger Type	EER (%)	TAR (%) @ FAR = 0.01%
Thumb	0.95	95.89
Index	0.48	98.31
LT + LI	0.00	99.77
RT + RI	0.00	99.74
RT + LT	0.00	99.80
RI + LI	0.00	99.89

TABLE X: Applying separate warping modules for thumbs vs. non-thumbs (TAR (%)) @ FAR=0.01%.

Method	Trained on thumbs and non-thumbs	Trained on non-thumbs	Trained on thumbs
ZJU non-thumbs	92.22	92.24	91.66
ZJU thumbs	80.66	77.50	76.80
ZJU all	86.42	84.80	84.46

DeepPrint stand-alone but at a search time against 1.1 million that is quite slow in comparison to DeepPrint. This motivates combining both approaches into the aforementioned two-stage search algorithm which outperforms Verifinger at Rank-1 and reduces the search time by 50 seconds. In short, our two stage search algorithm obtains high levels of search accuracy on a large-scale gallery at a significant search time savings.

F. Segmentation Evaluation

A successful segmentation algorithm for contactless fingerprint images must not only reliably detect the distal phalange of the contactless fingerprint, but also be robust to varying illumination, background, and resolution that is expected to occur in highly unconstrained capture environments. The method by Malhotra *et al.* [19] performed well on the ISDFPDv2 dataset using certain hyperparameters that were fit to this particular dataset; however, the authors did not evaluate it on unseen datasets. In contrast, our algorithm requires no hyperparameter tuning and still performs well across a variety of different evaluation datasets, both seen and unseen. Table XIII gives a comparison on the unseen ZJU dataset between our method and our implementation of the baseline approach of Malhotra *et al.*, which was trained on the ISPFDv2 dataset. For this evaluation, we manually marked the first contactless fingerprint image of each unique finger in the ZJU dataset with ground truth segmentation masks of the distal phalange,

TABLE XI: Improvement in minutiae correspondence without and with warping correction on ZJU dataset.

	Avg. Number of Paired Minutiae	Avg. Number of Missing Minutiae	Avg. Number of Spurious Minutiae	Goodness Index [48]
Without warping	28.06	67.95	71.13	-0.0167
With warping	30.11	65.00	69.20	-0.0157

TABLE XII: Search performance of the proposed matcher on the ZJU dataset with a gallery of 1.1 million.

Method	Rank (%)				Search Time (s)
	1	10	100	500	
DeepPrint	83.56	93.06	95.86	97.08	0.4
Verifinger 12.0	95.25	96.47	96.95	97.20	60.1
DeepPrint + Verifinger 12.0	95.49	96.10	96.95	97.08	10.5

DeepPrint + Verifinger 12.0 refers to indexing the top-500 candidates with DeepPrint and then re-sorting those 500 candidates using a fusion of the Verifinger and DeepPrint score.

TABLE XIII: Intersection Over Union (IOU) for segmentation $S(\cdot)$.

Method	IOU
Baseline [19]	0.747
Proposed	0.899

and then computed the Intersection Over Union (IOU) metric between the predicted segmentation masks of our algorithm and our implementation of the benchmark algorithm in [19]. Our method does not require any hyper-parameter tuning and still achieves higher IOU compared to [19].

A qualitative analysis of our segmentation network (see Figure 5) shows our algorithm is robust to varying illumination, background, and resolution and generalizes across multiple datasets of contactless fingerprints. However, as seen in Figure 5 (b), the network may still fail in extremely challenging background and illumination settings. An additional consideration, which is of importance for real-time deployment, is the processing speed of the segmentation network. Our segmentation algorithm is extremely fast compared to existing methods - requiring just 12.6ms to segment a (900×1200) resolution image. In contrast, our parallel implementation of the baseline approach of Malhotra *et al.* requires 3s per image.

V. DISCUSSION

Despite the low error rates achieved across each dataset, there are many factors that complicate the cross-matching performance and lead to both type I (false rejects) and type II (false accepts) errors. Many of the type I and type II errors are attributed to a failure to correctly segment and scale only the distal phalange of the input contactless fingerprint. Incorrect segmentation can lead the large amounts of the image containing background rather than the relevant fingerprint region. Other errors can be attributed to the inherent low-contrast of the contactless fingerprints, despite any effort of contrast or resolution enhancement. The only way to mitigate these types of failures is to include a quality assurance algorithm at the point of capture of the contactless fingerprint images. Lastly, minimal overlap in the fingerprint ridge structure between genuine probe and gallery fingerprint images is the cause of many false rejections, whereas very similar ridge structure between imposter fingerprint pairs leads to a number of false accepts. This challenge is present in contact-contact matching; however, is exaggerated in C2CL because of the unconstrained pose variance of the finger in 3D space.

The potential for greater variance in the capture conditions when capturing contactless fingerprint images necessitates more robust preprocessing to reliably match contactless fingerprints. Thus, performance will likely be markedly lower in unconstrained scenarios compared to highly controlled capture environments that employ dedicated hardware for the image acquisition, such as the PolyU and UWA datasets. However, C2CL has pushed the SOTA forward both in matching more unconstrained fingerphotos and the more constrained dedicated-device captured contactless fingerprints. Additionally, one might consider acquiring multiple image views of the same finger to build a complete 3D model of the finger to guide the preprocessing stage; however, this would add additional computational costs and latency to the acquisition process. Furthermore, in some capture scenarios, certainly the setup employed by our capture app, this process may be ergonomically challenging for the user.

As highlighted in the ablation study of Table V, most of the improvement in interoperability between contactless and contact-based fingerprints is due to appropriate 500 ppi scaling of the contactless prints; however, incorporating a deformation correction module is also shown, with statistical significance⁷, to further improve the compatibility. Figure 7 aims to highlight this fact through an overlay of the fingerprint ridge structure of one pair of corresponding contact and contactless fingerprints before and after applying the deformation correction. Additionally, Table XI shows the average number of paired minutiae, missing minutiae, spurious minutiae, and Goodness Index [48] without and with the warping correction on the ZJU dataset. The GI, ranging from -1 to 3, is a combined measure of paired, missing, and spurious minutiae. The warping module improved the GI by 5.99%. Thus, the improved alignment indubitably leads to better minutiae-based and texture-based matching, as verified by our experiments.

Lastly, in order to utilize a large CNN, such as DeepPrint, for the task of contact-contactless fingerprint matching, we leveraged a large dataset [2] from a related domain of contact-contact fingerprint matching to pretrain our DeepPrint network. Since this dataset is not currently publicly available, we have repeated the verification experiments when pretraining DeepPrint on the publicly available NIST N2N dataset [50] (see Table XIV). Due to the smaller dataset, we experience a slight degradation in the DeepPrint performance on some of the evaluation datasets; however, further data augmentation and incorporation of other publicly available datasets can be used to improve the performance.

VI. COMPUTATIONAL EFFICIENCY

Our system architecture consists of a variety of deep networks (segmentation network, deformation correction and scaling STN, DeepPrint CNN feature extractor) and a minutiae feature extractor. The inference speeds of the segmentation network, STN, and DeepPrint are approximately 12.6ms,

⁷The Mann-Whitney rank test [49] was used to compute the statistical significance between the ROC curves of $S+E+T_s$ and $S+E+T_s+T_d$. For all four datasets, the p value is smaller than 0.05, indicating that the difference is statistically significant to reject the hypothesis that the two curves are similar with a confidence of 95%.

TABLE XIV: DeepPrint performance pretrained with NIST N2N [50] dataset vs. Longitudinal dataset referenced in Yoon and Jain [2].

Dataset	Pretrained on NIST N2N Dataset (publicly available)		Pretrained on Longitudinal Dataset	
	EER (%)	TAR (%) @ FAR = 0.01%	EER (%)	TAR (%) @ FAR = 0.01%
PolyU	2.04	71.30	2.37	72.07
UWA	5.62	56.99	5.29	83.40
ISPFDv2	2.60	81.83	2.33	84.33
ZJU [†]	3.09	77.92	2.08	86.42

[†] Cross-database evaluation, *i.e.*, not seen during training.

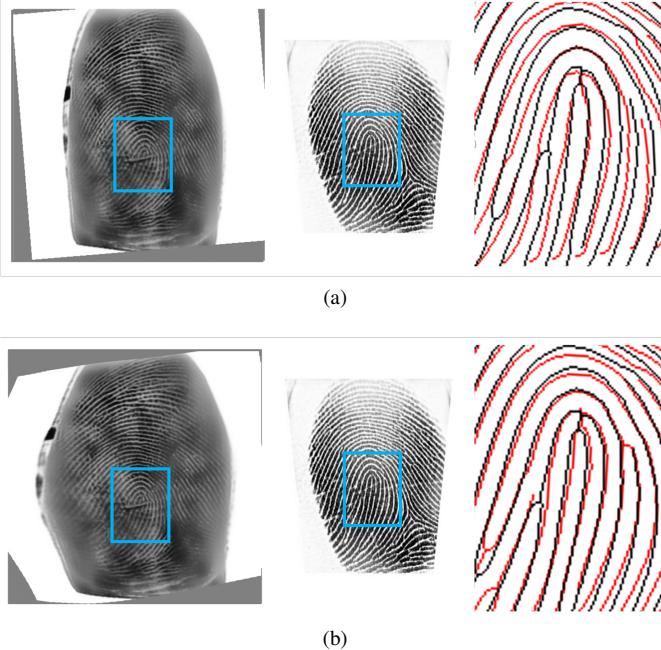


Fig. 7: Comparison of ridge overlap (a) with and (b) without the unwarping module. Use of the unwarping module results in better ridge alignment between contactless and contact-based images.

6.2ms, and 26.3ms using a single NVIDIA GeForce RTX 2080 Ti GPU and 143.8ms, 19.5ms, and 120.2ms on an Intel Core i7-8700X CPU @ 3.70GHz, respectively. The Verifinger 12.0 feature extractor requires 600ms on an Intel Core i7-8700X. In total, the inference speed of the end-to-end network is \approx 643.6ms with an NVIDIA GeForce RTX 2080 Ti GPU or \approx 883.5ms on an Intel Core i7-8700X CPU.

The deep network components of our algorithm are capable of very fast inference per input image; however, the system as a whole consumes a large amount of memory (400 MB). To fit into a resource constrained environment, such as a mobile phone, further optimization to the system architecture can easily be implemented with very little, if any, performance drop. First, the intermediate step of generating a scaled image prior to the deformation correction is not required for deployment and was just included for the ablation study. Instead, we can remove the affine transformation layer of our STN and directly scale and warp the input images in one step. As it stands, the main components of the algorithm, DeepPrint and Verifinger, require \approx 1s and \approx 1.2s on a mobile phone (Google Pixel 2), respectively. Thus, the inference time is estimated to be \approx 2 seconds. However, to further boost the speed,

rather than rely on a COTS system for minutiae extraction and matching, we can directly use the minutiae sets output by DeepPrint and a computationally efficient minutiae matcher to obtain the minutiae match scores, such as MSU’s Latent AFIS Matcher [51]. Porting the model with these optimizations to a mobile phone remains as a point of future work.

VII. CONCLUSION AND FUTURE WORK

We have presented an end-to-end system for matching contactless fingerprints (*i.e.*, finger photos) to contact-based fingerprint impressions that significantly pushes the SOTA in contact-contactless fingerprint matching closer to contact-contact fingerprint matching. In particular, our contact to contactless matcher achieves less than 1% EER across multiple datasets employing a variety of contactless and contact-based acquisition devices with varying background, illumination, and resolution settings. Critical to the success of our system is our extensive preprocessing pipeline consisting of segmentation, contrast enhancement, 500 ppi scale normalization, deformation correction, and our adaptation of DeepPrint for contact-contactless matching. Our cross-database evaluations and large-scale search experiments are more rigorous evaluations than what is reported in the open literature, and it enables us to confidently demonstrate a step toward a contact-contactless fingerprint matcher that is comparable to SOTA contact-contact fingerprint matching accuracy.

ACKNOWLEDGMENT

This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation under Grant No. 1841517. The authors would like to thank Debayan Deb for his help in developing the mobile phone contactless fingerprint capture application, Dr. Eryun Liu’s research group at Zhejiang University for overseeing the data collection effort for this project, and the various research groups who have shared their datasets that were used in this study.

REFERENCES

- [1] S. Pankanti, S. Prabhakar, and A. K. Jain, “On the individuality of fingerprints,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1010–1025, 2002.
- [2] S. Yoon and A. K. Jain, “Longitudinal study of fingerprint recognition,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 28, pp. 8555–8560, 2015.
- [3] A. M. Bazen and S. H. Gerez, “Fingerprint matching by thin-plate spline modelling of elastic deformations,” *Pattern Recognition*, vol. 36, no. 8, pp. 1859–1867, 2003.

[4] R. Cappelli, D. Maio, and D. Maltoni, "Modelling plastic distortion in fingerprint images," in *International Conference on Advances in Pattern Recognition*. Springer, 2001, pp. 371–378.

[5] A. Ross, S. Dass, and A. Jain, "A deformable model for fingerprint matching," *Pattern Recognition*, vol. 38, no. 1, pp. 95–103, 2005.

[6] G. Parziale, "Touchless fingerprinting technology," in *Advances in Biometrics*. Springer, 2008, pp. 25–48.

[7] K. Okerefor, I. Ekong, I. O. Markson, and K. Enwere, "Fingerprint biometric system hygiene and the risk of covid-19 transmission," *JMIR Biomedical Engineering*, vol. 5, no. 1, p. e19623, 2020.

[8] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, and C. Busch, "Mobile touchless fingerprint recognition: Implementation, performance and usability aspects," *arXiv preprint arXiv:2103.03038*, 2021.

[9] Y. Song, C. Lee, and J. Kim, "A new scheme for touchless fingerprint recognition system," in *Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, (ISPACS)*. IEEE, 2004, pp. 524–527.

[10] C. Lee, S. Lee, and J. Kim, "A study of touchless fingerprint recognition system," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2006, pp. 358–365.

[11] G. Parziale and Y. Chen, "Advanced technologies for touchless fingerprint recognition," in *Handbook of Remote Biometrics*. Springer, 2009, pp. 83–109.

[12] B. Y. Hiew, A. B. Teoh, and Y.-H. Pang, "Touch-less fingerprint recognition system," in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*. IEEE, 2007, pp. 24–29.

[13] A. Kumar and Y. Zhou, "Contactless fingerprint identification using level zero features," in *CVPR 2011 Workshops*. IEEE, 2011, pp. 114–119.

[14] X. Yin, Y. Zhu, and J. Hu, "Contactless fingerprint recognition based on global minutia topology and loose genetic algorithm," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 28–41, 2019.

[15] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh, "On smartphone camera based fingerphoto authentication," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2015, pp. 1–7.

[16] A. Malhotra, A. Sankaran, A. Mittal, M. Vatsa, and R. Singh, "Fingerphoto authentication using smartphone camera captured under varying environmental conditions," in *Human Recognition in Unconstrained Environments*. Elsevier, 2017, pp. 119–144.

[17] C. Stein, C. Nickel, and C. Busch, "Fingerphoto recognition with smartphone cameras," in *International Conference of Biometrics Special Interest Group (BIOSIG)*. IEEE, 2012, pp. 1–12.

[18] P. Birajdar, M. Haria, P. Kulkarni, S. Gupta, P. Joshi, B. Singh, and V. Gadre, "Towards smartphone-based touchless fingerprint recognition," *Sādhanā*, vol. 44, no. 7, pp. 1–15, 2019.

[19] A. Malhotra, A. Sankaran, M. Vatsa, and R. Singh, "On matching finger-selfies using deep scattering networks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 4, pp. 350–362, 2020.

[20] D. Deb, T. Chugh, J. Engelsma, K. Cao, N. Nain, J. Kendall, and A. K. Jain, "Matching fingerphotos to slap fingerprint images," *arXiv preprint arXiv:1804.08122*, 2018.

[21] C. Lin and A. Kumar, "Matching contactless and contact-based conventional fingerprint images for biometrics identification," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2008–2021, 2018.

[22] W. Zhou, J. Hu, I. Petersen, S. Wang, and M. Bennamoun, "A benchmark 3d fingerprint database," in *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2014, pp. 935–940.

[23] C. Lin and A. Kumar, "A cnn-based framework for comparison of contactless to contact-based fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 662–676, 2019.

[24] P. Wild, F. Daubner, H. Penz, and G. F. Domínguez, "Comparative test of smartphone finger photo vs. touch-based cross-sensor fingerprint recognition," in *2019 7th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2019, pp. 1–6.

[25] A. Dabouei, S. Soleymani, J. Dawson, and N. M. Nasrabadi, "Deep contactless fingerprint unwarping," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–8.

[26] B. Dorizzi, R. Cappelli, M. Ferrara, D. Maio, D. Maltoni, N. Houmani, S. Garcia-Salicetti, and A. Mayoue, "Fingerprint and on-line signature verification competitions at ICB 2009," in *International Conference on Biometrics*. Springer, 2009, pp. 725–732.

[27] C. I. Watson, G. P. Fiumara, E. Tabassi, S. L. Cheng, P. A. Flanagan, and W. J. Salamon, "Fingerprint vendor technology evaluation," 2015.

[28] J. J. Engelsma, K. Cao, and A. K. Jain, "Learning a fixed-length fingerprint representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[29] L. Ericson and S. Shine, "Evaluation of contactless versus contact fingerprint data phase 2 (version 1.1)," DOJ Office Justice Programs, I. ManTech Adv. Syst. Int., Fairmont, WV. Tech. Rep. 249552, 2015.

[30] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *arXiv preprint arXiv:1506.02025*, 2015.

[31] A. Dabouei, H. Kazemi, S. M. Iranmanesh, J. Dawson, and N. M. Nasrabadi, "Fingerprint distortion rectification using deep convolutional neural networks," in *2018 International Conference on Biometrics (ICB)*. IEEE, 2018, pp. 1–8.

[32] A. Ross, S. C. Dass, and A. K. Jain, "Fingerprint warping using ridge curve correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 19–30, 2005.

[33] A. W. Senior and R. M. Bolle, "Improved fingerprint matching by distortion removal," *IEICE Transactions on Information and Systems*, vol. 84, no. 7, pp. 825–832, 2001.

[34] X. Si, J. Feng, J. Zhou, and Y. Luo, "Detection and rectification of distorted fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 555–568, 2015.

[35] C. Lin and A. Kumar, "Contactless and partial 3d fingerprint recognition using multi-view deep representation," *Pattern Recognition*, vol. 83, pp. 314–327, 2018.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[37] K. Wada, "labelme: Image Polygonal Annotation with Python," <https://github.com/wkentaro/labelme>, 2016.

[38] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[39] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 6, pp. 1206–1213, 2016.

[40] L. Lugini, E. Marasco, B. Cukic, and I. Gashi, "Interoperability in fingerprint recognition: A large-scale empirical study," in *43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W)*, 2013, pp. 1–6.

[41] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," in *Biometric Authentication*, D. Maltoni and A. K. Jain, Eds. Springer, 2004, pp. 134–145.

[42] F. Alonso-Fernandez, R. N. Veldhuis, A. M. Bazen, J. Fierrez-Aguilar, and J. Ortega-Garcia, "Sensor interoperability and fusion in fingerprint verification: A case study using minutiae-and ridge-based matchers," in *International Conference on Control, Automation, Robotics and Vision*. IEEE, 2006, pp. 1–6.

[43] H. AlShehri, M. Hussain, H. AboAlSamh, and M. AlZuair, "A large-scale study of fingerprint matching systems for sensor interoperability problem," *Sensors*, vol. 18, no. 4, p. 1008, 2018.

[44] E. Marasco, L. Lugini, B. Cukic, and T. Bourlai, "Minimizing the impact of low interoperability between optical fingerprints sensors," in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, 2013, pp. 1–8.

[45] J. Jang, S. J. Elliott, and H. Kim, "On improving interoperability of fingerprint recognition using resolution compensation based on sensor evaluation," in *Advances in Biometrics*, S.-W. Lee and S. Z. Li, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 455–463.

[46] A. Ross and R. Nadgir, "A calibration model for fingerprint sensor interoperability," in *Biometric Technology for Human Identification III*, vol. 6202. International Society for Optics and Photonics, 2006, p. 62020B.

[47] S. A. Grosz, T. Chugh, and A. K. Jain, "Fingerprint presentation attack detection: A sensor and material agnostic approach," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2020, pp. 1–10.

[48] N. K. Ratha, S. Chen, and A. K. Jain, "Adaptive flow orientation-based feature extraction in fingerprint images," *Pattern Recognition*, vol. 28, no. 11, pp. 1657–1672, 1995.

[49] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, pp. 837–845, 1988.

[50] G. Fiumara, P. Flanagan, M. Schwarz, E. Tabassi, and C. Boehnen, "Nist special database 301," 2018.

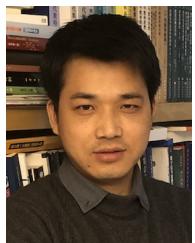
[51] K. Cao, D.-L. Nguyen, C. Tymoszek, and A. K. Jain, “End-to-end latent fingerprint search,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 880–894, 2019.



Steven A. Grosz received his B.S. degree with highest honors in Electrical Engineering from Michigan State University, East Lansing, Michigan, in 2019. He is currently a doctoral student in the Department of Computer Science and Engineering at Michigan State University. His primary research interests are in the areas of machine learning and computer vision with applications in biometrics.



Joshua J. Engelsma graduated magna cum laude with a B.S. degree in computer science from Grand Valley State University, Allendale, Michigan, in 2016. He is currently working towards a PhD degree in the Department of Computer Science and Engineering at Michigan State University, East Lansing, Michigan. His research interests include pattern recognition, computer vision, and image processing with applications in biometrics. He won best paper award at the 2019 IEEE ICB, and the 2020 MSU College of Engineering Fitch Beach Award.



Eryun Liu received his Bachelor degree in electronic information science and technology from Xidian University, Xi'an, Shaanxi China, in 2006 and Ph.D. degree in Pattern Recognition and Intelligence System from the same university in 2011. He was affiliated with Xidian University as an assistant professor from 2011 to 2013. He was a Post Doctoral Fellow in the Department of Computer Science & Engineering, Michigan State University, East Lansing, USA. He is currently affiliated with the Department of Information Science & Electronic

Engineering (ISEE), Zhejiang University, Hangzhou, China, as an associate professor. His research interests include biometric recognition, point pattern matching and information retrieval, with a focus on fingerprint and palmprint recognition.



Anil K. Jain is a University distinguished professor in the Department of Computer Science and Engineering at Michigan State University. He served as the editor-in-chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence and was a member of the United States Defense Science Board. He has received Fulbright, Guggenheim, Alexander von Humboldt, and IAPR King Sun Fu awards. He is a member of the National Academy of Engineering and foreign fellow of the Indian National Academy of Engineering and Chinese Academy of Sciences.