Criminal and victim identification based on deep and large feature sets from hand biometrics

Wojciech Michal Matkowski

2019

Wojciech Michal Matkowski. (2019). Criminal and victim identification based on deep and large feature sets from hand biometrics. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/137049>

<https://doi.org/10.32657/10356/137049>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 18:48:28 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**CRIMINAL AND VICTIM IDENTIFICATION BASED
ON DEEP AND LARGE FEATURE SETS FROM HAND
BIOMETRICS**

WOJCIECH MICHAL MATKOWSKI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

2019

CRIMINAL AND VICTIM IDENTIFICATION BASED ON DEEP AND LARGE FEATURE SETS FROM HAND BIOMETRICS

by

WOJCIECH MICHAL MATKOWSKI

Supervisor: Prof. Adams Wai Kin Kong

A thesis submitted to the Nanyang Technological University

in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

School of Computer Science and Engineering

Nanyang Technological University

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

29 Aug 2019

.....
Date

Matkowski

.....
Wojciech Michal Matkowski

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

24/Aug/2019
Date

Adams Wai Kin Kong
Adams Wai Kin Kong

Authorship Attribution Statement

This thesis contains material from 3 papers published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Chapter 3 is published as Wojciech Michal Matkowski, Frodo Kin Sun Chan and Adams Wai Kin Kong. A Study on Wrist Identification for Forensic Applications. *Image and Vision Computing*, vol. 88, August 2019, pp 96-112. <https://doi.org/10.1016/j.imavis.2019.05.005>

The contributions of the co-authors are as follows:

- Prof Adams Wai Kin Kong provided initial project ideas, provided collected images and supervised the project and helped to edit the manuscript.
- Frodo Kin Sun Chan shared his codes and edited the draft of the manuscript.
- I developed and implemented the wrist identification algorithm.
- I prepared and annotated the wrist image database based on the collected images provided by Prof Adams Wai Kin Kong.
- I collected the wrist images from the Internet.
- I wrote the drafts, revised the manuscript, wrote the final manuscript and prepared all the figures.
- I performed all the experiments.
-

Chapter 4 is published as Wojciech Michal Matkowski, Tingting Chai and Adams Wai Kin Kong. Palmprint Recognition in Uncontrolled and Uncooperative Environment. *IEEE Transactions on Information Forensics and Security*, 2019 (Early Access). 10.1109/TIFS.2019.2945183

The contributions of the co-authors are as follows:

- Prof Adams Wai Kin Kong provided initial project ideas, supervised the project and helped to edit the manuscript.
- Tingting Chai collected, developed and cleaned part of NTU Palmprints from the Internet database.

- Tingting Chai helped to collect codes of other palmprint recognition methods used in the comparison and run some of them.
- Tingting Chai extracted and provided ROI palmprint images from CASIA palmprint database.
- I proposed and implemented the algorithm.
- I performed the experiments.
- I checked the databases and annotated hand landmarks in NTU Palmprints from the Internet database.
- I wrote the draft, the final version of the manuscript and prepared all figures.

Chapter 5 is accepted at a conference as Wojciech Michal Matkowski, Krzysztof Matkowski, Adams Wai Kin Kong and Cory Lloyd Hall. The Nipple-Areola Complex for Criminal Identification. In *International Conference on Biometrics (ICB)*, June 2019.

The contributions of the co-authors are as follows:

- Prof Adams Wai Kin Kong provided initial project ideas, provided collected images.
- Krzysztof Matkowski prepared and drafted the review of related nipple-areola complex studies.
- Cory Lloyd Hall edited the drafted manuscript and provided comments.
- I prepared and implemented the experimental evaluation.
- I prepared and annotated the nipple-areola complex image database based on the collected images provided by Prof Adams Wai Kin Kong.
- I wrote the draft and the final manuscript and prepared all the figures.
- I performed all the experiments.

29 Aug 2019
.....
Date

Matkowski
.....
Wojciech Michal Matkowski

Acknowledgements

It has been over four years since I started my life in Singapore, seven hours ahead in time and almost ten thousand kilometers to the East in the distance from my hometown. Life full of images, coding, papers, smaller and bigger ups and downs. Here, I would like to express my gratitude to the people who have been around during this time.

I would like to thank my supervisor, Professor Adams Wai Kin Kong, for his great guidance. It is hard to imagine a better supervisor. He gave me lot of inspiration and freedom to keep my mind open for a broad spectrum of ideas. He showed me what it means to do meaningful research. It definitely, gave me the necessary motivation for the past four years.

I am grateful to my parents, brother and grandparents for their support. It was more than helpful to Skype with them, visit them two times in a year, have good time and food together. I also thank my parents for visiting me here in Singapore. It was like home sweet home coming here.

I would like to thank my friends who visited me here in Singapore. It felt like the good, old times arriving here, bringing some necessary balance and freshness.

I am thankful to my friends, which I met in Singapore, for parties, trips, lunches, dinners, sports, evening beers, coffees, conversations, board games, BBQ sessions, random meetings on campus or bus.

I am really grateful to Soohyun for her patience, support, food and always positive attitude.

Contents

Abstract	xvi
1 Introduction	1
1.1 Biometrics	1
1.2 Motivation	6
1.3 Research Objectives	9
1.4 Research Contributions	9
1.5 Organization of the Thesis	10
2 Background and Related Work	11
2.1 Hand Biometrics	11
2.1.1 Overview	11
2.1.2 Geometry and Shape	13
2.1.3 Vein	15
2.1.4 Dorsal Surface	17
2.1.5 Skin Texture	19
2.1.6 Palmprint	20
2.2 Palmprint Recognition Studies	23
2.2.1 ROI Extraction	23
2.2.2 Feature Extraction and Comparison	25
2.2.3 Coding-based Methods	26

2.2.4	Local Descriptors Methods	27
2.2.5	Convolutional Neural Networks	28
2.2.6	Experiments and Results in Palmprint Recognition Studies	31
2.2.6.1	Performance Evaluation	31
2.2.6.2	Baseline Selection	33
2.2.6.3	Current State	34
3	Wrist Identification for Forensic Application	36
3.1	Overview	36
3.2	NTU Wrist Image Database	37
3.3	The Proposed Wrist Identification Algorithm	42
3.3.1	Wrist Segmentation	42
3.3.2	ROI Extraction	43
3.3.3	Feature Extraction	51
3.3.4	Wrist Matching	52
3.3.5	Meta-Recognition	54
3.4	Experiments	55
3.4.1	Recognition Methods	55
3.4.2	Implementation Details	59
3.4.3	Computational Cost	61
3.4.4	Evaluation of the Proposed Wrist Identification Algorithm	61
3.4.5	Dimension Reduction and Feature Selection	68
3.4.6	Deep vs Hand Crafted Features	69
3.5	Summary	72
4	Palmprint Recognition in Uncontrolled and Uncooperative Environment	79
4.1	Overview	79
4.2	Databases for Palmprint Recognition	81

4.2.1	Selected Contactless Palmprint Databases	81
4.2.2	NTU Palmprint Databases	82
4.3	The Architecture Design	85
4.3.1	Hand Landmarks	86
4.3.2	Spatial Transformer	87
4.3.3	A Building Block for Palmprint Alignment and Feature Extraction	88
4.3.4	The Proposed Architecture	90
4.4	Experiments	92
4.4.1	Evaluation Protocols	92
4.4.2	Implementation Details	93
4.4.3	Training Strategy	95
4.4.4	Localization Network Training	96
4.4.5	Experiments on NTU Palmprints from the Internet	97
4.4.6	Experiments on CASIA, IITD, and PolyU Contactless Palmprint Benchmark Databases	98
4.4.7	Comparison with the State-of-the-Art Palmprint Recognition Meth- ods	101
4.4.8	Performance Analysis	104
4.5	Summary	107
5	The Nipple-Areola Complex for Offender Identification	113
5.1	Overview	113
5.2	Nipple-Areola Complex	114
5.3	NTU-Nipple-v1 dataset	115
5.4	Experiments	117
5.4.1	Evaluation Protocol	117
5.4.2	Recognition Algorithms	117
5.4.3	Implementation Details	118

5.4.4	Deep features	118
5.4.5	Fine-tuning the last layer	119
5.4.6	Fine-tuning fully connected layers of AlexNet	120
5.4.7	Modified AlexNet vs hand-crafted features	121
5.5	Summary	122
6	Discussion	125
6.1	Summary of Main Contributions	125
6.2	Accuracy	127
6.3	Database	127
6.3.1	Size	127
6.3.2	Collection	129
6.4	Application	130
6.5	Future Work	130
	Bibliography	133

List of Tables

2.1	The highest rank-1 accuracy (%) and the lowest EER (%) from different studies on four contactless palmprint databases.	35
2.2	Rank-1 accuracy (%) reported in two studies for four overlapping methods and the proposed by the authors methods on HFUT contactless palmprint database.	35
3.1	Details of the datasets in the NTU-Wrist-Image-Database-v1.	41
3.2	Number of images for one wrist in SET1, SET2 and SET3.	41
3.3	Recognition and Meta recognition systems details.	56
3.4	Details of three different experiments.	59
3.5	Identification accuracy (%) of the proposed WMM and WMFA algorithms and selected palmprint recognition methods.	67
3.6	Identification accuracy (%) differences between PCA compressed features and the original high-dimensional features.	69
3.7	Identification accuracy (%) differences between features selected by mRMR and the original high-dimensional features.	69
3.8	Identification accuracy (%) of AlexNet and VGG-16.	72
3.9	Identification accuracy (%) - deep and hand crafted features comparison. .	73
4.1	The Details and comparison of benchmark contactless palmprint databases and NTU Palmprint Databases.	85

4.2	Rank-1 and Rank-30 accuracy (%) of different training strategies S and classifiers on NTU Palmprints from the Internet database.	100
4.3	Rank-1 and Rank-30 accuracy (%) of the selected training strategies S and classifiers on IITD, PolyU and CASIA contactless palmprint databases. .	101
4.4	Rank-1 and Rank-30 accuracy (%) of state-of-the-art palmprint and deep learning recognition methods on NTU-PI-v1 database.	105
4.5	Rank-1 accuracy (%) and EER (%) of state-of-the-art palmprint and deep learning recognition methods on NTU-CP-v1, IITD, CASIA, PolyU databases.	106
5.1	The rank accuracies (%) of the deep learning architectures using softmax and SVM for identification.	120
5.2	The rank accuracies (%) of the AlexNet architectures and the hand-crafted methods.	123

List of Figures

1.1	Schematic illustration of a typical biometric system. The ID is or is not input to the system depending on the recognition task.	2
1.2	An illustration of common biometric recognition tasks and questions using face as an example. The true decisions (answers) and probes (queries) are highlighted in red and blue, respectively.	3
1.3	Example of hand images as a biometric characteristic captured in different acquisition environments. In the uncontrolled and uncooperative environment more variations, e.g., in posing, illumination, gesture and resolution are visible.	4
1.4	Examples of common biometric systems' applications and their schematic relations to biometric characteristic, recognition tasks and acquisition environment. This figure can be read: [Application] is [Biometric characteristic] [Recognition task] in the [Acquisition environment] environment.	5
1.5	Examples of images where subjects faces are covered, masked or not visible at all. Child sexual abuse images also contain hands; however they cannot be put here due to legal reasons.	8
2.1	Pipeline of a biometric recognition algorithm. Dashed boxes indicate that these steps are optional.	13

2.2	Example of a contact-based hand geometry biometric system. (a) user places hand onto a device. (b) the captured hand image where fingers positions are constrained by pegs. (c) examples of the extracted measurements. Images adapted from [Sanchez-Reillo et al., 2000].	14
2.3	Examples of hand biometric characteristics: (a) wrist vein, (b) palm vein (c) finger vein. (a), (b), (c) images reprinted from [Nikisins et al., 2018], [Kang and Wu, 2014] and HKPU database [Kumar and Zhou, 2012], respectively.	15
2.4	Example of visualized veins. (a) original color image, (b) NIR image, (c) clear, visualized veins based on image (a). NIR image (b) is unavailable for uncovering the image (c). (d) images without clear visualized veins. Images in (a-c) and (d) reprinted from [Tang et al., 2011] and [Chan et al., 2017], respectively.	17
2.5	Example of a contact-based dorsal finger biometric system. (a) CCD camera with a light source: user places hand on the support in contact-based manner. (b) the captured dorsal fingers image. Images reprinted from [Kumar and Ravikanth, 2009].	19
2.6	Example of a contact-based finger-knuckle biometric system. (a) CCD camera with LED light source: user places finger into a bracket in a contact-based manner. (b) the captured finger-knuckle image. Images reprinted from [Zhang et al., 2010].	19
2.7	Examples of skin textures from different studies. Images (a), (b), (c) and (d) reprinted from [Cula et al., 2005], [Xie et al., 2012], [Kumar and Xu, 2016] and [Chan et al., 2017], respectively.	21

2.8	Example of contact-based palmprint biometric system. (a) CCD based palmprint scanner: user places hand into a device in a contact-based manner. (b) the captured palmprint image. Images reprinted from [Kong et al., 2009].	22
2.9	Example of a contactless (touchless) palmprint and palm vein biometric system with multiple cameras. (a) multicamera acquisition device: a user places hand over a device in a contactless manner. (b) captured images from left to right: IR, color palmprint, grayscale palmprint and palm vein. Images reprinted from [Liang et al., 2019].	22
2.10	Right: Example of detected key points based on boundary and left: extracted palmprint ROI. Image adapted from [Kong et al., 2009].	24
2.11	Example of encoded features using OrdinalCode.	27
2.12	Example of matched SIFT descriptors (a) from the same palmprint and (b) from two different palmprints.	29
2.13	Illustration of forming a feature vector from LBP histograms extracted from blocks.	29
3.1	Examples of corresponding wrist images in the NTU-Wrist-Image-Database-v1: (a) and (c) Standard pose gallery images, (b) Standard pose probe images, (d) Non-standard pose probe images, (e) Internet gallery images and (f) Internet probe images. Original scales are preserved in (e) and (f). Images in (a), (b), (c) and (d) were resized down for better visualization. (a) and (b), (c) and (d), (e) and (f) show image pairs from the same wrists. .	39
3.2	Examples of resized down original Internet images from the NTU-Wrist-Image-Database-v1. Wrists are highlighted in the original images, cropped and enlarged for better visualization. The scale is preserved between corresponding wrist images.	40
3.3	The wrist segmentation scheme.	43

3.4	Illustration of graph edges and weights between vertices determined based on the gradient image.	46
3.5	(a), (d) gradient maps from two wrist images. (b), (e) key points extracted using Proc2 from (a), (d). (c), (f) key points extracted using Proc2/3 from (a), (d). Red dots are boundaries while green dots are extracted wrinkles. .	50
3.6	(a) Heat map representing sum of the key points from SET1 images. (b) wrist template is built by taking the reference of the heat map.	50
3.7	ROI extraction steps using both Proc2 and Proc2/3 in order to produce respectively ROI#1 and ROI#2.	51
3.8	Seven different grids (a) b_1 , (b) b_2 , (c) b_3 , (d) b_4 , (e) b_5 , (f) b_6 , (g) b_7	53
3.9	The schematic diagram of the whole Wrist Meta Matching system (WMM) and other Meta-WMFA systems. Each RS block represents a WMFA algorithm.	56
3.10	CMC curves for EXP1 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS1} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.	64
3.11	CMC curves for EXP2 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS1} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.	65
3.12	CMC curves for EXP3 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS2} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.	66

3.13	Top 10 matches returned by the WMM algorithm in EXP3 when the tail size is 50% of the gallery size (b) and 10% (c). The genuine match is highlighted with a red box. Probe and gallery Internet images taken outdoor containing the Baltimore rioter are shown in (a). Wrist images are resized for better visualization. The scale is indicated above each image. .	74
3.14	Top 10 matches returned by the WMM algorithm in EXP3. The genuine matches are highlighted with a red box. Wrist images are resized for better visualization. The scale is indicated above each image.	77
4.1	Different types of palmprints studied by the biometric community. (a) Contact-based, (b) contactless, (c) latent, and (d) 3D palmprints. (a), (c) and (d) reprinted from [Kong et al., 2009], [Jain and Feng, 2009] and [Fei et al., 2019a], respectively.	81
4.2	Examples of images from (a) CASIA, (b) IITD, (c) PolyU contactless palmprint databases.	82
4.3	Sample images from (a) NTU-PI-v1 and (b) NTU-CP-v1 databases. The images in the same column in (a) and (b) are from the same palm. Images are resized to the same size.	83
4.4	Examples of original images downloaded from the Internet. Hands are highlighted in the original images.	84
4.5	The proposed initial configuration of hand landmarks.	87
4.6	Palmprint ROI extraction using TPS and a sampler. Sampling grid T_θ is created by TPS based on the hand landmarks in a hand image I . The sampler takes T_θ and I , and returns palmprint ROI I_{ROI} by sampling I into a regular grid G_{ROI} . The red dots are a schematic visualization of the sampling grid.	89
4.7	The ROI-LAnet architecture. The input is the hand image I resized to $h \times w$ pixels and the output is palmprint ROI image I_{ROI}	91

4.8	The proposed End-to-End Palmprint Recognition Network (EE-PRnet) architecture.	92
4.9	Illustration of the in-network ROI augmentation. Top row: examples of different hand landmarks $\hat{\theta}$ (red dots) detected by the Localization Network in the same hand image I . Bottom row, the same column: the corresponding palmprint ROI images I_{ROI} , returned by ROI-LAnet. Zoom in recommended for the best view.	96
4.10	The proposed network architectures and EE-PRnet training schemes. . . .	99
4.11	CMC curves of the proposed algorithm and state-of-the-art palmprint recognition and deep learning methods on NTU-PI-v1 database.	104
4.12	The performance of the proposed algorithm as a function of (a) a number of available training samples per palmprint, (b) full hand image width/height in pixels.	108
4.13	Examples of top-10 identification results. The probe images are in the yellow boxes in the original images. The true matches in the gallery are in the red boxes. In each subfigure, the relative scales of the hand images are preserved for comparison.	112
5.1	Left: Nipple-areola complex with approximately marked nipple (red solid), areola (blue dashed). Right: close-up image with marked tubercles (green solid).	115
5.2	Examples of cropped NAC images. NAC are highlighted in the original chest images. The NAC are from the same subject from two different sessions.	116
5.3	Examples of different NAC from NTU-Nipple-v1 dataset from session 1 (top) and session 2 (bottom). Images in the same column are from the same subject's NAC.	116

5.4	Histograms showing (a) the distribution of images of the same single NAC, (b) the width/height of NAC images.	117
5.5	CMC curve of different deep learning architectures where (a) L2 distance and (b) softmax and SVM are used for identification.	119
5.6	Schematic illustration of the spatial data augmentation. During the training, the bigger NAC image is randomly cropped within the red box and input to the network.	121
5.7	(a) CMC curve comparing AlexNet (Sections 5.4.4 and 5.4.5) and the AlexNet with fully connected layers fine-tuned and the augmentation applied (Section 5.4.6). (b) CMC curve comparing the best AlexNet from 5.4.5 and 5.4.6 and the Modified AlexNet (Section 5.4.7), and hand-crafted LBP and SIFT based methods.	122
5.8	Rows: examples of top-10 identification results using Modified AlexNet- SVM. The match between each probe and the rank-ordered gallery images is in the blue boxes.	124

List of Publications

Wojciech Michal Matkowski, Frodo Kin Sun Chan and Adams Wai Kin Kong. A Study on Wrist Identification for Forensic Applications. *Image and Vision Computing*, vol. 88, pp 96-112, August 2019.

Wojciech Michal Matkowski, Tingting Chai and Adams Wai Kin Kong. Palmprint Recognition in Uncontrolled and Uncooperative Environment. *IEEE Transactions on Information Forensics and Security*, vol. 15, pp 1601-1615, October 2019.

Wojciech Michal Matkowski, Krzysztof Matkowski, Adams Wai Kin Kong and Cory Lloyd Hall. The Nipple-Areola Complex for Criminal Identification. In *International Conference on Biometrics (ICB)*, June 2019.

Wojciech Michal Matkowski, Frodo Kin Sun Chan and Adams Wai Kin Kong. A Study of Wrist Identification for Forensic Applications. Oral presentation and abstract in *American Academy of Forensic Sciences (AAFS), 71st Annual Meeting*, February 2019.

Frodo Kin Sun Chan, Xiaojie Li, **Wojciech Michal Matkowski** and Adams Wai Kin Kong. An Investigation of Distinctiveness of Skin Texture for Forensic Applications. Oral presentation and abstract in *American Academy of Forensic Sciences (AAFS), 71st Annual Meeting*, February 2019.

Wojciech Michal Matkowski, Adams Wai Kin Kong, Han Su, Peng Chen, Rong Hou and Zhihe Zhang. Giant Panda Face Recognition Using Small Dataset. In *International Conference on Image Processing (ICIP)*, September 2019.

Zhihe Zhang, Wai Kin Adams Kong, Rong Hou, **Matkowski Wojciech Michal**, Xingpeng Xu, Han Su, Peng Chen, Kai Cui. "Panda Face Recognition System Based on Limited Training Data", China Patent application number 201811500473.2, filled at 29/12/2018.

Abstract

In forensics, criminal and victim identification based on digital evidence images is highly challenging because the face or other obvious characteristics such as tattoos are occluded, covered, or not visible at all. Existing recognition methods, which make use of biometric characteristics, such as vein, skin mark, height, skin color, weight, race, etc., are considered for solving this problem. The soft biometric traits, including skin color, gender, height, weight and race, provide useful information but not distinctive enough. Veins and skin marks are limited to high-resolution images and some body sites may neither have enough skin marks nor clear veins. Regardless of the availability of these characteristics, other ones, e.g., hands can be used to support the evidence or provide some useful clues to the investigator. Terrorists and rioters tend to expose their hands, including palms and wrists in a gesture of triumph, greeting or salute, while child sexual offenders usually show them when touching their victims.

Wrists, in particular, were neglected by the biometric community for forensic applications. To study this problem, a wrist identification algorithm, which includes skin segmentation, key point localization, image to template alignment, large feature set extraction, classification and post-recognition score analysis has been proposed. The proposed algorithm has been evaluated on a new NTU-Wrist-Image-Database-v1, which consists of 3945 images from 731 different wrists, including 205 pairs of wrist images collected from the Internet, taken under uneven illuminations with different poses and resolutions. In the experiments, the proposed algorithm has been compared to palmprint recognition methods. The extracted large feature sets have been studied and compared with selected deep features and feature selection and reduction schemes. The experimental results indicated that wrist is a useful clue for criminal and victim identification.

Online palmprint recognition and latent palmprint identification are two branches of palmprint studies. The former uses middle-resolution images collected by a digital camera

in a well-controlled or contact-based environment with user cooperation for commercial applications and the latter uses high-resolution latent palmprints collected in crime scenes for forensic investigation. Nevertheless, the merit of contactless palmprint recognition in an uncontrolled and uncooperative environment for forensic investigation is not fully exposed yet. To study this problem, a new palmprint database is established and an end-to-end deep learning algorithm has been proposed. The new database named NTU Palmprints from the Internet (NTU-PI-v1) contains 7881 images from 2035 palms collected from the Internet. The proposed algorithm consists of an alignment network, feature extraction network, in-network data augmentation scheme, and is end-to-end trainable. The proposed algorithm has been compared with the state-of-the-art online palmprint recognition methods and evaluated on three publicly available contactless palmprint databases, IITD, CASIA, and PolyU and two new databases, NTU-PI-v1 and NTU contactless palmprint database. The experimental results showed that the proposed algorithm outperforms the existing palmprint recognition methods.

In child sexual abuse images, the chest of the offender can be sometimes visible or even captured simultaneously with the victim's hand. Thus, the nipple-areola complex (NAC) is proposed for offender identification. Popular deep learning and hand-crafted methods are evaluated on a newly established NTU-Nipple-v1 database, which contains 2732 images from 428 different male's NAC. Experimental results indicate that the proposed NAC can be useful for offender identification.

Chapter 1

Introduction

1.1 Biometrics

Biometrics or biometric recognition is an automated recognition of individuals based on biological or behavioural characteristic [ISO/IEC, 2017]. Many types of biological or behavioural characteristics, also called biometric characteristics or traits, e.g., face, fingerprint, palmprint, DNA, iris, gait, tattoo, voice, signature, ear, hand geometry, veins, keystroke, etc. have been subject of research studies for over five decades [Jain et al., 2016]. A biometric characteristic is intrinsically associated with an individual. Thus, it can be used by a biometric system to recognize an individual in various applications such as surveillance, searching suspect, border or bank account access control, etc.

Biometric systems, which process biometric characteristics are used to recognize individuals automatically or semi-automatically and meet the application requirements. A typical biometric system operates in enrollment and recognition stage (see Figure 1.1). During the enrolment, the system acquires a biometric characteristic of an individual through a capture device, processes it to a desired representation and stores it in a database called **gallery**. During the recognition the system also acquires the biometric characteristic, which is called a **probe** or query. Then, the probe is processed to a desired representation and compared (matched) with the characteristics already enrolled in the gallery returning a

Figure 1.1: Schematic illustration of a typical biometric system. The ID is or is not input to the system depending on the recognition task.

comparison score for each comparison. These scores are used to recognize an individual from the probe and return the decision.

The aim of a biometric system is to make a decision about an individual's identity (ID) based on biometric characteristic. In other words, it is to answer a question about an individual's ID. Usually, these questions are: "who is this individual?", "is this individual the claimed one?", "are these two individuals the same one?". Because humans can recognize faces relatively well, let us consider the above questions based on face images and list down (bold) the requirements for answering them. These common questions, which correspond to different biometric recognition tasks are also illustrated in Figure 1.2.

- Who is this individual? - to answer this question, the **probe** face sample is compared with the **gallery** face samples with assigned **IDs** to determine the probe face ID. This task is called identification or search and it requires **one-to-many comparison**.
- Is this individual the claimed one? - the **probe** face sample with the claimed **ID** is compared with the **gallery** face sample associated with the corresponding **ID** to answer whether they are from the same ID. This task is called verification and it requires **one-to-one comparison** of samples from known ID.
- Are these two individuals the same one? - a pair of **probe** face samples is compared to answer whether they are from the same ID. This task is called pair-wise matching

Figure 1.2: An illustration of common biometric recognition tasks and questions using face as an example. The true decisions (answers) and probes (queries) are highlighted in red and blue, respectively.

and it requires **one-to-one comparison**. Contrary to verification, it does not require IDs.

From the above and based on Figure 1.2, one can see that these biometric recognition tasks correspond to different questions and have different requirements.¹ Therefore, they have different applications which specify the design of a biometric system.

Each biometric system is designed to solve a specific recognition task, e.g., identification or verification using a specific characteristic, for a specific application. Many biometric systems using characteristics such as face, fingerprint, palmprint, iris, gait, voice, and handwriting have been proposed. Some of them, e.g., fingerprint and iris have already achieved very high accuracy [Jain et al., 2016] and have been commercially deployed, e.g., for access control at the airports. Many face recognition methods are already close to human-level performance [Taigman et al., 2014, Schroff et al., 2015]. Law enforcement agencies have been using fingerprints for searching suspects from the early 20th century [Paulino et al., 2013] and nowadays, tattoo recognition technologies are gaining attention [Ngan et al., 2018]. Iris, voice and palmprint recognition also perform very well [Jain et al., 2016]. Each biometric system is designed to operate on the characteristics acquired

¹These recognition tasks are common ones. Obviously, there can be other tasks such as identification/search of a probe with assigned ID in a gallery without assigned IDs, i.e., unsolved latent prints.

Figure 1.3: Example of hand images as a biometric characteristic captured in different acquisition environments. In the uncontrolled and uncooperative environment more variations, e.g., in posing, illumination, gesture and resolution are visible.

under a specific environment. In a controlled and cooperative environment, control over some of data acquisition parameters and user cooperation is assumed. This results in good quality samples and high recognition performance. However, in an uncontrolled and uncooperative environment, there is no such assumption. Thus, samples usually are poor quality, resulting in significantly lower recognition performance. Figure 1.3 illustrates the visual difference of samples from different environments. Although several biometric areas are very successful and various research studies have been done, the recognition in the uncontrolled and uncooperative environment is still very challenging. In some applications, e.g., forensics, the uncontrolled and uncooperative environment and poor quality samples are very common. Therefore, biometric system for such applications should be designed to operate also on poor quality samples.

To sketch relations among different applications, biometric recognition tasks, the acquisition environments and plausible biometric characteristics, let us consider a basic example using fingerprint, tattoo and DNA in their common applications. This example is also schematically illustrated in Figure 1.4. Fingerprint is used in various scenarios and applications, e.g., in automatic airport access control, which is a verification task, in a controlled and cooperative environment. Using tattoos or DNA for access control at the airport does not seem like a reasonable idea because of, e.g., privacy issues and user's convenience. Fingerprint, tattoo and DNA can be well used for searching suspect in a database based on evidence from the crime scene, which is always an uncontrolled

Figure 1.4: Examples of common biometric systems' applications and their schematic relations to biometric characteristic, recognition tasks and acquisition environment. This figure can be read: [Application] is [Biometric characteristic] [Recognition task] in the [Acquisition environment] environment.

environment. DNA can also be used to confirm parenthood in a well-controlled laboratory environment. Fingerprint or DNA recognition systems work as verification or identification systems in various environments depending on the application. However, tattoo recognition system, which is used for searching suspect in a database has to work as an identification, not a verification system because it will be unlikely used in any verification scenario [Ngan et al., 2018].

Now, we can briefly see how different applications indicate a specification of a biometric system. In particular, what the biometric system primarily should do, what it does not have to do and under what conditions. In this work, I investigate the identification problem (see Figure 1.2) rather than verification or others because the application is forensic, i.e., criminal and victim identification or searching suspect in a database with known identities (see Figure 1.4). In such applications, a human (forensic expert) is also involved in a post-analysis of the identification results returned by a biometric system [Best-Rowden et al., 2014].

1.2 Motivation

Biometric characteristics previously mentioned in Section 1.1 such as DNA, face, fingerprint and tattoo are commonly used in forensic applications [Champod and Tistarelli, 2017]. Latent fingerprint or DNA are physically collected at a crime scene. Nonetheless, sometimes the crime scene location may remain unknown or the crime is unreported making the physical examination and collection of such evidence impossible. Nowadays, more evidence is also in the form of digital images, where, e.g., face or tattoo can be visible. Face and tattoo have some characteristic and obvious features, which allow identification or at least to narrow down the list of suspects.

An important part of forensic investigation is criminal and victim identification based on evidence images. In some cases, digital images can be the only evidence available. How to identify a person if none of the biometric characteristics from Figure 1.4 such as face, tattoo, DNA and fingerprint are available? For example, how to identify persons in Figure 1.5?. Criminals cover or hide their faces or tattoos, to avoid identification and potentially prosecution when forensic investigators or law enforcement agencies take over the digital images. The identification from evidence images is very troublesome and challenging if no obvious characteristics such as face or tattoos are available. Terrorists, rioters, child sexual offenders usually hide their faces or tattoos to avoid identification, but other body parts, e.g., hands, forearms, back, chest, and legs can be still visible. Sometimes, criminals also try to hide the identity of their victims.

Terrorists and rioters usually cover their faces with masks or clothes, but at the same time they may expose their forearms and hands, e.g., to express a triumph, greeting and salute or while holding weapons, etc. Images containing rioters usually have higher resolution than terrorist images, if they are taken by press photographers such as the case of the masked Baltimore rioter who was photographed more than twice by Associated Press

photographer exposing his wrist² (see Figure 1.5, highlighted in red). Terrorist images are more challenging because they often are low resolution. These images are captured for communication, threatening, or propaganda purpose and can be uploaded on social media.

Another serious issue is child pornography, whose official term in the Interpol is child sexual abuse materials. The National Center for Missing & Exploited Children (NCMEC) has reviewed more than 283 million abusive images and videos and law enforcement identified more than 17,000 child victims since 2002³. Offenders take pictures for sale, personal records or exchange with other offenders. A lot of these deals take place in the Dark web - study finds that over 80% of visits are related to child pornography⁴. To avoid recognition, offenders almost always hide or blur their faces, tattoos and other traits such as scars, deformities in the abusive videos and images [ECPAT International, 2018]. Nevertheless, other body parts such as back, chest, thigh, arm or hand [Spaun, 2007] are still visible [INTERPOL, 2018]. Some of these images can be close-up images with good quality and relatively high resolution.

The images without available face and tattoo have been considered in some recent studies to develop biometric characteristics such as vein or blood vessels [Tang et al., 2011], skin mark [Nurhudatiana et al., 2013], androgenic hair [Su and Kong, 2014], skin texture [Chan et al., 2017], v-sign hand geometry [Hassanat et al., 2016], finger-knuckle [Kumar, 2014] for forensic applications. Veins and skin marks require high-resolution image [Tang et al., 2011]. Some of the body parts have no or not enough skin marks or hair and skin texture may not be discriminative enough if only small skin region is available [Chan et al., 2017]. Other soft biometric characteristics such as skin/hair color, height, weight, age or clothing can also provide useful information [Nixon et al., 2015, Dantcheva

²<http://www.dailymail.co.uk/news/article-3058359/Masked-rioter-caught-live-TV-cutting-hose-firefighters-tried-blaze-broke-pharmacy-looting.html>

³<http://www.missingkids.com/KeyFacts>

⁴<https://www.wired.com/2014/12/80-percent-dark-web-visits-relate-pedophilia-study-finds/>

Figure 1.5: Examples of images where subjects faces are covered, masked or not visible at all. Child sexual abuse images also contain hands; however they cannot be put here due to legal reasons.

et al., 2016]. However, they are not distinctive enough or sometimes hard to estimate from an image. Other related biometric characteristics include, e.g., finger surface [Woodard and Flynn, 2005], finger-knuckle [Zhang et al., 2010, Kumar, 2014] and periocular [Proenca and Neves, 2018, Tiong et al., 2019], which could be applied to some images, where the particular regions are visible. Regardless of the availability of these characteristics, other ones can be used to search suspect, support the evidence or provide some useful clues for the forensic investigator.

Hand regions can be observed in some of the evidence images [Spaun, 2007, Champod and Tistarelli, 2017], particularly when the subjects salute, wave, raise their hands, hold a gun, touch the victim or the abuser, or try to cover the camera (see Figure 1.5). Hand biometrics is well studied in the literature, mostly for commercial applications such as access control in well-controlled and cooperative environments. However, their forensic applications such as criminal and victim identification or uncontrolled and uncooperative environments, except latent fingerprint and palm print, are not well exposed yet.

1.3 Research Objectives

In this work, I investigate criminal and victim identification based on deep and large feature sets from hand biometrics. The objectives of this work are as follows:

- To establish biometric databases containing images, which can be used for studying forensic applications such as criminal and victim identification.
- To propose and develop identification algorithms employing deep and large features sets.
- To evaluate the proposed algorithms and relevant existing methods on the established databases.

1.4 Research Contributions

In this work, I selected two hand regions: palmar side of a wrist (Chapter 3) and palmprint (Chapter 4) images to study them for forensic applications including challenging imaging environments. According to my best knowledge, there were no publicly available studies of wrist identification for forensic applications until my work [Matkowski et al., 2019b], and no studies of contactless palmprint in uncontrolled and uncooperative environment have been available before. In addition, I also proposed to use the nipple-areola complex (NAC) for sexual offender identification (Chapter 5). The main contributions of this research are as follows:

- Development of three new biometric databases: 1) wrist image database for forensic applications; 2) palmprint database from uncontrolled and uncooperative environment, which is the largest palmprint database in terms of subjects and palms; 3) NAC database for offender identification.
- Development of two identification algorithms for forensic applications including: 1)

wrist identification algorithm which bases on hand-crafted large feature sets; 2) an end-to-end deep neural network for palmprint alignment and feature extraction.

- The extensive experimental results show the effectiveness of the proposed algorithms comparing to the representative, state-of-the-art palmprint and deep learning recognition methods on relevant databases.

1.5 Organization of the Thesis

The rest of this thesis is organized as follows. Chapter 2 provides more detailed background and related studies of hand biometrics. Chapter 3 and 4 present a study of wrist and palmprint identification for forensic applications, respectively. In both Chapters 3 and 4, the corresponding new databases, the proposed algorithms and the experimental evaluations are provided. Chapter 5 proposes to use nipple-areola complex for offender identification. Chapter 6 gives a discussion and suggests future work.

Chapter 2

Background and Related Work

2.1 Hand Biometrics

2.1.1 Overview

A typical pipeline of a biometric recognition algorithm operating on hand images consists of a sequence of modules responsible for feature extraction, comparison and decision making. Before feature extraction module, it may also consist of pre-processing, segmentation, alignment or region of interest (ROI) extraction modules to produce a suitable image. Figure 2.1 shows the pipeline. The modules use image processing and pattern recognition methods to extract useful information at each stage and forward it to the next module. To successfully process the information, in most of the existing algorithms, the modules are carefully hand-crafted using human knowledge about the hand and also the image acquisition settings.

In the literature, various approaches have been proposed for low or middle-resolution hand image biometric recognition, mostly for commercial or governmental applications such as access control or time and attendance in a well-controlled and cooperative environment¹. They can be categorized based on the characteristics they use and the image

¹High-resolution fingerprint and palmprint have been proposed also for forensic applications. However, they are out of scope in this study because they contain high-resolution characteristics such as minutiae and

acquisition environment. Hand biometric characteristic types include:

- Hand shape/geometry [Sanchez-Reillo et al., 2000, Zheng et al., 2007, Ferrer and Morales, 2011] (in Section 2.1.2, Figure 2.2).
- Vein: finger vein [Hong et al., 2017, Das et al., 2019], palm vein [Kang and Wu, 2014], dorsum vein [Wang et al., 2008] or wrist vein [Kabacinski and Kowalski, 2011, Nikisins et al., 2018] (in Section 2.1.3, Figure 2.3).
- Dorsal surface of the hand: finger-knuckle/surface [Woodard and Flynn, 2005, Zhang et al., 2010, Kumar, 2014] (in Section 2.1.4, Figures 2.6 and 2.5), hand back/dorsum [Kumar and Xu, 2016].
- Skin texture: hand [Cula et al., 2005], hand back/dorsum [Xie et al., 2012, Kumar and Xu, 2016] (in Section 2.1.5, Figure 2.7).
- Palmar surface of the hand: palmprint [Zhang et al., 2003, Kong and Zhang, 2004, Lu et al., 2008, Wu and Zhao, 2015, Genovese et al., 2019] (in Section 2.1.6), digitprint or full fingerprint [Ribaric and Fratric, 2005, Savič and Pavešić, 2007].
- Combinations of the above [Yörük et al., 2006, Kanhangad et al., 2011a, Liang et al., 2019] (Figure 2.9).

These characteristics can be captured in a well-controlled, contact-based or contactless, pose free or constrained by pegs acquisition environments, mostly with user cooperation and explicit intention of recognizing an individual. The image capture devices typically are CCD or CMOS (consumer camera), multispectral, near-infrared (NIR) or infrared (IR) cameras, 3D digitizers or multiple cameras. Following figures illustrate examples of biometric systems: Figure 2.2: contact-based hand geometry, Figure 2.5: contact-based dorsal finger, Figure 2.6: contact-based finger-knuckle, Figure 2.9: contactless palmprint/palm vein, Figure 2.8: contact-based palmprint. Figure 2.3 shows wrist vein, palm vein and ridges, which are not visible in low and middle-resolution images considered in this study.

Figure 2.1: Pipeline of a biometric recognition algorithm. Dashed boxes indicate that these steps are optional.

finger vein images.

2.1.2 Geometry and Shape

Hand geometry or hand shape based biometric recognition refers to technologies which use geometric information from hand such as finger lengths and widths, areas, ratios, angles, finger shapes or entire hand contour, etc. to recognize an individual for commercial or governmental applications. According to [Jain et al., 2004] and [Duta, 2009], the distinctiveness of hand geometry is medium. Figure 2.2 shows an example of hand geometry biometric system. Typically, a hand image is captured by a camera in a well-controlled and cooperative environment.

In [Sanchez-Reillo et al., 2000], the authors proposed a contact-based system, in which a user places a hand on a platform with pegs to control fingers and hand positions. A CCD camera captures the dorsal side of a hand on a uniform background. For pre-processing, simple arithmetic operations and Sobel filter are used to segment the hand and extract its contour. Feature vector consists of 25 hand measurements including widths, heights, deviations and angles which are classified using Gaussian Mixture Models (GMM). Yoruk et al. [Yoruk et al., 2006] employed Hausdorff distance between aligned hand contours and features from independent component analysis (ICA) of segmented hand images taken in a contact-based, without pegs environment. In some approaches, the contour alignment can be performed for each finger separately or for the whole hand [Duta, 2009]. Zheng et

Figure 2.2: Example of a contact-based hand geometry biometric system. (a) user places hand onto a device. (b) the captured hand image where fingers positions are constrained by pegs. (c) examples of the extracted measurements. Images adapted from [Sanchez-Reillo et al., 2000].

al. [Zheng et al., 2007] proposed to use projective invariants for hand shape recognition in a contactless, without peg environment, where user can pose relatively freely. However, as the authors point out, their method is not applicable when the fingers bend or stay apart. Ferrer and Morales [Ferrer and Morales, 2011] proposed an acquisition device which captured dorsal hands on a platform without pegs, under the visible and IR light. Svoboda et al. [Svoboda et al., 2015] employed a 3D camera for contactless hand geometry recognition. Researchers also combined hand geometry with other characteristics such as palmprint [Yörük et al., 2006, Kanhangad et al., 2010] to increase the distinctiveness.

Hand is a very flexible object with 27 degrees of freedom. Hand geometry and shape features usually suffer due to 3D posing variations, hand flexions, finger movement or bending [Duta, 2009, Kanhangad et al., 2010] which always occur in an uncontrolled environment. Extracting accurate hand shape/geometry information from an evidence image which was taken in uncontrolled and uncooperative without pegs environment is problematic or unlikely to obtain, which according to [Duta, 2009] may limit its forensic applications.

Figure 2.3: Examples of hand biometric characteristics: (a) wrist vein, (b) palm vein (c) finger vein. (a), (b), (c) images reprinted from [Nikisins et al., 2018], [Kang and Wu, 2014] and HKPU database [Kumar and Zhou, 2012], respectively.

2.1.3 Vein

Blood vessels form a network-like structure under the skin, which transports blood within a body. There are three major types of blood vessels: arteries, capillaries and veins. They form distinctive patterns among different individuals including identical twins [Kang and Wu, 2014]. According to [Jain et al., 2004], their distinctiveness is medium. In the natural lighting environment, blood vessels situated under the skin are almost invisible for a human eye, which is sensitive to 390-700 nm (visible light) light spectrum. Therefore, they are hard to steal. Within the NIR spectrum, human tissue is a highly scattering medium, whereas hemoglobin absorption within veins is sufficient enough to result in high contrast. Thus, veins can be seen (see Fig. 2.3) and captured under NIR light spectrum (700-1000 nm). According to [Zharov et al., 2004], the best contrast occurs at approximately 920 nm. Therefore, a typical biometric system operating on veins use NIR camera to capture vein images.

A number of researchers addressed palm, finger, dorsum and wrist vein recognition based on images for commercial or governmental applications. The vein images are mostly captured under NIR spectrum in a well-controlled, contact-based or contactless environment. Accurately extracting or enhancing veins from these images is extremely important

for recognition performance. Zhou and Kumar [Zhou and Kumar, 2011] proposed a method for the feature extraction that enhances palm veins based on Radon Transform or Hessian Phase. The authors reported 0.51% and 1.44% equal error rate (EER) performance for the left and right hand images taken under 850 nm wavelength. Kang and Wu [Kang and Wu, 2014] proposed a method for palm vein texture extraction, which uses gradient-based maximal principal curvature (MPC) and k-means clustering to segment veins. For matching, the mutual foreground between two images, local binary patterns (LBP) and χ^2 distance were used. The authors reported very low EER of 0.267% for the left hand images taken under 940 nm wavelength (see Figure 2.3b). The authors also suggested that the proposed method can be suitable for palmprint recognition. Wang et al. [Wang et al., 2008] proposed to use minutiae features from hand back vein patterns acquired under far-infrared light. The authors reported 0% EER using modified Hausdorff distance. Huang et al. [Huang et al., 2015] proposed a method which detects and matches key points in dorsum vein images focusing not only on vein pattern but also surrounding tissue characteristic. Raghavendra and Busch [Raghavendra and Busch, 2016] proposed a device with NIR camera and pegs for wrist vein recognition in a contact-based environment. The authors evaluated popular feature extraction schemes on their database and reported encouraging results. Nikisins et al. [Nikisins et al., 2018] proposed a system, which is capable of compensating for scale, rotation and translation between two wrist vein patterns, demonstrating the promising verification performance on the largest wrist vein image database. Recently, some researchers proposed to use convolutional neural networks (CNN) for vein images taken under NIR spectrum in a contactless environment. Hong et al. [Hong et al., 2017] used a pre-trained VGG-16 model and fine-tuned it using finger vein images. Qin and El-Yacoubi [Qin and El-Yacoubi, 2017] proposed a CNN to predict vein and background pixels in finger NIR image. The authors also developed a fully convolutional network (FCN) for recovering missing vein patterns. Das et al. [Das et al., 2019] proposed a CNN architecture for finger vein identification and trained it from scratch with softmax loss.

Figure 2.4: Example of visualized veins. (a) original color image, (b) NIR image, (c) clear, visualized veins based on image (a). NIR image (b) is unavailable for uncovering the image (c). (d) images without clear visualized veins. Images in (a-c) and (d) reprinted from [Tang et al., 2011] and [Chan et al., 2017], respectively.

NIR evidence images are not available in forensic applications. However, some researchers showed that some veins are hidden in color images and can be uncovered and visualized for forensic [Tang et al., 2016] or medical applications [Song et al., 2015]. The visualized vein highly depends on image compression, resolution, skin clarity and concentration of fat or melanin [Tang et al., 2011]. In other words, if image quality is bad or the subject's skin does not fall into the desired criteria, it can be hard to uncover and visualize veins from color images (see Figure 2.4). Thus, in such forensic scenarios, vein may not have sufficient discriminatory power [Chan et al., 2017].

2.1.4 Dorsal Surface

Dorsal surface of the hand or hand back biometric studies mostly focus on finger surface, finger-knuckle and hand back/dorsum skin texture (see Section 2.1.5) recognition in contact-based or contactless environments for commercial or governmental applications. Figures 2.6 and 2.5 show examples of finger-knuckle and dorsal finger biometric systems, respectively.

Woodard and Flynn [Woodard and Flynn, 2005] studied finger surface recognition

based on range and color hand images taken in the contactless environment on a black background. Color images were only used to segment hands and find fingers. Range images were used to extract feature vectors from segmented fingers based on principal curvatures. For recognition, the authors utilized correlation coefficient between two feature vectors, reporting promising results. Kumar and Ravikanth [Kumar and Ravikanth, 2009] developed an acquisition system without pegs for dorsal finger image collection in a well-controlled and cooperative environment (see Figure 2.5). To represent the fingers' back, the authors proposed to use geometric features from four fingers and texture features (eigenknuckle) from the surface of four finger-knuckles. Zhang et al. [Zhang et al., 2010] developed another acquisition system with a bracket to capture a finger-knuckle image in a contact-based environment (see Figure 2.6). Comparing to the system from [Kumar and Ravikanth, 2009], which captures four fingers, Zhang et al.'s [Zhang et al., 2010] system is smaller in size but it captures only one finger-knuckle image at a time. Moreover, the acquisition environment in [Zhang et al., 2010] is very well-controlled and highly constrained by the bracket, which is also reflected in high recognition performance. In [Kumar and Ravikanth, 2009] and [Zhang et al., 2010], only the surface of one finger-knuckle from each finger was studied and named major finger-knuckle and finger-knuckle-print (FKP), respectively. However, each finger has three knuckles (joints). Thus, Kumar [Kumar, 2014] and Kumar and Xu [Kumar and Xu, 2016] also studied other finger-knuckles, named by the authors, minor and second minor finger-knuckle, respectively. In these studies, dorsal finger images were captured in a similar setting to [Kumar and Ravikanth, 2009], which is a relatively well-controlled environment without viewpoint or resolution variations. In the experiments, the authors evaluated popular feature descriptors and different combinations of finger-knuckles and hand dorsum (see Section 2.1.5). According to the authors of [Kumar, 2014, Kumar and Xu, 2016], the accuracy are quite poor and both studies are for forensic applications.

Figure 2.5: Example of a contact-based dorsal finger biometric system. (a) CCD camera with a light source: user places hand on the support in contact-based manner. (b) the captured dorsal fingers image. Images reprinted from [Kumar and Ravikanth, 2009].

Figure 2.6: Example of a contact-based finger-knuckle biometric system. (a) CCD camera with LED light source: user places finger into a bracket in a contact-based manner. (b) the captured finger-knuckle image. Images reprinted from [Zhang et al., 2010].

2.1.5 Skin Texture

Skin texture for human recognition was proposed over a decade ago. Even though skin is on face, hands, etc., in this context, the skin is considered as a piece of skin from a body site without any obvious characteristics. Cula et al. [Cula et al., 2005] modeled the hand and face skin texture using bidirectional feature histogram in a well-controlled environment based on high-resolution images (see Figure 2.7a) taken by a CCD camera. Their method used a filter bank to filter the images and create a texton library using k-means clustering. For recognition, the authors employed principal component analysis (PCA), and nearest neighbor classifier, achieving 73% recognition accuracy. Hand back

skin texture recognition was also studied in [Xie et al., 2012] and named by the authors (HBST). Xie et al. [Xie et al., 2012] developed an acquisition device with a CCD camera and a light source that captures high-resolution (450 dpi) hand back images (see Figure 2.7b) in a well-controlled environment. The authors proposed to use a filter bank and learn a sparse texon library for feature extraction. For recognition, χ^2 distance was used, achieving 86% and 98% accuracy for identity and gender recognition, respectively. Kumar and Xu [Kumar and Xu, 2016], also investigated dorsal surface of the hand (see Figure 2.7c) and its combination with different finger-knuckles (see Section 2.1.4).

In [Cula et al., 2005, Xie et al., 2012, Kumar and Xu, 2016] studies, images were taken under well-controlled environment and had relatively high resolution. A recent study [Chan et al., 2017] of skin texture based on low resolution (from 6.7 to 30.6 dpi) forearm and thigh images from uncontrolled environment showed promising results for forensic application such as searching suspect in a database. Figure 2.7 visually compares skin textures from these different studies. [Chan et al., 2017] proposed a skin texture recognition algorithm and compared the distinctiveness of skin texture with blood vessels which are regarded as a hard biometric trait. The authors reported that the skin texture in low-resolution images outperformed blood vessels. In addition, they concluded that the skin texture may not be discriminative enough if only the small region of skin is available.

2.1.6 Palmprint

Palm characteristics, including flexion creases, wrinkles, ridges, and minutiae are located on a palmar side of a hand and are considered to be permanent and unique to an individual [Kong et al., 2009]. Palmprint has been studied for over two decades. Figures 2.8 and 2.9 show examples of a contact-based and contactless palmprint² biometric systems, respectively. The previous studies can be classified into contact-based, contactless, latent and 3D

²Note that Figure 2.9 shows multicamera system. Thus, except color and grayscale images of palmprint it also contains IR and NIR images of palm vein images, which are not the subject of the conventional palmprint studies.

Figure 2.7: Examples of skin textures from different studies. Images (a), (b), (c) and (d) reprinted from [Cula et al., 2005], [Xie et al., 2012], [Kumar and Xu, 2016] and [Chan et al., 2017], respectively.

palmprint. Except for the latent palmprint, which has been studied for forensic applications, all studies have been for commercial or governmental applications. Many methods have been proposed, achieving high recognition performance, especially on publicly available palmprint databases, established in environments, where image acquisition parameters such as illumination, background, type of camera, etc. can be controlled [Fei et al., 2019a].

So far, there are no publicly available studies of palmprint recognition in an uncontrolled and uncooperative environment and there were no publicly available studies of wrist identification for forensic applications until my work [Matkowski et al., 2019b]. The most related problem to both of them is contactless (contact-free, touchless) palmprint recognition, which has been addressed by a number of researchers [Kanhangad et al., 2011b, Wu et al., 2014, Kang and Wu, 2014, Aykut and Ekinici, 2015, Michael et al., 2012, Zhang et al., 2017, Liang et al., 2019]. In a contactless setting, the image is taken by a camera without physical contact between the hand and any surface of the acquisition device. This setting usually results in some rotation, translation, scale and illumination

Figure 2.8: Example of contact-based palmprint biometric system. (a) CCD based palmprint scanner: user places hand into a device in a contact-based manner. (b) the captured palmprint image. Images reprinted from [Kong et al., 2009].

Figure 2.9: Example of a contactless (touchless) palmprint and palm vein biometric system with multiple cameras. (a) multicamera acquisition device: a user places hand over a device in a contactless manner. (b) captured images from left to right: IR, color palmprint, grayscale palmprint and palm vein. Images reprinted from [Liang et al., 2019].

variations, depending on the acquisition device. The acquired image, consisting of a full or partial palmar side of a hand is input and processed by a recognition method.

Originally, to capture and extract the palmprint features, coding-based methods such as PalmCode [Kong and Zhang, 2002], CompCode [Kong and Zhang, 2004], and OrdinalCode [Sun et al., 2005] were proposed for the contact-based environment (see Figure 2.8). However, e.g., according to [Fei et al., 2019a], they usually do not perform well³ on

³It may not be an accurate statement especially if the procedure of horizontal and vertical shifting is applied to the features from coding-based methods.

contactless palmprints mostly because of image misalignment, which is caused by less constrained environment. Thus, some studies suggested using more robust features such as Scale Invariant Feature Transform (SIFT) [Wu et al., 2014] and Local Binary Pattern (LBP) based features [Kang and Wu, 2014, Fei et al., 2017] or Local Micro-structure Tetra Pattern [Li and Kim, 2017]. Recently, several new coding-based methods such as double orientation code (DOC) [Fei et al., 2016b] and difference of normals (DON) [Zheng et al., 2016] which extracts 3D information from a 2D image were also proposed. Some researchers also proposed to use conventional CNN [Svoboda et al., 2016, Ramachandra et al., 2018, Fei et al., 2019a] or non-conventional CNN [Minaee and Wang, 2017, Meraoumia et al., 2017, Genovese et al., 2019] based methods for palmprint recognition. More detailed description about palmprint recognition approaches and current studies is given in Section 2.2

2.2 Palmprint Recognition Studies

2.2.1 ROI Extraction

In the related palmprint recognition literature, the proposed methods operate on hand images [Kanhangad et al., 2011a, Zheng et al., 2016, Genovese et al., 2019, Liang et al., 2019] or already extracted palmprint ROI images [Wu et al., 2014, Luo et al., 2016, Zheng et al., 2016, Jia et al., 2017], which are provided in palmprint databases. Thus, the methods that take a hand image as an input apply some pre-processing, segmentation and alignment steps to extract palmprint ROI. The goal of pre-processing and segmentation is to extract the hand and fingers contours and localize the hand landmarks or key points, which are used to define palmprint ROI. Usually, two key points situated between 1) middle and index fingers and 2) ring and little fingers are used.

In a well-controlled environment where the background has single color and fingers are separated, Gaussian smoothing, thresholding, edge detectors and boundary tracking

Figure 2.10: Right: Example of detected key points based on boundary and left: extracted palmprint ROI. Image adapted from [Kong et al., 2009].

methods give satisfactory pre-processing and segmentation results. Then, to align palmprints into the same coordinate system, the key points are used to extract square palmprint ROI [Kong and Zhang, 2002]. Figure 2.10 shows extracted palmprint ROI based on two key points. In the literature, this approach is called tangent-based [Kong et al., 2009]. To extract palmprint ROI, most of the current palmprint recognition studies use this approach [Zhang et al., 2017, Genovese et al., 2019] or some variation of this approach, e.g., using additional information from hand contour, fingertips, etc. [Kanhagad et al., 2011a, Ito et al., 2015, Liang et al., 2019].

Very few works studied palmprint ROI extraction in less controlled environment [Ong et al., 2008, Ito et al., 2015, Chai et al., 2016], where the background is more complex or fingers are closed and touch each other. Firstly, such environment requires more sophisticated segmentation methods because of more complex background. Secondly, it requires different key point detection methods because the boundary tracking may fail if there are no clear boundaries between closed fingers. Ong et al. [Ong et al., 2008], in their palmprint recognition pipeline, proposed to use skin color model to segment hand in complex background. Ito et al. [Ito et al., 2015] studied ROI extraction in opened and

closed fingers scenarios. The authors compared and analyzed three different methods for ROI extraction, including the most popular one from [Kong and Zhang, 2002], showing the failure cases due to closed fingers or hand pose variations. The authors also proposed a method for key points detection and selection, which combines segmented and edge images to handle the closed finger scenarios. The ROI was extracted using Kong and Zhang's [Kong and Zhang, 2002] method, based on the detected key points. Bao and Guo [Bao and Guo, 2016] used a CNN for key points detection and the same method as in [Kong and Zhang, 2002] to extract the ROI based on two detected key points. However, their CNN outperformed the whole ROI extraction method proposed in [Kong and Zhang, 2002] only when the Gaussian noise was added to hand images. Moreover, the evaluation was performed on images from CASIA-Palmprint database which have a black background.

2.2.2 Feature Extraction and Comparison

Most of palmprint recognition approaches extract features from palmprint ROI using coding (Section 2.2.3), local descriptors (Section 2.2.4) or recently also CNN (Section 2.2.5) based methods. The comparison between feature vectors is typically performed using a nearest neighbor distance-based classifier, e.g., Hamming or sometimes hyperplane classifier, e.g., support vector machine (SVM) or neural network, which return comparison scores. In most of the biometric recognition studies, the comparison score is considered as a match

- in verification, if it exceeds the predefined threshold,
- in closed-set identification, if it is the highest/lowest score,
- in open-set identification, if it is the highest/lowest score and exceeds the predefined threshold.

2.2.3 Coding-based Methods

Coding-based approaches apply filters to an image to get filter responses, which capture local line orientation information. Typically, Gabor or Gaussian 2D filters are used. Filter responses' magnitudes and phases are encoded and used as features. The features are typically in the form of a matrix. The comparison scores are calculated using Hamming or other distances between two feature matrices. To compensate spatial misalignments between features extracted from two images, vertical and horizontal shifts and rotations are applied and multiple comparison scores are calculated. The lowest normalized comparison score is considered as a match.

Kong and Zhang [Kong and Zhang, 2002] applied a single Gabor filter to palmprint images to encode the features. This method is called PalmCode and it was derived from the IrisCode proposed by Daugman [Daugman, 1993]. To extract orientation information more accurately than PalmCode, Kong and Zhang [Kong and Zhang, 2004] proposed Competitive Code (CompCode), which uses multiple Gabor filters with six different orientations and encoding scheme based on a dominant orientation. To compare two features, the authors proposed angular distance. Sun et al. [Sun et al., 2005] proposed OrdinalCode features, which use Gaussian filters with three orientations and encoding scheme based on the difference between a filter response and an orthogonal filter response. The authors used Hamming distance to calculate comparison scores between two features. Figure 2.11 shows an example of OrdinalCode feature matrices. Jia et al. [Jia et al., 2008] proposed RLOC, which is an improved version of CompCode, which uses modified Radon transform (MFRAT) to extract palmprint lines. To further improve CompCode, Guo et al. [Guo et al., 2009] proposed BOVC method, which encodes not only dominant orientation but also other orientations. Zhang et al. [Zhang et al., 2012b] proposed extended BOVC (E-BOVC) to discard the fragile bits from BOVC proposed in [Guo et al., 2009]. Fei et al. [Fei et al., 2016b] and Xu et al. [Xu et al., 2018] pointed out that most of the coding-based approaches consider only one dominant orientation. Thus, Fei et al. [Fei et al.,

Figure 2.11: Example of encoded features using OrdinalCode.

2016b] proposed double orientation code (DOC), which considers two most dominant orientations. Xu et al. [Xu et al., 2018] proposed discriminative and robust competitive code (DRCC), which uses 12 instead of 6 Gabor filters with different orientations and a Gaussian filter to consider contributions of neighborhood pixels' orientations. To encode 3D information from a 2D palmprint image, Zheng et al. [Zheng et al., 2016] proposed difference of normals (DoN) method, which partially describes the difference between the normal vectors of two neighborhood pixels.

2.2.4 Local Descriptors Methods

In palmprint studies, researchers proposed to use local descriptors to handle palmprint ROI misalignments and illumination variations caused by the contactless acquisition environment. Local descriptors transform an image into another representation by processing it some algorithm. Local general purpose descriptors such as SIFT [Lowe, 1999] (see Figure 2.12), LBP [Ojala et al., 2002] (see Figure 2.13) or some variations of them have been commonly applied to extract features from contactless palmprint. The processed image is divided into small blocks (or local regions) to preserve spatial information. Typically, histograms from these blocks are used as feature vectors. The comparison scores are calculated using some distance measure, key point matching scheme or a trained classifier.

Ong et al. [Ong et al., 2008] proposed to apply LBP to palmprint. The authors divided palmprint ROI into nine blocks and extracted LBP histograms. Figure 2.13 shows an example of extracting LBP histograms. For recognition, the authors used a probabilistic neural

network. Morales et al. [Morales et al., 2010] applied modified SIFT and combined it with OrdinalCode (SIFT_OLOF) approach introduced in [Sun et al., 2005]. The comparison score was calculated as a combination of the SIFT score (number of matched key points between two palmprints) and a Hamming distance between two feature matrices. Wu et al. [Wu et al., 2014] proposed to apply SIFT to detect key points in a filtered palmprint ROI. Then, to refine the detected key points, the authors used the RANSAC algorithm (SIFT_RANSAC). Figure 2.12 shows an example of matched SIFT descriptors on palmprints. Jia et al. [Jia et al., 2014] proposed histogram of lines (HOL) descriptor, which uses responses from Gabor filters or MFRAT to calculate histograms in small blocks. For recognition, the authors used Euclidean distance after applying PCA or linear discriminant analysis (LDA) on the extracted feature sets. Luo et al. [Luo et al., 2016] proposed local line directional pattern (LLDP), a variant of LBP, which operates on edge filter responses. Jia et al. [Jia et al., 2017] proposed a complete direction representation (CDR) framework, which utilizes a number of filters, e.g., Gabor, Gaussian, MFRAT, etc. with different scales and orientations. The authors proposed a training scheme to select a subset of these filters. Fei et al. [Fei et al., 2019b] proposed a discriminant direction binary code (DDBC) method. DDBC is the unsupervised learning method to learn the global projection matrix from a training set of filtered palmprint ROIs. To extract features, the projection matrix was applied to a sample and the histograms from small regions were calculated and concatenated. Comparison scores were calculated using χ^2 distance between two feature vectors.

2.2.5 Convolutional Neural Networks

The success of deep learning, including convolutional neural networks (CNN) [Krizhevsky et al., 2012] in computer vision, has been encouraging researchers to shift from the traditional feature-engineering approach to the deep learning approach. In the deep learning approach, rather than designing features in each module separately, the network architecture

Figure 2.12: Example of matched SIFT descriptors (a) from the same palmprint and (b) from two different palmprints.

Figure 2.13: Illustration of forming a feature vector from LBP histograms extracted from blocks.

is designed and during the network training, the hierarchical data representations are being learned, typically using backpropagation with stochastic gradient descent (SGD) optimization [LeCun et al., 2015]. A typical CNN consists of multiple convolutional, activation, pooling and fully connected layers. The success of deep learning highly depends on the amount of training data. In the image-based biometrics, the shift occurred especially in the face recognition domain where large datasets are already available and many deep learning methods [Taigman et al., 2014, Schroff et al., 2015, Parkhi et al., 2015, Hu et al., 2015] outperformed traditional face recognition methods on very challenging unconstrained benchmarks such as Labelled Faces in the Wild (LFW) or YouTube Faces (YTF) with significant margins. In addition to face, CNNs were also applied to fingerprint [Cao and Jain, 2015], iris [Zhao and Kumar, 2017] and palmprint recognition [Svoboda et al.,

2016, Minaee and Wang, 2017, Ramachandra et al., 2018, Genovese et al., 2019].

To the best of my knowledge, there are very few systematic studies on deep learning application to palmprint recognition. Svoboda et al. [Svoboda et al., 2016] proposed to use a Siamese network, with an architecture similar to AlexNet [Krizhevsky et al., 2012] for palmprint recognition. During training, the network is fed with triplets of palmprint ROIs (128 by 128 pixels) from different images and outputs 32-dimensional feature vectors for calculating the d-prime loss. This method requires a predefined triplet generation scheme and at least two samples collected from each subject's palmprint because each triplet contains two different samples coming from the same palm and one sample from a different palm. During the recognition, the comparison scores are determined by the distances between the feature vectors of two input samples. Svoboda et al. evaluated the proposed architecture on two contactless palmprint benchmarks and compared it with OrdinalCode and CompCode. Minaee et al. [Minaee and Wang, 2017] employed a deep scattering convolutional network (DSCN) [Bruna and Mallat, 2013] for feature extraction with PCA SVM for recognition, achieving very good results. However, the scattering network consists of fixed, non-trainable filters and the evaluation was performed on a contact-based dataset. In the recent palmprint recognition survey [Fei et al., 2019a], Fei et al. evaluated a number of methods for contactless palmprint recognition, including four deep learning architectures, AlexNet, VGG-16 [Simonyan and Zisserman, 2015], GoogLeNet [Szegedy et al., 2015] and Res-Net-50 [He et al., 2016]. These networks were pre-trained using the ImageNet dataset and then palmprint ROI images were used to fine-tune the networks. The authors concluded that deep learning methods achieve comparable or even higher performance than conventional (coding or local descriptor based) palmprint methods. Ramachandra et al. [Ramachandra et al., 2018] investigated contactless palmprint recognition of newborn babies using a pre-trained AlexNet. 2000 newborns' palmprints were taken by a smartphone camera in two sessions with around 30 hours time interval. The authors manually cropped palmprint ROIs to fine-tune AlexNet

and trained SVM using features from the last fully connected layer. For recognition, the sum rule of scores produced by softmax and SVM classifiers was used. Genovese et al. [Genovese et al., 2019] proposed a 3-layer network called PalmNet, which combines Gabor filters and PCA. During training, in the first layer, a subset of Gabor filters is selected based on the highest magnitude responses of the most frequent orientations. In the second layer, PCA was used based on the responses from the first layer. In the third layer, a threshold was applied to the responses from the second layer (binarized). For feature extraction, the binarized responses from the third layer were converted into decimal numbers and histograms from non-overlapping blocks were calculated. For recognition, Euclidean distance between two feature vectors was used. The authors compared their method with local descriptors based methods on four contactless palmprint databases.

Other works [Dian and Dongmei, 2016], [Jalali et al., 2015] performed a rather ambiguous evaluation. Dian and Dongmei [Dian and Dongmei, 2016] used AlexNet for extracting features from palm ROIs and Hausdorff distance for matching but they did not compare with other palmprint recognition methods. Jalali et al. [Jalali et al., 2015] tried to address deformation invariance by using a small CNN but the evaluation was only done on a very small dataset of 200 palmprint images (28 by 28 pixels) from 10 persons only and the CNN was trained using 85% of the dataset.

2.2.6 Experiments and Results in Palmprint Recognition Studies

2.2.6.1 Performance Evaluation

There are no standard, unified benchmark evaluation protocols for palmprint recognition algorithms and databases. Thus, researchers evaluate algorithms using different protocols. Moreover, to compare with other algorithms, they need to select suitable, representative algorithms and often need to re-implement them if no source codes are available.

The outcome of the absence of unified protocol is that very different results are reported for the same algorithm evaluated on the same database. For example, let us consider results

reported for CompCode on IITD touchless palmprint database in different studies. In the recent palmprint review, Fei et al. [Fei et al., 2019a] reported rank-1 accuracy and EER of 77.79% and 0.101%. However, in [Fei et al., 2016b] and [Fei et al., 2019b], Fei et al. reported better results. Zheng et al. [Zheng et al., 2016] reported rank-1 accuracy and EER of 98.95% and 1.0%, respectively. Luo et al. [Luo et al., 2016] reported rank-1 accuracy and EER of 88.16% and 7.72%, respectively. It is hard to explain these differences.

It is even harder to explain the same results for the same methods on the same database when different protocols are used. For example, Fei et al. [Fei et al., 2016b] performed verification experiment on IITD touchless palmprint database reporting 2300 hand images in the database. In the evaluation protocol, the authors reported that there were 4600 genuine comparisons and 100,970 imposter comparisons. Xu et al. [Xu et al., 2018] also performed verification experiment on IITD touchless palmprint database reporting 2600 (300 more than in [Fei et al., 2016b]) hand images in the database. In the evaluation protocol, the authors reported that there was the same number of genuine comparisons and 2,369,250 (2,268,280 more than in [Fei et al., 2016b]) imposter comparisons. Nevertheless, the EERs in [Fei et al., 2016b] and [Xu et al., 2018], for six different methods (CompCode, OrdinalCode, FusionCode, BOVC, DOC and RLOC) are the same.

Another reason for the differences in experimental results is discarding images from databases. For example, Zheng et al. [Zheng et al., 2016] and [Kumar, 2019] did not use 239 (4% of images in the database) images from CASIA database in their experiments, without indicating it and providing an explanation. Genovese et al. [Genovese et al., 2019], explicitly indicated the percentage of discarded ROIs due to the ROI extraction failure. Depending on the database, from approximately 0.5% even up to 13% of images were not used in the performance evaluation.

in the same study, researchers sometimes use different protocols for identification and verification experiments. For example, verification in [Fei et al., 2016b, Fei et al., 2016a, Xu et al., 2018] was performed using leave-one-out procedure, which means that

all samples from a palmprint database except the probe are enrolled in the gallery. For identification, the authors [Fei et al., 2016b, Fei et al., 2016a, Xu et al., 2018] used from 1 to 4 samples for enrollment. Genovese et al. [Genovese et al., 2019] did the opposite. For verification, the authors used 4 samples for enrolment and identification was performed as a leave-one-out procedure. Some studies do not report verification [Zhang et al., 2017] or identification results [Morales et al., 2010, Wu et al., 2014] without an explanation.

2.2.6.2 Baseline Selection

Different studies use different contactless palmprint databases and baseline algorithms for comparison. Kumar et al. [Kumar, 2019] selected only two methods for comparison on IITD, CASIA and newly introduced PolyU-IITD contactless palmprint databases. Zheng and Kumar [Zheng et al., 2016] selected IITD, CASIA and PolyU 2D contactless palmprint databases and three coding-based methods for comparison. [Zheng et al., 2016] Kumar [Kumar, 2019] and Zheng and Kumar [Zheng et al., 2016] selected 1150 and 1300 right hand images from IITD database, respectively. Genovese et al. [Genovese et al., 2019] selected four databases IITD, CASIA, REST and Tongji. The authors did not compare with any coding-based methods. Tongji database has 12,000 images, but Genovese et al. [Genovese et al., 2019] claimed it has 6000 images and used only half of Tongji database in experiments. In [Fei et al., 2016b, Xu et al., 2018], the authors used one contactless palmprint database (IITD) and compared with only coding-based methods. Luo et al. [Luo et al., 2016] used IITD and Cross Sensor contactless palmprint databases. Jia et al. [Jia et al., 2017] used one contactless palmprint database (HFUT). In the survey, Fei et al. [Fei et al., 2019a] claimed to select 13 state-of-the-art methods and reported their identification results on 3 contactless palmprint databases. In [Fei et al., 2019b], Fei et al. claimed to select 12 state-of-the-art methods and reported their identification results on 4 contactless palmprint databases. In [Fei et al., 2019a] and [Fei et al., 2019b] only

4 methods (CompCode, OrdinalCode, E-BOCV and LLDP) and 2 databases (IITD and CASIA) overlapped.

2.2.6.3 Current State

It is hard to assess the current state of contactless palmprint recognition studies because of the reasons listed in Sections 2.2.6.1 and 2.2.6.2. According to Fei et al. [Fei et al., 2019a], contact-based palmprint recognition have already achieved high accuracy comparing to contactless palmprint recognition and further research on palmprint recognition should focus on the contactless one. Nevertheless, from the accuracy perspective, rank-1 and EERs reported in various studies indicate that the palmprint recognition methods have already achieved very high recognition accuracy also in contactless environment. Table 2.1 aims to show the current state and progress from the accuracy perspective. Table 2.2 shows an example of the performance differences, which may affect the assessment of the current state and progress.

Table 2.1 shows the highest rank-1 accuracy and the lowest EERs reported in different studies which performed the evaluation on contactless palmprint databases.

Table 2.2 shows differences in reported rank-1 accuracy of the proposed and overlapping methods (OrdinalCode, CompCode, LLDP, HOL) on HFUT contactless palmprint database in two recent studies that claimed the state-of-the-art performance. EERs are not given in Table 2.2 because [Fei et al., 2019b] did not report EERs. [Jia et al., 2017] and [Fei et al., 2019b] used respectively 4 and 10 samples from the first session to form a gallery and all samples from the second session were used as probe samples. Note that in [Fei et al., 2019b] more samples are used to create a gallery and the same number of samples are used as probes. First, in [Jia et al., 2017] the rank-1 accuracy of overlapping methods is higher than the corresponding ones in [Fei et al., 2019b]. Second, the proposed method rank-1 accuracy in [Fei et al., 2019b] is lower than the proposed one in [Jia et al., 2017]. Third, the rank-1 accuracy differences between methods in a given study are smaller

Table 2.1: The highest rank-1 accuracy (%) and the lowest EER (%) from different studies on four contactless palmprint databases.

Study	Method	IITD		CASIA		Tongji		PolyU 2D	
		rank-1	EER	rank-1	EER	rank-1	EER	rank-1	EER
[Genovese et al., 2019]	PalmNet	99.37	0.52	99.77	0.72	99.83	0.16	-	-
[Fei et al., 2019b]	DDBC	96.4373	-	96.4085	-	98.7333	-	-	-
[Fei et al., 2019a]	review	96.22	0.028	95.21	0.029	-	-	-	-
[Kumar, 2019]	proposed*	-	1.17	-	0.126	-	-	-	-
[Xu et al., 2018]	DRCC	90.87	0.0548	-	-	-	-	-	-
[Zhang et al., 2017]	CR-CompCode	-	-	-	-	98.78	-	-	-
[Luo et al., 2016]	LLDP	92.00	4.07	-	-	-	-	-	-
[Zheng et al., 2016]	DoN	99.15	0.68	99.3	0.53	-	-	100	0.22
[Svoboda et al., 2016]	proposed*	-	1.64	-	1.86	-	-	-	-
[Fei et al., 2016a]	LMDP	96.20	0.0264	-	-	-	-	-	-
[Aykut and Ekinci, 2015]	proposed*	99.4	0.03	-	-	-	-	-	-
[Wu et al., 2014]	SIFT_RANSAC	-	0.4252	-	0.3969	-	-	-	-
[Kanhagad et al., 2011a]	CompCode	-	-	-	-	-	-	98.3	1.22
[Morales et al., 2010]	SIFT_OLOF	-	0.2	-	-	-	-	-	-

*The methods proposed by the authors of the corresponding study. No name was given to a method. Thus, in this table, "proposed" is used.

than rank-1 accuracy differences between these two studies.

Table 2.2: Rank-1 accuracy (%) reported in two studies for four overlapping methods and the proposed by the authors methods on HFUT contactless palmprint database.

Study	OrdinalCode	CompCode	LLDP	HOL	Proposed
[Jia et al., 2017]	99.600	99.637	99.888	99.275	99.913
[Fei et al., 2019b]	97.0625	96.9750	97.2500	95.9000	97.8625

Chapter 3

Wrist Identification for Forensic Application¹

3.1 Overview

Criminal and victim identification based on evidence images is an important part of forensic investigation. Although digital images are a useful clue for law enforcement agencies, suspects can intentionally hide their faces and tattoos to prevent identification. Existing identification methods based on vein, skin mark, height, skin color, weight, race, etc., have been considered for solving this problem. Nevertheless, as each biometric trait, they have some limitations. The soft biometric traits, including skin color, gender, height, weight and race, provide useful information but not distinctive enough. Veins and skin marks are limited to high-resolution images and some body sites may neither have enough skin marks, hair nor clear veins.

Even though hands are observable in some evidence images [Spaun, 2007], hand biometrics has been mostly studied for commercial applications in well-controlled environments. In this chapter, I propose to use wrist for criminal and victim identification. Terrorists and rioters tend to expose their wrists in a gesture of triumph, greeting or salute.

¹This chapter is based on [Matkowski et al., 2019b]

Child sexual offenders may expose their wrists when using a hand to abuse a victim in child sexual abuse materials. However, wrists were neglected by the biometric community for forensic applications. According to my best knowledge, it is the first study on wrist identification for forensic application based on images.

In this chapter, a wrist identification algorithm, which includes skin segmentation, key point localization, image to template alignment, large feature set extraction, and classification, is proposed. The proposed algorithm is evaluated on the collected NTU-Wrist-Image-Database-v1, which consists of 3945 images from 731 different wrists, including 205 pairs of wrist images collected from the Internet, taken under uneven illuminations with different poses and resolutions.

The rest of this chapter is organized as follows. In Section 3.2, a new database NTU-Wrist-Image-Database-v1 is presented. In Section 3.3, the proposed algorithm is provided in details. In Section 3.4, different experiments are performed to evaluate the proposed algorithm. In Section 3.5, the summary is given.

3.2 NTU Wrist Image Database

Wrist images collected in Singapore were taken during two occasions with more than one week time interval. The images were cropped such that nothing but wrist area from the palmar side is visible². The widths and heights of the wrist images vary and on average they have 525 by 320 pixels. They are from left and right wrists of Asians - Chinese, Indian, Malay and also some Caucasian and Eurasian subjects. To increase the number of images in experiments, right hand wrists images were flipped such that they emulate left hand wrists from different subjects. The dataset was divided into three exclusive subsets. Gallery set SET1 includes 1948 images from 526 different wrists from 320 subjects taken

²Note, that the wrist detection is out of scope in this study. There are some studies focusing on body parts detection in controlled [Huynh et al., 2014] or uncontrolled environments [Oliveira et al., 2016]. However, this work aims to show that wrist is a useful clue for identification.

during the first session. Probe set SET2 includes 1452 images from 397 different wrists from 282 subjects, taken during the second session. In both sessions, wrist images were taken in a frontal (palmar) pose without strict pose requirements and they are considered as standard pose images. Non-standard set SET3 includes 135 images from 133 different wrists of 84 subjects taken in varying conditions with viewpoint, pose and out of focus (OOF) variations. Images in SET2 and SET3 are a subset of images in SET1. Images in SET1, SET2 and SET3 were taken by Canon EOS 500D or NIKON D70s cameras. To evaluate the proposed algorithm in more challenging cases, which are more similar to forensic environment, two additional datasets SET4 and SET5 were built. Gallery SET4 includes all images from SET1 and 205 additional images downloaded from the Internet (SET1P). Probe set SET5 contains 205 corresponding Internet wrist images. The wrists in the Internet images usually have low resolution and their sizes vary from 21 by 12 to 196 by 126 pixels. Most of them have less than 100 by 50 pixels. Some examples of the standard, non-standard pose and Internet images are shown in Figure 3.1 and the dataset details are given in Tables 3.1 and 3.2. SET4 and SET5 are omitted in Table 3.2 because SET4 is SET1 with additional 205 different wrist images and SET5 contains corresponding 205 wrist images. Figure 3.2 shows examples of original images downloaded from the Internet and the corresponding wrists. The NTU-Wrist-Image-Database-v1 will be available online³ for research purpose.

³https://github.com/BFLTeam/NTU_Dataset

Figure 3.1: Examples of corresponding wrist images in the NTU-Wrist-Image-Database-v1: (a) and (c) Standard pose gallery images, (b) Standard pose probe images, (d) Non-standard pose probe images, (e) Internet gallery images and (f) Internet probe images. Original scales are preserved in (e) and (f). Images in (a), (b), (c) and (d) were resized down for better visualization. (a) and (b), (c) and (d), (e) and (f) show image pairs from the same wrists.

Figure 3.2: Examples of resized down original Internet images from the NTU-Wrist-Image-Database-v1. Wrists are highlighted in the original images, cropped and enlarged for better visualization. The scale is preserved between corresponding wrist images.

Table 3.1: Details of the datasets in the NTU-Wrist-Image-Database-v1.

Set	No. images	No. wrists	No. subjects	Description
SET1	1948	526	320	Gallery standard pose images
SET2	1452	397	282	Probe standard pose images
SET3	135	133	84	Non-standard images with illumination, pose, OOF variations
SET4	2153	731	505	SET1 + Internet images SET1P (205 images from 205 wrists)
SET5	205	205	185	Internet images

Table 3.2: Number of images for one wrist in SET1, SET2 and SET3.

Images Wrists	1	2	3	4	5	6	7	8
SET1	6	86	37	361	5	29	0	2
SET2	4	62	23	295	5	7	0	1
SET3	131	2	0	0	0	0	0	0

3.3 The Proposed Wrist Identification Algorithm

The WMFA (Wrist Matcher for Forensic Applications) algorithm consists of four main steps. First, for image pre-processing, a skin segmentation step employing superpixels and ensemble of decision trees is applied (Section 3.3.1). Then, wrist region of interest (ROI) is found by using a ROI extraction scheme (Section 3.3.2). Then, large feature sets are extracted (Section 3.3.3). Next, a one against all classifier-based matching scheme is used (Section 3.3.4). On top of these, there is a post-recognition score analysis which aims to improve the overall accuracy by combining four WMFAs with different settings to build the whole WMM (Wrist Meta-Matching) system (Section 3.3.5).

3.3.1 Wrist Segmentation

The WMFA algorithm is fully automated except segmentation for non-standard images and does not require any manual and time consuming human supervision. For establishing large forensic datasets, automatization is required to effectively collect and process images from prisoners or suspects, but for crime scene images, manual correction is acceptable. Currently, law enforcement agencies also use a semi-automatic approach to handle latent prints. Since the region of interest (ROI) is a wrist, wrist skin pixels are segmented for further analysis. Segmentation based on a simple threshold may not be ideal because the skin of different population has diverse color. Even in the same wrist image, skin color can be different significantly because palms and upper wrist regions have less concentration of melanin, comparing with lower wrist regions and forearms, especially for brown, dark brown and black skin persons, i.e., Fitzpatrick scale V and VI. Spatial arrangement and patch based information can increase the accuracy of skin segmentation [Kruppa et al., 2002] and therefore, superpixel methods, which group pixels into perceptually meaningful patches, are used. Simple linear iterative clustering (SLIC) is a popular and fast method that has desirable properties: superpixels adhere well to boundaries and they improve segmentation results [Achanta et al., 2011]. Thus, SLIC is employed for this study. Before

applying SLIC, images are resized such that their heights are fixed at 200 pixels and their widths are varied to preserve aspect ratio. Then, SLIC method is used to generate 200 superpixels for each image. Within each superpixel and 8 neighbors, mean and standard deviation (statistics) values from RGB, HSV, LAB, YCbCr, YIQ, normalized RGB color spaces and seven gradient maps (Sobel in two directions, Prewitt in two directions, Laplacian, Difference of Gaussians, Laplacian of Gaussians) are extracted to form 450-dimensional feature vectors⁴. To segment skin superpixels, an ensemble of decision trees (EoDT) is trained with the bagging method. EoDT classifies each superpixel into either skin or non-skin class, resulting in a segmented wrist image. The wrist segmentation scheme is illustrated in Figure 3.3.

Figure 3.3: The wrist segmentation scheme.

3.3.2 ROI Extraction

This section aims to extract a common ROI from the segmented wrists. At the beginning, it should be emphasized that the wrist area is not well defined. It is located in between hand and forearm. Human hand has distinctive and common key points, i.e. hand shape allowing efficient palm ROI extraction for commercial palmprint and palm vein applications [Wu

⁴Dimension of one superpixel description: $[6 \text{ color spaces}] \times [3 \text{ channels}] \times [2 \text{ statistics}] + [7 \text{ gradient maps}] \times [2 \text{ statistics}] = 50$. Because each superpixel is represented by itself and 8 neighbors then the feature vector size of one superpixel is: $50 + 50 \times 8 = 450$.

et al., 2014, Kang and Wu, 2014]. But this work is for forensics so most of the hand information is assumed not available which makes ROI extraction more challenging. In real applications, hands of criminals can be occluded by weapons, e.g., guns, when they hold them. Pose variation is caused by different posing among different subjects and scale variation is caused by uncontrolled imaging environments.

To address these pose and scale issues, a robust ROI extraction scheme is developed. First, since the images are segmented in the previous section, lower and upper boundaries (see Figure 3.3), which will serve as key points, are easily extracted. Based on the observation of the wrists, they usually have around two prominent wrinkles. The ROI of the wrists is bounded by the lower and upper boundaries in the horizontal direction and has two wrinkles in the vertical direction (see Figure 3.5). To find ROI, a two stage scheme is proposed for detecting the wrinkle and boundary points. In the first stage, to extract vertical lines, the segmented wrist image channels I_c , where c index denotes R, G and B channels of the segmented image I are convolved with the Sobel operator in the vertical direction obtaining G_{rx} , G_{gx} , G_{bx} , respectively. Note that pixel values in G_{rx} , G_{gx} , G_{bx} are normalized between 0 and 1. The final inverted gradient image is obtained by selecting maximum response for each pixel as follows $G_x = 1 - \max(|G_{rx}|, |G_{gx}|, |G_{bx}|)$. This operation enhances vertical wrinkles as shown in Figures 3.5a and 3.5d. Pre-processed image G_x is resized to have 40-pixel height. Now, a small directed sparse graph $H = (V, E)$ is built with Moore neighbourhood connectivity (8 neighbours) within the wrist, which is actually a quasi-King's graph. The graph is based on the gradient image of wrist pixels. In other words, a graph node exists at some position (i, j) if there is a skin pixel at that position represented by mask image $M(i, j) = 1$. The distance to reach the node from its neighbour is equal to a gradient image value at the position. To regulate distances in diagonal directions in point 6 of the Algorithm 1, assuming that one pixel is a unit square and following Pythagorean Theorem, $r = \sqrt{2}$ is used. Figure 3.4 illustrates the graph. This stage is described in detail in Algorithm 1.

Algorithm 1 Build graph

1: **procedure** BUILDGRAPH(I, M)

2: $G_{cx} \leftarrow \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \star I_c$ ▷ Convolve images

3: $G_x \leftarrow 1 - \max(|G_{rx}|, |G_{gx}|, |G_{bx}|)$

4: Resize G_x to 40 pixel height

5: $H = (V, E)$, it has vertices in the foreground: $V \in \{v_{i,j} | M(i, j) = 1\}$. It has edges between eight closest pixel neighbours: $E\{(v_{i,j}, v_{i+m,j+n})\}, m, n = -1, 0, 1 \wedge m \neq n \neq 0$ ▷ Define connections

6:

$$w(v_{i,j}, v_{i+m,j+n}) = \begin{cases} G_x(i+m, j+n) \cdot r & |m| = |n| \wedge m \neq n \neq 0 \\ G_x(i+m, j+n) & otherwise \end{cases}$$

▷ Define weights

7: **return** H ▷ Sparse graph H

8: **end procedure**

Figure 3.4: Illustration of graph edges and weights between vertices determined based on the gradient image.

Next stage aims to find two prominent wrinkles. Given that the graph represents an energy map, wrinkles are believed to be localized such that they contain minima points between two boundaries. By taking advantage of the small sparse graph representation generated by Algorithm 1, the shortest paths P_1 and P_2 between the upper boundary b_{up} and the lower one b_{down} are found using Johnson's algorithm, which searches the shortest paths between all pairs of vertices and is applicable for sparse graphs [Cormen et al., 2009]. The upper boundary b_{up} determines a set of start nodes S and the lower boundary b_{down} determines destination nodes F . The first shortest path is found in the original graph H . Since the second shortest path is expected not to contain the nodes in P_1 or its closest neighbourhood, P_2 is found in a subgraph. The subgraph is obtained by subtracting P_1 and its von Neumann neighbours (four adjacent nodes) from the graph H , to make sure that P_1 and P_2 do not overlap. This stage is described in Algorithm 2. Algorithm 2 may be run with (*adjust* in stage 4 and 10 of Algorithm 2 is set to be true) or without the procedure described in Algorithm 3 inside, which are denoted respectively by Proc2/3 and Proc2. Proc2 has a hard constraint on the start point and destination point which have to be localized on boundaries b_{up} and b_{down} . However, observing wrist wrinkles, one can

say that they are not always situated from boundary to boundary. To detect such wrinkles, Algorithm 3 is also proposed. Proc2/3 employing both Algorithms 2 and 3 aims to relax the constraint by introducing the normalized path and searching for the minimal one between shifted down the start nodes and shifted up the destination nodes which are controlled by a parameter a . Parameter a determines the length of the wrinkles, such that higher a results in shorter wrinkles. The difference between wrinkle-like key points returned by Proc2 and Proc2/3 is presented in Figures 3.5 and 3.7.

Algorithm 2 Detect wrinkles

```

1: procedure DETECTWRINKLES( $H, b_{up}, b_{down}, a, adjust$ )
2:    $S \in \{s_i | s_i \in b_{up}\}, F \in \{f_j | f_j \in b_{down}\}$       ▷ Define nodes based on boundaries
3:   Apply Johnson's algorithm on  $H$  to get all shortest paths  $P_{All}$ 
4:   Among all the shortest paths  $P_{All}$  find the shortest one  $P_1$  between  $S$  and  $F$ 
5:   if  $adjust = \text{true}$  then
6:      $P_1 \leftarrow \text{ADJUSTPATH}(P_1, P_{All}, b_{up}, b_{down}, a)$       ▷ Run Algorithm 3
7:   end if
8:    $V_{sub} \leftarrow ((V - P_1) \ominus B) \cup S \cup F$ 
9:    $H = (V_{sub}, E)$       ▷ Graph modification
10:  Apply Johnson's algorithm on  $H$  to get all shortest paths  $P_{All}$ 
11:  Among all the shortest paths  $P_{All}$  find the shortest one  $P_2$  between  $S$  and  $F$ 
12:  if  $adjust = \text{true}$  then
13:     $P_2 \leftarrow \text{ADJUSTPATH}(P_2, P_{All}, b_{up}, b_{down}, a)$       ▷ Run Algorithm 3
14:  end if
15:  return  $P_1, P_2$       ▷ Two wrinkles
16: end procedure

```

In the Procedures, a structuring element B is set up to be a 3x3 cross. Parameter a is set up to 0.2. Operator $\ominus$ denotes morphological erosion, $\cup$ is a union of sets, $\#$ is the number

Algorithm 3 Adjust the path

```
1: procedure ADJUSTPATH( $P_a, P_{All}, b_{up}, b_{down}, a$ )
2:    $d_p \leftarrow \frac{1}{n} \sum_{v \in P_a} w(v_k, v_{k+1}), k = 1, 2, \dots, n + 1. \#P_a = n + 1$  ▷ Define
   normalized path length
3:   for  $i \leftarrow 1$  to  $\lfloor n \cdot a \rfloor$  do
4:     for  $j \leftarrow 1$  to  $\lfloor n \cdot a \rfloor$  do
5:       Among all the shortest paths  $P_{All}$  find the shortest one  $P_s$  between  $b'_{up,i}$ 
       and  $b'_{down,j}$ 
6:       if  $\frac{1}{n-i-j} \sum_{v \in P_s} w(v_k, v_{k+1}) < d_p, k = i + 1, \dots, n - j$  then
7:          $d_p \leftarrow \frac{1}{n-i-j} \sum_{v \in P_s} w(v_k, v_{k+1})$ 
8:          $P_a \leftarrow (v_{i+1}, \dots, v_{n-j})$ 
9:       end if
10:    end for
11:  end for
12:  return  $P_a$  ▷ Adjusted path
13: end procedure
```

of elements in the set and $\lfloor \cdot \rfloor$ is an operator of a floor function. Points $b'_{up,i}$ and $b'_{down,j}$ are start and destination nodes shifted in vertical direction by respectively i pixels down and j pixels up. Using Proc2, a wrist template is estimated based on all extracted key points from the training images in SET1. More precisely, a heat map is obtained by finding the boundaries b_{up} and b_{down} and the two vertical wrinkles P_1 and P_2 in each training image and summing them up. Using the heat map (Figure 3.6a) as a reference, template (Figure 3.6b) for ROI extraction is proposed. By representing the template consisting of extracted wrist boundaries and wrinkles as 2D points, Coherent Point Drift (CPD) registration method, which fits the Gaussian mixture model centroids from one set to another set of points by maximizing likelihood, is employed to find the correspondence between template key points and wrist key points in input images under affine transformation. Detailed description of the CPD method can be found in [Myronenko and Song, 2010]. After applying CPD to align images to the template, a ROI is found by cropping columns of pixels from both, left and right sides of the image. In the next step, a ROI is constrained by two horizontal lines, whose positions are defined based on the number of foreground pixels in a row. If more than three quarters of pixels in a row are foreground pixels, then the entire row is considered as a foreground, otherwise a background. As it was mentioned before, there are two procedures Proc2 and Proc2/3 that produce slightly different key points and therefore different alignments and ROIs are expected. ROI#1 and ROI#2 denote the ROIs extracted by Proc2 and Proc2/3, respectively. These two procedures relax the assumptions on the wrist wrinkles and provide more flexible representation because each wrist can be now represented by ROI#1 and ROI#2. Extracted key points are shown in Figures 3.5b, 3.5c, 3.5e, 3.5f. The ROI extraction step is illustrated in Figure 3.7, which also presents how one can obtain different ROIs using different key point localization procedures.

Figure 3.5: (a), (d) gradient maps from two wrist images. (b), (e) key points extracted using Proc2 from (a), (d). (c), (f) key points extracted using Proc2/3 from (a), (d). Red dots are boundaries while green dots are extracted wrinkles.

Figure 3.6: (a) Heat map representing sum of the key points from SET1 images. (b) wrist template is built by taking the reference of the heat map.

Figure 3.7: ROI extraction steps using both Proc2 and Proc2/3 in order to produce respectively ROI#1 and ROI#2.

3.3.3 Feature Extraction

After image alignment is performed and ROI is obtained, features, including local binary patterns (LBP) in each RGB channel, Gabor orientation field histograms and dense scale invariant feature transform (DSIFT) on a grayscale image, are extracted. These features have been successfully applied to biometric applications [Ahonen et al., 2006, Su and Kong, 2014, Wang et al., 2010a]. LBP features, including $LBP_{D,R}^{riu2}$ and uniform $LBP_{D,R}^{u2}$, are useful for texture recognition. $LBP_{D,R}^{riu2}$ is invariant to rotation and monotonic grey level changes. Uniform $LBP_{D,R}^{u2}$ has at most two bitwise transitions from 0 to 1 or 1 to 0 when considered as circular patterns and is invariant to monotonic grey level changes [Ojala et al., 2002]. Gabor filters capture line segments [Su and Kong, 2014] and DSIFT describes shape related information [Wang et al., 2010a]. Raw $LBP_{D,R}$ is obtained as:

$$LBP_{D,R} = \sum_{p=0}^{D-1} \xi(g_p - g_c) 2^p \quad (3.1)$$

$$\xi(g_p - g_c) = \begin{cases} 1 & g_p \geq g_c \\ 0 & g_p < g_c \end{cases} \quad (3.2a)$$

$$(3.2b)$$

where D is the number of sampling points, R is the radius of sampling circle, g_c is the centre pixel value and g_p is its neighbour pixel value. Gabor orientation field features are

defined the same as in [Su and Kong, 2014]. These features are obtained by convolving image I with Gabor filters G as follows:

$$O(x, y) = \arg_{\theta_k} \max_{m,k} |G_{\lambda_{mk}, \theta_l, \sigma_m, \lambda} \star I| \quad (3.3)$$

where $G_{\lambda_{mk}, \theta_l, \sigma_m, \lambda}$ is a Gabor filter with an orientation $\theta_k = k\pi/8$, sinusoidal component of the wavelength λ_{mk} , standard deviation of the elliptical Gaussian window σ_m , spatial aspect ratio λ , scale indices m and orientation indices k . DSIFT features are extracted by a SIFT descriptor on a regular grid controlled by pixel step, spatial bin size and area where it should be extracted. Detailed SIFT description can be found in [Lowe, 2004]. To preserve spatial information, ROI is divided into many blocks forming a grid structure. Seven different grids $b_1, b_2, b_3, b_4, b_5, b_6, b_7$ shown in Figure 3.8 are proposed to extract Gabor and LBP histograms from each block. Each DSIFT feature is extracted from a specified fixed block. First, a part of the feature vector f_{LBP} is extracted on R, G and B channels and consists of the concentrated histograms from $LBP_{8,1}^{riu2}$ on b_1, b_2 and $LBP_{8,2}^{u2}$ on b_3, b_4, b_5, b_6, b_7 . Two different LBP features are used because applying $LBP_{8,2}^{u2}$ on the first two grids which have small blocks would make the histogram sparse. On the other hand, applying rotational invariant $LBP_{8,1}^{riu2}$ on bigger blocks would cause the loss of discriminative directional information in already aligned images. Gabor filters with 16 different orientations and 4 scales $s \in \{0.2, 0.5, 0.7, 0.9\}$ on all grids are used to extract the second part of the feature vector f_{Gabor} . Moreover, DSIFT features, which are extracted with 16 pixel step and spatial bin size of 16 are added as feature vector f_{DSIFT} . Feature vectors $f_{LBP}, f_{Gabor}, f_{DSIFT}$ dimensions are 13074, 2112 and 1280 respectively. The final 16466-dimensional feature vector $f = \{f_{LBP}, f_{Gabor}, f_{DSIFT}\}$ is used to represent each ROI of the wrist image.

3.3.4 Wrist Matching

Partial least squares (PLS) regression is a statistical method that employs latent variables to find the relations between two mean centred and variance scaled matrices X and y . In

Figure 3.8: Seven different grids (a) b_1 , (b) b_2 , (c) b_3 , (d) b_4 , (e) b_5 , (f) b_6 , (g) b_7 .

this study, y is in fact a vector. It uses the properties of nonlinear iterative partial least squares (NIPALS) algorithm to calculate the principal components iteratively [Geladi and Kowalski, 1986]. PLS decomposes matrices X and y into:

$$X = TP^{\top} + E \quad (3.4)$$

$$y = UQ^{\top} + F \quad (3.5)$$

where T and U are score matrices, P and Q are loading matrices and F and E are residual matrices and $^{\top}$ denotes transpose operator. In order to build classifiers, the regression coefficients β are found such that matching score can be obtained by $\hat{y} = \beta^{\top}x + \bar{y}$ where $\hat{y}$ is a classifier response or matching score, x is an input feature column vector and $\bar{y}$ is the sample mean. Detailed description of PLS and the computation of β coefficients using NIPALS algorithm can be found in [Geladi and Kowalski, 1986] and [Rosipal and Krämer, 2006] respectively.

Wrist identification can be performed as a multiclass classification. One-against-all classifiers are trained, which means that every wrist in the gallery set has its own unique classifier. All images in the gallery set are used to build the classifiers. In the training, label $y_i = 1$ when the images belong to a particular wrist and $y_i = -1$ otherwise. y_i is

the i^{th} element of y . In the experiments, PLS with $k = 5$ latent dimensions is used and compared to Support Vector Machine (SVM).

3.3.5 Meta-Recognition

To boost the matching performance by taking advantage of the two ROIs, a Meta-recognition approach is used. Here, the notion of a recognition system refers to the WMFA algorithm under a specific set of parameters described in Section 3.3.2 while the notion of Meta-recognition refers to analyse matching scores to select the best system for each input wrist. Because there are two different procedures, Proc2 and Proc2/3 for ROI extraction, each image after alignment has ROI#1 and ROI#2, which are represented by two different feature vectors. It allows building two different PLS as well as two different SVM classifiers. In practice, it means that each gallery wrist has four different classifiers with input features extracted from the two ROIs. In other words, there are four different recognition systems RS_{PLS1} , RS_{PLS2} , RS_{SVM1} , and RS_{SVM2} . One could also consider different techniques to utilize the four different recognition systems, such as ensemble learning or score fusion. Differences in particular steps in the recognition systems are presented in Table 3.3.

The Meta-recognition step analyses matching scores from each of the recognition systems and decides which one of them is the most likely correct. Scheirer et al. proposed a Meta-recognition system based on Extreme Value Theorem (EVT) also known as the Fisher-Tippett Theorem [Scheirer et al., 2011]. The key idea behind assumes that there is a sufficient number of nonmatch samples in order to model the non-matched score distribution and the matched score is regarded as an outlier of the distribution. They also showed that modelling the nonmatch scores in a tail is an extreme value problem. Thus, the tail of the distribution can be modelled by one of extreme value distributions. According to [Scheirer et al., 2011], Weibull distribution is the most suitable one for modelling nonmatch distribution in statistical Meta-recognition. Its cumulative distribution function is expressed

by: $F(x) = 1 - \exp(-(\frac{x}{a})^b)$ for $x \geq 0$ and $F(x) = 0$ for $x < 0$. where b is a shape parameter and a is a scale parameter. In the experiment, there are always $m - 1$ nonmatch responses from m different wrist classifiers and one match response from the correct one. Since it satisfies the assumption for modelling nonmatch distribution, Weibull-Based Statistical Meta-Recognition is adapted to run on top of our four different RS_i . In addition, there are three Meta-WMFA recognition systems for further comparison: RS_{PLS} , RS_{SVM} and $RS_{(PLS+SVM)} = WMM$ which details are presented in Table 3.3. System RS_{PLS} refers to WMFA algorithms with PLS-based matching step; RS_{SVM} refers to the WMFA algorithms with SVM-based matching step and WMM is the Meta-recognition system that considers all combinations of the WMFA algorithms. A schematic diagram of the systems is presented in Figure 3.9. To choose the best RS_{best} from four RS_i for each input wrist image, I propose a modification of Rank-1 Statistical Meta-Recognition Algorithm from [Scheirer et al., 2011] as shown in Algorithm 4. It fits Weibull distribution to $l_t - 1$ highest matching scores from each recognition system skipping the top matching score $\hat{y}_{i,1}$ and determinates its parameters. Then, from the cumulative distribution function (CDF) F_i , one can answer the question what is the probability that the top matching score does not come from the distribution. Based on that, the preferable RS_{best} would be the one with the highest value of its CDF calculated at the top matching score $\hat{y}_{best,1}$. Thus, the final prediction scores s of the system are the scores returned by RS_{best} . In experiments, the length of the tail l_t is experimentally set up to be a function of the gallery size $l_t = 0.5 \cdot m$, where m is the number of the classifiers in the gallery.

3.4 Experiments

3.4.1 Recognition Methods

The WMFA algorithms and Meta-WMFA systems were evaluated under several aspects and compared with six representative biometrics recognition methods designed for palm-

Algorithm 4 Modified Rank-1 Statistical Meta-Recognition

```

1: procedure METARECOGNITION( $\hat{Y}_i$ )           ▷ Sorted in descending order scores
    $\hat{Y}_i = (\hat{y}_{i,1}, \dots, \hat{y}_{i,m})$  from  $n$  recognition systems  $RS_1, \dots, RS_n \in RS$ , where  $1 \leq i \leq n$ 
2:   for  $i \leftarrow 1$  to  $n$  do
3:     Fit Weibull distribution  $F_i$  to  $l_t - 1$  largest scores:  $\hat{y}_{i,2}, \dots, \hat{y}_{i,l_t}$ 
4:   end for
5:    $k \leftarrow \underset{i}{\operatorname{argmax}} [F_i(y_{i,1})]$            ▷ Find  $RS_{best}$  index  $k$ 
6:   return  $s \leftarrow \hat{Y}_k$            ▷ Matching scores
7: end procedure
  
```

Table 3.3: Recognition and Meta recognition systems details.

Meta-WMFA	Meta-WMFA	WMFA	ROI	Matching
WMM $RS_{PLS+SVM}$	RS_{PLS}	RS_{PLS1}	ROI#1	PLS
		RS_{PLS2}	ROI#2	PLS
	RS_{SVM}	RS_{SVM1}	ROI#1	SVM
		RS_{SVM2}	ROI#2	SVM

 Figure 3.9: The schematic diagram of the whole Wrist Meta Matching system (WMM) and other Meta-WMFA systems. Each RS block represents a WMFA algorithm.

print, palm vein or fingerprint, OrdinalCode [Sun et al., 2005], double-orientation code (DOC) [Fei et al., 2016b], SIFT-based method (SIFT_RANSAC) [Wu et al., 2014], Mutual Foreground Based LBP with χ^2 distance (MF_LBP) [Kang and Wu, 2014], Deep Scattering Convolutional Network (DSCN) and Translation Invariant Scattering Network (TISN). The methods were selected for the following reasons. 1) According to the best of our knowledge, this is the first work on wrist identification based on color images. 2) Palmprint nearby the wrist also has wrinkles and seems to be more similar than other body parts. 3) Sun et al.'s Ordinal Code is a well-known method coding-based method which uses Hamming distance for comparison. 4) Fei et al.'s DOC method achieves the highest accuracy among coding based approaches in palmprint verification and identification. 5) Wu et al.'s SIFT-based method is applicable to contactless palmprint verification and does not require contact image acquisition devices, which has some similarity to our study. 6) Kang et al.'s Mutual Foreground (MF)-Based LBP with χ^2 distance is developed for contactless palm veins. It uses a mutual foreground between two images which is extracted based on maximal principal curvatures and similarly to WMFA algorithm, LBP is employed as features. Note that in this work, the feature vector from LBP features is different than in (MF_LBP) [Kang and Wu, 2014] because it uses different grids, whereas in [Kang and Wu, 2014], it uses only one grid and the mutual foreground. 7) Mianee et al.'s TISN [Minaee and Wang, 2015] and DSCN [Minaee and Wang, 2017] are methods which employ deep features - a similar feature representation as deep convolutional networks (DCN) and achieve very promising results using principal component analysis (PCA) and multiclass SVM in fingerprint and palmprint recognition. Note that although in forensic applications, imaging environments are also contactless, this study and the previous studies are under very different assumptions. In Wu et al. and Kang et al.'s studies, the contactless imaging environment, including illumination setting, is under control and users cooperate with the imaging devices. In forensic applications, imaging environments are out of control and there is no user cooperation. Thus, large pose and illumination variations are expectable.

All the experiments were performed on the NTU-Wrist-Image-Database-v1 using different combinations of the sets. Three experiments EXP1, EXP2, and EXP3 are considered as a representation of different forensic cases. Description of these experiments is presented in Table 3.4. Detailed quantitative description of the image sets is given in Tables 3.1 and 3.2. Note that the gallery (training set) and the probe (testing set) are exclusive sets of images and each probe wrist has its corresponding wrist in the gallery. The four WMFA recognition systems, RS_{PLS1} , RS_{PLS2} , RS_{SVM1} and RS_{SVM2} and Meta-recognition systems, RS_{PLS} , RS_{SVM} and WMM were compared and the contribution of LBP, Gabor and DSIFT features for the proposed WMFA algorithm was evaluated. An algorithm is considered to be a superior or better when it achieves higher rank-1 accuracy. The results of the experiments are presented as cumulative match characteristic (CMC) curves in Figures 3.10, 3.11, 3.12, summarized in Table 3.5 and described in Section 3.4.4. Additionally, in Section 3.4.5, PCA and minimal-redundancy-maximal-relevance (mRMR) criterion [Peng et al., 2005] are considered to reduce the high-dimensional feature vectors and the results are reported in Tables 3.6 and 3.7.

The features employed in the proposed algorithm: LBP, Gabor and DSIFT are well known features, successfully applied in many domains, including biometrics [Chan and Kong, 2015, Ahonen et al., 2006, Su and Kong, 2014, Wang et al., 2010a, Chan et al., 2017]. Recently researchers tend to move towards deep features, which are believed to be more powerful than hand crafted features. For example, in [Minaee and Wang, 2015] and [Minaee and Wang, 2017] the authors apply deep features to fingerprint and palmprint recognition achieving very good results in controlled environment. These deep features are obtained from the layers of the scattering network [Bruna and Mallat, 2013]. There are also some recent advancements in hand crafted features e.g. multichannel decoded local binary pattern (mdLBP)[Dubey et al., 2014], which can be considered to enhance matching performance. However, the aim and focus of this work are not to find the best representation for a wrist but to show that wrist is a useful clue for criminal and victim

identification. Additional experiments are performed to compare LBP with mdLBP and the hand crafted features with the deep features. The results of the experiments are summarized in Table 3.9 and described in Section 3.4.6.

Table 3.4: Details of three different experiments.

Exp	Gallery	Probe	Forensic cases related to:
EXP1	SET1	SET2	Child sexual abuse. Good image quality, no much pose or viewpoint variation.
EXP2	SET1	SET3	Child sexual abuse, rioter images. With sufficient resolution but with pose, viewpoint, illumination variations.
EXP3	SET4	SET5	Terrorist images. Low resolution, pose, viewpoint, illumination, shadowing variations.

3.4.2 Implementation Details

The proposed WMM algorithm is implemented in MATLAB and the code is publicly available⁵. Before classifiers training and wrist matching were performed, standard pose images were segmented automatically, while the Internet images were segmented manually, which is acceptable in forensic cases. Two ensembles of 300 decision trees, which served as skin superpixel classifiers, were trained by using bagging method. There were over 500,000 skin superpixels and 125,000 non-skin superpixels used to build the two-class classifiers. 3400 standard pose images were divided into two separate folds F_1 and F_2 , which were

⁵<https://github.com/matkowski-voy/Wrist-Identification-for-Forensic-Investigation>

used to train classifiers C_1 and C_2 . Then, the segmentation results were obtained using C_1 (C_2) to classify superpixels of F_2 (F_1). Next, the standard pose SET1 and SET2, the non-standard pose SET3 and the Internet images SET5 and SET1P were automatically aligned to the template and the proposed features were extracted from ROI#1 and ROI#2, such that each image was represented by two feature vectors. Segmentation and alignment results were checked and no obvious errors were noted. Finally, for each wrist in the gallery set, two PLS and two SVM classifiers were trained using one-against-all approach.

Other recognition methods are also implemented in Matlab. In the OrdinalCode [Sun et al., 2005], the default parameters are used. In the DOC [Fei et al., 2016b], the code provided by authors and the default parameters are used. In the DSCN [Minaee and Wang, 2017] and TISN [Minaee and Wang, 2015], the deep scattering features are extracted using the implementation provided in ScatNet toolbox⁶, and for the recognition, one-against-all linear SVM is used. In the MF-LBP [Kang and Wu, 2014], the uniform LBP [Ojala et al., 2002] with radius 2 and 8 neighbours and χ^2 distance are used. In the SIFT-IRANSAC [Wu et al., 2014], the parameters of the Gabor filter are $\omega = 5$, $\sigma = 7.25$ and $F = 0.03$ and the SIFT descriptors are detected using the VL_FEAT library⁷ with the threshold set to 100 and matched with the threshold 1.6. For mdLBP [Dubey et al., 2014] and mRMR [Peng et al., 2005], the codes provided by the authors are used.

Note that when applying the recognition methods from [Sun et al., 2005, Wu et al., 2014, Kang and Wu, 2014, Fei et al., 2016b, Minaee and Wang, 2015, Minaee and Wang, 2017], input ROI images are always extracted using the proposed algorithms (see Sections 3.3.1 and 3.3.2), because none of these methods can handle the wrist segmentation, alignment and ROI extraction. Thus, the comparisons among WMM, WMFAs and other methods are at the feature and recognition levels. Moreover, if there is no Meta-recognition step on the top of a matching algorithm and ROIs are not indicated, the results are always presented for the ROI that gives higher result in terms of rank-1 accuracy.

⁶<https://www.di.ens.fr/data/software/scatnet/download/>

⁷<http://www.vlfeat.org/overview/sift.html>

3.4.3 Computational Cost

The proposed algorithm run on a PC with Xeon E5 3.5 GHz CPU. During the pre-processing stage, the average time of skin segmentation (Section 3.3.1) is 5 seconds/image, which gives 0.025 seconds for classification of one superpixel. Note that this step is run only once because after skin segmentation step, all segmented images are stored. For ROI extraction (Section 3.3.2) and feature extraction (Section 3.3.3), the average time cost is 0.03 and 0.28 seconds/image respectively. Matching one probe image with one gallery image is very efficient because it is performed as a dot product of a feature vector and regression coefficients (Section 3.3.4), which takes 0.0004 seconds/image. In the experiment with the largest number of wrists in the gallery set (EXP3), the average time of Meta-recognition stage (Section 3.3.5) is 0.005 seconds/image. The computational complexity of the WMM is linear with respect to the number of gallery wrists because it requires n comparisons of probe and gallery wrists, where n is a number of gallery wrists.

3.4.4 Evaluation of the Proposed Wrist Identification Algorithm

In the first experiment EXP1, both sets are standard pose images. Gallery SET1 is used to build 526 classifiers from 1948 images and probe SET2 contains 1452 images for testing. The gallery set is larger than the probe set in terms of classes. It has 526 different classes while the probe set has 397 corresponding classes. The experimental results in Figure 3.10a show that the WMM and WMFA algorithms outperform the state-of-the-art palmprint and palm vein matching methods by achieving accuracy of 91.94% at rank-1 and 97.80% at rank-30. Feature contribution is evaluated for two different WMFAs, RS_{PLS1} and RS_{SVM1} because they achieve better result than RS_{PLS2} and RS_{SVM2} . The most discriminative features are LBP, then SVM's DSIFT and Gabor (Figure 3.10b). The four WMFAs RS_{PLS1} , RS_{PLS2} , RS_{SVM1} and RS_{SVM2} are compared to evaluate PLS and

SVM classifiers and the ROI extraction procedures. Figure 3.10c shows that WMFAs with SVMs generally perform worse than their corresponding WMFAs with PLS and the highest rank-1 accuracy of 86.16% and 85.33% is achieved by respectively RS_{PLS1} and RS_{SVM1} indicating that ROI#1 is more robust than ROI#2. Comparing with the four WMFAs, the proposed WMM ($RS_{PLS+SVM}$) provides 5.78% improvement at rank-1, which demonstrates the effectiveness of the Meta-recognition scheme based on EVT. Figure 3.10d compares WMM, RS_{PLS} and RS_{SVM} , all based on the Meta-recognition scheme, showing that RS_{PLS} outperforms RS_{SVM} with a margin of 2.35% at rank-1.

In experiment EXP2 the gallery set remains the same as in EXP1, but 135 images from SET3 are used as a probe set. The probe set contains 133 unique wrists which means that only 2 wrists have 2 images and the rest have one image per wrist. EXP2 is more challenging than EXP1 because the probe images usually have variation due to illumination, pose or out of focus. Some of the example images can be found in Figure 3.1d. As shown in Figure 3.11a, the proposed WMM algorithm also significantly outperforms the state-of-the-art palmprint and palm vein matching methods with a margin of 16.67% in terms of rank-1 accuracy. It achieves rank-1 accuracy of 38.64% and rank-30 accuracy of 68.18%. Similarly to EXP1, the most discriminative features are LBP and ROI#1 as shown in Figure 3.11b. Comparison of WMFAs, including RS_{PLS1} , RS_{PLS2} , RS_{SVM1} and RS_{SVM2} , is presented in Figure 3.11c and similarly to EXP1 it shows that ROI#1 is more robust than ROI#2. Figure 3.11d shows that WMM algorithm, which considers all RS , achieves the highest accuracies for rank-1: 38.64%. Unlike in EXP1, at rank-15, WMM performs 0.75%, 1.51% worse than RS_{PLS1} and RS_{SVM1} respectively which is presented in Table 3.5. The Meta-recognition scheme considers the top matching score and therefore, it can provide improvement at rank-1. However, there is no theoretical guarantee that it would improve accuracy of other ranks.

In experiment EXP3, the standard pose gallery set is enlarged by adding 205 Internet images to create gallery SET4, which contains 2153 images from 731 different wrists. 205

probe images in SET5 are only Internet images and each of them is from a corresponding wrist in the gallery. This experiment is the most challenging one because the Internet images were taken under extremely uncontrolled environments. Their variations come from uneven illumination, shading, pose, out of focus, point of view, and low resolution. Some examples are given in Figures 3.1e and 3.1f. Moreover, each of the Internet image classifiers was trained based on one positive sample and 2152 negative samples coming from the standard pose and Internet images. PLS is capable to handle this imbalanced classification problem [Qu et al., 2010]. As mentioned before, the standard pose images have more than one positive sample image to train the classifiers. Figure 3.12a shows that the WMM algorithm achieves rank-1 accuracy of 24.88% and rank-15 accuracy of 50.73%, which significantly outperforms the other methods with a margin of 18.02% at rank-1 accuracy. For the best WMFAs - RS_{PLS2} and RS_{SVM1} , feature channels are evaluated and similarly to EXP1 and EXP2, the most discriminative features are LBP and Gabor features perform the worst (Figure 3.12b). Comparison of different WMFAs shows again that PLS achieve better results (Figure 3.12c). Figure 3.12d shows that Meta-recognition step increases rank-1 accuracies for both RS_{PLS} and RS_{SVM} but neither improves nor degrades WMM, comparing with RS_{PLS} at most of the ranks. WMM performs the same as RS_{PLS} until rank-12, though WMM performs 0.5% better at rank-30.

Some of the top 10 matches are presented in Figures 3.13 and 3.14. The proposed algorithm is able to retrieve the masked Baltimore rioter within top 10 ranks (see Figure 3.13). The resolution of the masked Baltimore rioter in the gallery set is 198 by 342 pixels, while the resolution of the corresponding probe image is only 19 by 45 pixels (see Figure 3.13a). As shown in Figures 3.13b and 3.13c, WMM matches them at rank-8 and rank-1 when 50% and 10% of the tail of the matching scores are respectively used to estimate the parameters of the Weibull distribution. However decreasing the tail size from 50% to 10% slightly decreases general performance of the WMM achieving accuracy of 24.39%, 46.83% and 53.17% at respectively rank-1, rank-15 and rank-30. RS_{PLS1} ,

RS_{PLS2} , RS_{SVM1} and RS_{SVM2} rank Baltimore rioter image at respectively the 4th, 8th, 1st, and 3rd positions. This example exposes the potential of wrists for criminal and victim identification.

Figure 3.10: CMC curves for EXP1 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS1} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.

Figure 3.11: CMC curves for EXP2 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS1} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.

(a)

(b)

(c)

(d)

Figure 3.12: CMC curves for EXP3 of (a) the WMFA algorithm and six different palmprint and palm vein matching methods, (b) particular feature using RS_{PLS2} and RS_{SVM1} , (c) different recognition systems and (d) Meta-recognition systems.

Table 3.5: Identification accuracy (%) of the proposed WMM and WMFA algorithms and selected palmprint recognition methods.

	EXP1			EXP2			EXP3		
Method	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
WMM	<u>91.94</u>	<u>96.97</u>	<u>97.80</u>	<u>38.64</u>	63.64	68.18	<u>24.88</u>	50.73	56.59
RS_{PLS}	91.74	96.49	97.59	37.88	63.64	68.18	<u>24.88</u>	50.73	56.10
RS_{SVM}	89.39	96.69	97.59	32.58	56.82	62.88	23.90	46.83	57.58
RS_{PLS1}	86.16	95.18	96.56	36.36	64.39	69.70	18.54	43.90	55.61
RS_{PLS2}	84.16	93.18	95.18	33.33	56.06	64.39	22.93	<u>51.71</u>	55.61
RS_{SVM1}	85.33	95.73	96.97	34.09	<u>65.15</u>	<u>71.21</u>	20.00	44.88	56.10
RS_{SVM2}	82.85	94.28	96.07	25.76	53.03	60.61	16.59	47.32	<u>58.05</u>
SIFT_RANSAC	78.79	88.50	90.29	12.88	24.24	33.33	5.88	8.33	9.80
TISN	67.91	89.60	91.87	21.97	48.48	56.82	4.39	18.05	28.78
DSCN	62.40	84.50	87.67	19.70	39.39	45.45	2.44	20.00	29.76
DOC	47.31	58.20	62.53	17.78	34.07	39.26	3.42	9.76	10.73
MF_LBP	43.80	76.31	83.54	5.30	31.82	42.42	6.86	32.84	39.71
OrdinalCode	39.46	49.93	53.51	10.61	20.45	23.48	4.41	8.33	10.29

The highest percentage accuracy is highlighted.

3.4.5 Dimension Reduction and Feature Selection

This subsection aims to compare the high-dimensional (16466-dimensional) and lower dimensional feature vectors extracted from ROI#1 and ROI#2, which are used to train PLS and SVM classifiers. However, it should be pointed out that PLS regression performs a supervised dimension reduction internally [Rosipal and Krämer, 2006]. Thus it can be also considered as an internal dimension reduction scheme, controlled by the number of the latent components. Additional experiments are performed to investigate the effect of lower dimensional (200-dimensional) features. For dimension reduction, PCA with 200 components in four different RS settings denoted as $PCA + PLS1$, $PCA + PLS2$, $PCA + SVM1$ and $PCA + SVM2$ is applied. For feature selection, minimal-redundancy-maximal-relevance (mRMR) criterion [Peng et al., 2005] is used to select 200 features for $mRMR + PLS1$, $mRMR + PLS2$, $mRMR + SVM1$ and $mRMR + SVM2$. In Table 3.6, the results are reported as a difference between $PCA + PLS1$, $PCA + PLS2$, $PCA + SVM1$, $PCA + SVM2$ and the corresponding RS_{PLS1} , RS_{PLS2} , RS_{SVM1} , RS_{SVM2} , respectively. In Table 3.7, the results are presented in the same way as in Table 3.6, but for the mRMR feature selection scheme. The negative and positive numbers indicate the performance drop and gain respectively. For example, -4.14 for $PCA + PLS1$ means that after applying PCA on the high-dimensional feature vector extracted from ROI#1 and using PLS, the performance drops by 4.14% with respect to RS_{PLS1} . Table 3.6 shows that applying PCA with 200 components mostly decreases the identification performance, especially for PLS in EXP1 and EXP2 and SVM in EXP3. In our experiments, similar performance degradation is also observed for a different number of principal components ranging from 100 to 1000 and thus the results with 200 components are provided. Table 3.7 shows that the performance drop is even more significant, when mRMR is used to select 200 features. In [Peng et al., 2005], the authors used up to 50 selected features, even for 9703-dimensional feature vectors. However in our experiments with such low-dimensional feature vectors selected by mRMR, the performance degradation in EXP1,

EXP2 and EXP3 is even higher. Thus 200 features are used for the fair comparison with PCA. Note that I do not claim that the high-dimensional features used in the WMFAs are the best wrist representation and other representations can be also considered. Nevertheless, in this study, the main focus is to show that wrist can be useful for forensic investigation.

Table 3.6: Identification accuracy (%) differences between PCA compressed features and the original high-dimensional features.

	EXP1			EXP2			EXP3		
Method	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
<i>PCA + PLS1</i>	-4.14	-1.04	-0.77	-5.30	-3.03	-3.80	-1.47	-1.47	-1.47
<i>PCA + PLS2</i>	-6.75	-3.24	-2.69	-12.88	-6.82	-2.76	-4.40	-2.93	2.92
<i>PCA + SVM1</i>	-0.97	-1.45	-0.83	2.27	-7.58	-6.82	-9.27	-11.71	-12.20
<i>PCA + SVM2</i>	-0.42	-1.45	-2.07	-0.01	-0.76	-3.80	-5.42	-7.32	-9.76

Table 3.7: Identification accuracy (%) differences between features selected by mRMR and the original high-dimensional features.

	EXP1			EXP2			EXP3		
Method	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
<i>mRMR + PLS1</i>	-22.52	-8.47	-5.24	-22.72	-23.48	-18.18	-7.81	-8.30	-10.74
<i>mRMR + PLS2</i>	-26.52	-8.47	-5.24	-22.72	-23.48	-18.18	-7.81	-8.30	-10.74
<i>mRMR + SVM1</i>	-14.46	-8.06	-4.75	-19.70	-19.70	-18.94	-11.22	-7.32	-9.76
<i>mRMR + SVM2</i>	-18.32	-9.36	-6.61	-8.34	-17.42	-20.46	-6.83	-12.20	-8.78

3.4.6 Deep vs Hand Crafted Features

This set of experiments aims to evaluate the deep features (DF) from [Minaee and Wang, 2015] and [Minaee and Wang, 2017] denoted as df1 and df2 respectively, mdLBP from

[Dubey et al., 2014] and their fusion with the hand crafted features employed in WMFAs and WMM denoted as HC. The HC consists of LBP, Gabor and DSIFT described in Section 3.3.3. The fusion is indicated by ”+” sign. For example, df1+ and mdLBP+ mean that the feature vector consists of df1 with HC and mdLBP with HC respectively. The same set of experiments, EXP1, EXP2 and EXP3 with different settings on the features is performed. Note that these experiments do not aim to compare different ROIs. Thus as mentioned before, for each feature and experiment, the results for the ROI with the highest rank-1 accuracy are given. In addition, the results of two representative deep learning methods AlexNet [Krizhevsky et al., 2012] and VGG-16 [Simonyan and Zisserman, 2015] are also reported in Table 3.8.

The results of LBP, mdLBP and mdLBP+ are reported in Table 3.9a. Employing mdLBP (34864-dimensional) or mdLBP+ (51330-dimensional) generally neither improves the performance nor outperforms LBP (13074-dimensional) and HC (16466-dimensional). Note that similarly to LBP, mdLBP features are also extracted within each block using the seven grids shown in Figure 3.8. There are three cases when mdLBP+ with SVM outperforms LBP or HC. In EXP1, for rank-15 and rank-30 mdLBP+ performs 0.55% and 0.62% higher than LBP and for rank-30 mdLBP+ achieves 0.34% higher accuracy than HC (RS_{SVM1}).

The deep features df1 (782-dimensional) and df2 (12512-dimensional) are extracted from the first and the second layer of the scattering network which is a DCN that instead of learning the filters uses predefined wavelets [Bruna and Mallat, 2013]. In [Minaee and Wang, 2015] the mean and variance over the whole scattering transformed images are calculated to form df1, whereas in [Minaee and Wang, 2017] the same statistics are calculated from 16 blocks (4x4 grid) over the transformed images to form df2. In these experiments, no PCA is used because PCA compressed features with 200 principal components from [Minaee and Wang, 2015] and [Minaee and Wang, 2017], perform worse than df1 and df2, respectively (see Table 3.5 and 3.9b). Note that the features from [Minaee

and Wang, 2015] and [Minaee and Wang, 2017] without PCA are equivalent to df1 and df2, respectively. The comparison of DF, HC and the fusion (df1+, df2+) using SVM and PLS is presented in Table 3.9b. In all the experiments HC outperforms DF. Given the same type of classifier, df2 is superior to df1. Employing DF with SVM always gives better results than PLS. The feature fusions of DF with HC: df1+ (17248-dimensional) and df2+ (28978-dimensional) always outperform df1 and df2. Generally, df1+'s performance is higher than df2+ but for rank-1s, HC still achieves higher accuracies than df1+. In EXP2 and EXP3, for rank-15s and rank-30s, df1+ achieves around 1% and 3% higher accuracy than HC respectively. As shown in Table 3.9c if there is the Meta-recognition step on top of the systems, HC outperforms df1+ and df2+, except for EXP1 rank-15 and EXP3 rank-30, where they perform the same.

It should be emphasized that in the deep scattering network, the filters are not learnt but fixed. Thus, in the experiments, there is no learning for df1 and df2. Training deep learning architectures requires sufficient number of training examples, whereas the number of wrist examples in NTU-Wrist-Image-Database-v1 is not very large (see Table 3.2). Mostly, there are four images per wrist in the standard pose SET1 and SET2 whereas the Internet images in SET1P and SET5 have only one image per wrist. Nevertheless, two well-known deep learning architectures AlexNet [Krizhevsky et al., 2012] and VGG-16 [Simonyan and Zisserman, 2015] are selected to investigate whether the training datasets in NTU-Wrist-Image-Database-v1 are able to support these deep learning methods. The proposed segmentation and ROI extraction methods are used to pre-process the images before training and testing. Additionally, the extracted ROI (ROI#1) images are padded with zeros to make them square because AlexNet and VGG-16 require fixed size, square images. First, the networks from MatConvNet library [Vedaldi and Lenc, 2015], pretrained on ImageNet dataset are selected. Then, the networks are fine-tuned until convergence which typically occurs after 25 epochs, with batch size 128, momentum 0.9, no weight decay and ADAM optimizer [Kingma and Ba, 2014]. The learning rate is 0.001 for the last

layer and 10 times smaller for other layers. To prevent overfitting, 0.5 dropout [Srivastava et al., 2014] is applied before the last fully connected layer. The rank-1, rank-15 and rank-30 accuracy are presented in Table 3.8. The results show that all the proposed WMFAs and WMM perform significantly better than both deep learning methods. In addition, AlexNet and VGG-16 are outperformed by almost all other methods used for comparison (see Tables 3.8 and 3.5). NTU-Wrist-Image-Database-v1 is relatively small comparing to other databases used for deep learning. Thus, it may not be able to support these deep learning methods which require larger number of training examples.

Table 3.8: Identification accuracy (%) of AlexNet and VGG-16.

Method	EXP1			EXP2			EXP3		
	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
AlexNet	23.21	58.06	69.15	2.27	14.39	25.00	5.58	24.39	32.68
VGG-16	43.04	79.89	86.85	4.54	16.67	28.79	5.36	25.37	36.10

3.5 Summary

In this chapter, wrist identification for forensic is studied. A new wrist image database named NTU-Wrist-Image-Database-v1 containing 3945 images from 731 wrists from 505 subjects is established. The wrist identification algorithm, including skin segmentation, key point localization, image to template alignment, matching schemes, and post-recognition score analysis is proposed. In the pre-processing stage, the segmentation scheme using an ensemble of decision trees provides reliable results. The wrinkles detected by the shortest paths in the graph and the boundaries are used to determine two ROIs. LBP, Gabor and DSIFT features are extracted to represent wrists as 16466-dimensional feature vectors. For each wrist, PLS and SVM classifiers are built using the one-against-all approach. In order to select the best ROI and matching methods and increase the accuracy of the whole

Table 3.9: Identification accuracy (%) - deep and hand crafted features comparison.

(a) LBP VS mdLBP

F	C	EXP1			EXP2			EXP3		
		rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
LBP	P	<u>85.12</u>	94.63	96.14	<u>37.88</u>	62.88	69.70	<u>22.93</u>	<u>48.78</u>	<u>56.10</u>
LBP	S	83.82	95.11	96.69	32.58	<u>64.39</u>	<u>71.97</u>	20.00	43.41	54.14
mdLBP	P	70.94	91.25	94.01	25.00	52.27	62.12	13.66	37.07	47.80
mdLBP	S	77.20	94.56	96.49	27.27	62.88	68.94	12.20	37.07	49.27
mdLBP+	P	82.78	94.70	96.14	26.53	49.24	56.06	16.10	<u>48.78</u>	54.63
mdLBP+	S	83.47	<u>95.66</u>	<u>97.31</u>	26.53	53.03	60.61	15.61	41.96	54.63

(b) Deep vs hand-crafted features in WMFA

F	C	EXP1			EXP2			EXP3		
		rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
df2	S	78.65	91.94	93.53	30.30	55.30	58.33	7.70	32.30	43.90
df2	P	61.91	87.60	91.12	12.12	41.67	48.48	5.85	32.68	44.88
df1	S	73.48	89.53	91.05	25.00	47.73	53.79	6.82	29.76	39.51
df1	P	39.26	81.61	88.02	4.54	25.00	41.67	2.43	20.49	33.17
df2+	S	83.26	95.25	96.69	33.33	57.58	65.91	17.07	40.49	50.24
df2+	P	80.51	94.49	95.94	35.61	64.39	70.73	18.05	46.34	55.12
df1+	S	84.30	95.80	96.69	32.58	<u>65.91</u>	71.97	20.00	44.88	<u>59.02</u>
df1+	P	84.92	94.97	96.83	35.61	65.15	<u>72.73</u>	20.98	<u>52.68</u>	56.59
HC	S	85.33	<u>95.73</u>	<u>96.97</u>	34.09	65.15	71.21	20.00	44.88	56.10
HC	P	<u>86.16</u>	95.18	96.56	<u>36.36</u>	64.39	69.70	<u>22.93</u>	51.57	55.61

(c) Deep vs hand-crafted features in WMM

F	C	EXP1			EXP2			EXP3		
		rank 1	rank 15	rank 30	rank 1	rank 15	rank 30	rank 1	rank 15	rank 30
df2+	M	87.26	95.45	96.35	37.88	58.33	66.67	18.05	44.88	<u>56.59</u>
df1+	M	91.18	<u>96.97</u>	97.59	37.12	56.82	63.64	23.90	47.32	55.61
WMM	M	<u>91.94</u>	<u>96.97</u>	<u>97.80</u>	<u>38.64</u>	<u>63.64</u>	<u>68.18</u>	<u>24.88</u>	<u>50.73</u>	<u>56.59</u>

F - features, C - classifier, P - PLS, S - SVM, M - Meta-recognition.

(a)

(b)

(c)

Figure 3.13: Top 10 matches returned by the WMM algorithm in EXP3 when the tail size is 50% of the gallery size (b) and 10% (c). The genuine match is highlighted with a red box. Probe and gallery Internet images taken outdoor containing the Baltimore rioter are shown in (a). Wrist images are resized for better visualization. The scale is indicated above each image.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 3.14: Top 10 matches returned by the WMM algorithm in EXP3. The genuine matches are highlighted with a red box. Wrist images are resized for better visualization. The scale is indicated above each image.

system, Meta-recognition step is used on top of WMFAs. For wrist matching, PLS is compared with linear SVM and generally performs slightly better especially in lower ranks whereas SVM tends to achieve more accurate results in higher ranks. Kernel SVM, which projects features to a high dimensional space, is not employed because the dimension of the feature vectors is already high. LBP are the most discriminative features in the proposed algorithm; DSIFT features are the second one while the Gabor features are the last. PCA dimension reduction and feature selection based on mRMR criterion do not improve the performance of PLS and SVM classifiers. Thus the high-dimensional feature vectors are used to keep the discriminatory power. Moreover, the experimental results show that LBP over each RGB channel performs better than inter-channel based descriptor such as mdLBP. The comparison of HC and DF from [Minaee and Wang, 2015, Minaee and Wang, 2017] demonstrates that the proposed HC outperforms the selected DF. The Meta-recognition step based on EVT improves rank-1 accuracy. The experimental results show that the proposed WMM algorithm significantly outperforms the representative palmprint and palm vein recognition methods and two deep learning methods. The proposed WMM algorithm successfully matches the masked Baltimore rioter at the 1st rank and the 8th rank depending on the algorithm parameter. It clearly shows that wrist is possible to be used for forensic investigation.

Chapter 4

Palmprint Recognition in Uncontrolled and Uncooperative Environment¹

4.1 Overview

The previous palmprint research focused mostly on online palmprint recognition for commercial applications and latent palmprint identification in crimes such as robbery and homicide for forensic applications. The biometric community neglected the uncontrolled and uncooperative palmprint recognition problem and the merit of palmprint recognition for forensic investigation based on digital images is not fully exposed yet. Online palmprint recognition was studied using contact-based, contactless or 3D palmprint images taken in well-controlled environments. Contrary to the online palmprint recognition, the uncontrolled and uncooperative palmprint recognition problem has different requirements, especially in terms of the imaging environment and the recognition algorithm has to be applicable to images taken under uncontrolled and contactless environment without subject's cooperation. Comparing with latent palmprint recognition, the problem studied in this chapter has no high-resolution features, e.g., minutia and ridges. Figure 4.1 shows examples of different palmprints studied in the literature. Most of the previous studies

¹This chapter is based on [Matkowski et al., 2019a]

use the feature-engineering approach and the application of deep learning to palmprint recognition received moderate attention from the biometrics community. Even though the previous works [Svoboda et al., 2016, Minaee and Wang, 2017, Ramachandra et al., 2018, Genovese et al., 2019] already applied CNNs, their palmprint recognition pipelines were not trained end-to-end because the ROI was extracted by some traditional method or even manually [Ramachandra et al., 2018] or some features were hand-crafted [Genovese et al., 2019].

This chapter aims to address the uncontrolled and uncooperative palmprint recognition problem by establishing a new palmprint database named NTU Palmprints from the Internet database (NTU-PI-v1) containing 7881 images from 2035 palms collected from the Internet and proposing an end-to-end deep learning algorithm composed of an alignment network, a feature extraction network and an in-network data augmentation scheme. The proposed algorithm is compared with the state-of-the-art online palmprint recognition methods and evaluated on three publicly available contactless palmprint databases, IITD, CASIA, and PolyU and two new databases, NTU-PI-v1 and NTU contactless palmprint database (NTU-CP-v1).

The rest of this chapter is organized as follows. In Section 4.2, related contactless palmprint databases: CASIA, IITD and PolyU, and the new NTU-PI-v1, which is the largest palmprint database are described. In Section 4.3, the proposed network architecture design is provided. In Section 4.4, experiments and comparison with the state-of-the-art contactless palmprint recognition methods and deep learning methods are given. In Section 4.5, the summary are given.

Figure 4.1: Different types of palmprints studied by the biometric community. (a) Contact-based, (b) contactless, (c) latent, and (d) 3D palmprints. (a), (c) and (d) reprinted from [Kong et al., 2009], [Jain and Feng, 2009] and [Fei et al., 2019a], respectively.

4.2 Databases for Palmprint Recognition

4.2.1 Selected Contactless Palmprint Databases

No one studies palmprint recognition in an uncontrolled and uncooperative environment, till now. The most related one is contactless palmprint recognition, which has been addressed by a number of researchers. For contactless palmprint recognition, CASIA², IITD³ and PolyU⁴ databases are widely used and representative benchmarks. Figure 4.2 shows sample images from these three databases. CASIA-Palmprint database consists of 5502 grayscale hand images from 618 palms and is the largest publicly available contactless palmprint database in terms of the number of palms until now. The images were captured during one session and contain the palmar side of the hand with partial fingers on a black background. IITD Touchless Palmprint database provides 2601 colour hand images from 460 palms and also the corresponding ROI images. The images were taken during one session on the same background and have good illumination quality.

²<http://biometrics.idealtest.org/dbDetailForUser.do?id=5>

³http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Palm.htm

⁴<http://www4.comp.polyu.edu.hk/~csajaykr/myhome/database.request/>

[3dhand/Hand3D.htm](#)

(a)

(b)

(c)

Figure 4.2: Examples of images from (a) CASIA, (b) IITD, (c) PolyU contactless palmprint databases.

PolyU Contactless 2D Palmprint database was collected from 177 hands. Each hand has ten colour images, taken during two sessions (5 images/session), under various illumination conditions with a black background. CASIA, IITD and PolyU databases are more suitable to study contactless palmprint recognition for commercial and governmental applications, rather than uncontrolled and uncooperative or forensic applications because the diversity of the images is not very large; there was some control over image acquisition parameters and high degree of user cooperation and the images were taken with an explicit goal of personal identification (see details in Table 4.1). Thus, a new database is needed for this study.

4.2.2 NTU Palmprint Databases

For studying the uncontrolled and uncooperative palmprint recognition, the NTU Palmprints from the Internet (NTU-PI-v1) database, which consists of 7781 hand images

(a)

(b)

Figure 4.3: Sample images from (a) NTU-PI-v1 and (b) NTU-CP-v1 databases. The images in the same column in (a) and (b) are from the same palm. Images are resized to the same size.

collected from 2035 different palms of 1093 subjects with different ethnicity, sex and age, is established. According to our best knowledge, it is the largest publicly available palmprint database in terms of both palms and subjects. The images were collected and downloaded from the Internet galleries (see Figure 4.4); then hand regions were manually cropped using a square bounding box. It is the same practice as in IARPA Janus Benchmark-A, one of the unconstrained face recognition benchmark [Klare et al., 2015], to avoid a bias of the database towards a "detector". Note that hand detection is out of scope in this study and one can employ F-RCNN or other detection methods for this task. There is no publicly available palmprint database for the target forensic application. Thus, the uncontrolled and uncooperative images downloaded from the Internet can be considered as a substitution to simulate the forensic environment. The database aims to reflect palm's diversity in the scenarios, where there is no control over image acquisition

Figure 4.4: Examples of original images downloaded from the Internet. Hands are highlighted in the original images.

parameters and no subject's cooperation and the images are taken without any intention of palmprint recognition. The diversity is represented by significant differences in hand gestures, viewpoint, illumination, background, image quality and resolution (see Figures 4.3b and 4.4). The hand image sizes vary from 30 by 30 pixels to 1415 by 1415 pixels and the median size is 115 by 115 pixels. Note that the sizes represent full hand images with fingers, not only palmprint ROI, which is likely to be two times smaller. In addition to the images, this database also provides annotated hand landmarks (see Section 4.3.1) and segmented hand images. Another new palmprint database is NTU Contactless Palmprint (NTU-CP-v1) database, which contains 2478 images from 655 palms of 328 subjects, mainly Asian – Chinese, Indian, Malay, and some Caucasian and Eurasian persons. The database was collected in Singapore during two sessions without strict pose requirements in a contactless environment. The images were taken by Canon EOS 500D or NIKON D70s cameras and the hand regions were cropped. The hand image sizes vary from 420 by 420

Table 4.1: The Details and comparison of benchmark contactless palmprint databases and NTU Palmprint Databases.

	NTU-PI-v1	NTU-CP-v1	IITD	PolyU	CASIA
No of palms	2035	655	460	177	618
No of images	7781	2478	2601	1770	5520
No of sessions	NA	2	1	2	1
Color image	Yes	Yes	Yes	Yes	No
ROI provided	No	No	Yes	Yes	No
Image size (pixels)	Mdn. 115x115	Mdn. 1373x1373	1600x1200	640x480	640x640
ROI image size (pixels)	NA	NA	150x150	128x128	NA
All fingers visible	Usually	Yes	Yes	Yes	No
Space between fingers	Sometimes	Usually	Yes	Yes	Yes
Posing variations*	Large	Small	Small	Small	Small
Complex background*	Yes	No (bright)	No (dark)	No (black)	No (black)
Illumination variations*	Large	Medium	Small	Medium	Small
Taken by one camera	No	No	Yes	Yes	Yes
Taken for palmprint recognition	No	Yes	Yes	Yes	Yes
Hand landmarks provided	Yes	No	No	No	No

*based on my visual assessment (see Figures 4.2 and 4.3).

The indicators of the uncontrolled and uncooperative environment are highlighted.

pixels to 1977 by 1977 pixels and the median size is 1373 by 1373 pixels. Some sample images from the two NTU databases are shown in Figures 4.3 and 4.4. The details and comparison of the palmprint databases are given in Table 4.1. NTU-PI-v1 and NTU-CP-v1 will be available online⁵ for research purpose.

4.3 The Architecture Design

In this section, the design of the network architecture is described. At the beginning, the motivations for particular choices in the proposed architecture such as the proposed hand landmarks initialization and configuration (Section 4.3.1), type of transformation (Section

⁵https://github.com/BFLTeam/NTU_Dataset

4.3.2), and a CNN for feature extraction (Section 4.3.3) are discussed. Then, the details of the proposed, end-to-end architecture are given (Section 4.3.4).

4.3.1 Hand Landmarks

Usually, in conventional palmprint recognition methods (see Section 4.1), the ROI is defined based on the landmarks on the extracted hand contour, in particular in-between fingers. In the controlled environments, these landmarks are well-defined, stable, and allow to handle affine deformation of hand images because the hand contour can be extracted easily without error. However, such ROI extraction is sensitive to 3D pose variations, elastic palm deformations, unclear hand contours and complex background, which always appear in the uncontrolled environment. In other words, the traditional methods based on thresholding, edge detectors and boundary tracking work well only if the background and the hand have the significant color difference and there are clear spaces between fingers. Nonetheless, in the uncontrolled environment, such assumptions do not hold and the traditional methods will most likely fail in the detection. Thus, a CNN for hand landmark detection (see Section 4.3.4 and Figure 4.7: Localization Network), additional landmarks and non-affine transformation (see Section 4.3.2) are considered in this study. In this study, landmarks are used to correct elastic deformation caused by different hand poses. Hence, the landmarks must be more than three in order to be able to parametrize non-rigid transformations. Note that three points can only determine affine transformations. Furthermore, the landmarks have to span the entire palm because more skin area carries more discriminative information [Chan et al., 2017]. Hence, the initial landmarks are defined at the four corners of the palm, which are labeled as L1, L3, L7 and L9 in Figure 4.5. The landmarks have their corresponding points in a template. If there are only four landmarks, spanning the palm area, the line between L1 and L3 will cut across the palm and some palm information will be lost and the clear landmark L2 will be ignored (see red dashed line in Figure 4.5). However, if L2 is used as a landmark, the two issues do

Figure 4.5: The proposed initial configuration of hand landmarks.

not hold anymore and transformations with higher degrees of freedom can be defined for more accurate alignment. L1-L3 are well-defined, situated at the bottom of the fingers, whereas the exact locations of L7 and L9 are not always very clear. L7 and L8 are placed approximately at the border between the palm and wrist. In addition to these landmarks, the middle points of the lines, L1, L7, L7, L9 and L3 L9 are automatically extracted and labelled respectively as L4, L8 and L6. The middle point between L8 and L2 is also extracted and labelled as L5. In this study, nine landmark configuration is proposed. Note that the proposed hand landmark configuration is used for the Localization Network training (see Section 4.4.4).

4.3.2 Spatial Transformer

The aim of palm ROI extraction is to align each palmprint into the same coordinate system. A module that enables efficient spatial image manipulation within a neural

network is a Spatial Transformer [Jaderberg et al., 2015]. Thus, in this work, the Spatial Transformer is used for the ROI extraction (see Section 4.3.4). The module consists of a trainable Localization Network, which regresses transformation parameters θ , a sampling grid generator and a sampler. It can be used to implement any transformation T_θ that is differentiable with respect to its parameters θ , e.g., affine, plane projective or thin plate spline (TPS) transformations. The module is differentiable and can be put into any place of a CNN architecture forming a Spatial Transformer Network, which can be trained with a standard backpropagation. Jaderberg et al. [Jaderberg et al., 2015] showed that the Spatial Transformer Network with TPS [Bookstein, 1989] is the most powerful for elastically deformed digits. Also, some recent state-of-the-art feature matching and alignment methods use Spatial Transformer Network with TPS [Kanazawa et al., 2016, Ufer and Ommer, 2017, Rocco et al., 2018]. Another reason for using the TPS transformation is that it is non-rigid and can be parametrized by control points in the Cartesian coordinate system, which are x, y coordinates of hand landmarks in this study. Thus, in this work, hand landmark coordinates θ are used to parametrize the TPS deformation of the sampling grid. The illustration of the TPS palm transformation based on the proposed hand landmarks is shown in Figure 4.6.

4.3.3 A Building Block for Palmprint Alignment and Feature Extraction

The convolutional layers from the VGG-16 network are used as a basic building block for palmprint alignment and feature extraction. The details of the VGG-16 can be found in [Simonyan and Zisserman, 2015]. The proposed alignment scheme does not use the entire VGG-16 network. The VGG-16 is first pre-trained on ImageNet and then the top layers are pruned. The rationale and motivation of such choice are that the palmprint databases are relatively small and training deep networks from scratch is very sensitive to overfitting. However, using a pre-trained network and reducing its depth can alleviate

Figure 4.6: Palmprint ROI extraction using TPS and a sampler. Sampling grid T_θ is created by TPS based on the hand landmarks in a hand image I . The sampler takes T_θ and I , and returns palmprint ROI I_{ROI} by sampling I into a regular grid G_{ROI} . The red dots are a schematic visualization of the sampling grid.

the data demand. Furthermore, palmprint consists of lines and texture. These lower level features also appear in natural images and can be extracted by the first several layers of a network trained on natural images [Yosinski et al., 2014]. On the other hand, features from the top layers carry more semantic meanings and therefore, the top layers trained on natural images may not accurately represent human hands. The modified pre-trained VGG is used as a building block (e.g., see Figure 4.7: Feature Extraction CNN) in the Localization and Feature Extraction and Recognition Networks (see Figure 4.8). Note that these two building blocks in the proposed algorithm do not share parameters, because one operates on full hand images, which contain fingers and background whereas the other one operates on ROIs, which contain palmprint only.

4.3.4 The Proposed Architecture

The proposed End-to-End Palmprint Recognition Network (EE-PRnet) consists of two main networks, ROI Localization and Alignment Network (ROI-LAnet) and Feature Extraction and Recognition Network (FERnet). Figures 4.7 and 4.8 illustrate the architecture of the ROI-LAnet and EE-PRnet, respectively. To align all palmprints into the same coordinate system and localize their ROIs, ROI-LAnet takes an original hand image I resized to $h \times w$ pixels as an input and outputs palmprint ROI image I_{ROI} . The first part of the ROI-LAnet is the modified VGG-16 discussed in Section 4.3.3. More precisely, the pre-trained VGG-16 network [Simonyan and Zisserman, 2015] which is pruned after the layer *pool3* with local response normalization (LRN) on top and is used as a feature extractor in the ROI-LAnet. This setting produces L2-normalized feature maps f_h , which retain the spatial information in the hand image. The feature maps f_h are connected to the second part of the ROI-LAnet, which is a fully connected Regression Network with two hidden layers *fc1* and *fc2* with 512 and 128 neurons, respectively. Both layers are followed by ReLU activations and dropout to avoid neurons co-adaptation [Srivastava et al., 2014] and serve as a ROI augmentation mechanism in the training (see Section 4.4.3). The

Figure 4.7: The ROI-LANet architecture. The input is the hand image I resized to $h \times w$ pixels and the output is palmprint ROI image I_{ROI} .

Feature Extraction Network together with the Regression Network form the Localization Network (see Figure 4.7), which outputs normalized coordinates $\hat{\theta}$ of the hand landmarks (see Section 4.3.1). The normalization range is between -1 and 1 and the normalized coordinates $\hat{\theta}$ are forwarded to the grid generator which transforms a regular, square grid G to a deformed grid $T_{\hat{\theta}}(G)$ based on $\hat{\theta}$ (see). A bilinear sampler takes the deformed grid $T_{\hat{\theta}}(G)$ as an input and samples the original hand image I (not the resized I inputted to the ROI-LANet) to form a regular grid of $h_{ROI} \times w_{ROI}$ pixels, which is the ROI image, I_{ROI} (see Figure 4.7). Note that because of the coordinate normalization, the sampler can take an image of any size.

The ROI-LANet is connected to the Feature Extraction and Recognition Network (FERnet) (see Figure 4.8), which is responsible for palmprint feature extraction and recognition. To extract palmprint features, the ROI image I_{ROI} , is passed to another independent CNN, which is also a VGG-16 pruned after *pool3* layer with the LRN on top. Even though its structure and the structure of the feature extraction network in the ROI-LANet are same, there is no weight sharing between them. The output from the LRN layer, f_{sROI} is a $\frac{h_{ROI}}{8} \times \frac{w_{ROI}}{8} \times 256$ spatial representation of the palmprint ROI. Note that the spatial dimensions $h_{ROI} \times w_{ROI}$ are defined in the grid generator in the ROI-LANet. Then, the LRN is connected to a dropout layer and an embedding layer $fc4$, which outputs the final 512 dimensional palmprint descriptor f_{PD} . The vector f_{PD} is passed to another dropout layer and then fully connected layer $fc5$, which returns the palmprint labels PL . In the training, L2 loss is used to train the Localization Network in the ROI-LANet before

Figure 4.8: The proposed End-to-End Palmprint Recognition Network (EE-PRnet) architecture.

connecting it to FERnet and Softmax loss is used to train the FERnet and the EE-PRnet (see Sections 4.4.3 and 4.4.4).

4.4 Experiments

4.4.1 Evaluation Protocols

In all experiments, the databases are divided into training and testing sets. Only the images in the training set are used to train the networks. To increase the number of possible comparisons, the left hands are flipped into the right ones. The evaluation protocol for the CASIA and IITD databases is the same as in the recent palmprint survey [Fei et al., 2019a], where the first four gallery images are for training and the rest for testing. The evaluation protocol for the PolyU and NTU-CP-v1 databases is the same as in [Zheng et al., 2016], which images from the first session are for training and from the second session for testing.

The NTU Palmprints from the Internet (NTU-PI-v1) database is divided into training/gallery and testing/probe sets as follows: if there is an even number of images from the same palm, then the images are randomly split into 50% for training and 50% for testing; if there is an odd number of images from the same palm, then the one more image goes to the training set. In this split, each palm in the testing set has a corresponding palm in the training set and in the case of odd number of samples, the training set has one more sample. Splitting NTU-PI-v1 database in this manner results in 3380 images from 1805 palms for testing and 4501 images from 2035 palms for training. No outside palmprint is used for

training. The images in the training set which have no corresponding images in the testing set can be considered as distractors, which make the recognition more difficult. As with the protocol in IARPA Janus Benchmark-A [Klare et al., 2015], which was designed to study unconstrained face recognition for a particular group of subjects, e.g., identification of terrorists or gangsters in a watch list, this protocol allows performing training on the gallery set. Cumulative match characteristic (CMC) curve is used as an evaluation metric, which is common for forensic applications [Jain and Feng, 2009, Paulino et al., 2013, Klare et al., 2015, Chan et al., 2017]. In addition, rank-1 and rank-30 identification accuracies are also used as evaluation metrics.

4.4.2 Implementation Details

The proposed architecture is implemented in MATLAB using MatConvNet library [Vedaldi and Lenc, 2015] and the code is publicly available⁶. The pre-trained convolutional blocks from the VGG-16 are used (see Section 4.3.3). The proposed networks are trained with ADAM optimizer [Kingma and Ba, 2014] with momentum 0.9, learning rate 0.001, no weight decay and batch size 128. The average R, G and B values are subtracted from each channel in input images I . In the ROI-LAnet, the input images I are resized to 56×56 pixels and thus the size of the feature f_h is $7 \times 7 \times 256$. The ROI image I_{ROI} size is set to 112×12 pixels and thus the f_{sROI} dimension is $14 \times 14 \times 256$. The dropouts $Drop1$, $Drop2$, $Drop3$ and $Drop4$ are set to 0.2, 0.1, 0.5 and 0.5, respectively. The ReLU units in the Regression Network are leaky ReLU activations with the leak value of 0.1. In the newly initialized layers $fc1$ - $fc5$, the weights and biases are randomly sampled from a Gaussian distribution with 0 mean and 0.001 variance. The number of transformation parameters $L = 18$ because there are nine landmarks (see Section 4.3.1) and each is defined by x, y coordinates. The output size N_{class} of the last $fc5$ layer is the number of different palmprints in a training set. For recognition, the softmax, one-against-all partial

⁶<https://github.com/matkowski-voy/Palmprint-Recognition-in-the-Wild>

least square regression (PLS) [Rosipal and Krämer, 2006] and linear SVM are compared in Section 4.4.5. When using PLS and SVM, the palmprint descriptors f_{PD} extracted from EE-PRnet are standardized (z-score) as input features. If the features are from the target palmprint of the PLS or SVM, its label is assigned to 1, otherwise -1 . The number of latent components in PLS is 25 in Sections 4.4.5 and 4.4.6 and 50 in the final setting in Section 4.4.7. The training is run on a PC with a single GPU card GeForce GTX TITAN X.

Other palmprint recognition methods are also used in the evaluation. CompCode, OrdinalCode and DoN are run using the programs provided by the authors in [Zheng et al., 2016]. The parameters in these methods are the same as in [Zheng et al., 2016] for the CASIA, IITD and PolyU 2D databases and for NTU-PI-v1 and NTU-CP-v1 are the same as those used for PolyU 2D contactless database. In the DOC [Fei et al., 2016b], the code provided by authors and the default parameters are used. In the DSCN [Minaee and Wang, 2017], the deep scattering features are extracted using the implementation provided in ScatNet toolbox⁷, and for the recognition, one-against-all linear SVM is used. In the MF-LBP [Kang and Wu, 2014], the uniform LBP [Ojala et al., 2002] with radius 2 and 8 neighbours and χ^2 distance are used. In the SIFT-IRANSAC [Wu et al., 2014], the parameters of the Gabor filter are $\omega = 5$, $\sigma = 7.25$ and $F = 0.03$ and the SIFT descriptors are detected using the VL_FEAT library [VLFeat,] with the threshold set to 100 and matched with the threshold 1.6. PalmNet [Genovese et al., 2019] and PCANet [Chan et al., 2015] which was also used for palmprint recognition in [Meraoumia et al., 2017] are run using publicly available codes with default parameters. CR-CompCode [Zhang et al., 2017] and LLDP [Luo et al., 2016] with Gabor filter are run using codes provided by the authors with default parameters. The method from [Ramachandra et al., 2018] employs fine-tuning pre-trained AlexNet and sum rule of SVM and softmax scores for recognition and in the comparison is referred as ANSS. Three state-of-the-art deep

⁷<https://www.di.ens.fr/data/software/scatnet/download/>

learning architectures VGG-16 [Simonyan and Zisserman, 2015], ResNet-50 [He et al., 2016] and GoogLeNet [Szegedy et al., 2015] are trained for the comparison. All these networks are first pre-trained on the ImageNet dataset and then further fine-tuned on a palmprint database with ADAM optimizer, momentum 0.9, no weight decay and batch size 128, until convergence. The ROI images are resized to the default 224×224 pixels. For all these networks, the learning rate 0.001 is applied to the classification (last) layers only, whereas the lower layers have 100, 10 and 10 times smaller learning rates for VGG-16, ResNet-50 and GoogLeNet, respectively. VGG-16 and GoogLeNet are fine-tuned for 20 epochs, whereas the convergence of ResNet-50, which is very deep, takes 60 epochs.

4.4.3 Training Strategy

In the experiments, different training and data augmentation strategies are investigated to analyze the components of the proposed networks. As the starting point for all the training stages, not to destroy the pre-trained convolutional layers, only the randomly initialized top layers are trained. Then, when the top layers are well initialized, lower layers can also be trained. In Stage I, the Localization Network is trained (Section 4.4.4) to initialize the ROI-LAnet, which is used to extract palmprint ROIs. In Stage II, the FERnet is trained based on palmprint ROIs. In Stage III, to investigate the end-to-end training of the EE-PRnet, the initialized ROI-LAnet from Stage I is connected to the FERnet and trained together. Note that the FERnet in Stage III is not the trained FERnet from Stage II.

In Stage II, the input is the palmprint ROI whereas in Stage III, the full hand image. When training the EE-PRnet, the palmprint ROI inputted to the FERnet depends on the input hand image I and the ROI-LAnet. The results from Stage II and Stage III are expected to be different because 1) the dropouts $Drop1$, $Drop2$; 2) the parameters updates in the Localization Network in the ROI-LAnet and 3) on-the-fly data augmentation of hand images I , which will allow that in the same hand image I , the different hand landmarks $\hat{\theta}$ can be estimated. Thus, different palmprint ROIs I_{ROI} can be extracted in training. In

Figure 4.9: Illustration of the in-network ROI augmentation. Top row: examples of different hand landmarks $\hat{\theta}$ (red dots) detected by the Localization Network in the same hand image I . Bottom row, the same column: the corresponding palmprint ROI images I_{ROI} , returned by ROI-LAnet. Zoom in recommended for the best view.

other words, the ROI-LAnet performs on-the-fly, in-network ROI augmentation, whose contribution is analyzed in experiments in Section 4.4.5. Figure 4.9 illustrates the ROI augmentation based on the dropouts in the Localization Network.

4.4.4 Localization Network Training

The Localization Network is trained on annotated hand images from the NTU-PI-v1 (see Section 4.2.2 and 4.3.1) with the L2 loss (see Figure 4.10). The network is fed with the original and segmented hand images from the training set. For data augmentation three hand images rotated by 90, 180 and 270 degrees and three randomly rotated segmented images are used. At first, to train the Regression Network and not to destroy the pre-trained convolutional blocks, only the parameters from layers $fc1$, $fc2$ and $fc3$ denoted as block A (Figure 4.10: Stage I-A) are being trained for 10 epochs. Then, when the Regression Network is well-initialized, the convolutional layers denoted as block B are unfrozen and trained together for the next 15 epochs (Figure 4.10: Stage I-AB). Such pre-trained Localization Network is inserted into the ROI-LAnet and used for the ROI extraction.

The accuracy of ROI extraction is measured by normalized mean error (NML) [Jourabloo and Liu, 2015] of the detected hand landmarks, which is $1.99 \pm 0.36\%$ after the Localization Network training. It should be emphasized that the objective is palmprint recognition and the Localization Network training is for initializing the network with reasonably good weights before the EE-PRnet training. Moreover, some landmarks such as L7 and L9 are not well defined and their ground truth annotations are expected to have higher errors than other landmarks. Thus, the Localization Network is further trained in the EE-PRnet with Softmax loss.

4.4.5 Experiments on NTU Palmprints from the Internet

This subsection aims to analyze the components of the proposed EE-PRnet by employing different training strategies S on the NTU-PI-v1 and three different classifiers for recognition. The strategies details are given in Figure 4.10. In strategies $S1$ - $S5$, the ROI-LAnet and the FERnet are connected and the EE-PRnet is trained based on hand images, whereas in $S0$, only FERnet is trained separately based on ROI images. In the beginning, for all the strategies, only the top layers $fc4$ and $fc5$ denoted as block C (see Figure 4.10) are trained. Then, when the network converges which typically occurs after 20 epochs, the Localization Network can also be fine-tuned and the training is stopped at epoch $N = 40$. For the color data augmentation denoted as ct , the saturation and contrast of each image are randomly shifted on-the-fly.

In the first and most basic strategy $S0$, the ROI images are extracted by the ROI-LAnet and input to train the FERnet (Stage II-C). During the training using the strategy $S1$, all the layers in the ROI-LAnet and the *ConvBlock1-3* in the FERnet are frozen and only the parameters in block C are being updated for N epochs, while the dropouts *Drop1*, *Drop2* are switched off (Stage III-C-off). The strategy $S2$ is the same as $S1$, but the dropouts *Drop1*, *Drop2* are switched on (Stage III-C-on). The strategies $S3$ and $S4$ are the same as the $S2$, but the block D (see Figure 4.10) in the ROI-LAnet is fine-tuned after epoch 20

with the learning rate 0.0001 and the dropouts *Drop1*, *Drop2* are switched off for the *S3* (Stage III-CD-off) and on for the *S4* (Stage III-CD-on). Finally, the strategy *S5* is the same as *S4* but after epoch 35 (from Stage III-CD-on) the dropouts *Drop1*, *Drop2* are switched off (Stage III-CD-on/off). Additionally, the results, when no color augmentation *ct* is used in the strategy *S0*, are also provided (*S0nct*). To show the merit of the ROI extraction by the ROI-LAnet, hand images are input directly to the FERnet (Stage II-C), and the results are reported as *S0h*. The comparison of the proposed training strategies and three classifiers is presented in Table 4.2.

The results show that connecting two networks ROI-LAnet and FERnet into EE-PRnet, benefits the overall performance (see *S0* vs *S1*) and PLS outperforms both Softmax and SVM. The dropout based in-network ROI augmentation (see Figure 4.9) by the ROI-LAnet also positively contribute (*S1* vs *S2*) and its significance is even more noticeable when the Regression Network (block D) is fine-tuned (*S3* vs *S4*). The dropouts' decay for the last 5 epochs while fine-tuning also can improve the performance (*S4* vs *S5*). In summary, when training on NTU-PI-v1 images with PLS as a classifier, the proposed strategy with connected networks outperforms the separated networks setting (*S0* vs *S5*) with a margin of 3.11% and 3.67% at rank-1 and rank-30, respectively. Moreover, the ROI extraction by the ROI-LAnet is an essential step (*S0h* vs *S0*) and without it, the performance drops by 23.13%.

4.4.6 Experiments on CASIA, IITD, and PolyU Contactless Palm-print Benchmark Databases

The top three strategies *S2*, *S4* and *S5* from Section 4.4.5 are used to train the EE-PRnet for the CASIA, IITD and PolyU databases. To compare the ROI extraction performance, the strategy *S0* is also applied when the databases, i.e., the IITD and PolyU databases provide extracted ROI. Because the proposed EE-PRnet requires square image input, some simple automatic pre-processing steps are applied on the CASIA, IITD and PolyU images. Note

Figure 4.10: The proposed network architectures and EE-PRnet training schemes.

that the network architecture, training and testing policies are fixed and exactly the same as described in Section 4.4.5.

The images in the CASIA database, which have a black background are padded with zeros to make them square. In the IITD database, only the fixed central part is cropped. In the PolyU database, the images except for hands also contain a significant portion of a black background. To detect the hands only, Otsu thresholding and the square bounding box around the biggest connected component are used.

Such pre-processed images are used to train the EE-PRnet for $N = 40$ epochs. In

Table 4.2: Rank-1 and Rank-30 accuracy (%) of different training strategies S and classifiers on NTU Palmprints from the Internet database.

S	PLS		Softmax		SVM	
	Rank-1	Rank-30	Rank-1	Rank-30	Rank-1	Rank-30
$S0$	37.69	60.53	24.05	50.05	28.67	52.69
$S1$	39.32	62.04	24.94	52.04	28.52	53.52
$S2$	39.82	63.52	28.14	56.33	33.64	59.14
$S3$	38.96	62.25	24.97	51.86	28.76	53.43
$S4$	40.80	63.70	29.05	57.04	33.67	59.02
$S5$	40.80	64.20	29.70	57.96	33.85	58.96
$S0nct$	37.07	60.18	23.46	49.08	26.12	51.80
$S0h$	14.56	32.19	10.12	30.18	10.30	27.57

The top 3 strategies are highlighted.

strategies $S4$ and $S5$, the learning rate in block D is set to 0.001, which is 10 times higher than in Section 4.4.5, because the block D is a part of the Localization Network, which is pre-trained on the NTU-PI-v1 (Section 4.4.4). To compare the FERnet trained on ROIs extracted by traditional methods with the EE-PRnet, the ROIs provided in the IITD and PolyU databases are used to train the FERnet with the strategy $S0$. The results are presented in Table 4.3. The results show that even though the Localization Network is pre-trained on the different database (NTU-PI-v1), it can generalize well and can be fine-tuned to a given database ($S2$ vs $S4/S5$).

However, we observe slightly lower performance on the CASIA database (see the CASIA results in the brackets in Table 4.3) which can be caused by the discrepancy between the CASIA and NTU-PI-v1 (see Figures 4.3b and 4.2 and Table 4.1). Contrary to NTU-PI-v1, the hand images in the CASIA database are grayscale and usually, the full fingers and thumbs are not visible. The visual difference between the CASIA database and the other databases can be seen in Figure 4.1. Thus, a different augmentation technique

Table 4.3: Rank-1 and Rank-30 accuracy (%) of the selected training strategies S and classifiers on IITD, PolyU and CASIA contactless palmprint databases.

S	IITD		PolyU		CASIA	
	Rank-1	Rank-30	Rank-1	Rank-30	Rank-1	Rank-30
$S0$	96.19	99.08	98.98	99.55	-	-
$S2$	97.11	99.08	99.55	99.89	94.64 (94.14)*	97.78 (97.55)*
$S4$	98.69	99.47	99.66	99.89	96.82 (94.44)*	99.40 (98.21)*
$S5$	98.66	99.61	99.66	99.89	97.52 (94.91)*	99.54 (98.31)*

*The numbers in the brackets for the CASIA database are the results without using data augmentation at .

is applied to refine the training strategy for this database. Rather than using color data augmentation scheme ct on CASIA greyscale images taken under good illumination, a more aggressive spatial data augmentation scheme denoted as at is used because the images have significant scale and translation variations. In the augmentation scheme at , each input hand image is randomly either rotated ($\pm 5, 20, 90$ deg.) or scaled (0.8, 0.9, 1, 1.1, 1.2) or translated ($\pm 0.1, 0.15, 0.2$) on-the-fly. The results (see Table 4.3: CASIA) show that the proposed strategies combined with the augmentation scheme at benefit the performance on the CASIA which is visually quite distant from the target database NTU-PI-v1, for which the algorithm is designed. In the next subsection, the training strategy $S5$ is combined with the augmentation scheme at for all the databases and the final results of the proposed algorithm are reported.

4.4.7 Comparison with the State-of-the-Art Palmprint Recognition Methods

In this section, the comparison of the proposed algorithm with the state-of-the-art palmprint recognition and deep learning methods on the representative contactless palmprint benchmarks (Section 4.2.1) and the NTU-Palmprint databases (Section 4.2.2) is given.

For each database, the proposed EE-PRnet is trained for $N = 60$ epochs with the $S5$ strategy. Data augmentation scheme ct is used for all the databases, except the CASIA, in which the scheme at is used. After 40 epochs the augmentation at (see Section 4.4.6) is also applied to all databases. The evaluation protocols are described in Section 4.4.1. In the experiments on NTU-PI-v1 and NTU-CP-v1, the ROIs extracted by the pre-trained ROI-LAnet (Section 4.4.4) are input to CompCode [Kong and Zhang, 2004], OrdinalCode [Sun et al., 2005], DOC [Fei et al., 2016b], DoN [Zheng et al., 2016], DSCN [Minaee and Wang, 2017], SIFT-IRANSAC [Wu et al., 2014], MF-LBP [Kang and Wu, 2014], PalmNet [Genovese et al., 2019], PCANet [Chan et al., 2015], CR-CompCode [Zhang et al., 2017], LLDP [Luo et al., 2016], ANSS [Ramachandra et al., 2018], VGG-16 [Simonyan and Zisserman, 2015], ResNet-50 [He et al., 2016] and GoogLeNet [Szegedy et al., 2015] because their preprocessing steps cannot handle images in these two databases, which have complex backgrounds or no gaps between the fingers. In other words, if the proposed ROI-LAnet is not used to "help" other palmprint recognition methods to extract ROI, all of them would fail. The implementation details of these methods are described in Section 4.4.2. The ROIs provided in the IITD and PolyU and the ROI extracted from CASIA databases by us using the method from [Zhang et al., 2003] are used as inputs for all the methods, except for the proposed algorithm.

Table 4.4 and Table 4.5 compare the performance of the 11 methods on the NTU-PI-v1 and the four contactless palmprint databases, respectively. To compare the methods on the CASIA and IITD databases, the numbers with * are taken from the most recent palmprint survey [Fei et al., 2019a]. The difference between the numbers with * taken from [Fei et al., 2019a] and the numbers from our experiments can be caused by different parameters and fine-tuning techniques. These parameters and fine-tuning settings are not provided in [Fei et al., 2019a]. Thus theirs and our results are given. To compare the performance on the PolyU database, the numbers with ** are taken from [Zheng et al., 2016]. The numbers without any star indicator are from our experiments. Note that some results from

our experiments are obtained by using the code or program provided by the authors (see Section 4.4.2). Figure 4.11 shows CMC curves for the NTU-PI-v1. On the NTU-PI-v1 database, the proposed algorithm achieves 41.92%, 59.88% and 64.73% at rank-1, rank-15 and rank-30, respectively. The results show that the proposed algorithm outperforms the existing methods on the NTU-PI-v1 with a margin of 12.58% and 14.79% at rank-1 and rank-30, respectively. To visualize the characteristic of the proposed algorithm performance, examples of top-10 retrievals on the NTU-PI-v1 database are shown in Figure 4.13. The proposed algorithm also achieves the highest identification performance on the CASIA and IITD databases. In [Fei et al., 2019a], the authors showed that the deep learning based methods can achieve higher performance than conventional palmprint recognition methods. More clearly, the highest rank-1 accuracy on the CASIA and IITD databases was achieved by extracting ROI with a conventional method and then fine-tuning the ResNet-50 and GoogLeNet, respectively. Nevertheless, the proposed algorithm outperforms the ResNet-50 and GoogLeNet results reported in [Fei et al., 2019a] with a margin of 2.44% and 3.39%, respectively. The proposed algorithm is unable to achieve the rank-1 accuracy of 100% on the PolyU database as it is reported for DoN in [Zheng et al., 2016]. It achieves 99.77% rank-1 accuracy, which means that only 2 images are not retrieved at rank-1. It should be emphasized that in the experiment, the training/testing policy and hyper-parameters of the proposed algorithm are fixed for all databases whereas in [Zheng et al., 2016], besides a carefully designed palmprint descriptor, the authors also carefully tuned the algorithm parameters for a given database, especially the offset which defines the horizontal and vertical translations to improve the alignment between two images. In other words, each image from the testing set is being translated n_T times and then compared with one training image resulting in n_T comparison scores for one testing image. However, in the proposed algorithm, there is no such post-processing.

Because the existing palmprint recognition methods and NTU-CP-v1, CASIA, IITD and PolyU contactless palmprint databases are designed for commercial applications, likely

Figure 4.11: CMC curves of the proposed algorithm and state-of-the-art palmprint recognition and deep learning methods on NTU-PI-v1 database.

used for verification, rather than identification, the equal error rates (EER, the smaller, the better) which measure the verification performance are also reported in Table 4.5. The results show that the proposed algorithm outperforms all competing palmprint recognition methods on all the four databases. It achieves 0.76%, 0.26%, 0.15% and 0.73% EER on NTU-CP-v1, IITD, PolyU and CASIA databases, respectively, which are 11.19%, 0.25%, 0.07% and 0.06% less than the corresponding second best. It should be pointed out that the proposed algorithm, deep learning methods and DSCN are supervised learning method and they could "see" the potential imposters during training. Thus, their application in open-set scenarios should be further studied, e.g., employing models which target open-set recognition using supervised approach [Scheirer et al., 2013].

4.4.8 Performance Analysis

In this section, a performance analysis of the proposed algorithm on NTU-PI-v1 database is given. The comparison scores from Section 4.4.7 are used to analyze rank-1 identification

Table 4.4: Rank-1 and Rank-30 accuracy (%) of state-of-the-art palmprint and deep learning recognition methods on NTU-PI-v1 database.

Method	NTU-PI-v1	
	Rank-1	Rank-30
Proposed	41.92	64.73
CompCode	5.41	15.00
OrdinalCode	29.34	49.94
DoN	25.47	46.30
DOC	20.68	39.08
DSCN	15.47	33.70
SIFT-IRANSAC	10.14	21.24
MF-LBP	5.91	20.30
PalmNet	20.21	36.21
PCANet	20.56	45.62
CR-CompCode	23.37	45.27
LLDP	9.44	24.47
ANSS	11.78	29.85
ResNet-50	4.46	17.31
GoogLeNet	4.40	21.09
VGG-16	8.79	27.54

The highest percentage accuracy are highlighted.

Table 4.5: Rank-1 accuracy (%) and EER (%) of state-of-the-art palmprint and deep learning recognition methods on NTU-CP-v1, IITD, CASIA, PolyU databases.

Method	NTU-CP-v1		IITD		CASIA		PolyU	
	Rank-1	EER	Rank-1	EER	Rank-1	EER	Rank-1	EER
Proposed	95.34	0.76	99.61	0.26	97.65	0.73	99.75	0.15
CompCode	23.62	30.92	77.79*/97.76	1.39	79.27*/97.32	1.08	99.21**	0.68**
OrdinalCode	50.59	19.10	73.26*/97.50	2.09	73.32*/96.32	1.75	99.55**	0.23
DoN	52.62	19.51	98.16	1.45	96.89	1.16	100.00**	0.22
DOC	36.61	31.86	89.99*/90.01	16.79	78.51*/82.40	18.71	90.96	8.31
DSCN	49.15	11.95	97.10	0.51	95.23	0.79	82.25	6.21
SIFT-IRANSAC	74.75	15.26	91.46*/99.47	1.45	93.28*/96.69	3.46	88.58	15.74
MF-LBP	16.86	26.92	63.60	19.29	72.41	11.30	79.32	10.19
PalmNet	68.47	16.95	98.42	1.31	97.02	1.92	99.77	0.24
PCANet	72.03	7.76	97.37	1.18	95.53	1.46	99.66	0.45
CR-CompCode	52.63	18.19	98.16	1.74	91.73	3.18	97.18	1.02
LLDP	51.61	18.19	95.17*/98.16	1.74	93.00*/95.73	2.67	99.21	0.68
ANSS	31.86	17.72	97.24	0.92	92.76	1.79	70.96	14.91
ResNet-50	7.28	28.47	95.57*/76.74	3.68	95.21*/74.40	4.27	54.92	13.45
GoogLeNet	27.97	14.88	96.22*/87.91	1.45	93.84*/89.02	1.65	68.59	9.96
VGG-16	28.05	13.14	92.12*/86.47	1.97	94.01*/89.02	1.65	73.33	11.19

The numbers without any star indicator are from our experiments.

The numbers with * are from the recent review [Fei et al., 2019a].

The numbers with ** from the the DoN paper [Zheng et al., 2016].

accuracy (ACC) with respect to (a) a number of training palmprint samples and (b) the full hand image size. Thresholds of minimum and range for the number of training palmprint samples and image size are applied to calculate the respective accuracy. In other words, only the scores of probe images that satisfy the desired criteria are selected. Note that, the gallery size remains the same, which is 2035 palmprints. Figure 4.12a and 4.12b show $ACC@N(X)$, and $ACC@L(X)$, which are accuracy calculated from the rank-1 scores with X , and at least X training palmprint samples/pixels in the probe image, respectively. For example (see Figure 4.12a), if there are 4 training samples, $ACC@N(4) = 63.59\%$; if there are at least 4 training samples, $ACC@L(4) = 74.43\%$.

On average, each palmprint has only 2.88 training samples in NTU-PI-v1 database, whereas PolyU, CASIA, and IITD databases have at least 4 gallery samples. In NTU-PI-v1 database, adding training samples almost linearly increases the performance to reach 95.24% for 8 samples. In forensics, the training samples and the gallery can be created from multiple sources, e.g., Internet images, videos, images taken by prison imaging systems, etc. The image size also affects the performance. However, it is less significant compared with the number of training samples. Note that in the first two bins of the histogram in Figure 4.12b, the full image width/height ranges from 40 to 80 pixels meaning that the approximate palmprint ROI sizes have only 20×20 or 40×40 pixels. These show how challenging the NTU-PI-v1 database is. Nevertheless, to study forensic applications the database has to be challenging and reflect the variety of different scenarios.

4.5 Summary

In this chapter, the uncontrolled and uncooperative palmprint recognition problem is investigated. The new NTU Palmprints from the Internet (NTU-PI-v1) database is provided to study uncontrolled and uncooperative palmprint recognition for forensic applications. The database aims to simulate forensic scenarios and contains 7781 images from 2035 different palms collected from the Internet, which makes it the largest publicly available

Figure 4.12: The performance of the proposed algorithm as a function of (a) a number of available training samples per palmprint, (b) full hand image width/height in pixels.

palmprint database, in terms of palms and subjects. In addition to the NTU-PI-v1 database, the contactless palmprint database, NTU-CP-v1 is provided. In this chapter, the end-to-end trainable EE-PRnet, which consists of the alignment network, feature extraction network, and in-network ROI augmentation scheme based on dropout, is proposed. In the experiments, the network components of the proposed EE-PRnet and different training strategies are analyzed and their contributions to the performance are studied. The proposed EE-PRnet is compared with twelve representative, state-of-the-art palmprint recognition and three deep learning methods on five palmprint databases, including the uncontrolled and uncooperative NTU-PI-v1 and contactless NTU-CP-v1, CASIA, IITD and PolyU databases. The experimental results show that the proposed algorithm outperforms the considered palmprint and deep learning recognition methods. On the NTU-PI-v1 database, its performance is significantly higher even though the other methods have the "help" from the proposed ROI-LAnet, which targets to extract ROIs in images collected in the uncontrolled and uncooperative environment. The proposed algorithm on the NTU-PI-v1 database achieved the rank-1 and rank-30 accuracy of 41.92% and 64.73%, respectively.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

Figure 4.13: Examples of top-10 identification results. The probe images are in the yellow boxes in the original images. The true matches in the gallery are in the red boxes. In each subfigure, the relative scales of the hand images are preserved for comparison.

Chapter 5

The Nipple-Areola Complex for Offender Identification¹

5.1 Overview

Traditionally, in the investigation of child sexual assault, the investigators use biometric characteristics such as latent prints, or DNA from, e.g., body fluids collected at a crime scene [Young et al., 2006]. Nonetheless, sometimes the crime scene location may remain unknown or the crime is unreported, making the physical examination and collection of such evidence impossible. Nowadays, more evidence is also in the form of digital images, because the offenders take pictures for their personal record or to share with other pedophiles, e.g., on Dark Web. The offenders cover or hide their faces, to avoid personal identification and prosecution when forensic investigators or law enforcement agencies take over the child sexual assault materials. Some researchers proposed to use skin marks [Yu and Kong, 2018], skin texture [Chan et al., 2017], visualized veins [Zhang et al., 2012a] or androgenic hair [Su and Kong, 2014] to identify criminals and victims in images where no obvious traits such as face or tattoos are visible. In the previous chapters, studies of wrist and palmprint for forensic applications are presented. Some

¹This chapter is based on [Matkowski et al., 2019c]

of them may not be available or have limitations, i.e., not enough skin marks or hair, no clear veins, etc. make the identification challenging. Regardless of the availability of these characteristics, other ones can be used to support the evidence or provide some useful clues for the investigator. In this chapter, the male nipple-areola complex (NAC) is proposed for criminal identification. The offender's chest and thus NAC can be visible in some of the sexual abuse images. The offender's NAC can be captured simultaneously with the victim's wrist or palms, e.g., when the victim is touching the offender or being raped.

In the computer vision related literature, studies on nipples mostly aim nipple detection in color images for nude image detection [Wang et al., 2010b, Thaweekote et al., 2013] or nipple detection in medical images [Abdel-Nasser et al., 2016, Chae et al., 2017] According to my best knowledge, no studies of the NAC identification for forensic investigation were available before our study [Matkowski et al., 2019c].

In this chapter, a study of offender identification based on the NAC and NTU-Nipple-v1 dataset, which contains 2732 images of 428 different male nipple-areolae are presented. Popular deep learning and hand-crafted recognition methods are evaluated on the provided dataset. The results indicate that the NAC can be a useful characteristic for offender identification.

The rest of this chapter is organized as follows. In Section 5.2, basic information about the NAC morphology and related medical studies on the NAC properties are given. In Section 5.3, the new NTU-Nipple-v1 dataset is described. In Section 5.4, the experiments are performed. In Section 5.5, the summary is given.

5.2 Nipple-Areola Complex

NAC contains nipple (papilla), areola and tubercles. Figure 5.1 shows male NAC. The nipple is located approximately in the center of the NAC and surrounded by the areola. The areola's color is darker due to melanin accumulation and its structure is uneven due to the presence of tubercles [Nicholson et al., 2009, Wojcik et al., 2018]. The NAC, vary in

Figure 5.1: Left: Nipple-areola complex with approximately marked nipple (red solid), areola (blue dashed). Right: close-up image with marked tubercles (green solid).

shape, dimension, texture and color among different people. In addition, two NAC of the same person also vary in appearance [Zucca-Matthes et al., 2016]. In some cases of, e.g., breast surgery, massive weight loss, trauma, gender reassignment and gynecomastia affect the NAC’s morphology. The most information about the NAC visual appearance is found in the reconstructive and aesthetic surgery literature. The measurement studies focus on a position and configuration of the male NAC in the chest and its shape and dimensions [Beer et al., 2001, Shulman et al., 2001, Kasai et al., 2015, Tanini and Lo Russo, 2018]. Researchers also study the correlation between male NAC parameters and body weight, height, body mass index (BMI) [Shulman et al., 2001, Kasai et al., 2015] and the size of pectoral muscle [Tanini and Lo Russo, 2018]. However, the medical literature mostly focuses on female rather than male NAC [Kasai et al., 2015].

5.3 NTU-Nipple-v1 dataset

NTU-Nipple-v1 dataset was collected during two different sessions in a relatively well-controlled environment with at least a one-week time interval between sessions. The images were taken by Canon EOS 500D or NIKON D70s cameras. There were no strict posing requirements for subjects. The subjects are male and mainly Chinese, Indian, Malay

Figure 5.2: Examples of cropped NAC images. NAC are highlighted in the original chest images. The NAC are from the same subject from two different sessions.

Figure 5.3: Examples of different NAC from NTU-Nipple-v1 dataset from session 1 (top) and session 2 (bottom). Images in the same column are from the same subject's NAC.

and some Caucasian. The subjects' NAC are cropped using a square bounding box such that the nipple is approximately in the center of the image (see Figure 5.2). The dataset consists of 2732 images from 428 different NAC. Figure 5.3 shows examples of the NAC images from the dataset. Note, that the right and left NAC from the same subject are considered as a different NAC. The median NAC image size is 179 by 179 pixels and the median number of images of a single NAC is 6. Figure 5.4 shows some quantitative statistics of the dataset. NTU-Nipple-v1 dataset will be available online².

²<https://github.com/BFLTeam/NTU-Dataset>

Figure 5.4: Histograms showing (a) the distribution of images of the same single NAC, (b) the width/height of NAC images.

5.4 Experiments

5.4.1 Evaluation Protocol

NTU-Nipple-v1 dataset is split into training/gallery and testing/probe sets. The images in the gallery and probe are from different sessions. Each probe NAC has its corresponding one in the gallery and vice versa. The gallery and the probe sets consist of 1577 and 1155 NAC images, respectively. The experiments in the NAC identification (one-to-many comparison) scenario are performed. The results are presented in cumulative match characteristic (CMC) curves. CMC is a common evaluation metric for forensic applications. In the presented results, the CMC's ranks range from 1 to 30. The human expert may visually examine, e.g., top 30 ranks, which is a common practice in forensic investigation, e.g., in latent prints identification [Jain and Feng, 2009]. Therefore, not only rank-1 but also higher ranks such as rank-30 are important.

5.4.2 Recognition Algorithms

Five popular deep learning architectures AlexNet, VGG-16, VGG-Face, GoogLeNet and ResNet-50 are selected for evaluation. All these architectures except VGG-Face were

pretrained on the ImageNet dataset. The VGG-Face was pretrained on the face image dataset [Parkhi et al., 2015]. The pretrained models and fine-tuning approach are considered because NTU-Nipple-v1 dataset is relatively small and training from scratch is likely to overfit on such dataset. In addition, two popular hand-crafted features LBP with the matching method from [Ahonen et al., 2006] and scale invariant SIFT with L2 feature matching are also used in the evaluation.

5.4.3 Implementation Details

The deep learning algorithms are implemented using MatConvNet library and the pretrained models³. These pretrained models require fixed sized input images. Thus NAC images are resized to 224 by 224 pixels (227 by 227 for AlexNet). The last softmax layer is fine-tuned (Section 5.4.5) using ADAM [Kingma and Ba, 2014] optimizer with 0.001 learning rate, 0.0005 weight decay, 0.9 momentum and 128 batch size. When fine-tuning lower layers (Section 5.4.6) their learning rate is set to 10 times smaller, which is 0.0001. For hand-crafted features (Section 5.4.7), the color images are first transformed into the grayscale. Uniform LBP image is used with radius 2, 8 neighbors and the LBP histograms are extracted in blocks from a 7 by 7 uniform grid and matched (compared) using χ^2 distance. The SIFT descriptors are detected using VL_FEAT library⁴ with the threshold set to 100 and matched (compared) with the threshold set to 1.6.

5.4.4 Deep features

In this experiment, the deep features extracted from the one before the last layer are evaluated. The feature vectors are L2-normalized and the 1-NN L2 distance is used to obtain comparison scores. The feature vector sizes for AlexNet, VGG-16, VGG-Face, GoogLeNet and ResNet-50 are 4096, 4096, 4096, 2048, 1024, respectively. The results are

³<http://www.vlfeat.org/matconvnet/>

⁴<http://www.vlfeat.org/overview/sift.html>

Figure 5.5: CMC curve of different deep learning architectures where (a) L2 distance and (b) softmax and SVM are used for identification.

given in Figure 5.5a. AlexNet, which is the "shallowest" among the selected architectures, achieves the highest performance. Its rank-1 and rank-30 accuracies are 46.58% and 89.44%, respectively.

5.4.5 Fine-tuning the last layer

In the previous Section 5.4.4, the 1-nearest neighbor in the Euclidean space (L2) is used for identification. In this experiment, the softmax and linear support vector machine (SVM) are employed to investigate the parametric classifiers. SVM is used in the one-against-all setting, the feature vectors are standardized (z-score) before training, and the positive labels are set to 1 and the negative ones to -1. The results are presented in Figure 5.5b. For all five deep learning architectures, the performance increases when using softmax or SVM comparing to the performance in Section 5.4.4. The order of the five deep architectures with respect to accuracy remains the same as in Section 5.4.4. In addition, SVM generally achieves higher results. The rank-1, rank-15 and rank-30 accuracy are given in Table 5.1.

Table 5.1: The rank accuracies (%) of the deep learning architectures using softmax and SVM for identification.

Architecture	Classifier	Rank-1	Rank-15	Rank-30
AlexNet	SVM	59.57	86.92	91.60
AlexNet	Softmax	58.00	88.57	92.12
VGG-16	SVM	48.31	81.12	86.49
VGG-16	Softmax	46.58	82.33	89.43
VGG-Face	SVM	45.71	81.88	86.06
VGG-Face	Softmax	47.27	82.68	88.31
GoogLeNet	SVM	29.26	68.13	75.06
GoogLeNet	Softmax	27.01	66.14	75.75
ResNet-50	SVM	22.51	57.49	65.80
ResNet-50	Softmax	16.36	50.21	59.82

5.4.6 Fine-tuning fully connected layers of AlexNet

In this experiment, the aim is to fine-tune the fully connected layers. AlexNet architecture is selected because it achieves the highest performance in previous experiments in Sections 5.4.4 and 5.4.5. To avoid, fast overfitting, the dropout layer with the dropout rate set to 0.5 is put before the last fully connected layer and the data augmentation on the fly is used. For augmentation, the contrast and the saturation random shifts are applied with the shift factor set to 0.5. Additionally, aggressive spatial random vertical and horizontal translations and scaling are applied. To avoid bordering effects, the 1.3 bigger NAC image is translated and scaled and then cropped as shown in Figure 5.6. The crop is always within the image boundaries. The results are presented in Figure 5.7a. In the figure legend, the "FTFC" indicates the results obtained in this experiment. The highest rank-1 accuracy of 62.16% is achieved by FTFC-AlexNet-L2. This result indicates that the employed data augmentation

Figure 5.6: Schematic illustration of the spatial data augmentation. During the training, the bigger NAC image is randomly cropped within the red box and input to the network.

can boost performance. Especially, when 1-nearest neighbor L2 is used for identification, the rank-1 accuracy increases by 15.58% (AlexNet-L2 vs FCFT-AlexNet-L2).

5.4.7 Modified AlexNet vs hand-crafted features

In this experiment, the comparison of deep learning and hand-crafted based methods is given. The hand-crafted LBP and SIFT are used as described in Section 5.4.3. LBP and SIFT achieve 62.51% and 57.5% rank-1 accuracy, respectively. These results are higher than all deep learning results from Section 5.4.4. Note that, LBP and SIFT also use the same type of classifier (1-nearest neighbor L2 distance) as is used in Section 5.4.4. In addition, LBP outperforms at rank-1 all the results from Section 5.4.5, where the SVM and softmax classifier are used for identification. SIFT performance at rank-1 is only worse than AlexNet. These results indicate that more specialized deep learning architectures and training techniques are needed to increase the identification accuracy.

The additional experiment, using AlexNet is performed to give some insights and further investigate the possibility of boosting deep learning performance. The AlexNet's fully connected layers (which contain over 54 million parameters) are replaced with one fully connected (4.7 million parameters) layer with 512 neurons and softmax layer, with two dropouts put in-between with 0.5 dropout rate. The same augmentation as in Section

5.4.6 is used. The results are presented in Figure 5.7b and Table 5.2. Modified AlexNet and L2 distance are able to achieve the highest performance among the selected methods. It achieves 70.2% and 95.67% rank-1 and rank-30 accuracy, respectively. Employing SVM further increases the identification performance to 75.76% and 96.54%, rank-1 and rank-30 accuracy, respectively. Fig. 10 shows the examples of top-10 matches returned by the Modified AlexNet-SVM.

Figure 5.7: (a) CMC curve comparing AlexNet (Sections 5.4.4 and 5.4.5) and the AlexNet with fully connected layers fine-tuned and the augmentation applied (Section 5.4.6). (b) CMC curve comparing the best AlexNet from 5.4.5 and 5.4.6 and the Modified AlexNet (Section 5.4.7), and hand-crafted LBP and SIFT based methods.

5.5 Summary

In this chapter, a single NAC is proposed to identify child sexual offenders in the images where chest and NAC are visible. NTU-Nipple-v1 dataset which contains 428 different male NAC is provided. Five popular deep learning architectures and two well-known hand-crafted methods are employed in the evaluation. The experimental results indicate that the NAC can be a useful characteristic for offender identification. In the experiments,

Table 5.2: The rank accuracies (%) of the AlexNet architectures and the hand-crafted methods.

Method	Rank-1	Rank-15	Rank-30
Modified AlexNet-SVM	75.76	94.46	96.54
Modified AlexNet-L2	70.30	91.95	95.67
LBP	62.51	84.33	94.72
FTFC-AlexNet-L2	62.16	89.96	86.49
AlexNet-SVM	59.57	86.93	91.60
SIFT	57.40	82.42	86.41

the highest reported rank-1, rank-15 and rank-30 identification accuracies are 75.76%, 94.46% and 96.54%, respectively.

Figure 5.8: Rows: examples of top-10 identification results using Modified AlexNet-SVM. The match between each probe and the rank-ordered gallery images is in the blue boxes.

Chapter 6

Discussion

In this chapter, I summarize the main contributions of this thesis, discuss their significance and suggest directions, which could be explored in the future.

6.1 Summary of Main Contributions

In this thesis, I studied problems of criminal and victim identification, which are not well-explored by the biometric community, yet. I presented these problems and proposed effective algorithms and databases useful for forensic applications. The following is the summary of the main technical contributions of this thesis:

- New biometric database, NTU-Wrist-Image-Database-v1 to study wrist identification for forensic applications. The database consists of 3945 images from 731 different wrists, including 205 pairs of wrist images collected from the Internet, taken under uncontrolled and uncooperative environment.
- The wrist identification algorithm (WMM) which includes skin segmentation, key point localization based on the shortest paths, image to template alignment, large feature set extraction, matching schemes, and post-recognition score analysis to select the best alignment and matching scheme.

- In the experiments, rank-1 accuracies for non-standard and Internet images are 38.64% and 24.88%, respectively, and rank-15 accuracies are over 60% and 50%, respectively. The proposed algorithm matches the masked Baltimore rioter within top-10 ranks. He is retrieved at the 1st rank and the 8th rank depending on the algorithm parameter. It shows that wrist is possible to be used for forensic investigation.
- New biometric database, NTU Palmprints from the Internet (NTU-PI-v1) to study uncontrolled and uncooperative palmprint recognition for forensic applications. The database aims to simulate forensic scenarios and contains 7781 images from 2035 different palms collected from the Internet, which makes it the largest publicly available palmprint database, in terms of palms and subjects.
- An end-to-end trainable deep neural network (EE-PRnet) which consists of the localization and alignment network to extract palmprint ROI, feature extraction network, and in-network ROI augmentation scheme based on dropout. Different training strategies for the EE-PRnet are proposed.
- The proposed EE-PRnet is compared with twelve representative palmprint recognition and three deep learning methods on five palmprint databases. The experimental results show that the proposed algorithm, which is designed for forensic applications, could also be applied to well-controlled, closed-set contactless palmprint recognition and outperforms the twelve considered palmprint recognition and three deep learning methods. The proposed algorithm on the NTU-PI-v1 database achieved the rank-1 and rank-30 accuracy of 41.92% and 64.73%, respectively.
- New biometric database, NTU-Nipple-v1 to study the proposed nipple-areola complex for offender identification. The dataset consists of 2732 images from 428 different NAC, taken under a relatively well-controlled environment.
- Evaluation of popular deep learning and hand-crafted based methods on NTU-Nipple-v1 database. In the experiments, the highest reported rank-1, rank-15 and

rank-30 NAC identification accuracies are 75.76%, 94.46% and 96.54%, respectively. These preliminary results indicate that NAC can be useful for offender identification.

6.2 Accuracy

From the accuracy perspective, one may say that some of the obtained accuracies, especially in uncontrolled and uncooperative environments do not seem spectacular. The poor quality and uncontrolled images are common in forensic investigation, which makes the identification rates considerably lower than in cooperative and well-controlled environments. For latent palmprints of ugly quality in [Jain and Feng, 2009], Jain et al. showed in the CMC that the rank-1 accuracy is around 25% whereas rank-20 accuracy is slightly above 40%. In other uncontrolled and uncooperative areas, i.e., person re-identification from surveillance cameras, Wang et al. [Wang et al., 2018] reported rank-1 accuracies of 34.8-58.2%. Even well-studied and mature fingerprint recognition methods applied to bad quality images achieve low accuracy, e.g., the rank-1 accuracy of 21.2-34.1% for commercial fingerprint matchers on fingerprint images with ugly quality reported by Paulino et al. [Paulino et al., 2013]. In forensic investigation, not only rank-1 but also higher ranks, e.g., rank-30 are valuable because they can reduce the suspect list and the number of manual searches by forensic investigators. According to [Jain et al., 2012], in case of face recognition, even up top-200 retrievals can be used for manual analysis by forensic examiners.

6.3 Database

6.3.1 Size

The databases provided in this thesis have 2035 (7781 images), 731 (3945 images) and 428 (2732 images) different palms, wrists and NAC, respectively. Databases or galleries in

forensic are not always very large. The list of suspects (gallery size) can be narrowed down using some hints, such as age, sex, ethnicity, location area, date, etc. This information can be obtained from evidence images or other related sources.

For example, let us consider a case of a child sexual abuse at a certain area. Investigators may only need to search suspects in a sex offender registry. In a study by Harris et al. [Harris et al., 2014] the number of registered sex offenders (RSO) in the US, estimated by the NCMEC, is 236 per 100,000 citizens, which is 0.236%. In the NTU-PW-v1 database, there are 2035 different palms in the gallery. Given this gallery size and using the above statistic: $2035/0.00236 = 847,457$, which means that the gallery size is equivalent to the total number of sexual offenders in a city with approximately 850 thousand citizens. In other words, assuming that there is 0.236% of sex offenders in a population, its size can approximate the gallery in each individual city, e.g., in Ontario except for Toronto and Ottawa, in California, except for San Diego, San Jose, San Francisco and LA, in Poland, except for Warsaw. It can be used in each individual city in countries such as Norway, Denmark, Finland, Switzerland, Portugal¹. Finally, it can be even used in each individual country such as Luxembourg, Andorra, Liechtenstein, Monaco and Maldives². The population of my hometown, Opole is around 130,000 citizens, which means that using the above statistics the gallery would have around 310 suspects, which is far less than in all databases provided in this thesis.

Taking another example, investigators would like to search a US suspect who is involved in a child sexual abuse case in Cambodia. From the Exif of evidence photos, investigators know when the abuse happened. Using this information, investigators only need to search all US tourists in Cambodia on a particular date. Statistics show that 153,953 US citizens visited Cambodia in 2011 (the most updated information on the website³). Assume that each of them stayed 5 days. The average number of US visitors in Cambodia per day was

¹<https://www.wolframalpha.com/> - search: largest city in Portugal

²<https://www.wolframalpha.com/> - search: Maldives population

³https://en.wikipedia.org/wiki/Tourism_in_Cambodia

only 2108. It is only slightly larger than the database containing 2035 different palms. If the investigators can narrow down the suspect list, e.g., filtering by sex (since women are less likely to be in child abuse images [INTERPOL, 2018]), families, etc. the number of suspects can be reduced drastically. Note that Cambodia is a hot spot of child sexual abuse and sexual exploitation of children in the context of travel and tourism.

Take Jihadi John, the terrorist, who beheaded James Foley, Steven Sotloff, David Haines and Alan Henning as another example. Because of his London accent, investigators only needed to search London citizens in the Middle East. Thus, the number of suspects was greatly reduced.

6.3.2 Collection

The databases provided in this thesis were collected from two types of sources 1) Internet images considered as an uncontrolled and uncooperative environment, and 2) indoor sessions considered as a well-controlled and cooperative environment without strict posing requirements.

In forensic practice, the evidence images are always from uncontrolled and uncooperative environments, e.g., child sexual abuse images on the Internet or hard drives or terrorist propaganda images on social media, rioter image taken by press photographers, etc. No public databases for target forensic applications are available. Therefore, uncontrolled and uncooperative images from Internet galleries are considered to simulate evidence images in very challenging scenarios.

The images from the indoor session should be considered as less challenging scenarios, where for example, one probe or gallery image is taken, e.g., during the suspect booking and another evidence has relatively good quality. Currently, during bookings, police departments collect images of face (mugshots), high-resolution fingerprints and palm prints, tattoo, marks and scars to create a database. In case of sexual assault, the crime scene guideline provided by [National Forensic Science Technology Center, 2013] also says:

”Suspects should be photographed in the clothing that they are wearing, in both overall and midrange photographs. A separate photograph of the face as well as the hands should be taken. Any scar, marks, tattoos and injuries, or lack thereof, should be photographed close-up.” Thus, it is likely that some of the hand images are already available to create galleries. In addition to bookings, law enforcement may also create galleries of already known subjects from the Internet, social media, hard drives, relatives, etc. In such cases, subjects may not be even aware of the collection process.

6.4 Application

Typically, biometric systems for commercial applications such as access control or time and attendance perform real-time verification or identification in the open-set scenario. Thus, in such applications due to security reasons, it is critical to reject imposter users, automatically. However, the forensic applications studied in this thesis do not have such strict requirements. First, the real-time requirement is not necessary. Second, the open-set identification, which involves verification and rejecting imposters is also not necessary, because the candidate list, e.g., top-30 candidates returned by the system are assumed to be manually reviewed and analyzed by a human expert or experts. In other words, even when the subject is not in the gallery, the system returns a candidate list of the most similar subjects from the gallery. The same assumptions are also in related forensic applications such as latent prints [Jain and Feng, 2009], tattoos [Ngan et al., 2018] and even face [Best-Rowden et al., 2014] recognition, except watch list surveillance, which may require real-time operation and accurate rejection of subjects that are not in the watch list.

6.5 Future Work

Accurate identification is the first and the main goal. Therefore, the studies should be very much oriented in the direction of improving the accuracy of the algorithms. Suggested

potential future directions based on this thesis are as follows:

- Development of databases. Increasing the database size in terms of images and subjects, collecting more images taken in the uncontrolled and uncooperative environment, e.g., from the Internet. Not only the subject identity labels, but also other labels such as gender, ethnicity, age, sample quality, etc. could be considered.
- Currently the proposed algorithms require re-training classifiers when a new sample is enrolled in the gallery. In cases of not very large galleries, it may not be an issue because the classifier training is fast for considered database sizes. In larger-scale biometric problems, this scalability issue is typically solved by distance-based classifiers, which do not require training, e.g., nearest neighbor. Other approach could be designing or employing discriminative classifiers which target open-set recognition. On the other hand, larger-scale databases and more training samples may improve performance of deep learning algorithms.
- Fusion of wrist and palmprint could be also considered. Nevertheless, it should be pointed out that in the provided databases palmprint and wrist are not always observable together, i.e. fist, wristband, watch, bracelet, holding object, gesture, finger bending, etc. may affect the fusion of wrist and palmprint.
- Adding full fingers or considering full palmar side of the hand is another option. However, some of the issues listed in the point above also apply. Thus many fingers will not be available because of occlusions and posing. Moreover, to accurately localize separate fingers, finger joints, tips and valleys, could require additional annotation efforts.
- Current wrist identification algorithm is entirely hand-crafted. A deep learning version should be developed. In particular, wrist key points detected by current approach could be considered as ground truth labels for localization network training. Generally, it could be interesting to study localization network properties during

end-to-end training using identities instead of key point as training signals. Similarly, in the proposed localization network for palmprint, it is unclear how the localization network should be trained (with identity signals) without the risk of collapsing into a solution which produces transformations far from acceptable. Current solution is to set relatively small learning rates in the localization network. Another solution could be to constrain the domain of possible transformations.

- Current EE-PRnet uses pre-trained, pruned VGG-16 and is trained using softmax loss. Thus, different architectures, training from scratch and other loss functions including auxiliary loss which could employ, e.g., gender or ethnicity information, could be studied.
- The proposed NAC was studied using popular existing recognition methods. The NAC recognition algorithm including the alignment or for example normalization using polar coordinates should be proposed. The fusion of two NAC is also an option.

Bibliography

- [Abdel-Nasser et al., 2016] Abdel-Nasser, M., Saleh, A., Moreno, A., and Puig, D. (2016). Automatic nipple detection in breast thermograms. *Expert Syst. Appl.*, 64:365–374.
- [Achanta et al., 2011] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2011). SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE TPAMI*, 34(11):2274–2282.
- [Ahonen et al., 2006] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE TPAMI*, 28(12):2037–2041.
- [Aykut and Ekinici, 2015] Aykut, M. and Ekinici, M. (2015). Developing a contactless palmprint authentication system by introducing a novel ROI extraction method. *Image Vis. Comput.*, 40:65–74.
- [Bao and Guo, 2016] Bao, X. and Guo, Z. (2016). Extracting region of interest for palmprint by convolutional neural networks. In *2016 Sixth Int. Conf. Image Process. Theory, Tools Appl.*, pages 1–6. IEEE.
- [Beer et al., 2001] Beer, G. M., Budi, S., Seifert, B., Morgenthaler, W., Infanger, M., and Meyer, V. E. (2001). Configuration and localization of the nipple-areola complex in men. *Plast. Reconstr. Surg.*, 108(7):1947–52; discussion 1953.

- [Best-Rowden et al., 2014] Best-Rowden, L., Han, H., Otto, C., Klare, B. F., and Jain, A. K. (2014). Unconstrained Face Recognition: Identifying a Person of Interest From a Media Collection. *IEEE Trans. Inf. Forensics Secur.*, 9(12):2144–2157.
- [Bookstein, 1989] Bookstein, F. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *TPAMI*, 11(6):567–585.
- [Bruna and Mallat, 2013] Bruna, J. and Mallat, S. (2013). Invariant Scattering Convolution Networks. *IEEE TPAMI*, 35(8):1872–1886.
- [Cao and Jain, 2015] Cao, K. and Jain, A. K. (2015). Latent orientation field estimation via convolutional neural network. In *ICB*, pages 349–356.
- [Chae et al., 2017] Chae, S.-H., Jeong, J.-W., Choi, J.-H., Chae, E. Y., Kim, H. H., Choi, Y.-W., and Lee, S. (2017). Fully automated nipple detection in digital breast tomosynthesis. *Comput. Methods Programs Biomed.*, 143:113–120.
- [Chai et al., 2016] Chai, T., Wang, S., and Sun, D. (2016). A palmprint ROI extraction method for mobile devices in complex environment. In *2016 IEEE 13th Int. Conf. Signal Process.*, pages 1342–1346. IEEE.
- [Champod and Tistarelli, 2017] Champod, C. and Tistarelli, M. (2017). Biometric Technologies for Forensic Science and Policing: State of the Art. In *Handb. Biometrics Forensic Sci.*, pages 1–15. Springer, Cham.
- [Chan and Kong, 2015] Chan, F. K. S. and Kong, A. W. K. (2015). Using Hair Follicles with Leg Geometry to Align Androgenic Hair Patterns. *2015 Eur. Intell. Secur. Informatics Conf.*, pages 137–140.
- [Chan et al., 2017] Chan, F. K. S., Li, X., and Kong, A. W. K. (2017). A Study of Distinctiveness of Skin Texture for Forensic Applications Through Comparison With Blood Vessels. *TIFS*, 12(8):1900–1915.

- [Chan et al., 2015] Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., and Ma, Y. (2015). PCANet: A Simple Deep Learning Baseline for Image Classification? *TIP*, 24(12):5017–5032.
- [Cormen et al., 2009] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). Introduction to Algorithms. chapter VI Graph A, pages 701–704. The MIT Press, third edit edition.
- [Cula et al., 2005] Cula, O. G., Dana, K. J., Murphy, F. P., and Rao, B. K. (2005). Skin Texture Modeling. *Int. J. Comput. Vis.*, 62(1-2):97–119.
- [Dantcheva et al., 2016] Dantcheva, A., Elia, P., and Ross, A. (2016). What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. *IEEE Trans. Inf. Forensics Secur.*, 11(3):441–467.
- [Das et al., 2019] Das, R., Piciuccio, E., Maiorana, E., and Campisi, P. (2019). Convolutional Neural Network for Finger-Vein-Based Biometric Identification. *IEEE Trans. Inf. Forensics Secur.*, 14(2):360–373.
- [Daugman, 1993] Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1148–1161.
- [Dian and Dongmei, 2016] Dian, L. and Dongmei, S. (2016). Contactless palmprint recognition based on convolutional neural network. In *2016 IEEE 13th Int. Conf. Signal Process.*, pages 1363–1367. IEEE.
- [Dubey et al., 2014] Dubey, S. R., Singh, S. K., and Singh, R. K. (2014). Rotation and Illumination Invariant Interleaved Intensity Order-Based Local Descriptor. *IEEE TIP*, 23(12):5323–5333.
- [Duta, 2009] Duta, N. (2009). A survey of biometric technology based on hand shape. *Pattern Recognit.*, 42(11):2797–2806.

- [ECPAT International, 2018] ECPAT International (2018). TRENDS IN ONLINE CHILD SEXUAL ABUSE MATERIAL. Technical report.
- [Fei et al., 2016a] Fei, L., Jie Wen, Zheng Zhang, Ke Yan, and Zuofeng Zhong (2016a). Local multiple directional pattern of palmprint image. In *2016 23rd Int. Conf. Pattern Recognit.*, pages 3013–3018. IEEE.
- [Fei et al., 2019a] Fei, L., Lu, G., Jia, W., Teng, S., and Zhang, D. (2019a). Feature Extraction Methods for Palmprint Recognition: A Survey and Evaluation. *IEEE Trans. Syst. Man, Cybern. Syst.*, 49(2):346–363.
- [Fei et al., 2016b] Fei, L., Xu, Y., Tang, W., and Zhang, D. (2016b). Double-orientation code and nonlinear matching scheme for palmprint recognition. *PR*, 49:89–101.
- [Fei et al., 2017] Fei, L., Xu, Y., Teng, S., Zhang, W., Tang, W., and Fang, X. (2017). Local Orientation Binary Pattern with Use for Palmprint Recognition. In *Chinese Conf. Biometric Recognit.*, pages 213–220. Springer, Cham.
- [Fei et al., 2019b] Fei, L., Zhang, B., Xu, Y., Guo, Z., Wen, J., and Jia, W. (2019b). Learning Discriminant Direction Binary Palmprint Descriptor. *TIP*, 28(8):3808–3820.
- [Ferrer and Morales, 2011] Ferrer, M. A. and Morales, A. (2011). Hand-Shape Biometrics Combining the Visible and Short-Wave Infrared Bands. *IEEE Trans. Inf. Forensics Secur.*, 6(4):1305–1314.
- [Geladi and Kowalski, 1986] Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Anal. Chim. Acta*, 185(C):1–17.
- [Genovese et al., 2019] Genovese, A., Piuri, V., Plataniotis, K. N., and Scotti, F. (2019). PalmNet: Gabor-PCA Convolutional Networks for Touchless Palmprint Recognition. *TIFS*, 14(12):3160–3174.

- [Guo et al., 2009] Guo, Z., Zhang, D., Zhang, L., and Zuo, W. (2009). Palmprint verification using binary orientation co-occurrence vector. *Pattern Recognit. Lett.*, 30(13):1219–1227.
- [Harris et al., 2014] Harris, A. J., Levenson, J. S., and Ackerman, A. R. (2014). Registered Sex Offenders in the United States. *Crime Delinq.*, 60(1):3–33.
- [Hassanat et al., 2016] Hassanat, A. B. A., Alhasanat, M. B., Abbadi, M. A., Btoush, E., Al-Awadi, M., and Tarawneh, A. S. (2016). Victory Sign Biometric for Terrorists Identification. page 7.
- [He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *CVPR*, pages 770–778. IEEE.
- [Hong et al., 2017] Hong, H. G., Lee, M. B., and Park, K. R. (2017). Convolutional Neural Network-Based Finger-Vein Recognition Using NIR Image Sensors. *Sensors (Basel)*, 17(6).
- [Hu et al., 2015] Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., Li, S. Z., and Hospedales, T. (2015). When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition. In *ICCV*.
- [Huang et al., 2015] Huang, D., Tang, Y., Wang, Y., Chen, L., and Wang, Y. (2015). Hand-Dorsa Vein Recognition by Matching Local Features of Multisource Keypoints. *IEEE Trans. Cybern.*, 45(9):1823–1837.
- [Huynh et al., 2014] Huynh, N. Q., Xu, X., Kong, A. W. K., and Subbiah, S. (2014). A preliminary report on a full-body imaging system for effectively collecting and processing biometric traits of prisoners. In *2014 IEEE Symp. Comput. Intell. Biometrics Identity Manag.*, pages 167–174. IEEE.
- [INTERPOL, 2018] INTERPOL (2018). Towards a Global Indicator on Unidentified Victims in Child Sexual Exploitation Material. Technical report.

- [ISO/IEC, 2017] ISO/IEC (2017). ISO/IEC 2382-37:2017(en), Information technology — Vocabulary — Part 37: Biometrics. Technical report.
- [Ito et al., 2015] Ito, K., Sato, T., Aoyama, S., Sakai, S., Yusa, S., and Aoki, T. (2015). Palm region extraction for contactless palmprint recognition. In *2015 Int. Conf. Biometrics*, pages 334–340. IEEE.
- [Jaderberg et al., 2015] Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). Spatial Transformer Networks. In *NIPS*.
- [Jain et al., 2004] Jain, A., Ross, A., and Prabhakar, S. (2004). An Introduction to Biometric Recognition. *IEEE Trans. Circuits Syst. Video Technol.*, 14(1):4–20.
- [Jain and Feng, 2009] Jain, A. K. and Feng, J. (2009). Latent Palmprint Matching. *IEEE TPAMI*, 31(6):1032–1047.
- [Jain et al., 2012] Jain, A. K., Klare, B., and Park, U. (2012). Face Matching and Retrieval in Forensics Applications. *IEEE Multimed.*, 19(1):20–20.
- [Jain et al., 2016] Jain, A. K., Nandakumar, K., and Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *PRL*, 79:80–105.
- [Jalali et al., 2015] Jalali, A., Mallipeddi, R., and Lee, M. (2015). Deformation Invariant and Contactless Palmprint Recognition Using Convolutional Neural Network. *Proc. 3rd Int. Conf. Human-Agent Interact.*, pages 209–212.
- [Jia et al., 2014] Jia, W., Hu, R.-X., Lei, Y.-K., Zhao, Y., and Gui, J. (2014). Histogram of Oriented Lines for Palmprint Recognition. *IEEE Trans. Syst. Man, Cybern. Syst.*, 44(3):385–395.
- [Jia et al., 2008] Jia, W., Huang, D.-S., and Zhang, D. (2008). Palmprint verification based on robust line orientation code. *Pattern Recognit.*, 41(5):1504–1513.

- [Jia et al., 2017] Jia, W., Zhang, B., Lu, J., Zhu, Y., Zhao, Y., Zuo, W., and Ling, H. (2017). Palmprint Recognition Based on Complete Direction Representation. *IEEE Trans. Image Process.*, 26(9):4483–4498.
- [Jourabloo and Liu, 2015] Jourabloo, A. and Liu, X. (2015). Pose-Invariant 3D Face Alignment. In *ICCV*, pages 3694–3702. IEEE.
- [Kabacinski and Kowalski, 2011] Kabacinski, R. and Kowalski, M. (2011). Vein pattern database and benchmark results. *Electron. Lett.*, 47(20):1127.
- [Kanazawa et al., 2016] Kanazawa, A., Jacobs, D. W., and Chandraker, M. (2016). Warp-Net: Weakly Supervised Matching for Single-view Reconstruction. In *CVPR*.
- [Kang and Wu, 2014] Kang, W. and Wu, Q. (2014). Contactless Palm Vein Recognition Using a Mutual Foreground-Based Local Binary Pattern. *TIFS*, 9(11):1974–1985.
- [Kanhanged et al., 2010] Kanhanged, V., Kumar, A., and Zhang, D. (2010). Human hand identification with 3D hand pose variations. In *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work.*, pages 17–21. IEEE.
- [Kanhanged et al., 2011a] Kanhanged, V., Kumar, A., and Zhang, D. (2011a). A Unified Framework for Contactless Hand Verification. *IEEE Trans. Inf. Forensics Secur.*, 6(3):1014–1027.
- [Kanhanged et al., 2011b] Kanhanged, V., Kumar, A., and Zhang, D. (2011b). Contactless and Pose Invariant Biometric Identification Using Hand Surface. *TIP*, 20(5):1415–1424.
- [Kasai et al., 2015] Kasai, S., Shimizu, Y., Nagasao, T., Ohnishi, F., Minabe, T., Momosawa, A., and Kishi, K. (2015). An Anatomic Study of Nipple Position and Areola Size in Asian Men. *Aesthetic Surg. J.*, 35(2):NP20–NP27.
- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. In *ICLR*.

- [Klare et al., 2015] Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Burge, M., and Jain, A. K. (2015). Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *CVPR*, pages 1931–1939. IEEE.
- [Kong and Zhang, 2004] Kong, A. and Zhang, D. (2004). Competitive coding scheme for palmprint verification. In *ICPR*, pages 520–523. IEEE.
- [Kong et al., 2009] Kong, A., Zhang, D., and Kamel, M. (2009). A survey of palmprint recognition. *PR*, 42(7):1408–1418.
- [Kong and Zhang, 2002] Kong, W. K. and Zhang, D. (2002). Palmprint Texture Analysis Based on Low-Resolution Images for Personal Authentication. In *Proc. 16 th Int. Conf. Pattern Recognit.*, page 30807. IEEE Computer Society Press.
- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105. Curran Associates Inc.
- [Kruppa et al., 2002] Kruppa, H., Bauer, M. A., and Schiele, B. (2002). Skin Patch Detection in Real-World Images. In *Annu. Symp. Pattern Recognit. DAGM*, page 109f.
- [Kumar, 2014] Kumar, A. (2014). Importance of Being Unique From Finger Dorsal Patterns: Exploring Minor Finger Knuckle Patterns in Verifying Human Identities. *IEEE Trans. Inf. Forensics Secur.*, 9(8):1288–1298.
- [Kumar, 2019] Kumar, A. (2019). Toward More Accurate Matching of Contactless Palmprint Images Under Less Constrained Environments. *IEEE Trans. Inf. Forensics Secur.*, 14(1):34–47.
- [Kumar and Ravikanth, 2009] Kumar, A. and Ravikanth, C. (2009). Personal Authentication Using Finger Knuckle Surface. *IEEE Trans. Inf. Forensics Secur.*, 4(1):98–110.

- [Kumar and Xu, 2016] Kumar, A. and Xu, Z. (2016). Personal Identification Using Minor Knuckle Patterns From Palm Dorsal Surface. *IEEE Trans. Inf. Forensics Secur.*, 11(10):2338–2348.
- [Kumar and Zhou, 2012] Kumar, A. and Zhou, Y. (2012). Human Identification Using Finger Images. *IEEE Trans. Image Process.*, 21(4):2228–2244.
- [LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- [Li and Kim, 2017] Li, G. and Kim, J. (2017). Palmprint recognition with Local Micro-structure Tetra Pattern. *PR*, 61:29–46.
- [Liang et al., 2019] Liang, X., Zhang, D., Lu, G., Guo, Z., and Luo, N. (2019). A Novel Multicamera System for High-Speed Touchless Palm Recognition. *IEEE Trans. Syst. Man, Cybern. Syst.*, (early acc:1–15.
- [Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proc. Seventh IEEE Int. Conf. Comput. Vis.*, volume 2, pages 1150–1157. IEEE.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110.
- [Lu et al., 2008] Lu, G., Zhang, D., Kong, W. K., and Wong, M. (2008). A Palmprint Authentication System. In *Handb. Biometrics*, pages 171–187. Springer US, Boston, MA.
- [Luo et al., 2016] Luo, Y.-T., Zhao, L.-Y., Zhang, B., Jia, W., Xue, F., Lu, J.-T., Zhu, Y.-H., and Xu, B.-Q. (2016). Local line directional pattern for palmprint recognition. *PR*, 50:26–44.

- [Matkowski et al., 2019a] Matkowski, W. M., Chai, T., and Kong, A. W. K. (2019a). Palmprint Recognition in Uncontrolled and Uncooperative Environment. *IEEE Trans. Inf. Forensics Secur.*
- [Matkowski et al., 2019b] Matkowski, W. M., Chan, F. K. S., and Kong, A. W. K. (2019b). A study on wrist identification for forensic investigation. *Image Vis. Comput.*, 88.
- [Matkowski et al., 2019c] Matkowski, W. M., Matkowski, K., Kong, A. W.-K., and Hall, C. L. (2019c). The Nipple-Areola Complex for Criminal Identification. In *Int. Conf. Biometrics*.
- [Meraoumia et al., 2017] Meraoumia, A., Kadri, F., Bendjenna, H., Chitroub, S., and Bouridane, A. (2017). Improving Biometric Identification Performance Using PCANet Deep Learning and Multispectral Palmprint. In *Signal Process. Secur. Technol.*, pages 51–69. Springer International Publishing.
- [Michael et al., 2012] Michael, G. K. O., Connie, T., and Teoh, A. B. J. (2012). A contactless biometric system using multiple hand features. *J. Vis. Commun. Image Represent.*, 23(7):1068–1084.
- [Minaee and Wang, 2015] Minaee, S. and Wang, Y. (2015). Fingerprint recognition using translation invariant scattering network. In *2015 IEEE Signal Process. Med. Biol. Symp.*, pages 1–6. IEEE.
- [Minaee and Wang, 2017] Minaee, S. and Wang, Y. (2017). Palmprint Recognition Using Deep Scattering Convolutional Network. In *Int. Symp. Circuits Syst. IEEE*.
- [Morales et al., 2010] Morales, A., Ferrer, M. A., and Kumar, A. (2010). Improved palmprint authentication using contactless imaging. In *2010 Fourth IEEE Int. Conf. Biometrics Theory, Appl. Syst.*, pages 1–6. IEEE.
- [Myronenko and Song, 2010] Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *IEEE TPAMI*, 32(12):2262–2275.

- [National Forensic Science Technology Center, 2013] National Forensic Science Technology Center (2013). Crime Scene Investigation A Guide for Law Enforcement.
- [Ngan et al., 2018] Ngan, M., Grother, P., and Hanaoka, K. (2018). Tattoo Recognition Technology - Evaluation (Tatt-E) performance of tattoo identification algorithms. Technical report, National Institute of Standards and Technology, Gaithersburg, MD.
- [Nicholson et al., 2009] Nicholson, B. T., Harvey, J. A., and Cohen, M. A. (2009). Nipple-Areolar Complex: Normal Anatomy and Benign and Malignant Processes. *RadioGraphics*, 29(2):509–523.
- [Nikisins et al., 2018] Nikisins, O., Eglitis, T., Anjos, A., and Marcel, S. (2018). Fast cross-correlation based wrist vein recognition algorithm with rotation and translation compensation. In *2018 Int. Work. Biometrics Forensics*, pages 1–7. IEEE.
- [Nixon et al., 2015] Nixon, M. S., Correia, P. L., Nasrollahi, K., Moeslund, T. B., Hadid, A., and Tistarelli, M. (2015). On soft biometrics. *Pattern Recognit. Lett.*, 68:218–230.
- [Nurhudatiana et al., 2013] Nurhudatiana, A., Kong, A. W.-K., Matinpour, K., Chon, D., Altieri, L., Siu-Yeung Cho, and Craft, N. (2013). The Individuality of Relatively Permanent Pigmented or Vascular Skin Marks (RPPVSM) in Independently and Uniformly Distributed Patterns. *IEEE Trans. Inf. Forensics Secur.*, 8(6):998–1012.
- [Ojala et al., 2002] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *TPAMI*, 24(7):971–987.
- [Oliveira et al., 2016] Oliveira, G. L., Valada, A., Bollen, C., Burgard, W., and Brox, T. (2016). Deep learning for human part discovery in images. In *2016 IEEE Int. Conf. Robot. Autom.*, pages 1634–1641. IEEE.

- [Ong et al., 2008] Ong, M. G. K., Connie, T., and Jin, T. A. B. (2008). Touch-less palm print biometrics: Novel design and implementation. *Image Vis. Comput.*, 26(12):1551–1560.
- [Parkhi et al., 2015] Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep Face Recognition. In *BMVC*, pages 41.1–41.12.
- [Paulino et al., 2013] Paulino, A. A., Feng, J., and Jain, A. K. (2013). Latent fingerprint matching using descriptor-based hough transform. *TIFS*, 8(1):31–45.
- [Peng et al., 2005] Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE TPAMI*, 27(8):1226–1238.
- [Proenca and Neves, 2018] Proenca, H. and Neves, J. C. (2018). Deep-PRWIS: Periocular Recognition Without the Iris and Sclera Using Deep Learning Frameworks. *IEEE Trans. Inf. Forensics Secur.*, 13(4):888–896.
- [Qin and El-Yacoubi, 2017] Qin, H. and El-Yacoubi, M. A. (2017). Deep Representation-Based Feature Extraction and Recovering for Finger-Vein Verification. *IEEE Trans. Inf. Forensics Secur.*, 12(8):1816–1829.
- [Qu et al., 2010] Qu, H. N., Li, G. Z., and Xu, W. S. (2010). An asymmetric classifier based on partial least squares. *PR*, 43(10):3448–3457.
- [Raghavendra and Busch, 2016] Raghavendra, R. and Busch, C. (2016). A low cost wrist vein sensor for biometric authentication. In *2016 IEEE Int. Conf. Imaging Syst. Tech.*, pages 201–205. IEEE.
- [Ramachandra et al., 2018] Ramachandra, R., B. Raja, K., Venkatesh, S., Hegde, S., D. Dandappanavar, S., and Busch, C. (2018). Verifying the Newborns without Infection Risks Using Contactless Palmprints. In *ICB*, pages 209–216. IEEE.

- [Ribaric and Fratric, 2005] Ribaric, S. and Fratric, I. (2005). A biometric identification system based on eigenpalm and eigenfinger features. *IEEE TPAMI*, 27(11):1698–1709.
- [Rocco et al., 2018] Rocco, I., Arandjelovic, R., and Sivic, J. (2018). End-to-end weakly-supervised semantic alignment. In *CVPR*.
- [Rosipal and Krämer, 2006] Rosipal, R. and Krämer, N. (2006). Overview and Recent Advances in Partial Least Squares. In *Subspace, Latent Struct. Featur. Sel.*, pages 34–51. Springer Berlin Heidelberg.
- [Sanchez-Reillo et al., 2000] Sanchez-Reillo, R., Sanchez-Avila, C., and Gonzalez-Marcos, A. (2000). Biometric identification through hand geometry measurements. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1168–1171.
- [Savič and Pavešić, 2007] Savič, T. and Pavešić, N. (2007). Personal recognition based on an image of the palmar surface of the hand. *PR*, 40(11):3152–3163.
- [Scheirer et al., 2013] Scheirer, W. J., de Rezende Rocha, A., Sapkota, A., and Boulton, T. E. (2013). Toward Open Set Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(7):1757–1772.
- [Scheirer et al., 2011] Scheirer, W. J., Rocha, A., Micheals, R. J., and Boulton, T. E. (2011). Meta-Recognition: The Theory and Practice of Recognition Score Analysis. *IEEE TPAMI*, 33(8):1689–1695.
- [Schroff et al., 2015] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. In *CVPR*, pages 815–823.
- [Shulman et al., 2001] Shulman, O., Badani, E., Wolf, Y., and Hauben, D. J. (2001). Appropriate location of the nipple-areola complex in males. *Plast. Reconstr. Surg.*, 108(2):348–51.

- [Simonyan and Zisserman, 2015] Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*.
- [Song et al., 2015] Song, J. H., Kim, C., and Yoo, Y. (2015). Vein Visualization Using a Smart Phone With Multispectral Wiener Estimation for Point-of-Care Applications. *IEEE J. Biomed. Heal. Informatics*, 19(2):773–778.
- [Spaun, 2007] Spaun, N. A. (2007). Forensic Biometrics from Images and Video at the Federal Bureau of Investigation. In *BTAS*, pages 1–3. IEEE.
- [Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *JMLR*, 15:1929–1958.
- [Su and Kong, 2014] Su, H. and Kong, A. W. K. (2014). A Study on Low Resolution Androgenic Hair Patterns for Criminal and Victim Identification. *TIFS*, 9(4):666–680.
- [Sun et al., 2005] Sun, Z., Tan, T., Wang, Y., and Li, S. (2005). Ordinal Palmprint Representation for Personal Identification. In *CVPR*, pages 279–284. IEEE.
- [Svoboda et al., 2015] Svoboda, J., Bronstein, M. M., and Drahanaky, M. (2015). Contactless biometric hand geometry recognition using a low-cost 3D camera. In *2015 Int. Conf. Biometrics*, pages 452–457. IEEE.
- [Svoboda et al., 2016] Svoboda, J., Masci, J., and Bronstein, M. M. (2016). Palmprint recognition via discriminative index learning. In *ICPR*, pages 4232–4237. IEEE.
- [Szegedy et al., 2015] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going Deeper with Convolutions. In *CVPR*, pages 1–9. IEEE.

- [Taigman et al., 2014] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deep-Face: Closing the Gap to Human-Level Performance in Face Verification. In *CVPR*, pages 1701–1708. IEEE.
- [Tang et al., 2011] Tang, C., Kong, A. W. K., and Craft, N. (2011). Uncovering vein patterns from color skin images for forensic analysis. In *CVPR*, pages 665–672. IEEE.
- [Tang et al., 2016] Tang, C., Zhang, H., and Kong, A. W.-K. (2016). Using multiple models to uncover blood vessel patterns in color images for forensic analysis. *Inf. Fusion*, 32:26–39.
- [Tanini and Lo Russo, 2018] Tanini, S. and Lo Russo, G. (2018). Shape, Position and Dimension of the Nipple Areola Complex in the Ideal Male Chest: A Quick and Simple Operating Room Technique. *Aesthetic Plast. Surg.*, 42(4):951–957.
- [Thaweeakote et al., 2013] Thaweeakote, V., Songram, P., and Jareanpon, C. (2013). Automatic nipple detection based on face detection and ideal proportion female using random forest. In *2013 IEEE Int. Conf. Comput. Intell. Cybern.*, pages 11–15. IEEE.
- [Tiong et al., 2019] Tiong, L. C. O., Teoh, A. B. J., and Lee, Y. (2019). Periocular Recognition in the Wild with Orthogonal Combination of Local Binary Coded Pattern in Dual-stream Convolutional Neural Network. In *Int. Conf. Biometrics*.
- [Ufer and Ommer, 2017] Ufer, N. and Ommer, B. (2017). Deep Semantic Feature Matching. In *CVPR*.
- [Vedaldi and Lenc, 2015] Vedaldi, A. and Lenc, K. (2015). MatConvNet Convolutional Neural Networks for MATLAB. In *ACM Int. Conf. Multimed.*
- [VLFeat,] VLFeat. VLFeat - Tutorials: SIFT detector and descriptor.

- [Wang et al., 2018] Wang, J., Zhu, X., Gong, S., and Li, W. (2018). Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-identification. In *2018 IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 2275–2284. IEEE.
- [Wang et al., 2010a] Wang, J. G., Li, J., Yau, W. Y., and Sung, E. (2010a). Boosting dense SIFT descriptors and shape contexts of face images for gender recognition. In *2010 IEEE CVPR - Work.*, pages 96–102. IEEE.
- [Wang et al., 2008] Wang, L., Leedham, G., and Cho, D. S.-Y. (2008). Minutiae feature analysis for infrared hand vein pattern biometrics. *Pattern Recognit.*, 41(3):920–929.
- [Wang et al., 2010b] Wang, Y., Li, J., Wang, H., and Hou, Z. (2010b). Automatic Nipple Detection Using Shape and Statistical Skin Color Information. In *Adv. Multimed. Model.*, pages 644–649. Springer, Berlin, Heidelberg.
- [Wojcik et al., 2018] Wojcik, R., Wojcik, R., Adanczyk, K., Adamczyk, D., Kozak, K., Czerwinka, M., Dziedzic, M., and Madej–Czerwinka, B. (2018). Male breast hypertrophy - a review of modern methods of treatment. *J. Educ. Heal. Sport*, 8(8):180–191.
- [Woodard and Flynn, 2005] Woodard, D. and Flynn, P. (2005). Personal Identification Utilizing Finger Surface Features. In *CVPR*, volume 2, pages 1030–1036. IEEE.
- [Wu and Zhao, 2015] Wu, X. and Zhao, Q. (2015). Deformed Palmprint Matching Based on Stable Regions. *IEEE Trans. Image Process.*, 24(12):4978–4989.
- [Wu et al., 2014] Wu, X., Zhao, Q., and Bu, W. (2014). A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors. *PR*, 47(10):3314–3326.
- [Xie et al., 2012] Xie, J., Zhang, L., You, J., Zhang, D., and Qu, X. (2012). A Study of Hand Back Skin Texture Patterns for Personal Identification and Gender Classification. *Sensors*, 12(12):8691–8709.

- [Xu et al., 2018] Xu, Y., Fei, L., Wen, J., and Zhang, D. (2018). Discriminative and Robust Competitive Code for Palmprint Recognition. *IEEE Trans. Syst. Man, Cybern. Syst.*, 48(2):232–241.
- [Yörük et al., 2006] Yörük, E., Dutağaci, H., and Sankur, B. (2006). Hand biometrics. *Image Vis. Comput.*, 24(5):483–497.
- [Yoruk et al., 2006] Yoruk, E., Konukoglu, E., Sankur, B., and Darbon, J. (2006). Shape-based hand recognition. *IEEE Trans. Image Process.*, 15(7):1803–1815.
- [Yosinski et al., 2014] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *NIPS*.
- [Young et al., 2006] Young, K. L., Jones, J. G., Worthington, T., Simpson, P., and Casey, P. H. (2006). Forensic Laboratory Evidence in Sexually Abused Children and Adolescents. *Arch. Pediatr. Adolesc. Med.*, 160(6):585.
- [Yu and Kong, 2018] Yu, P. and Kong, A. W. K. (2018). A Meta-Recognition Based Skin Marks Matching Algorithm with Feature Fusion for Forensic Identification. In *2018 Int. Conf. Biometrics*, pages 111–117. IEEE.
- [Zhang et al., 2003] Zhang, D., Kong, W. K., You, J., and Wong, M. (2003). Online palmprint identification. *TPAMI*, 25(9):1041–1050.
- [Zhang et al., 2012a] Zhang, H., Tang, C., Kong, A. W. K., and Craft, N. (2012a). Matching vein patterns from color images for forensic investigation. In *BTAS*, pages 77–84.
- [Zhang et al., 2012b] Zhang, L., Li, H., and Niu, J. (2012b). Fragile Bits in Palmprint Recognition. *IEEE Signal Process. Lett.*, 19(10):663–666.
- [Zhang et al., 2017] Zhang, L., Li, L., Yang, A., Shen, Y., and Yang, M. (2017). Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach. *PR*, 69:199–212.

- [Zhang et al., 2010] Zhang, L., Zhang, L., Zhang, D., and Zhu, H. (2010). Online finger-knuckle-print verification for personal authentication. *PR*, 43(7):2560–2571.
- [Zhao and Kumar, 2017] Zhao, Z. and Kumar, A. (2017). Towards More Accurate Iris Recognition Using Deeply Learned Spatially Corresponding Features. In *ICCV*, pages 3829–3838.
- [Zharov et al., 2004] Zharov, V. P., Ferguson, S., Eidt, J. F., Howard, P. C., Fink, L. M., and Waner, M. (2004). Infrared imaging of subcutaneous veins. *Lasers Surg. Med.*, 34(1):56–61.
- [Zheng et al., 2007] Zheng, G., Wang, C.-J., and Boulton, T. (2007). Application of Projective Invariants in Hand Geometry Biometrics. *IEEE Trans. Inf. Forensics Secur.*, 2(4):758–768.
- [Zheng et al., 2016] Zheng, Q., Kumar, A., and Pan, G. (2016). A 3D Feature Descriptor Recovered from a Single 2D Palmprint Image. *TPAMI*, 38(6):1272–1279.
- [Zhou and Kumar, 2011] Zhou, Y. and Kumar, A. (2011). Human Identification Using Palm-Vein Images. *IEEE TIFS*, 6(4):1259–1274.
- [Zucca-Matthes et al., 2016] Zucca-Matthes, G., Urban, C., and Vallejo, A. (2016). Anatomy of the nipple and breast ducts. *Gland Surg.*, 5(1):32–6.