Contributions to Unconstrained Palmprint Recognition on Smartphones

Adrian-Stefan Ungureanu

College of Engineering and Informatics
National University of Ireland, Galway

This dissertation is submitted for the degree of
Doctor of Philosophy

Supervisor: Prof. Peter Corcoran

May 2020

“He who has a why to live for can bear almost any how.”

~Friedrich Nietzche, *Twilight of the idols*

“El mundo nada puede contra un hombre que canta en la miseria.”

~Ernesto Sabato, *La resistencia*

Table of contents

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Adrian-Stefan Ungureanu
May 2020

Acknowledgements

I wish to thank everyone that helped me, in some way or another, to reach the position of submitting this PhD thesis. It represents the culmination of many years of hard work.

First I would like to thank my supervisor, Prof. Peter Corcoran. For his guidance, helpful support and opportunities given for my becoming an independent researcher. I thank Prof. Christopher Dainty and Alexandru Drîmbărean for their useful feedback throughout my PhD.

A special thank you goes to Dr. Shejin Thavalengal for not only being a great friend and a gym-buddy, but a true mentor in research. I am grateful for the chance of working with you as a colleague!

Thank you to Dr. Hossein Javidnia, Dr. Tudor Nedelcu and Dr. Shabab Bazrafkan for all the wonderful times we spent together, the funny lunches and conversations. I will miss our group very much!

To Dr. Joe Lemley for all the support he gave me when I was breaking into the wonderful (and challenging) world of Deep Learning with Keras. To Snail, his wife, for the help she gave with proof-reading and the world's best vegan snacks.

To Dr. Asma Khatoon and Mary Raymond for all the nice lunches and tea breaks. To Dr. Faisal Khan for the foosball sparring, and to Dr. Ali Farooq for the time spent at University as deskmates. To Dr. Shubhajit Basak for being so enthusiastic about work!
To Dr. Viktoras Varkarakis and Aoife McDonagh, for making me want to become better at table football.

To Dr. Timothee Cognard, for being a true camarade, my confident and longstanding friend. We have shared many challenges, as well as wonderful moments together. I am grateful for being flatmates for more than 4 years, for being there for each other regardless of the situation. You will always be part of my family!

To Dr. Claudia Costache and Dr. Saqib Salahuddin, for being the best postdocs in the world. For always being there with a kind word and helpful advice, technical or otherwise. Thank you to Joe Desbonnet for all the support with hardware!

Thank you to all the employees of Fotonation (Galway, Bucharest and Brasov offices), for the constructive conversations and all the collaborations throughout the years. Special thanks go to Dr. Gabriel Costache, Satish Mangapuram, Dr. Shejin Thavalengal, Alexandru

Drîmbărean, Arpad Zoldi and Cosmin Rotariu.
I want to thank Dr. Brian O'Sullivan for all the walks and friendship, including the salsa trips.
A special thank you goes to Alin and Diana Tănasă, for being the most fun Romanian colleagues in the Fotonation office. For their brilliant company, their unconditional support and for taking the decision to move to Galway, thus having our lives intersect for a year.

I am grateful for being able to participate in the Youth Academy program which was directed by Geraldine Marley. It was a real privilege to be in the position of a teacher. I have enjoyed all the classes I taught throughout the years!

I want to thank the entire group in the Electronic Engineering department in NUIG, especially: Dr. Evismar Andrade, Dr. Richard Hart, Dr. Dean Sweeney and Dr. Barbara Oliveira.

A warm thank you to my Spanish group, old members and new ones: Anne-Lise, Javier, Chiara, Fernando, Eduardo, Claudia, Lucie, Sabine, Murchadh, Ana and Imke, Laura, Seosamh, Erika, Mary, Maura, Jorge, Laura, and Cristina (Toffee and Johan!), along with all the other members with whom I shared countless happy hours conversing in Spanish.

A joyous thank you to my extended salsa familia: Maestro Pedro Alberto and Maestra Consuelo Padilla for being such inspiring salsa teachers, along with the rest of the salsa familia: Thais, Kinga, Louis, Vernon, Tushar and Nikhil, Marta, Maria and all the salseros in Galway.

A special thank you to Dr. Ali Reza Abaei, my very close friend. I will miss our conversations and coffee breaks, as well as the unrelenting thirst for living every moment. I hope to maintain our friendship over the years!

Thank you to Father Tudor Ghiță, the central figure of the Orthodox spiritual life for the Romanian in Galway. We are all extremely grateful for having you in our midst, so far away from home!

I cannot thank enough my family: my mother Rusăndica and father Lucian-Octav, my brothers Matei-Eugen and Andrei, as well as my aunt Camelia Rîpeanu and my great aunt Geta Paraschivoiu. Being away from you for these past 5 years has been very difficult, I can only hope that my actions have brought pride in your hearts.

Last but not least, the most special and heart-warming acknowledgment goes to Dr. Juhi Samal, for being there for me in some of my toughest moments. The last few months have been especially rough, but at the same time sweet! I will always be grateful for participating in the Thesis Bootcamp. I hope to soon be by your side soon!

Abstract

This thesis investigates the suitability of unconstrained palmprint as a biometric modality for handheld devices equipped with a camera, such as smartphones. A detailed literature survey is provided, covering existing datasets, methods for region of interest extraction (ROI) extraction and feature extraction from palmprints. Following a series of exploratory experiments, a novel dataset of palmprints from 81 subjects and acquired using 5 different smartphone cameras is developed. Details are provided of initial data acquisitions, the final acquisition and management protocol and the associated Ethics Application. The dataset was collected in several acquisition phases over a period of 8 months. A set of baseline matching experiments is also detailed and manual mark-up of the palmprint data is included with the dataset. The accurate extraction of palmprint ROI was identified as a key component in the biometric recognition pipeline but the mark-up used for ROI extraction is not available in palmprint datasets. Thus, a second dataset was acquired using a 3D sensor and aligned camera with suitable mark-up data. Over 25,000 images were acquired from 26 subjects over the course of 1 year. Corresponding experiments were designed to evaluate a range of machine learning approaches to ROI extraction and a new quality measure was developed to compare the accuracy of ROI extraction for palmprints. Detailed experiments compared various ROI extraction techniques and have demonstrated that unconstrained palmprint can serve as a practical means of biometric authentication using standard smartphone cameras and without a need for specialized fingerprint or 3D face sensors.

List of figures

List of tables

Nomenclature

Acronyms/Abbreviations

EER	Equal Error Rate
FAR	False Acceptance Rate
FP	False Positive
GAR	Genuine Acceptance Rate
TP	True Positive
FRR	False Rejection Rate
ROI	Region of Interest
SVM	Support Vector Machine
LDA	Linear Discriminant Analysis
CRC	Collaborative Representation Classifier
SRC	Sparse Representation Classifier
KNN	K Nearest Neighborhood Classifier
CNN	Convolutional Neural Network
ML	Machine Learning
DL	Deep Learning
DNN	Deep Neural Network
ReLu	Rectified Linear Unit

PCK Percentage of Correct Keypoint

IoU Intersection over Union

GT Ground truth

ROC Receiver Operating Characteristic

CD Cross-device

DS Device-specific

LM Leap Motion

Chapter 1

Introduction

1.1 Palmprint Recognition as a Smartphone Biometric

As biometric recognition technologies gain a more central role in human society, they have to adapt to novel challenging environments. These are defined by many factors, including socioeconomic ones (affordability), but also legal (acceptability/reliability) but especially technological [2].
In this context, palmprint recognition has considerable potential in being a main biometric mode used to secure various web services on a large scale, especially when deployed on consumer devices such as smartphones. In the following chapters the challenges associated to unconstrained recognition of palmprints using smartphones is explored.

This chapter is an overview of the thesis' contents and rationale, with an introduction into the background of palmprint recognition and state-of-art being presented in Chapters 2 and 3.

1.2 Objectives and Scope of the Work

This thesis discusses the deployment of palmprint recognition as a viable solution for authentication on smartphones.

1.2.1 Today's Smartphone

When the work contained in this thesis was started, biometric recognition on smartphones was in its infancy, with the Android system having an easily circumventable face recognition application, and only few smartphones supporting fingerprint recognition [3].

The first mobile device to integrate fingerprint recognition was the iPAQ Pocket PC h5400 [4], launched in 2002. It was fitted with a fingerprint sensor requiring the user to swipe the finger in a certain direction, which in theory limited the reading errors.

The first smartphone to have a capacitive fingerprint sensor was the iPhone 5s, which was released in September of 2013 [5]. The sensor acquired images of 500 pixels per inch and thanks to its matching algorithms, the matching was rotationally invariant. Thanks to its ease of use, it became an overnight success [3] and was adopted by many manufacturers ever since.

1.2.2 Integration into consumer devices

Using the camera (either rear or frontal) for the acquisition of a palmprint image was considered to be a good alternative to relying on a scanner, thus providing a biometric solution for authentication on any consumer device (fitted with a camera). This idea was validated by the projects of companies such as Redrock Biometrics, who have released an SDK for their palmprint solution called PalmID in 2017 [6].

This approach can be deployed on any consumer device, an especially interesting use case being the seamless integration into how the Augmented Reality (AR) headsets can be unlocked. This was achieved in 2018 by developing an authentication protocol with the PalmID [6] software deployed on EPSON's MOVERIO BT300 AR headset [7].

1.2.3 Re-emergence of feature phones

The year 2017 saw the introduction of an operating system (entitled KaiOS [8]) that brings many smartphone-specific applications and functionalities (e.g. 4G connection, GPS, Whatsapp, Google maps, Google assistant, etc.) to more affordable feature phones, thus opening up many exciting possibilities in emerging markets such as Africa or India [9]. For such devices existing biometric solutions can be deployed using the rear camera (which generally reaches 2 Mega Pixels). In this context, palmprint recognition remains a valid option for biometric authentication, as demonstrated previously by Jia *et al.* [10] when using devices with similar specifications.

1.3 Contributions

The main contributions of this thesis can be summarized as follows:

1. A state of the art literature review on the use of palmprint recognition, with an emphasis being placed on the solutions implemented on mobile devices.

2. Introduced the first multi-smartphone database of palmprint images acquired in realistic and unconstrained conditions of authentication (NUIG_Palm1).

3. Implemented baseline matching scenarios, exploring the difficulties of unconstrained palmprint acquisition

4. Introduced an auxiliary database of palmprints aimed at providing support for the training of Convolutional Neural Networks. It was collected with a webcam and a Leap Motion device, which allowed the on-site labeling of hand images with specific keypoints (NUIG_Palm2).

5. Introduced the first palmprint Region of Interest (ROI) extraction approach for unconstrained hand pose (based on a Convolutional Neural Network using concepts from hand pose regression).

6. Investigated the effect of palmprint ROI misalignment on recognition accuracy.

1.3.1 List of publications

The work presented in this thesis resulted in the following journal papers, of which one was referenced in IEEE biometrics virtual journal - *IEEE Biometrics Compendium*:

- Adrian-Stefan Ungureanu, Shejin Thavalengal, Timothee Elie Cognard, Claudia Costache and Peter Corcoran, "Unconstrained palmprint as a smartphone biometric", in *IEEE Transactions on Consumer Electronics*, vol. 63(3), pp. 334–342 (August 2017). (Referenced in *IEEE Biometrics Compendium, issue 31, March 2018.*)

- Adrian-Stefan Ungureanu, Joseph Lemley and Peter Corcoran, "Unconstrained palmprint ROI extraction using CNN and Leap Motion", (*under review*).

- Adrian-Stefan Ungureanu, Saqib Salahudin and Peter Corcoran, "Towards Unconstrained Palmprint Recognition on Consumer Devices: a Literature Review", (*accepted for publication in IEEE Access*).

One conference paper was also published based on this work:

- Adrian-Stefan Ungureanu, Hossein Javidnia, Claudia Costache and Peter Corcoran, "A review and comparative study of skin segmentation techniques for handheld imaging

devices" in Proceedings of the 2016 *IEEE International Conference on Consumer Electronics (ICCE), Jan. 2016, Las Vegas, USA*.

Also, one article was also published in a IEEE Magazine:

- Adrian-Stefan Ungureanu, Claudia Costache and Peter Corcoran, "Palm print as a smartphone biometric: another option for digital privacy and security" in *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 71-78, 2016.

Delivered several oral or poster presentations:

- (Poster) "Palm-print Recognition for Authentication on Smartphones.", at *5th Annual Postgraduate Research Day*, National University of Ireland, Galway, April 2015, Galway, Ireland.
- (Poster) "Unconstrained palmprint recognition on smartphones", at *13th IAPR/IEEE/Eurasip International Summer School on Biometrics*, June 2016, Alghero, Italy.
- (Presentation) "L'identification biometrique de les impressions de paume dans le contexte des smartphones" at *Threesis in French*, Trinity College Dublin, March 2018, Dublin, Ireland.

1.4 Organization of the thesis

The rest of the thesis is organized as follows:

Chapter 2 introduces key concepts of biometric recognition, as well as statistical tools that are used to evaluate such recognition systems. The palmprint, as a biometric trait is defined and its use is discussed.

Chapter 3 contains an extensive overview of palmprint recognition, from databases to palmprint Region of Interest (ROI) extraction, as well as feature extraction and matching approaches.

Chapter 4 introduces two novel palmprint databases (along with an initial Proof of Concept database) that were collected to support the ongoing research of biometric recognition using smartphone cameras. These databases have been released to the research community.

Chapter 5 represents a preliminary stage of research, where the focus was placed on pre-processing stages of palmprint recognition. The Proof of Concept database of palmprints was used to provide an initial direction of research.

Chapter 6 introduces matching experiments performed on the palmprint database NUIG_Palm1. The unconstrained nature of the conditions of acquisition proved to be

especially challenging, as there were many factors that were not controlled (e.g. lighting, background, hand pose, etc.).

Chapter 7 uses the images from NUIG_Palm2 to train a Convolutional Neural Network that predicts keypoints used for the extraction of palmprint ROI. Several state-of-the-art architectures are compared and compared and evaluated with specific point prediction and object detection methods.

Chapter 8 then investigates the influence of misalignment introduced by the ROI extraction algorithm using images from the palmprint database NUIG_Palm1.

Chapter 9 outlines the main conclusions and future work based on the work contained in this thesis.

Chapter 2

Background

The general problem of recognizing (from the Latin *re*, again and *cognoscere*, know or learn) an individual ultimately comes down to assessing a potential threat. Another organism not sharing the same physiological traits is most likely part of a different species which may or may not be a predator looking for its next meal. In human history (and prehistory) this is further refined to recognizing members of the same tribe from those of other tribes.

With the advent of agricultural societies, this problem became known as identity management since every citizen of any sociopolitical group is expected to pay taxes and have certain legal requirements and responsibilities. Therefore, it is essential to distinguish between individuals regardless of their cultural background.

Throughout history several strategies were employed to solve this need. These approaches can be broadly classified into three groups: (1) something that I have, (2) something that I know and (3) something that I am. The first category uses a surrogate object, which proves the claimed identity, for instance a state issued ID, passport or token. The second category relies on certain knowledge only that individual is expect to know, such as a PIN code to a savings account or an e-mail account. The third and final category is made up of any physiological and/or behavioral traits an individual possesses. The best example is probably the cue humans (and other primates) have been using for millennia for recognizing one another – the face, which is also the main reason why IDs usually contain a photograph.

2.1 Biometric Recognition

Biometric recognition (or simply *biometrics*) is defined as "the science of establishing the identity of an individual based on the physical and/or behavioral characteristics of the person either in a fully automated or a semi-automated manner" ([11] p.2). The term is derived from Greek *bios*, life and *metron*, to measure, reflecting the initial intention of providing a

statistical interpretation in the field of evolutionary biology used in the 19th century. However, the concept changed in late 20th century, to replace the previously used term of Automated Person Identification [11]. Any biometric system relies on the premise that a biometric trait or combination of traits is able to provide a unique identification key.

2.1.1 What makes a Biometric Characteristic

A biometric trait needs to fulfill a set of requirements, as were defined by Jain *et al* [12]:

- Universality: all individuals should have the trait,
- Distinctiveness: two individuals should display differences with respect to the trait,
- Permanence: the trait should be stable and remain consistent over a period of time,
- Collectability: the trait should be measured in a quantitative way.

Furthermore, any biometric recognition system is expected to fulfill a number of conditions:

- Performance: achieving a targeted accuracy and speed for comparing the biometric trait,
- Acceptability: the willingness of individuals to provide their biometric trait,
- Circumvention: the difficulty with which the system is fooled by using a forged biometric trait.

2.1.2 Classifying Biometric Traits

After describing what makes a trait also a biometric trait, it is necessary to provide a classification of biometric traits. The most common distinction is made in terms of their origin - either physiological or behavioral.
Examples from the former group are the traits most commonly associated with biometric recognition - face, fingerprint, palmprint, iris, ear, and just about any trait which exploits a physically observable trait. Examples from the latter group include voice recognition but also gait, keystroke, dynamic signature recognition, etc. Recognition systems based on these traits also include a time component.

An evaluation of several biometric characteristics is presented in Fig. 2.1, based on the conditions defined in Subsection 2.1.1. As there is no optimal biometric characteristic (i.e. outperform all others in all operational environments), there is a context where one trait may be preferred to others.

The biometric characteristics that are highlighted belong to the category of hand biometrics. Compared to the other categories, these also offer the possibility of fusion, thus improving both performance and robustness to potential attacks.

Biometric identifier	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

Fig. 2.1 Classification of biometric traits, as evaluated by Jain *et al* [12]. H, M and L correspond to High, Medium and Low. The highlighted characteristics correspond to hand biometrics. (Source: [12] ©2004 IEEE)

2.1.3 Generic Biometric System - Components

Any biometric recognition system (BRS) that uses a vision-based approach is loosely made up of the following blocks (as represented in Fig. 2.2):

1. **Sensor**, encompassing both the environment of operation as well as a protocol regulating the means of acquisition. A protocol is designed to take into account the strengths and weaknesses of a given system and attempt to improve its performance.

2. **Feature extractor**, containing sub-modules such as Quality evaluation, ROI extraction and Feature encoding.

- **Quality evaluation** assesses the potential use for any acquisition. It needs to pass quality assurance criteria, such as image quality, blur evaluation, etc. Standard image processing techniques are applied to compensate noise levels if required (in the case of imaging sensors). Furthermore, some sort of liveness detection needs to be considered, as measure to reduce the circumventability of the system.
- **ROI extraction**, where the biometric modality has to be located and separated from the background.
- **Feature extraction** which encodes the information contained in the previously extracted ROI. This can also be called biometric sample.

3. A **Database** containing the list of all registered users and their stored biometric samples (templates). It is initially accessed during the enrollment stage, where several biometric samples are saved locally. During testing, the stored template is selected and provided for comparison.

4. The **Comparison Trial** stage (Matcher) is used during testing to determine a similarity score between a presented biometric trait and the stored sample in the database associated to a claimed identity. Based on this score, a decision is made to either give or deny access (in the case of an access control application) using a specific threshold.

2.1.4 Functionalities of a Biometric Recognition System

A biometric recognition system has two modes of operation: verification and identification. Following an initial stage of enrollment of biometric samples associated to a set of identities (classes), then any test samples presented to the system are compared with the stored samples, following either the protocol of verification or identification [11].

Verification

This mode essentially answers the question 'Are you who you claim you are?', thus requiring two specific inputs to the system: a biometric sample and a claimed identity (class). It is therefore required to only compare the test sample (query) with all biometric samples associated to one identity (a one-to-one comparison).
Formally, this is formulated by defining a set of identities Φ_i with i being total number of classes. Every class is defined by a set of samples $\phi^D_{i,j}$, where j represents the total number of samples associated to a class Φ_i. Having a query sample ϕ^T_i for class i, a similarity score

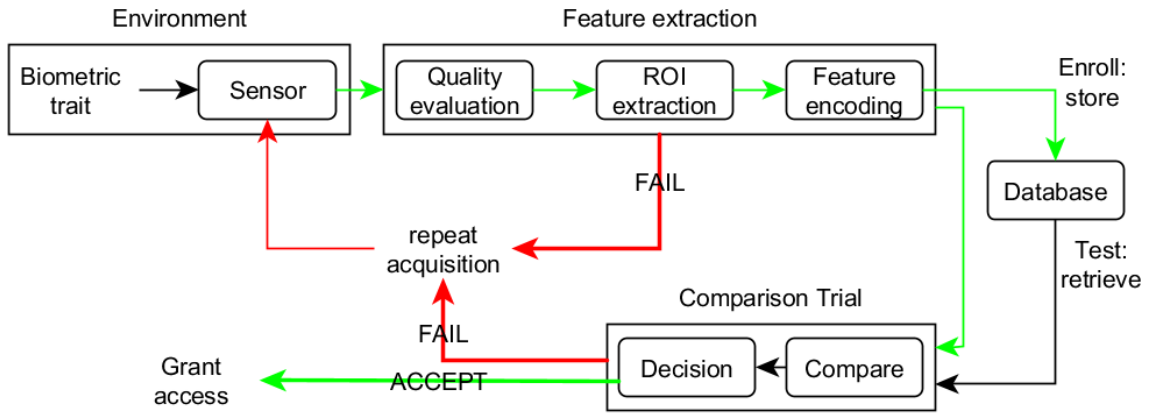


Fig. 2.2 Modes of operation for a generic biometric recognition system. Whereas the enrollment phase stores an extracted biometric sample, the testing phase (verification or identification) compares a query sample to the samples stored in the local database. (Adapted from *Introduction to Biometrics* [11] ©2011 Springer)

s is computed using an approach L to assess if the query sample does belong to that class:

$$L\left(\Phi_i, \phi_i^T\right) = \begin{cases} genuine, & if\ s \geq \tau \\ impostor, & if\ s < \tau \end{cases} \tag{2.1}$$

where the score s ranges from 0 to 1 and confirms adherence to class Φ_i the closer it is to 1, and τ corresponds to a predetermined threshold used when training the biometric recognition system. If a distance is used (instead of a similarity score), then the conditions for the threshold τ are reversed, as a computed distance between two samples from the same class is expected to be close to 0.
A **match** is defined as the positive comparison decision, whereas a **non-match** is the negative comparison decision.

Identification

This scenario attempts to answer the question 'Are you someone who is registered in the system?'. The system will therefore compare the query sample with all samples stored in the database in order to determine the best comparison (one-to-many). The system contains N+1 total set of identities (classes), where N=number of users, with the N+1 class generally kept for 'impostors'. If the best similarity score falls below a decided threshold (τ), then it can be considered that the identity cannot be found in the database of stored samples.
Formally, this problem can be formulated extending the condition in Equation 2.1 to include

an $arg\,max_s$ operation determining query class Φ^T based on the similarity scores s obtained with function L:

$$\Phi^T = arg\,max_s(L(\Phi_i, \phi_i^T)), \quad for\ i = 1..Nclasses \tag{2.2}$$

After determining the class which best comparisons with the query ϕ_i^T, the score needs to be validated using the pre-determined threshold τ, otherwise the query sample will be associated to class Φ_{N+1}, in which case it will be considered an impostor.

2.2 Metrics for Evaluation of Biometric Systems

2.2.1 Defining the Problem

One of the fundamental requirements for a trait to qualify as a biometric trait is that of *permanence*, or continuity across time. As biometric recognition systems are operating with traits from live organisms, a certain tolerance regarding the observed traits is expected.
The variation observed between the biometric samples corresponding to the same class (identity) is defined as intra-class, and is expected to be small (i.e. a chosen distance computed between these samples is supposed to be small). Conversely, the variation (distance) between the samples belonging to two different classes is expected to be high, and is called inter-class. Authors in the literature can also call intra-class and inter-class as *genuine* and *impostor* [12].

Any biometric system faces a generalized problem of classification. Given a collection Φ_K of one biometric feature with K classes, each class ϕ_i with several instances ϕ_j^i, the task of classification aims at constructing a hyperspace representation providing adequate separation between all classes. There exists a hypothetical η distance (or similarity score) threshold, which provides adequate separation (depending on the application) of intra-class from inter-class instances. An example of such a distribution is presented in Fig. 2.3.

The erroneous classification of a query sample to a class which it does not belong to, is called a False Acceptance. The rate of such errors is called False Acceptance Rate (FAR). The opposite, rejecting a sample although it belongs to that class is called a False Rejection. The rate of such errors is called False Rejection Rate (FRR). If we label the genuine and impostor distributions with ω_0 and ω_1 respectively, then the FAR and FRR can be defined mathematically as:

$$FAR(\eta) = p(s >= \eta|\omega_0) = \int_{\eta}^{+\infty} p(s|\omega_0)ds, \tag{2.3}$$

$$FRR(\eta) = p(s < \eta|\omega_1) = \int_{-\infty}^{\eta} p(s|\omega_1)ds \tag{2.4}$$

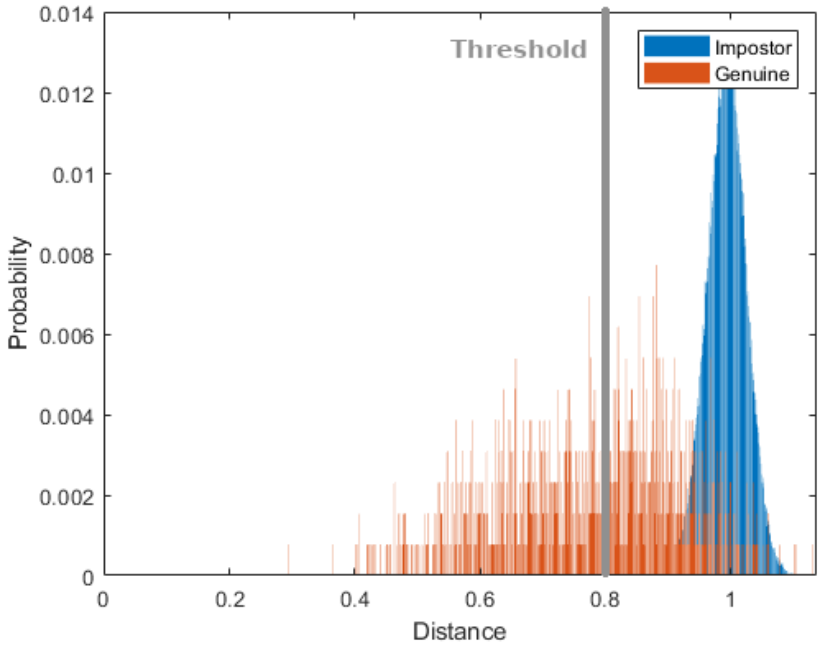


Fig. 2.3 Genuine and impostor distance distributions, originating from a Collaborative Representation Classifier (CRC) used in Section 6.3 of Chapter 6. The threshold η is used to determine the FAR and FRR rates.

To determine the Genuine Acceptance Rate (GAR), one only requires the FRR:

$$GAR(\eta) = 1 - FRR(\eta) \tag{2.5}$$

When comparing two biometric samples, the nature of the application is the one which dictates how important the FRR is. Forensic scenarios require that the FRR is low enough to include all potential suspects, regardless of the FAR. Conversely, the high security applications require very low FAR and can tolerate a high FRR. In contexts such as mobile phone authentication the FRR needs to be low enough, otherwise it risks affecting the user experience (an important factor when selling products).

2.2.2 Interpreting an ROC

A Receiver Operating Characteristic (ROC) is defined using the FAR and FRR values corresponding to a classifier where the threshold η is varied with a specific set, such that all similarity scores are being covered (from s_{min} to s_{max}).

Two system's ROCs can be used to compare their performance relative to each other. Depending on what is being plotted with respect to FAR (FRR or GAR), the closer the ROC

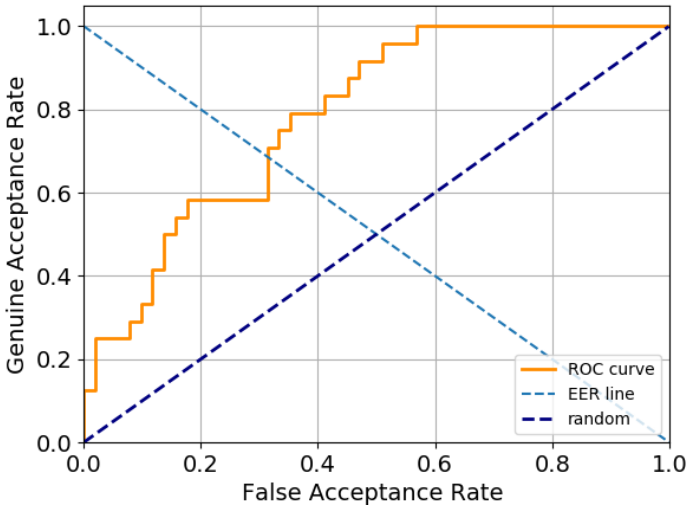


Fig. 2.4 Representation of a Receiver Operating Characteristic, applied to a Support Vector Machine used for classification on the scikit-learn [13] iris database. The value of EER corresponds to 31%, which is located at the intersection of the EER line with the ROC curve.

is to 1.0 or 0.0, the better the performance. An example ROC is provided in Fig. 2.4, where the dark blue line plots a random classification (correct classification is 50%).

2.2.3 Equal Error Rate

When evaluating two classifiers, a common metric used is the Equal Error Rate (EER). This is the point at which the FRR has the same value as FAR. The lower this value, the better the system. The EER can also be observed in the ROC, at the point of intersection with the EER-line.

It is important to note that EER is only to be used as a point of reference and rarely that threshold η is being deployed on a recognition system. As this threshold influences the FAR and FRR, it depends on the type of application, which dictates the requirements for a recognition system.

2.3 Palmprints

2.3.1 Definition and Features

The palmprint is defined in the "Encyclopedia of Biometrics" [14] using the two words it is being composed of *palm* and *print*:

> *The human palm is defined as the region on the inside part of the hand stretching from the base of the fingers to the wrist. The print is an impression made when the body part is pressed against a surface. A palmprint therefore illustrates the physical properties of skin patterns such as lines, minutiae and texture.*

Palmprints form concomitantly with fingerprints (during the embryonic skin development and are fully developed 6 months into the pregnancy) and are affected by the same genetic/behavioral factors. Therefore, they can be considered as fingerprints covering a greater surface that also include specific flexion lines [15].
The features that can be found on a palmar surface, as visualized in Fig. 2.5, are:

- Principal lines (major palmar flexion ridges): the most visible creases on the palm, include the life-line (radial transverse), the heart-line (distal transverse) and the head-line (proximal transverse). Their formation is directly linked to the hinging ability of the fingers.
- Datum points: represent the end points of principal lines, they can be used to align two palmprints.
- Palmar friction ridges: folded pattern of palm skin with sudoriferous glands but with no hair, having as main functionality the increase in grip. Their coarseness varies depending on the position inside the palm. Similar patterns are observed on the inside part of fingers that serve the same purpose.
- Minutiae: specific structures that can be observed in the palmar friction ridges. A dictionary of symbols is used to uniquely define symbols that can appear in such structures, allowing the comparison of two samples (e.g a ridge's bifurcation, ridge ending, ridge island, etc.).
- Wrinkles (minor palmar flexion ridges): thinner and more irregular lines as compared to the principal lines. They are found all over the body and allow the stretching of the skin in directions corresponding to the natural demands of each region. As the skin loses its elasticity (with age), permanent wrinkles form. Pronounced wrinkling is

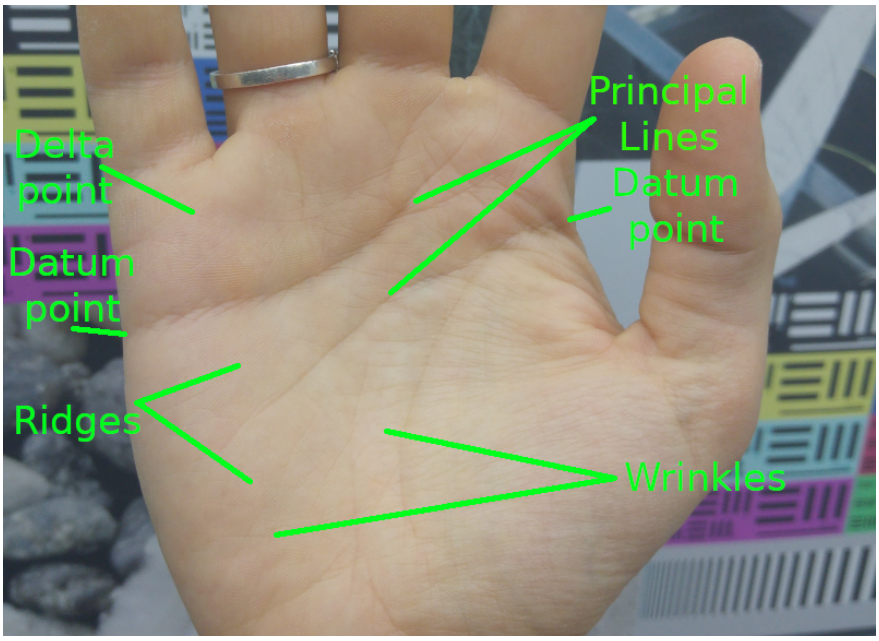


Fig. 2.5 Palmprint sample, with features.

usually due to muscle activity. These features are not visible under resolutions of 100 dpi.

- Delta points: specific patterns found close to the base of fingers, resembling a Δ symbol.

2.3.2 Why Palmprints

Palmprint recognition presents several advantages:

- As hands are used for many activities, and the widespread acceptance of fingerprint recognition, people are not hesitant to use palmprint recognition.
- A touchless approach can be used, not requiring users to touch certain surfaces (reducing hygiene concerns)
- The features contained by palmprints are stable throughout a person's life (principal lines, friction ridges especially).
- Contains a greater surface than fingerprints, offering a richer set of features to extract. Furthermore, having a greater surface provides some robustness to artifacts being present on the hand (dirt, grime or presents cuts).

- As a biometric trait, palmprints are not as public as faces, making them harder to spoof, especially when considering a constrained protocol of acquisition (requiring fingers to be spread).
- Can be deployed on consumer devices (e.g. smartphones) as no extra hardware is required. The rear (or frontal) camera can be used to acquire an image.
- Can be used in conjunction with other hand-based biometric traits (e.g. hand geometry, fingerprints, combining several hyperspectral bands, etc.), to provide a more robust comparison decision (multi-biometrics)
- Because the palmprint ridges and creases are formed during the first weeks of gestation, this makes palmprints capable of differentiating between monozygotic (identical) twins [16].

2.3.3 Palmprint Uses

The main uses of palmprints branch off into two main categories (as presented in Fig. 2.6): forensic and non-forensic, with access control functions. Palmprint recognition is used in forensic applications as a natural extension to fingerprints (they contain up to 10 times more minutiae when considering a full print [15]), as they leave longstanding marks on various surfaces (hence the use of 'latent'). Latent palmprints are collected on-site from a crime scene using specific techniques. These samples are then compared with inked records to find a potential match.

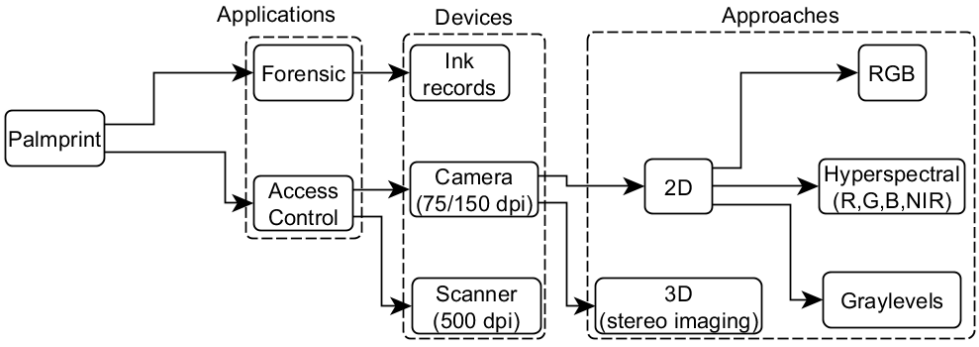


Fig. 2.6 Overview of applications using palmprint, with devices/methods used for acquisition and operating spectrum.

The other branch of applications revolves around Access Control but follows another separation regarding the imaging approaches used. For instance, Customs Control use

scanners to acquire high quality images (500 dots per inch) of palmprints alongside fingerprints and use fingerprint-specific approaches for comparing two samples (because of the level of detail and the richness of features). However, the most popular use of palmprint recognition relies on lower resolution cameras (i.e. up to 150 dpi). Such systems operate in a controlled environment for acquisition, with strict light conditions and fingers spread out, factors which facilitate pre-processing [17], [18].

The means of acquisition are typically optical, relying on either 2D or 3D imaging, visible or hyperspectral (specific spectral bands, including Near Infrared). Since there is no need for extra hardware, palmprint recognition was suggested also in the context of consumer devices (smartphones), to complement existing approaches at the starting time of the research contained in this thesis [19].

2.4 Summary

This chapter presented a brief introduction into the field of biometric recognition, with a focus on palmprints. A series of common metrics used for their evaluation was presented, as they will be used in the following chapters.

Chapter 3

Palmprint Biometrics: a Literature Review

This chapter covers a literature review focused on the pipeline of a palmprint recognition system. Section 3.1 describes existing databases of palmprint images, whereas Section 3.2 provides an overview of approaches developed for the palmprint ROI extraction from various palmprint databases. Section 3.3 presents an overview of approaches of feature extraction and classification algorithms. Section 3.4 introduces a discussion for each section and outlines several conclusions.

The material in this chapter has been accepted for publication in the *IEEE Access* as "Towards Unconstrained Palmprint Recognition on Consumer Devices: a Literature Review".

3.1 Palmprint Databases - a Literature Review

This section presents an overview of palmprint databases used for the recognition of palmprints in the visible spectrum (hyperspectral imaging at various wavelengths is not considered, nor 3D acquisition).

Current palmprint databases can be split into three categories, based on the restrictions imposed to the user during acquisition (as represented in Fig. 3.1):

1. Constrained environment of acquisition: This includes the most popular palmprint databases, which place the main focus on the feature extraction and comparison stages, simplifying the acquisition as much as possible (for the recognition system). Images tend to display hands with a specific hand pose (fingers straight and separated) against a uniform background with no texture, usually black.

2. Acquisition with at least one unconstrained element:

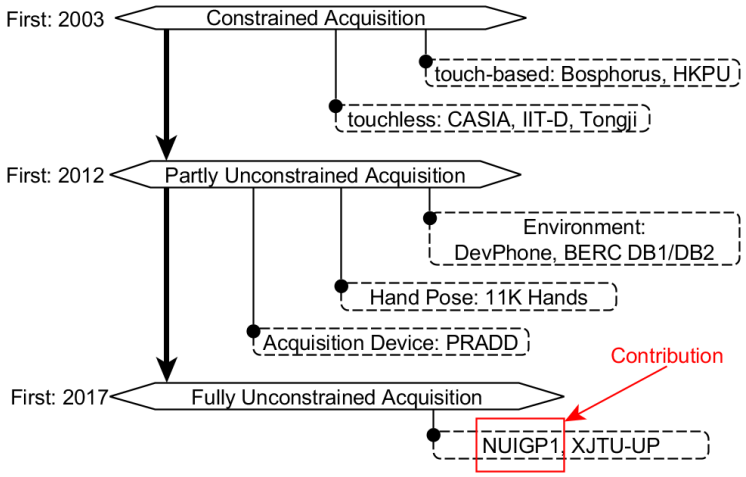


Fig. 3.1 Timeline overview of palmprint databases, based on how constrained their environment of acquisition is.

- **Unconstrained** environment: The background is unconstrained, which corresponds to the use case of consumer devices. The hand pose is required to follow a specific protocol, generally consisting of presenting the fingers spread out in front of the sensor (preferably the center of the image).
- **Unconstrained** hand pose: Allows the user to choose the pose of the hand during acquisition. This corresponds to the general expectations for consumer devices, which require a simplified (and intuitive) protocol of interaction.
- **Multiple devices** used for acquisition: Recognition of palmprint samples across several devices. Generally the other aspects of the acquisition process (hand pose and background) are constrained.

3. Fully unconstrained acquisition: **Unconstrained** environment and hand pose, this represents the most unconstrained scenario, where all conditions of acquisition are left to the choice of the user.

3.1.1 Constrained Palmprint Databases

This category includes the most popular palmprint databases, the Hong Kong Polytechnic University palmprint database (HKPU) [17] being the first to provide a large-scale database to compare recognition performance. The images were acquired using a scanner (**A1** in Table 3.1) having a cropped guide around the palm, reducing the impact of fingers' position (several samples provided in Fig. 3.2a). A similar approach to acquiring palmprints but

Table 3.1 Palmprint databases with constrained acquisition environments. The first column separates the touch-based (A1) acquisition systems from the touch-less ones (A2).

	Name	Year	Biometric	Acq. device*	Hands	Images	Description
A1	**HKPU** [17]	2003	palmprint	scanner	386	7,752	Cropped hand images, intended for performance comparison. Two sessions. **Graylevel** images with black background.
A1	**Bosphorus** [20]	2010	hand geometry + palmprint	scanner	1,560	4,846	**RGB** images of the entire hand.
A2	**CASIA**[21]	2005	palmprint	digital camera	624	5,502	Popular touch-less palmprint database. **Graylevel** images with black background. Uniform lighting. Cropped fingers.
A2	**IIT-D v1** [18]	2006	palmprint	digital camera	470	3,290	Popular touch-less palmprint database. **RGB** images of entire hand with black background. Uniform lighting.
A2	**COEP** [22]	2010	palmprint	digital camera	168	1,344	**RGB** images of palmprints against black background. Uses pegs to reduce variation in pose/scale.
A2	**GPDS-CL1** [23]	2011	hand geometry + palmprint	2 webcams	100	2,000	**RGB** images of the participants' right hand acquired in both visible and 850 nm.
A2	**Tongji** [24]	2017	palmprint	digital camera	1,200	12,000	**RGB** images of palmprints against black background. Large-scale database.
A2	**PolyU-IITD v3** [25]	2018	palmprint	2 digital cameras	700	-	**RGB** images of hands against black background. Contains images from 2 ethnicities. Significant physical variation considered, as well as long interval acquisition (15 years).
A2	**HFUT** [26]	2019	palmprint	digital camera	800	16,000	**RGB** images of entire hands (palm and wrist) against black background.

Acq. device = Acquisition device;
Hands = Number of hand classes (some databases only have images from one hand per participant)

including the entire hand can be found in the Bosphorus Hand Database [27] which provides RGB images with spread fingers (samples in Fig. 3.2b).
The earliest touch-less palmprint databases (**A2** in Table 3.1) were the ones released by the Chinese Academy of Sciences (CASIA) [21] and by the Indian Institute of Technology in Dehli (IIT-D) [18], of which several samples are provided in Fig. 3.2c and 3.2f respectively. Both used a digital camera for acquisition in an environment with uniform lighting. The main differences are the scale and color information contained in IIT-D. Hand images in CASIA are graylevel and have cropped fingers. The College of Engineering Pune (COEP) [22] released a touch-less database of palmprints, but the acquisition relied on pegs to direct the position of fingers relative to the camera (samples in Fig. 3.2e). Another touch-less database was released by Las Palmas de Gran Canaria University under the name GPDS [23]. They used two webcams to acquire palmprint images in two sessions. One of the webcams was adapted to acquire NIR images by removing its IR filter and replacing it with an RGB filter. The database is split into images acquired in visible range (GPDS-CL1) and in NIR (GPDS-CL2).
Zhang *et al.* [24] released in 2017 a large-scale database (12,000 images) of palmprints acquired with a dedicated device containing a digital camera (Tongji). The acquisition environment was very dark, a controlled light source illuminating the palm area (samples in Fig. 3.2h).
Recently Kumar [25] released a large-scale database of palmprints entitled PolyU-IITD Contactless Palmprint Database v3, introducing a variety of challenges. Firstly, it contains hand images from two ethnicities (Chinese and Indian). Secondly, the palmprints are acquired from both rural and urban areas. The physical appearance of the hands varies significantly, there being instances of birth defects, cuts and bruises, callouses from manual labour, ink stains and writing, jewelry and henna designs (samples provided in Fig. 3.2g). The database also contains a 2nd acquisition session after 15 years, for 35 subjects.
Several image samples from databases in Table 3.1 are displayed in Fig. 3.2.

[27].

3.1.2 Partly Constrained Palmprint Databases

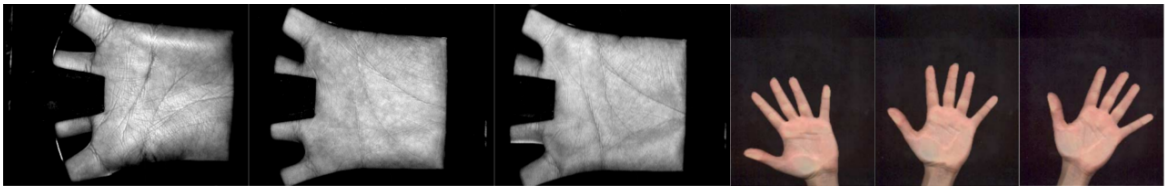
Moving away from constrained scenarios, several databases introduced at least one challenging factor in the context of palmprint recognition systems. An overview is presented in Table 3.2.

Considering an unconstrained environment for acquisition (**B1** in Table 3.2) leads to both variable background and lighting conditions. An initial step was made for palmprint recognition in the context of smartphones by Aoyama *et al.* [28] in 2013 with a small database of images (called DevPhone). Unfortunately, the conditions of acquisition are not

Table 3.2 Palmprint databases introducing at least one unconstrained aspect during acquisition: (B1) unconstrained environment/background, (B2) devices used during acquisition and (B3) unconstrained hand pose.

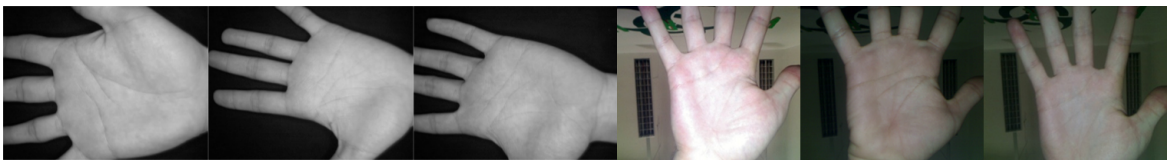
	Name	Year	Acquisition device	Hands	Images	Description
B1	**DevPhone** [28]	2013	1 smartphone camera	30	600	Controlled acquisition using a square guide displayed on the screen; no information about the environment of acquisition.
B1	**BERC DB1** [29] **DB2**	2015	1 smartphone camera	120	8,957 9,224	Unconstrained environment; controlled hand orientation using a visual guide on the device screen; uses flash illumination indoors (DB1) and outdoors (DB2).
B1	**Tiwari** *et al.* [30]	2016	1 smartphone camera	62	182 videos	Users had to position the palm in a circular guide displayed on the screen of the device. Each user filmed 3 short videos of the palm.
B1	**BMPD** [31]	2019	1 smartphone camera	41	1,640	Hand images against black background. The 2nd acquisition session used stronger angles relative to the hands.
B1	**SMPD** [31]	2019	1 smartphone camera	110	4,400	Hand images with black background and flashlight turned on. Included 4 scenarios of hand orientation, including tilted away from the camera.
B2	**Choras** *et al.* [32]	2012	3 smartphone cameras	84	252	RGB hand images with black background.
B2	**PRADD** [10]	2012	1 digital camera, 2 smartphones	100	12,000	Hand images against black background. Two lighting cases considered. Acquisition not performed by user.
B3	**11k Hands** [33]	2017	1 digital camera	380	11,076	RGB images of hands (palm and dorsal) against white background. Variable hand pose.
B3	**NUIGP2** (**Section 4.3 of Chapter 4**)	2020	1 webcam	52	24,631	RGB images of hands with variable hand pose and scales, with several backgrounds. Intended for training of palmprint ROI extraction algorithms.

Hands = Number of hand classes



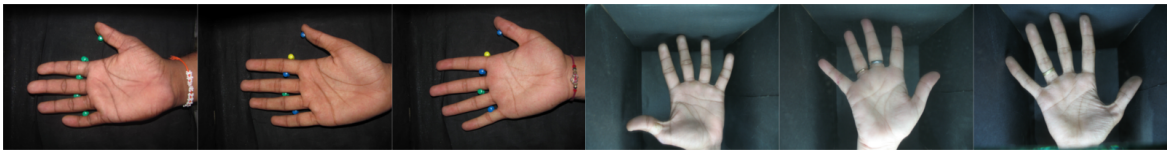
(a) HKPU [17]

(b) Bosphorus, reprinted from [27] ©2006 with permission from Elsevier



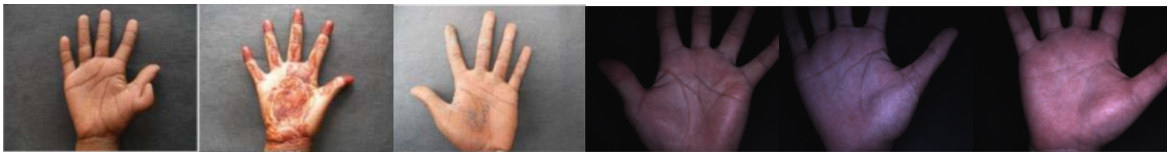
(c) CASIA [21]

(d) GPDS (source: [23] ©2011 IEEE)



(e) COEP [22]

(f) IITD [18]



(g) PolyU-IITD v3 (source: [25] ©2018 IEEE)

(h) Tongji [24]

Fig. 3.2 Hand image samples from several constrained databases in Table 3.1.

clear (how many backgrounds considered, if flashlight was enabled), besides the fact that users were required to use a square guide to align the palm with the center of the acquired image.

A much larger database was acquired by Kim *et al.* [29] both in-doors and out-doors (BERC DB1 and DB2). Both DB1 and DB2 included a scenario where the smartphone's flashlight was enabled. As in the case of DevPhone, the images in BERC DB1/DB2 contained hands with specific hand pose (open palm with spread fingers, to aid ROI extraction, as can be seen in Fig. 3.3a).

A different approach to acquisition was provided by Tiwari *et al.* [30] who recorded videos of palmprints with a smartphone, with the video centered on the user's palmprint.

Recently, Izadpanahkakhk *et al.* [31] introduced two palmprint databases acquired with a smartphone camera - Birjand University Mobile Palmprint Database (BMPD) and Sapienza

(a) BERC (Source: [29] ©2015IEEE)

(b) (Source: [32] ©2012Springer)

(c) 11K Hands [33]

(d) NUIG_Palm2

Fig. 3.3 Hand image samples from partly constrained databases from several databases in Table 3.2.

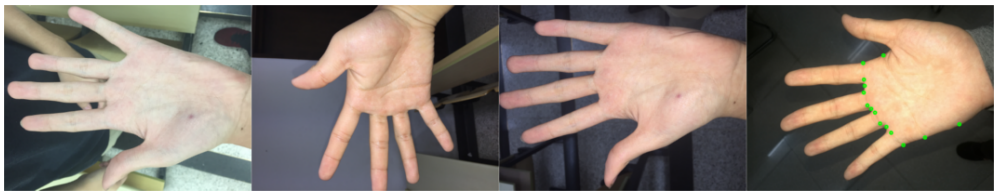
University Mobnile Palmprint Database (SMPD). The variation considered for investigation was the rotation of the hands (in both databases), both in-plane and out-of-plane rotation.

The first database of palmprints acquired with several devices (**B2** in Table 3.2), albeit of reduced size, was developed by Choras *et al.* [32] using three smartphones. Three smartphones were used for acquisition. Several samples are provided in Fig. 3.3b.
Jia *et al.* [10] developed a large database of images entitled Palmprint Recognition Accross Different Devices (PRADD) using two smartphones and one digital camera. The background used was a black cloth. The hand's posture was restricted. From the images provided by the authors [10], it appears that the acquisition was being performed by someone other than the participants.
Unfortunately, the databases developed by Choras *et al.* and Jia *et al.* are currently not available to the research community.

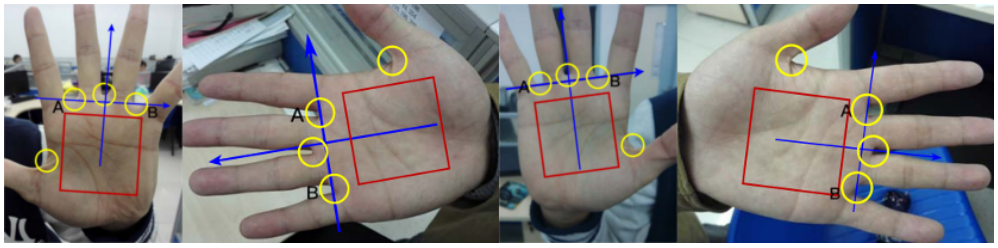
The first palmprint database to consider the hand pose variation (**B3** in Table 3.2), understood as open palms with spread fingers versus closed fingers, was collected by Afifi *et al.* and released under the name 11K Hands [33]. It contains over 11,000 images of hand images - both palmar and dorsal (each has about 5,500 images). The hands were acquired against a white background, using a digital camera (samples are provided in Fig. 3.3c).
An auxiliary palmprint database exploring various hand poses is introduced in Chapter 4 of this thesis, under the name NUIG_Palm2 (NUIGP2). NUIGP2 was designed to support the development of ROI extraction algorithms. Several samples are presented in Fig. 3.3d. The design and collection of the database is described in Section 4.3 of Chapter 4 of this thesis.

3.1.3 Fully Unconstrained Palmprint Databases

(a) NUIG_Palm1 [19]



(b) XJTU-UP (Source: [34] ©2019 IEEE)



(c) MPD (Source: [35] ©2019 IEEE)

Fig. 3.4 Hand image samples from fully unconstrained databases from databases in Table 3.3.

This category of palmprint databases attempts to bring to researchers conditions as close as possible to a realistic deployment of a palmprint recognition system on consumer devices. An overview is presented in Table 3.3.

The first database to provide such palmprint images was released in 2017 by Ungureanu *et al.* [19] under the name NUIG_Palm1 (NUIGP1). It contains images from several devices in unconstrained scenarios (both background and hand pose, as presented in Fig. 3.4a). The design of the database is described in Section 4.2 of Chapter 4 of this thesis.
Recently a large-scale database of palmprint images acquired in similar conditions to NUIGP1 was released by Shao *et al.*, entitled Xian Jiaotong University Unconstrained Palmprint database (XJTU-UP) [34]. The database contains 30,000+ images (200 hands) using five smartphones, making it the largest currently available palmprint database acquired with smartphone cameras. Several samples are provided in Fig. 3.4b.

Table 3.3 Fully unconstrained palmprint databases, as close as possible to the realistic deployment of a palmprint recognition system on smartphones (or similar consumer devices).

Name	Year	Acquisition device	Hands	Images	Description
NUIGP1 [19]	2017	5 smartphone cameras	81	1,816	**RGB** hand images with unconstrained background and hand pose. Two lighting conditions.
XJTU-UP [34]	2019	5 smartphone cameras	200	30,000+	**RGB** hand images with unconstrained background and hand pose. Scenario with enabled flashlight included.
MPD [35]	2019	2 smartphone cameras	200	16,000	**RGB** hand images with unconstrained background and lighting. Images include rotation but appear to have a single hand pose (spread fingers) to facilitate key-point detection.

Hands = Number of hand classes

Another large-scale palmprint database acquired with smartphones was released by Zhang *et al* [35]. They used two smartphones to collect 16,000 hand images in unconstrained conditions.

3.2 ROI Detection and Extraction - a Literature Review

This section presents a general overview of existing approaches for palmprint ROI extraction. The process of ROI extraction is an essential part of the palmprint recognition system, as any inconsistencies in ROIs will affect the recognition task.

Existing ROI extraction techniques can be grouped in four categories, based on the cues contained in the hand images (also represented in Fig. 3.5):

- Standard palmprint ROI extraction: algorithms based on separating the hand from the background (segmentation) and performing measurements to determine the landmarks (or palm region) required for ROI extraction. This family of techniques relies on accurate segmentation, as well as a specific hand pose (open palm with spread fingers).

- ROI extraction based on conventional Machine Learning (ML) algorithms: ML approaches used for the detection of palmprints or used for key-point regression. Key-point regression is understood as any algorithm receiving as input a hand image and returning a set of points used for ROI extraction.

- ROI extraction based on Deep Neural Networks (DNNs): Approaches relying on DNN soutions to perform detection or key-point regression task.

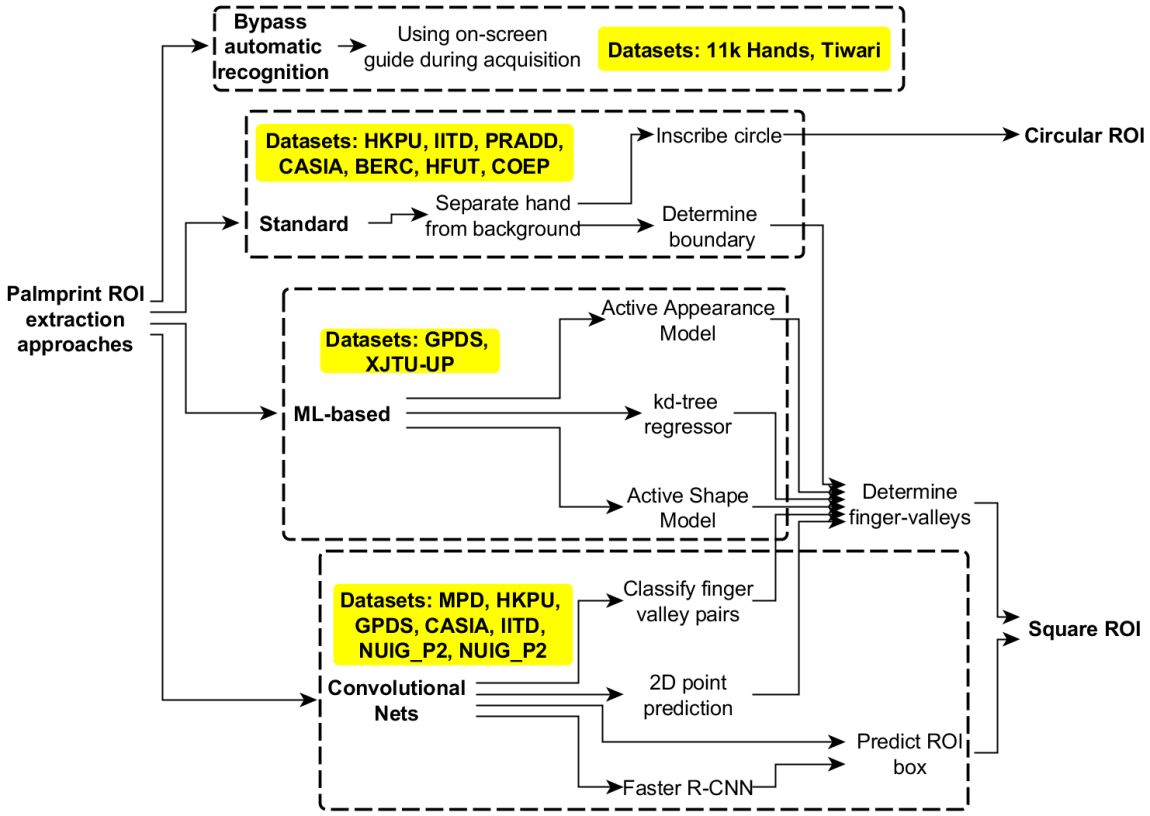


Fig. 3.5 Overview of approaches for palmprint ROI extraction, with four categories based on how constrained the databases are.

- Bypassing ROI detection altogether: based on specific acquisition protocols.

3.2.1 Standard Palmprint ROI Extraction

Standard palmprint ROI extraction algorithms rely on accurate segmentation of the hand region from the background. The most used approaches include using Otsu's thresholding method [36] applied to grayscale images, or using a skin-color model [37]. This is a pre-processing stage required to characterize the shape of the hand and find the key-points required for ROI extraction.

The most popular ROI extraction approach was introduced by Zhang *et al.* [38] in 2003, which relies on the constrained environment from images in databases such as the ones described in Table 3.1, either touch-based or touch-less. It relies on determining the tangent line between the two side finger valleys in order to normalize the palmprint's rotation and provide a reference point from which to extract a square region. This step is made possible thanks to the constrained environment of acquisition (black background, constant lighting),

characteristic of palmprint databases in Table 3.1.
Recently, Xiao *et al.* [26] proposed an approach based on the intersection of the binarized hand with lines of specific orientations, leading to having several candidate points for the finger valleys. They then use K-means clustering to obtain the center of each cluster.

A second category of approaches defines the contour of the extracted hand, and the distance from a point of reference (the geometric center [25, 39] or the wrist [40], etc) to the pixels found on the contour [28, 41–47]. Considering this distribution of distances, the peaks generally correspond to the tips of the fingers, while the local minimas correspond to the finger valleys. This approach is very sensitive to segmentation artifacts and generally includes a smoothing applied to the distribution of distances.

A third category traverses all the contour pixels and counts the pixels belonging to the hand region (a circle was considered for sampling). Balwant *et al.* [48] introduced specific rules to determine the finger valleys and finger tips, followed by the correct selection of finger valley points that form an isosceles triangle. Goh Kah Ong *et al.* [49] considered sampling with fewer points using 3 stages corresponding to circles with greater radius. The outliers resulting from segmentation artifacts were removed with specific rules. Franzgrote *et al.* [50] further developed this approach by classifying the angles of remaining lines in order to provide a rough rotation normalization step. The finger valley points are then determined with a horizontal/vertical line (depending on the orientation of the hand), which has 8 points of transition from non-hand region to hand region.
Morales *et al.* [51] fitted a circle inside the binarized hand, with its center found equidistantly from the finger valleys (previously determined with the center-to-contour distances).

A fourth category uses the convex hull to describe the concavity of the binarized hand map and finger valleys [52, 53].

The following are methods that are hard to classify into one category or another, as they either employ very different or combine several of the previously mentioned approaches together.
Khan *et al.* [54] determined the finger tips and the start of the palm by counting the hand-region pixels along the columns. After determining the pixels corresponding to finger valleys, several 2nd order polynomials were used to extrapolate the middle of the finger valleys. The palm's width was used to determine the size of the ROI (70% of palm size). This approach requires very specific hand pose, with hands always rotated towards the left with spread fingers.
Han *et al.* [55] successively cropped out of the binarized hand image regions corresponding to fingers (after rotation normalization with PCA) by determining the number of transitions from background to hand area. Leng *et al.* [37] determined the finger valleys by computing

differential maps upward, to the right and left. Onto these maps the AND operator was applied, which resulted in 4 regions corresponding to all finger valleys. Ito *et al.* [44] considered an approach based on line detection after determining the binarized hand region, and subtracting the major lines corresponding to finger edges. Then a distance is computed from center of the palm, allowing the detection of finger valleys even with closed fingers (not relying on spread fingers). They compared the effectiveness of this approach by considering the recognition rates resulted after using three other algorithms [37, 38, 55].
Liang *et al.* [56] used an ROI extraction approach loosely based on [38] and [57], where the tip of the middle finger was determined and then extended to the center of the palm 1.2 times. This point is then used as a reference to determine the distance to all contour points, allowing the detection of both finger valleys and tips.
Jia *et al.* [10] exploited the constrained nature of acquisition (hand position pose, scale and rotation) to base the ROI extraction on the accurate detection of the heart line's intersection with the edge of the hand (using the MFRAT defined in [58]), then performing specific pixel operations to decide on the ROI's center and size.
Kim *et al.* [29] combined several elements for ROI extraction, such as the use of a distance based on a YCbCr model, a specific hand pose (fingers spread) indicated by a guide displayed during acquisition, as well as validating finger valley points by sampling 10 pixels from the determined hand region.
Shang *et al.* [59] modified the original Harris corner detection algorithm [60] in order to locate the points at the middle of finger valleys. However, this approach relies on constrained acquisition, as the background is not overly complex. Another approach using Harris corners was proposed by Javidnia *et al.* [61]. After obtaining an initial candidate for the hand region based on skin segmentation, the palm region was located using an iterative process based on the strength of the Harris corners.

However, none of the standard approaches for palmprint ROI extraction can be used in circumstances where the background's color remotely resembles skin color or the hand's pose is not constrained (such as the databases in Table 3.3). Furthermore, one can point out the limitation of skin color segmentation regardless of the chosen color space, based on the inherent inability of classifying a pixel into skin or non-skin [62].

3.2.2 Palmprint ROI Extraction based on Conventional ML Algorithms

There are few approaches using ML algorithms for ROI extraction regressing either a predefined shape or a set of points.
Initially, Doublet *et al.* [63] considered to fit an Active Shape Model (ASM) to a number of points describing the shape of a hand (with spread fingers). The model regresses the output

of a skin segmentation step, after which the centers of the two finger valleys are used to normalize the hand's rotation. Ferrer *et al.* [23] used a similar ASM to extract the hand from the background in the GPDS-CL1 database.
Aykut *et al.* [64] considered an Active Appearance Model (AAM), which also considered the texture information from the hand's surface. They also provided the first evaluation of predicted key-points. Because the acquisition of images was performed in a considerably constrained environment, no normalization was required relative to the palmprint's scale, the authors preferring to report the error in terms of pixels (from the ground truth points).

Recently Shao *et al.* [34] employed a complex pipeline for ROI extraction for unconstrained palmprint recognition. It includes an initial stage of palmprint region detection using Histogram of Oriented Gradients (HOG) and a sliding window providing candidate regions at several scales to a pre-trained SVM classifier for palmprint detection. A tree regressor [65] (initially developed for face key-point detection) is then used for the landmark regression task applied to all 14 key-points. Unfortunately the authors did not provide details regarding the performance of their ROI extraction, how its accuracy influences the recognition task, or any comparison with prior algorithms.

3.2.3 Palmprint ROI Extraction based on Neural Networks

There have been only a handful of attempts to use CNNs for the ROI extraction, and most have consisted solely on experimenting on gray-level images. Bao *et al.* [66] used the CASIA palmprint database [21] to determine the positions of a hand's finger valley points. They use a shallow network composed of 4 Convolutional and 2 Fully-Connected layers, including several Dropout and MaxPooling layers. The proposed CNN architecture achieves results comparable to Zhang *et al.* [38] in very stable conditions, but surpasses it when noise is added. Whereas a CNN can adapt to noisy or blurred images, the pixel-based approach used by Zhang *et al.* is vulnerable to any kind of image quality degradation.

Izadpanahkakhk *et al.* [67] trained a similarly shallow network based on an existing model proposed by Chatfield *et al.* [68]. The network determines a point in the hand image and the corresponding width/height of the palmprint ROI. The network is composed of 5 Convolutional and 2 Fully-connected layers, including several MaxPooling layers and one Local Response Normalization Layer (LRN). The reported results are good for constrained images from HKPU [17], but the case of in-plane rotated hands is not considered.

Jaswal *et al.* [69] trained a Faster R-CNN [70] model based on Resnet-50 (87 layers) on three palmprint databases (HKPU, CASIA and GPDS-CL1). They reported lower Accuracy and Recall rates for CASIA (up to 5% less) than for HKPU and GPDS-CL1. This can be explained by slightly larger variation in rotation. Similar to [67], the predicted bounding

boxes (considered as ROIs) do not include measures for rotation normalization, which considerably affects the recognition rate for the scenario using images from CASIA, as they contain significant rotation variation. Comparatively, images from HKPU and GPDS-CL1 are already normalized rotation-wise.

Recently Liu *et al.* [71] similarly considered a Fast R-CNN [72] for palmprint ROI detection. They acquired several videos of palmprints in 11 environments (no other details provided) where the hand pose was varied (from spread to closed fingers, with several hand orientations). These acquisition sessions resulted in 30,000 images that were used for training and testing. The only method of evaluation the authors considered was the percentage of images above a given threshold for Intersection over Union (IoU).
Several important aspects were not covered: the number of subjects in the training set, the ROI being aligned with the hand (it is maintained vertical regardless of the hand's orientation) or how much an ROI having 60% IoU (with the ground truth) affects the recognition task.

3.2.4 Avoiding the ROI Detection Altogether

Tiwari *et al.* [30] provided a guide on the screen of the smartphone during acquisition, avoiding the need for an ROI step. Tiwari then used an algorithm to determine the best frames for feature extraction. Similar to Tiwari's approach, Leng *et al.* [73] presented a guide on the smartphone's screen, indicating a specific hand pose and orientation for the hand.

Afifi *et al.* [33] considered a different approach, having the entire image as the input to a CNN, thus removing any need for an ROI extraction phase. This approach is only feasible because all other parameters in the acquisition environment (background, lighting and hand orientation/scale) are constant.

3.3 Palmprint Feature Extraction and Classification - a Literature Review

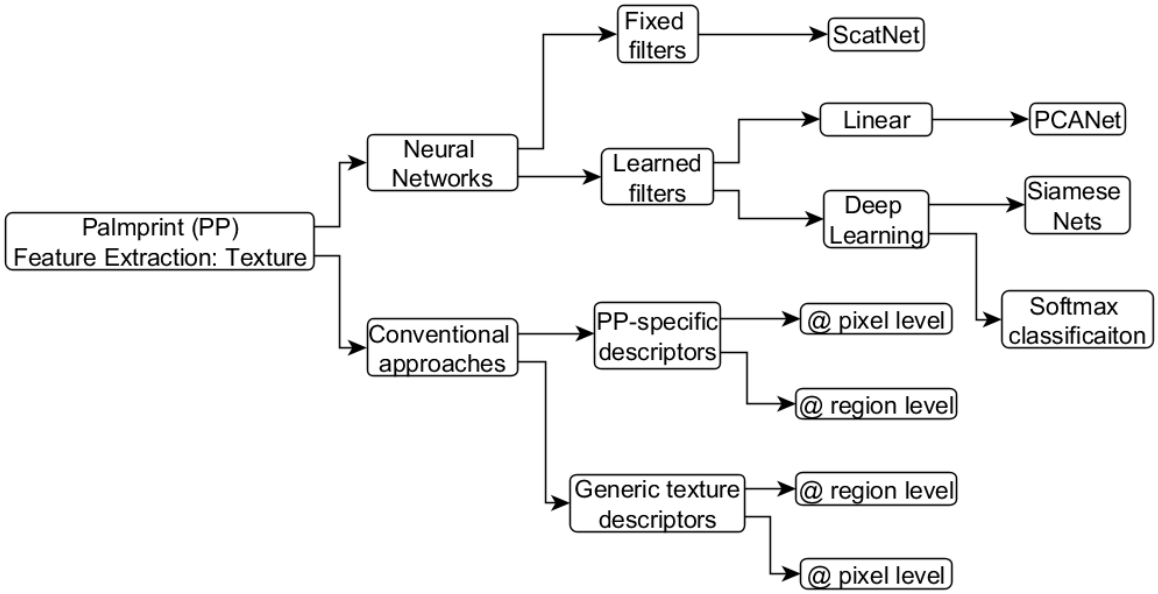


Fig. 3.6 Overview of existing palmprint feature extraction techniques.

This section presents a general overview of approaches used for palmprint feature extraction, with emphasis being placed on the more recent advancements. In this literature overview, the algorithms are split into two categories, based on how the kernels used for feature extraction were obtained (as visualized in Fig. 3.6):

1. Conventional approaches:
   (a) Encoding the line orientation **at pixel-level** with:
      i. Generic texture descriptors
      ii. Palmprint-specific descriptors.
   (b) Encoding the line orientation **at region-level**, with:
      i. Generic texture descriptors, a special category including descriptors such as SIFT, SURF and ORB, which are treated separately
      ii. Palmprint-specific descriptors.
2. Neural Networks approaches:
   (a) Having fixed kernels, such as ScatNet [74]

(b) Kernels learned based on a training distribution:

i. With no non-linearities, such as PCANet [75]

ii. Deep Learning approaches:

A. Classifying with softmax/cross-entropy

B. Using Siamese network architectures.

3.3.1 Palmprint Feature Extraction - Conventional Approaches

Conventional palmprint recognition approaches are mainly focused on line-like feature detection, subspace learning or texture-based coding. Of these, the best performing approaches have been the texture-based ones [76], which will represent the main focus of this overview. For a broader description of the other groups, please refer to the work of Zhang *et al.* [76], Kong *et al.* [77] and Dewangan *et al.* [78].

Jia *et al.* [79] defined a framework that generalized the palmprint recognition approaches. This framework is called Complete Directional Representation and is presented in Fig. 3.7. The stages of feature encoding are broken down and populated with various approaches. The following sub-sections describe these approaches and provide results in the form of either Equal Error Rate (EER) or Recognition Rate (RR) corresponding to popular palmprint databases such as HKPU [17], CASIA [21] or IITD [18].

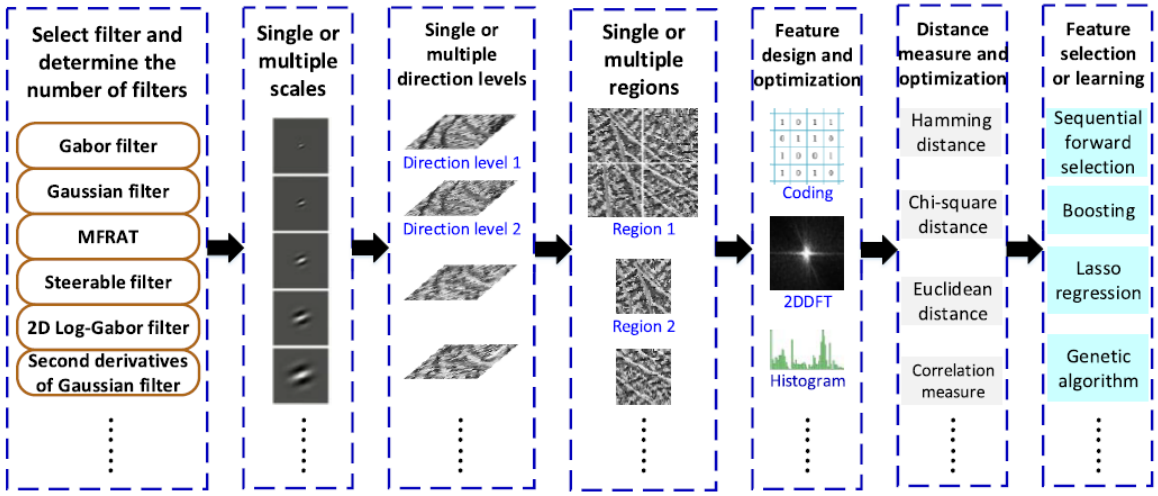


Fig. 3.7 Overview of the framework proposed by Jia *et al.* [79] regarding palmprint feature extraction. (Source: [79] ©2017 IEEE)

Table 3.4 Overview of general texture descriptors used for palmprint feature extraction.

Year	Acronym	Short Descr.	classifier	DB(s)	Best Result (RR/EER)
2010	SAX [80]	Discretization of a 2D grayscale image		HKPU[17] CASIA[21]	0.3% 0.9%
2014	BSIF [81]	Encoding filter responses from several BSIF filters	Sparse Repr. Classifier	HKPU[17] IITD[18]	6.19% (L) 0.42% (R) 1.31%
2014	HOG [82]	Histogram of Oriented Gradients	L2 distance	HKPU[17]	98.03%
2016	LDP [83]	Convolution with Kirsch edge masks	Manhattan + Chi-square	HKPU[17] IITD[18]	6.10% 10.42%
2016	DoN [84]	3D recovered descriptor from 2D image	Weighted sum of 3 scores	HKPU[17] IITD[18] CASIA[21]	0.033% 0.68% 0.53%
2017	LBP [85]	Local Binary Pattern		HKPU[17] IITD[18]	4.92% 9.71%
2017	LTrP [85]	Local Tetra Pattern		BERC1 BERC2[29] IITD[18]	1.49% 1.83% 0.94%

A. Extracting Palmprint Features with Texture Descriptors

Chen *et al.* [80] used a 2D Symbolic Aggregate approximation (SAX) to palmprint recognition. SAX represents a real valued data sequence using a string of discrete symbols or characters. Applied to Grayscale images, it encodes the pixel values, essentially performing a form of compression. Its low complexity and high efficiency makes is suitable for resource-constrained devices.

Ramachandra *et al.* [81] employed a series of BSIF filters that were trained for texture description on a large database of images. The ROI is convolved with the bank of filters and then binarized (using a specific threshold value), allowing for an 8-bit encoding.

Jia *et al.* [82] investigated the potential use of Histogram of Oriented Gradients (HOG) [86], which were successfully used in the past for robust object detection, especially pedestrians and faces. Furthermore, the Local Directional Pattern (LDP) [87] was evaluated in the context of palmprint feature extraction.

Zheng *et al.* [84] described the 2D palmprint ROI with a descriptor recovering 3D information, a feature entitled Difference of Vertex Normal Vectors(DoN). DoN represents the filter response of the palmprint ROI to a specific filter containing several sub-regions (of 1 or -1) intersecting in the center of the filter (borders are made up of 0s), with various orientations. To compare two DoN features, a weighted sum of AND, OR and XOR operators was used.

Li *et al.* [85] extracted the Local Tetra Pattern (LTrP) [88] from a palmprint image that was initially filtered with a Gabor [89] or MFRAT [58] filter. Only the real component from the Gabor convolution was taken into consideration, after the winner-take-all rule of arg_{min} was applied at pixel level between all filter orientations. Block-wise histograms of the LTrP values were then concatenated, to determine the final vector describing a palmprint image. Wang *et al.* [90] used the Local Binary Pattern (LBP), which encodes the value of a pixel based on a neighborhood around it [91]. Generally, the 3x3 kernel is used, allowing codes that range in value from 0 to 255.

B. Encoding Palmprint Line Orientation at Pixel Level

One of the first approaches to extract the palmprint features from an ROI relied on only one Gabor filter oriented at $\frac{\pi}{4}$, entitled PalmCode [38]. Three values were used in the comparison stage of PalmCode, namely the real, imaginary, as well as a segmentation mask to reduce the influence of poor ROI segmentation. Several approaches following a similar rationale were proposed in the following years after PalmCode, with the introduction of Competitive Code (CompCode) [89] and Robust Line Orientation Code (RLOC) [58]. Both CompCode and RLOC used a competitive rule (arg_{min}) between a bank of filters having 6 orientations. Every pixel from the palmprint ROI was considered to be part of a line, and as the lines in the palmprint correspond to black pixels, the minimum response was chosen. Whereas CompCode used the filter response from Gabor filters, RLOC used the filter response from a modified filter Jia *et al.* called MFRAT because it was inspired from the RADON transform. In the case of CompCode only the real component was used.

Gaussian filters were also used, either the derivative of two 2D Gaussian distributions (DoG [93]) or as the difference between two 2D orthogonal Gaussian filters (OLOF [92]).

Guo *et al.* [94] introduced Binary Orientation Co-occurence Vector (BOCV), obtained the filter response of a Gabor filterbank and encoded every pixel relative to a specific threshold (0 or another threshold, chosen based on the distribution of values after convolution with a specific filter). Every filter response was L1 normalized prior to the encoding, after which the thresholded values from each orientation were used to encode an 8-bit number corresponding to every pixel. An extension of this approach was introduced by Zhang *et al.* [97] with EBOCV, which includes masking the 'fragile' bits obtained after convolution with the Gabor filter-bank (as performed previously on IrisCode [102] in the context of iris recognition). In this context, a 'fragile' bit is interpreted as being the pixels close to 0 (after convolution).

Khan *et al.* [54] introduced ContourCode, obtained by convolving the input ROI in two distinct stages. Initially, the filter response corresponding to a Non-subsampled Contourlet Transform (uniscale pyramidal filter) is obtained, after which the ROI is convolved with a

Table 3.5 Overview of approaches encoding the orientation at pixel level.

Year	**Acronym**	**Short Descr.**	**Classifier**	**DB(s)**	**Best Result (RR/EER)**
2003	PalmCode [38]	Real and imaginary components of convolution with Gabor filter $\pi/4$	Normalized Hamming	HKPU-v1 [17]	0.60%
2004	CompCode [89]	Real components of convolution with Gabor filters (6 orientations)	Angular	HKPU-v1[17]	98% at 10^{-6} FAR
2005	OLOF [92]	Convolution with difference of orthogonal Gaussians	Hamming	HKPU-v1[17]	0.0%
2006	DoG [93]	Convolution with Derivative of Gaussians	Hamming	HKPU [17]	0.19%
2008	RLOC [58]	Convolution with 6 MFRAT filters		HKPU[17]	0.40%
2009	BOCV [94]	Thresholding Gabor filter response. Binary encoding	Hamming	HKPU[17]	0.019%
2011	Contour-Code [54]	Two-sequence convolution, followed by hashing	Hash table	HKPU-MS[95] CASIA-MS[96]	0.006% 0.3%
2012	E-BOCV [97]	Removing 'fragile' bits when comparing two samples	Fragile-bit pattern	HKPU[17]	0.0316%
2016	DOC [98]	Include the 2 strongest orientations at pixel-level	Non-linear angular distance	HKPU[17] IITD[18]	0.0092% 0.0622%
2016	Fast-RLOC [99]	Convolution with orthogonal pairs of Gabor/MFRAT filters	Hamming	HKPU[17]	0.041%
2016	Half-orientation Code [100]	Convolution with 2 pairs of half-Gabor filters. Using both halves during comparison		HKPU[17] IITD[18]	0.0204% 0.0633%
2017	COM [101]	Convolution with filters describing concavity	Angular Hamming distance	HKPU-v2 [17]	0.14%

HKPU-MS and CASIA-MS correspond to multispectral palmprint databases.

directional filter bank. The strongest sub-band is determined (arg_{max}) and the resulting code is binarized into a hash table structure.
Fei *et al.* [98] introduced the Double-orientation Code (DOC) which encodes the two lowest responses (to a Gabor filter bank). To compute the distance between two ROIs, a non-linear angular distance, measuring the dissimilarity of the two responses. Considering the 2nd best orientation improves the robustness of the feature.

Zheng *et al.* [99] investigated the effect of number of filter orientations on the efficiency of CompCode [89] and RLOC [58]. A single orthogonal pair of Gabor and MFRAT filters was found to perform better than when using 6 orientations. This encoding approach was called Fast-Compcode/Fast-RLOC due to its increase in speed, mostly due to a reduction in complexity.

An interesting approach was introduced by Tabejamaat *et al.* [101], who described the concavity of a 2D palmprint ROI by convolving it with several Banana wavelet filters [103]. Three pairs of filters (positive and negative concavity) were convolved with the ROI and a competitve rule (arg_{min}) was used for encoding. The joint representation was called Concavity Orientation Map (COM). An angular hamming distance was then used for comparing COMs.

C. Region-based Palmprint Line Orientation Encoding

Jia *et al.* [82] introduce an analysis of region-based methods applied to palmprint recognition. They extend the RLOC encoding capabilities to the region-level by using the histogram of dominant orientations (after the arg_{min} rule). The histograms of orientations were then concatenated. This approach essentially replaced the gradient information used in Histogram of Oriented Gradients (HOG) with the dominant MFRAT filter response. For comparing two extracted palmprint features, the L2 distance was used.
Zhang *et al.* [24] used a similar approach to retrieving block-wise histograms of CompCode orientations, but a Collaborative Representation Classifier (CRC) was used to perform the classification.

Kim *et al.* [29] used a modified version of CompCode, where a segmentation map was first determined by using the real values of the filter responses. This segmentation map was then used to compute the strongest gradients and compute the corresponding HOG. The Chi-square distance was used for comparing the extracted palmprint features.

Li *et al.* [85] extended the general approach of Local Tetra Patterns [88] by replacing the derivative along the width and length with the filter response to MFRAT [58] or Gabor [89] filter banks. Furthermore, the encoding method was modified to take into account the thickness of the palm lines. The image was then separated into regions and histograms are

Table 3.6 Overview of approaches encoding the orientation at region level.

Year	Acronym	Short Descr.	Classifier	DB(s)	Best Result (RR/EER)
2013	HOL [82]	Block-wise histogram of strongest orientation	L2 distance	HKPU[17] HKPU-MS(B) [95]	0.31% 0.064%
2015	[29]	Modified CompCode, with HOG	Chi-square	BERC1[29] BERC2[29] HKPU[17] IITD[18]	2.88% 3.15% 0.11% 5.19
2016	[104]	Neighboring direction indicator		HKPU[17] IITD[18]	0.0225% 0.0626%
2016	LLDP [83]	Extended encoding strategies to Gabor/MFRAT	Manhattan, Chi-square	HKPU[17] IITD[18]	0.021% 4.07%
2016	LMDP [105]	Block-wise encoding of multiple dominant orientations		HKPU-v2[17] IITD[18] GPDS[23]	0.0059% 0.0264% 0.1847%
2016	DRCC [106]	CompCode with side orientations (weighted)	Modified angular dist.	HKPU[17] IITD[18]	0.0189% 0.0626%
2017	LMTrP [85]	Local micro-tetra pattern	L2 distance	BERC1[29] BERC2[29] HKPU-MS [95] IITD[18]	1.11% 1.68% 0.0006% 0.87%
2017	CR-CompCode [24]	Block-wise histogram of CompCode	CRC	Tongji [24]	98.78%
2018	CDR [79]	Convolution with MFRAT at several scales (15) with 12 orientations. 6 overlapping regions	BLPOC	HKPU-v2[17] HFUT[26]	0.001% 0.0868%

HKPU-MS(B) corresponds to the subset of multispectral palmprints from the Blue wavelength.

computed for each region. Finally, they were concatenated and passed through a Kernel PCA filter to reduce the dimensionality of the extracted feature.

Luo *et al.* [83] introduced the Local Line Directional Pattern (LLDP), which represented an extension of general region encoding approaches (LDP [87], ELDP [107] and LDN [108]). The convolution stage replaced the use of Kirsch filters with Gabor or MFRAT filter banks. This step corresponds to replacing the general gradient information in a region with palmprint-specific line information. A similar approach was employed by Fei *et al.* [109] to encode the 2D information in the context of a 3D palmprint recognition system. The response to the Gabor bank of filters was encoded using the LBP [91] strategy. The system used a feature-level fusion technique.

Table 3.7 Overview of approaches based on rotation/scale invariant image descriptors.

Year	Acronym	Short Descr.	Classifier	DB(s)	Best Result (RR/EER)
2008	[111]	SIFT + SAX. Rank-level fusion		HKPU [17]	0.37%
2011	[112]	modified SIFT (OLOF)	Similarity + Hamming	IITD[18] GPDS[23]	0.21% 0.17%
2013	[113]	SIFT	Iterative RANSAC	IITD[18]	(L) 0.513% (R)0.552%
2014	[114]	RootSIFT	Hierarchical classification	CASIA-MS [96]	1.00%
2016	[30]	SIFT and ORB descriptors	Disimilarity index	Tiwari [30]	5.55%
2016	[47]	Sparse representation, fused at rank-level with SVM	SRC + SVM	REST[47] CASIA[21]	98.33% 99.72

Fei *et al.* [105] introduced the Local Multiple Directional Pattern (LMDP) as a way of representing two strong line orientations when these were present, instead of choosing only the dominant line orientation. The block-wise histograms of LMDP codes were computed and comparison was performed using the Chi-square distance. In a similar manner, Xu *et al.* [106] introduced SideCode as a robust form of CompCode, representing a combination of the dominant orientation with the side orientations in a weighted manner.
Fei *et al.* [104] used the Neighboring Direction Indicator (NDI) to determine the dominant orientation for each pixel, along with its relation to the orientations of the neighboring regions in the image.

Jia *et al.* [79] introduced the Complete Directional Representation (CDR) code, encoding the line orientation information at 15 scales with 12 MFRAT filters. From these images 6 overlapping regions were extracted, resulting in 1080 regions. These features were then compared using Band Limited Phase-only Correlation (BLPOC) [110]. This approach was based on the average cross-phase spectrum of the 2D Fast Fourier Transforms (FFT) corresponding to two palmprint samples. The impulse centered on (x_0, y_0) corresponds to the probability of the two samples belonging to the same class (large if intra-class, low if inter-class).

D. Image Descriptors used for Palmprint Feature Extraction

Image descriptors such as the Scale Invariant Feature Transform (SIFT) [115] represented a major breakthrough for object detection in unconstrained conditions because of the rotation and scale invariance of SIFT key-points. This brought much interest to SIFT descriptors, which were either applied directly to palmprint images, such as in [113], [30], [116] or with

certain modifications brought to one of its stages.
Morales *et al.* [112] replaced the Derivative of Gaussian (DoG) with the Ordinal Line Oriented Feature (OLOF) in the stage associated to key-point detection. Furthemore, the score determined from comparing SIFT descriptors was fused with the OLOF comparison decision, making the prediction more robust. Zhao *et al.* [113] improved the initial key-point detection stage by filtering the palmprint image with a circular Gabor filter. Then the corresponding SIFT descriptors were compared using a modified version of the RANSAC algorithm which used several iterations.

Kang *et al.* [114] introduced a modified SIFT which is more stable, called RootSIFT. Furthermore, histogram equalization of the graylevel image was added as a pre-processing stage. A mismatching removal algorithm (of SIFT descriptors) based on neighborhood search and LBP histograms further reduced the number of out-liers.

Charfi *et al.* [47] used a sparse representation of the SIFT descriptors to perform the comparison, as well as rank-level fusion with an SVM. Similarly, a rank-level fusion was performed by Chen *et al.* [111] comparing both SAX and SIFT descriptors.

Tiwari *et al.* compared SIFT and ORB [117] descriptors acquired using smartphone cameras. As with most other approaches using SIFT descriptors, a dissimilarity function was defined, counting the number of in-lier matches performed between two images. Srinivas *et al.* [118] used Speeded Up Robust Features (SURF) [119] to compare two palmprint ROIs. They further improved the comparison speed by only comparing the SURF descriptors extracted from specific subregions of the ROI, instead of the entire surface of the ROI.

3.3.2 CNN-based Approaches

One of the great advantages of using CNNs is that the filters are learned from a specific training distribution, which makes them relevant to the task of palmprint recognition. As opposed to traditional (crafted) features, the learned features are trained to describe any distribution. The main disadvantage of this approach lies in the requirement of abundant and accurately labeled training data, which generally is a problem.

The existing approaches for palmprint feature extraction relying on CNNs, can be split into three categories:

- Using pre-trained models (on ImageNet), the network's output is considered to be the extracted feature. Also relies on a classifier such as SVM.

- Networks of filters optimised using various approaches.

- Training from scratch (or using transfer-learning) of DNNs to determine embeddings that minimize intra-class distance and maximize inter-class distance.

A. Using pre-trained DNNs

Dian *et al.* [120] used AlexNet [121] pre-trained on ImageNet to extract deep features. These were then compared using the Hausdorff distance. In a similar fashion, Tarawneh *et al.* [122] used several networks pretrained on ImageNet (AlexNet, VGG16 [123] and VGG19). The extracted deep features from the images in two hand databases (COEP [22] and MOHI [124]) were then compared using a multi-class SVM.
Ramachandra *et al.* [125] used transfer-learning to compare palmprints acquired from infants. The class decision was obtained through a fusion rule, which also took into consideration the prediction from an SVM.

B. PCANet, ScatNet and PalmNet

Minaee *et al.* [74] employed a scattering network (ScatNet) that was first introduced by Bruna *et al.* [126] for pattern recognition tasks, especially for its invariance to transformations such as translation and rotation. ScatNet uses Discrete Wavelet Transforms (DWT) as filters and considers the output(s) at each layer as the network outputs (not just the last layer), providing information regarding the interference of frequencies in a given image [126]. Meraoumia *et al.* used a filter bank of 5 scales and 6 orientations, the network having an architecture composed of 2 layers. The palmprint ROIs were split into blocks of 32x32 pixels and passed through the network, resulting in 12,512 scattering features. PCA was applied to reduce the dimensionality, reducing it to the first 200 components. An SVM was then used for the classification task.

Chan *et al.* [129] initially introduced PCANet for general pattern recognition applications. Unlike DNNs which make use of the Rectified Linear Unit (ReLU), the PCANet does not contain any non-linearity. Instead, the filters are determined from a distribution of training images. Specifically, a series of overlapping blocks are extracted from every input image, after which the mean is removed. Based on the derived covariance matrix a number of Eigen vectors are extracted (after being sorted, the top 8) and considered as filters belonging to the first layer. The input to the second layer is the distribution of input images to the 1st layer, but convolved with the computed filters in layer 1. This process is repeated for any given number of layers, but generally architectures with 2 layers are commonplace.
PCANet was used for palmprint feature extraction by Meraoumia *et al.* [75] on two databases

Table 3.8 Pre-trained networks (C1), or linear Neural Networks (C2).

Year	Cat.	Acronym	Short Descr.	Classifier	DB(s)	Best Result (RR/EER)
2016	C1	[120]	AlexNet	Hausdorff dist.	HKPU[17] CASIA[21] IITD[18]	0.044% 0.0803% 0.1113%
2018	C1	[122]	AlexNet, VGG16/19	SVM	MOHI [124] COEP[22]	95.50% 98.00%
2018	C1	[125]	AlexNet	Fusion (+SVM)	CPNB [125]	0.310%
2016	C2	**ScatNet** [74]	Deep Scattering Network, fixed weights	SVM K-NN	HKPU [17]	100% 94.40%
2016	C2	**PCANet** [75]	Obtaining filters based on PCA and training distribution	SVM, RBF, RFC, KNN	HKPU-MS[95] CASIA-MS (wht) [96]	0.0% 0.29%
2019	C2	**PalmNet** [127]	Determinining Gabor filters best describing train distribution	1-NN, L2 dist.	CASIA[21] IITD[18] REST [47] Tongji [128]	0.72% 0.52% 4.50% 0.16%

- CASIA Multispectral and HKPU-MS. For classification, both SVM and KNN reported 100% recognition rates across all spectral bands.

Recently, Genovese *et al.* [127] expanded the PCANet approach to include convolutions with fixed-size and variable-sized Gabor filters. The described architecture entitled 'PalmNet' determines the most Gabor filters with the greates response, followed by a binarization layer. An alternative architecture is considered, entitled 'PalmNet-GaborPCA', where the filters of the first layer are configured using the PCA-based tuning procedure used in PCANet, whereas the kernels in the 2nd layer are configured using the Gabor-based tuning procedure. For classification a simple KNN classifier is used.
PalmNet represents an interesting approach for quickly training on large databases of palmprints, requiring few resources.

C. Training DNNs

The main distinction separating approaches in this category is the training strategy being used.
If the classification task is borrowed from the standard pattern recognition problem (like the ImageNet challenge), then the CNN is required to predict the class to which an input palm print belongs to. The network's last layer is fully connected with a number of units corresponding to the number of classes (in the form of a one-hot vector, depending on the size

Table 3.9 Training CNNs for palmprint feature extraction.

Year	Acronym	Short description	Classifier	DB(s)	Best Result (RR/EER)
2015	[130]	Shallow net	Softmax	HKPU-MS [96] Own	99.97% 93.4%
2015	[131]	Shallow net	Softmax	Own	90.60%
2017	[33]	Two-stream CNN	SVM	11KHands[33] IITD[18]	96.00% 94.80%
2018	[67]	Used CNN [68] pre-trained on ImageNet	KNN, SVM, RFC	HKPU[17] IITD[18]	100% 96.9%
2018	[132]	AlexNet, VGG16, InceptionV3 and ResNet50	Softmax	CASIA[21] IITD[18] GPDS-CL1[23]	3.78% 4.79% 4.69%
2019	JDCFR [133]	Shallow CNNs trained on each spectral band (53)	CRC	Own (MS)	0.01%
2019	[31]	Pre-trained VGG16, GoogLeNet, [68] on ImageNet	Softmax	SMPD[31]	93.40%

of the database), with the activation function being softmax (expressing the probability of that input image to belong to either class). In this case, the loss function is the cross-entropy. Example implementations include [31, 33, 67, 130–133].
Fei *et al.* [132] compared the performance of several networks like AlexNet, VGG16, InceptionV3 and ResNet50. Izadpanahkakhk *et al.* [31] trained and evaluated four networks (GoogLeNet, VGG16, VGG19 and a CNN developed by Chatfield *et al.* [68] for the ImageNet challenge) on two novel palmprint databases.
Alternatively, after training with cross-entropy loss, the output from the log-its layer (the layer preceding the softmax layer) can be considered as the extracted feature, which is then used to train a classifier such as SVM [33], Collaborative Representation Classifier (CRC) [133] or Random Forest Classifier (RFC) [67].

Another training approach is to use the Siamese architecture (overview presented in Table 3.10), characterized by two inputs (or several) resulting in two embeddings (usually 128 units corresponding to the last fully-connected layer) that are then compared with a loss function to determine how similar they are versus how similar they should be. This architecture, where the same network outputs the two embeddings, relies on a similarity estimation function, such as the Contrastive loss [134], or the Center loss [135], where the distance between inputs is minimized (intra-class) or increased (inter-class). When the three inputs (triplets) are considered, the distance between the anchor and the positive sample is reduced while increasing the distance between the anchor and the negative sample [136].

Table 3.10 Siamese approach to training CNNs for palmprint feature extraction.

Year	**Acronym**	**Short descr.**	**Classifier**	**DB(s)**	**Best Result (RR/EER)**
2016	[137]	Siamese network trained with d-prime loss		CASIA[21] IITD[18]	1.86% 1.64%
2018	RFN [71]	Soft-shifted Triplet loss		IITD[18] PolyU-IITD[25]	0.68% 0.15%
2018	[139]	VGG16 retrained last layers	Distance	HKPU[17], XJTU-UP[34]	0.2819% 4.559%
2018	Palm-RCNN [138]	Inception-ResNetV1 with Softmax and Center loss	SVM with L2	Tongji-MS	100%
2019	DHCN [34]	Binary Hashing, with Knowledge Distillation [140]		XJTU-UP[34] XJTU-kd [34]	0.60% 5.83%
2019	Deep-MV [35]	MobileNetV2 with secondary net	Softmax probability	MPD [35]	89.91%
2019	PalmGAN [141]	Cross-domain transformation		HKPU-MS [95] SemiU[141] Uncontr. [141]	0.0% 0.005% 5.55%
2019	Meta-siamese [142]	Siamese with secondary network. Few-shot training	Softmax probability	HKPU-MS[95] Pa, Pb Pc, Pd	99.4% 95.4%, 93.4% 98.8%, 96.4%

Svoboda *et al.* [137] introduced a loss function called 'discriminative index', aimed at separating genuine-impostor distributions. Zhang *et al.* [138] used a combination of softmax and center-loss functions during training for multi-spectral palmprint recognition. Zhong *et al.* [139] used transfer-learning based on VGG16 (initially trained on ImageNet) and contrastive loss.

Zhang *et al.* [35] used a Siamese architecture of two MobileNets [143] outputting feature vectors that are then fed to a sub-network tasked with the intra-class probability (between 0 for inter-class and 1 for intra-class, with 0.5 as a decision threshold). Du *et al.* [142] used a Siamese architecture including a sub-network (receiving as input the embedding vectors corresponding to two palmprint ROIs) trained using the few-shot strategy. In a similar fashion, Shao *et al.* [144] used the output of a 3-layer Siamese network, and compared the palmprints from two databases (HKPU-Multispectral and a database collected with a smartphone camera) with a Graph Neural Network (GNN). Unfortunately, the training details of the Siamese network are not clear.

Liu *et al.* [71] introduced the soft-shifted triplet loss as a 2D embedding specifically

developed for palmprint recognition (instead of a 1D embedding). Furthermore, translations on x and y axes were used to determine the best candidates for triplet pairs (at batch level). Recently, Shao *et al.* [34] introduced an approach based on hash coding, where the embeddings used to encode the palmprint classes are either 0 or 1. Furthermore, similar performance was obtained using a much smaller network, obtained via Knowledge Distillation [140]. These are worthwhile directions for development, as they represent solutions to the limitations of resource-constrained devices.

A promising strategy for cross-device palmprint recognition was recently proposed by Shao *et al.* [141] with PalmGAN, where a cycle Generative Adversarial Network (cycle GAN) [145] was used to perform cross-domain transformation between palmprint ROIs. A proof of concept was evaluated on the HKPU-Multispectral (HKPU-MS) palmprint database containing palm images acquired at several wavelengths, as well a semi-unconstrained database acquired with several devices.

3.4 Discussion and Chapter Conclusions

3.4.1 Palmprint Databases

The advancement of palmprint recognition relies on the release of relevant databases which reflect specific sets of requirements. Initially the main focus was placed on recognition, allowing little to no flexibility in terms of interaction with the system (e.g. HKPU [17]).
As the sensing technology progressed (and new consumer devices appeared on the market), there was more room for various aspects, i.e. contactless systems (IITD [18], CASIA [21]). Then invariance to various factors of the acquisition encouraged the introduction of databases like BERC [29] (background), or 11K Hands [33] (hand pose) and PRADD [10] (devices used for acquisition). Unfortunately, there are several databases that are no longer available to researchers, such as PRADD [10] or DevPhone [28]. Some recently introduced databases are yet to be released to the research community (e.g. HFUT [26], MPD [35] or XJTU-UP [34]).

Following the general trend of biometric recognition migrating to consumer devices, the last years have seen the introduction of several large-scale palmprint databases (e.g. XJTU-UP [34]) reflecting the challenging operating conditions brought by a mobile environment. These will be the most meaningful palmprint databases for the upcoming 5 years, anticipating the adoption of palmprint recognition on smartphones.
An overview of this transition was presented, the culmination of which is represented by the

fully unconstrained databases class, initiated with the introduction of NUIG_Palm1 [19] in 2017 (described in Section 4.2 of Chapter 4).

3.4.2 Palmprint ROI Extraction

The approaches used for palmprint region of interest extraction are linked directly with the operating conditions of devices used for acquisition. In palmprint databases where the background is fixed (e.g. HKPU, CASIA, IITD, COEP) the task of segmentation is a straightforward procedure. However, when the background is unconstrained such as is the case with images from BERC, skin color thresholding provides limited results, even when the skin model is computed for every image based on a distribution of pixels [29].

With the migration of palmprint recognition onto consumer devices, the general pipeline for ROI extraction needs to take into consideration more challenging factors such as lighting conditions, hand pose and camera sensor variation. It is in this context that more powerful approaches based on machine learning or deep learning can provide robust solutions without imposing strict protocols for acquisition onto the user of consumer devices.
A complete evaluation of these approaches is yet to be made in terms of:

1. The prediction error of the key-points used for ROI extraction/alignment. This seems to have been a commonly overlooked step in most research papers, with some exceptions (e.g. [54]).

2. Recognition rate and the main sources of error (from the ROI extraction) affecting recognition.

3. Running time and resource requirements, especially for CNN-based approaches. Low inference time is expected from all solutions running on consumer devices.

Furthermore, at the time of writing of this literature review, there are currently no CNN-based solutions to detect the palmprint in unconstrained environments, besides the Fast R-CNN approach demonstrated by Liu *et al.* [71] which does not normalize the extracted ROI's rotation. Also, there is a need for normalization of the extracted ROI using an affine-transformation based on 3D information recovered from the hand's pose, similar to the work of Kanhangad *et al.* [146], but in an unconstrained environment. This can be achieved starting from the CNN architecture recovering the 3D hand structure from a 2D image (as was developed by Mueller *et al.* [147]).

3.4.3 Palmprint Feature Extraction

Although palmprint recognition took off in early 2000's with the introduction of HKPU database, the pipeline stage that received the most attention from the research community has been the palmprint feature extraction.
As was the case for iris recognition, CNNs have become the current state of the art in palmprint recognition (Section 3.3.2). The general trend is to use Siamese networks (e.g. variant of triplet loss in RFN [71] or contrastive loss in DHCN [34], DeepMPV [35]) or non-linear networks (PCANet [75] and PalmNet [127]).

It is important to note that most of these works use in their training/evaluation scenarios images acquired with smartphones (on databases such as XJTU-UP [34] and MPD [35]). The cross-device training and recogniton will become a main focus especially for device-independent palmprint recognition solutions. This is first investigated in [19], with impressive results being obtained in [71]. The cross-domain conversion of a palmprint ROI using a generative approach [141] also represents a promising direction of research.

The complexity of architectures becomes an important factor to optimize as in DHCN [34], where the network is distilled (number of layers is reduced) and the network's output is a discrete hash code (binary values). This not only reduces the processing requirements (including comparison of two samples), but also the storage space necessary when dealing with a large number of classes.

As in the case of ROI extraction algorithms, the feature extraction approaches (especially the CNN-based solutions) require an evaluation in terms of processing time, as this aspect is only touched in few papers (e.g. [29] and [71]).

Chapter 4

Database(s) of Unconstrained Palmprints

The current chapter describes one of the contributions of this thesis, represented by the development of two publicly available databases of unconstrained palmprints.
Section 4.1 presents initial exploratory work regarding collecting palmprint images in 'the wild' (i.e. unconstrained conditions) for a proof of concept database. Section 4.2 describes in detail the planning and acquisition of the palmprint database entitled NUIG_Palm1 which was publicly released in August of 2017[1]. Section 4.3 describes an auxiliary database entitled NUIG_Palm2, used for developing a robust method for palmprint ROI extraction. NUIG_Palm2 was also made available to the research community in 2019.

4.0.1 Motivation

As mentioned previously in Section 3.4.1 of Chapter 3, the key challenge at the moment for biometric recognition systems is their migration onto consumer devices, which have specific constraints, generally associated with limited computing resources. This applies to both processing power, as well as sensor technology.
In this context, the use of palmprints for authentication presents much potential as it does not require extra hardware (e.g. iris recogniton requires a Near-infrared (NIR) camera, fingerprint recognition systems use dedicated scanners, etc.). An overview of current palmprint databases is presented in Section 3.1 of Chapter3. The most popular palmprint databases include the ones provided by the Hong Kong Polytechnic University (HKPU) [17], Chinese Academy of Sciences (CASIA) [21] and the Indian Institute of Technology Dehli (IIT-D) [18]. The main characteristics they have in common are the constrained nature of acquisition - both regarding the camera (lighting conditions and background), as well as the user's interaction

[1]The Unconstrained Palmprint Database NUIG_Palm1 is available upon contacting Prof. Peter Corcoran (peter.corcoran@nuigalway.ie). More details can be found on the C3Imaging website "http://C3imaging.org"

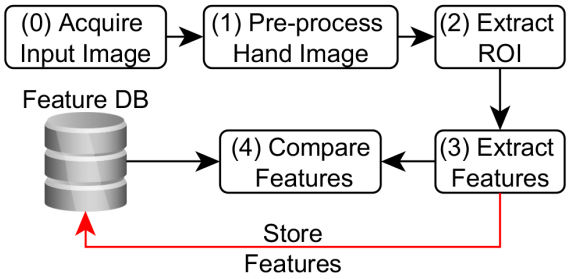


Fig. 4.1 Flowchart corresponding to the main stages of a palmprint recognition system.

with the device (specific hand pose is required). These conditions, however, do not emulate the use-case of consumer devices, which typically operate in unconstrained environments. The unconstrained database NUIG_Palm1 has started a new category of palmprint databases, defined by unconstrained acquisition applied to environment, devices and acquisition protocol (see Table 3.3 in Chapter3). Recently two new palmprint databases were introduced (XJTU-UP [34] and MPD [35]) which validate the research direction started in 2017 with the introduction of NUIG_Palm1.

4.1 Proof of Concept Database - 'Wild' Palmprints

To address this gap, an initial project was started to collect palmprint images using smartphone devices. This database, hereafter referred to as Proof of Concept (PoCDB), was the initial step towards launching a public database of unconstrained palmprints that were acquired with smartphones.

Before starting the collection, an exploratory phase was required, to consider the adequate number of variations, both in terms of utility for future experiments, as well as the steps performed in an actual use-case. Following the flowchart displayed in Fig. 4.1, the main sources of error were identified as being the acquisition and pre-processing stages. Firstly, the acquisition conditions are considered problematic because of the necessary stages requiring the user to perform when presenting their hand in front of the camera, but also because of the device's sensor behavior (and internal noise-reduction techniques). These are:

1. Lighting conditions which change the aspects of the presented palmprint. Strong shadows or intense light affecting the palm will affect both the ROI extraction algorithm's performance, as well as the feature extracting process (environment-related factor)

2. The color of the background, depending on whether or not the ROI extraction algorithm's strategy relies on separating the hand from the background (environment-related factor).

This either refers to a dark and uniform color (in the case of grayscale images) or to a background's whose color is as far from the hand's color as possible (in the case of RGB images)

3. Hand posture (aka pose) is also a critical factor, since most of the current ROI extraction algorithms rely on the user stretching out their fingers and presenting their palm as a flat surface, usually with a given orientation (the wrist is consistently in the same position). This step is essential for both scale and rotation normalization of the hand (user-related factor)

4. The behavioral component of the user's lifestyle affecting the quality of the biometric characteristic. This translates to subjects having skin conditions (dry skin) or marks from intense and prolonged manual labor (calloused hands) or pigment marks related to daily activities (ink stains, henna marks or tattoos). Also relevant are the various items that can be found on hands and fingers in particular - rings, bracelets. An algorithm that is user-friendly cannot be expected to require the removal of all such items. (behavioral-related factor)

5. Pixel size in the camera sensor, as well as the software-based noise reduction employed by the device. These aspects are important because there is a direct correlation between the size of a sensor's pixels and the amount of light which they can absorb (more light collected improves the quality of the image and reduces noise levels). The appearance of a given palmprint in low-light will differ significantly when compared to when it is exposed to normal (office) levels or very intense light (device-related factor)

6. Every device uses a different sensor (main reason for selecting devices from several manufacturers - including both budget and flagship devices), either from a different manufacturer or a more recent generation, ultimately giving to any image acquired with a specific device 'fingerprint' (or behavior). Although these differences are minor, they can be used to distinguish one device's behavior from another one. Put together with the previous point, the interoperability potential of a palmprint recognition system developed for smartphones becomes an important factor to be taken into consideration (device-related factor)

4.1.1 Acquisition Protocol

Based on these factors affecting the recognition system, an Acquisition Protocol was defined around three main variations: light conditions, background types and hand orientation, as

described in Table 4.1.

Table 4.1 Variations considered for the unconstrained PoC palmprint database.

Category	Variation	Specific value (if applicable)	Encoding
Light	Inside normal	7.5 EV or 450 lux	α=1
	Inside dark	6 EV or 160 lux	α=2
	Outside bright	13.5 EV or 28,900 lux	α=3
Background	Complex background-poster	-	β=1
	Real objects scene	-	β=2
	Wooden surface	-	β=3
Hand orientation	Horizontal	-	H_θ=1
	Vertical	-	H_θ=2
	Oblique	-	H_θ=3

To simplify the separation of images according to their combination of variations, a labeling system was developed. Three digits are defined, α, β and H_θ, where α corresponds to one of the three scenarios of light conditions, β corresponds to one of the three background types and H_θ corresponds to one of the three hand orientations considered. By considering α as the Most Significant digit (MS digit) and the H_θ as the Least Significant digit (LS digit) 27 possible combinations are defined, as displayed in Table 4.2.

Keeping in mind the potential for palmprint recognition to be deployed on as many smartphones as possible, 4 mid-to-upper range models are considered. These included recent smartphone-flagships (at the time) Samsung Galaxy S4, but also mid-range phones such as HTC Desire 610, Huawei P8 or the iPhone5. Relevant camera specifications are included in Table 4.3.

A naming convention was adopted for all the images acquired, where all subjects are given a uniquely identifying number, as well as the device's label (model). Alongside the labels encoding the acquisition conditions for a given image, certain metadata of the subject is also taken into consideration, such as age group (indicated by the 4 intervals: A) under 20 years old (y.o), B)20-30 y.o., C) 30-40 y.o. and d) 40+ y.o.), the participant's gender as Male or Female (as indicated by their ID cards), as well as the hand that is being acquired (left or right). This information is later used to provide an overview of the database.
Therefore the names of all image instances were standardized as <Participant ID><Participant gender><Participant Age Group>_<Device used>_<Scenario ID>. An example following this notation would be '001MBL_D610_S1' representing the palmprint image of subject ID 001, who is Male, is between 20 and 30 years old, acquiring images of their left hand

Table 4.2 Encoding procedure for scenario (light, background and hand orientation)

Light	Background	Hand orientation (H_θ)	Scenario encoding
α_1	β_1	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_1, S_2, S_3
	β_2	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_4, S_5, S_6
	β_3	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_7, S_8, S_9
α_2	β_1	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{10}, S_{11}, S_{12}
	β_2	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{13}, S_{14}, S_{15}
	β_3	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{16}, S_{17}, S_{18}
α_3	β_1	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{19}, S_{20}, S_{21}
	β_2	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{22}, S_{23}, S_{24}
	β_3	$H_{\theta_1}, H_{\theta_2}, H_{\theta_3}$	S_{25}, S_{26}, S_{27}

Table 4.3 Camera specifications for the smartphones considered in the Proof of Concept database.

Device name	Sensor size	Sensor Mega Pixels	Aperture	Launch date	Price (Euros)
iPhone 5	33mm (standard)	8 MP	f/2.4	Sept. 2012	340
Huawei P8	28mm (wide)	13 MP	f/2.0	April 2015	230
Samsung Galaxy S4	31mm (standard)	13MP	f/2.2	April 2013	320
HTC Desire 610	not available	8 MP	f/2.4	May 2014	210

using the HTC Desire 610 under the conditions encoded by S1 (α_1 lighting with H_{θ_1} hand orientation against background poster β_1).

All cameras were set to 'Auto' during the entire duration of the acquisition, without using the flash.

4.1.2 Initial Feedback and Contributions

Following the Protocol described in Section 4.1.1, several members of our research group participated in an initial acquisition session. A total of 9 individuals acquired images of their preferred hand.
The feedback they provided resulted into the following changes that were essential for the development of the public database of palmprints (NUIG_Palm1):

1. The procedure was considered to be too lengthy. Reduced lighting conditions to two options and allowed the user any preferred hand pose.

2. Reduced the number of backgrounds to two that could be made into posters. This allowed increased mobility for potential acquisitions.

(a) Background1 (b) Background2 (c) Background3

Fig. 4.2 Palmprint samples from the Proof of Concept database, according to the images' background. Lighting scenarios include: left) S_3, S_1, S_2; center) S_5, S_5, S_{13}; right) S_{23}, S_{27}, S_{19}.

3. Some palmprint images did not contain any of the landmarks traditionally used to extract the region of interest (ROI) such as fingertips or finger-valleys (see Fig. 4.2), which prompted us indicate during acquisition that some parts of the fingers should be included in the acquired image.

4. Decided to add more recent smartphone devices, as the models considered for the PoCDB raised concerns about the rapid obsolescence of the database.

Based on this initial database, a number of experiments implement the recognition system pipeline in Fig. 4.1 or address parts of the pipeline. These experiments and results are presented in Chapter 5.

4.1.3 Conclusions

A relevant gap was identified in existing literature around palmprint recognition on consumer devices. Collecting a small database of palmprint images in 'the wild' (i.e. in unconstrained conditions) allowed the a better understanding of what scenarios should be considered when developing a larger database of images.

4.2 Database of Unconstrained Palmprints - NUIG_Palm1

Introducing the changes listed in Section 4.1.2, a more suitable protocol was obtained to increase the capacity for acquisition.

4.2.1 Devices and Scenarios

The list of devices considered for acquisition was revised, adding one more model and replacing others. Therefore, the set of flagship-grade smartphones included the then recently launched Samsung Galaxy S6, as well as the iPhone 6S. The mid-range phones were replaced with the then recently launched LG G4 and the Huawey P8, leaving for the set of older models the iPhone5. The iPhone5 was taken into consideration in order to have a representative of older smartphone models, but also because of Apple's high standards when it comes to the technology placed in their products. A complete description of these models, as well as the technology they are equipped with, is provided in Table 4.4.

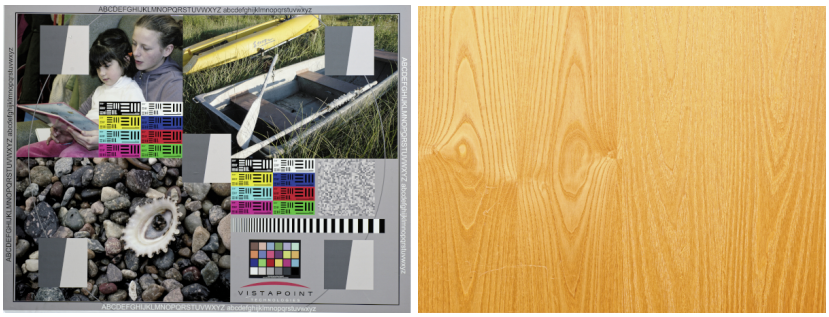
Table 4.4 Devices used for NUIG_Palm1 acquisition.

#	Device name	Sensor size	Stabilization	Sensor resolution	Aperture	Month of launch
1	LG G4 (G4)	1/2.6"	Yes, 3 axes	16 Mega-pixels	f/1.8	April 2015
2	Samsung Galaxy S6 (GS6)	1/2.6"	Yes, 3 axes	16 Mega-pixels	f/1.9	April 2015
3	Apple iPhone5 (i5)	1/3.2"	No	8 Mega-pixels	f/2.4	September 2012
4	Apple iPhone6S (i6S)	1/3"	Yes, 1 axis	12 Mega-pixels	f/2.2	September 2015
5	Huawey P8 (P8)	1/3.06"	Yes, 1 axis	13 Mega-pixels	f/2.0	April 2015

The scenarios in which environmental factors were varied were also reconsidered, choosing to place the main focus on the most challenging ones - 'inside (office) normal' versus 'inside dark'. Furthermore, the number of potential backgrounds was reduced to two posters representing either 'complex scene' or 'wooden texture', as presented in Fig. 4.3. Having these backgrounds in poster form allowed increased mobility for the acquisition phase, increasing the chances of collecting a larger set of images.

Table 4.5 Variation of parameters and encoding of scenarios for NUIG_Palm1. The hand's orientation is left to the decision of the user.

Background	Lighting	Scenario encoding
β_1	α_1, α_2	S_1, S_2
β_2	α_1, α_2	S_3, S_4



(a) Poster-complex scene (b) Poster-wood texture

Fig. 4.3 The two posters used as backgrounds for the acquisition of NUIG_Palm1 database.

4.2.2 Acquisition Protocol and Naming Convention

One of the key aspects of this database is that it not only contains images of palmprints, but that the subjects are the ones performing the acquisition. Making this process as user-friendly as possible meant that it would ease the interaction with the recognition system regardless of the level of tech know-how of the user. It was therefore important to impose as few restrictions as possible regarding the way in which photos were taken.
Subjects were as asked to hold the device with their main hand (usually the right) and take pictures of their other palm. The only requirement regarding the palm was that it remains clearly visible, with at least some part of the fingers. Various levels of bending of the palm were thus captured. Furthermore, the choice of hand pose and orientation was left to the choice of the subject, the most common pose having stretched fingers that are parallel with the floor.

The naming convention of the images was based on the one described for the PoC DB, but applied to the remaining two lighting scenarios and two background types. Therefore all scenario instances are named as <Participant ID><Participant Gender><Participant Age Group>_<Device used>_<Scenario ID>.

For every subject, 4 images were acquired per device (x5), which led to a total of 20 images per ID.

4.2.3 Database Design

To increase the chances for collecting a large database, an Ethics Application form for the collection of a publicly available database of palmprint images was developed over the course of 6 months, which allowed the acquisition of palmprint images from volunteers within the National University of Ireland, Galway (NUIG). The Ethics application form consisted of general background of the project, a detailed description of its purpose, legal background regarding the biometric data regulation in the European Union, as well as a risk assessment to participants. To this end a number of documents (used to provide information to potential participants, allow them to sign up for the project, but also including licensing documentation intended for the future dissemination of the database) were developed and submitted within the application, to the Research Ethics Committee from NUIG. The committee is required to evaluate all data collection procedures taking place within the university, ensuring that participants are protected during and after the procedures involved (precautions taken especially during medical trials that are potentially harmful), as well as the ethics involved in performing and collecting data. Other materials that were consulted included the Article 29 Data Protection Working Party (WP193) [148] which released guidelines for developments in biometric recognition technologies.

The main risk identified for participants was related to the potential connection being made between their identity and their biometric characteristics (present in the images). To address this concern the only link between a participant's ID and their name (and contact information) was kept on the Consent form that must be signed before any acquisition is carried out. These records will be physically destroyed 5 years after the start of the project (projected date is September 2019), to further protect the identities of participants.

4.2.4 Documents Required for Ethics Application

A number of documents were required to be included in the Ethics Application Form. These generally refer to how the project is advertising itself to potential participants, but also to establishing a legal base regulating the interaction between participants and the investigators performing the data collection.

A. Information Sheet

The information sheet is the document which delivered the project's core description, together with its potential risks as well as the participant's rights. Specifically, all participants were required to carefully read this document and ask the investigators for any necessary clarifications. The participant's right to be excluded from the study at any time (during or

after acquisition) was repeatedly stressed. It is standard GDPR (General Data Protection Regulation) procedure to always inform the individual of the manner in which their personal (and biometric) data is being handled.

B. Protocol

The acquisition protocol was presented in non-technical language that ensures a detailed understanding of the entire acquisition process.

C. Consent Form

Following an informing stage, the participant was asked to sign a consent form, which represents the only link between their ID (used for naming the hand images) and their identity. Special care was placed on handling these documents, which were kept secure in a locked cabinet in the office of the Principal Investigator.

D. License Agreement

A licensing document was required, in order to provide the contents of the palmprint database to any member of the research community based on a contract. The preparation of this document included the counseling of NUIG's Solicitor.

The full Ethics Application form is included in the Annex of this thesis.

4.2.5 Ethics Application - Feedback

It is the purpose of the Ethics Committee to asses and correct the interaction between the project's investigators and the project's participants. The feedback received regarding the collection of the public database of palmprint images indicated three points:

- A more detailed description of the participant's interaction with the smartphone devices during acquisition (in the Participant Information sheet)
- Allow more time to participants (24 hours) before deciding to participate in the study. This extended period of time is aligned with WP193's indications on consent [148].
- Specify potential future dangers of participation, namely that one's data may be misused at some later date.

These points were addressed by improving the contents of the documents mentioned (Information Sheet and Protocol), as well as including the destruction of all records linking

any participant's identity with their images in the database. Furthermore, a revision of the Licensing Agreement was added, asking for the document include the signature of the head of department the licensee researchers belong to (formalization of the contract).
Once these points were corrected, the collection phase was started.

4.2.6 Collection Stages

The collection phase was initiated in February 2016, initially within the research group in NUIG (both undergraduate and postgraduate students), followed by the FotoNation office in Galway, Ireland. A total of **45** subjects participated in this collection phase.
A second phase of acquisition took place on April 2016 in the FotoNation office in Bucharest, Romania. This increased the number of participants to **60**.
Finally, a 3rd and final stage was organized on August 2016 in the FotoNation office based in Brasov, Romania. This phase increased the number of participants to **81**.

Fig. 4.4 displays the S1-S4 images from the 5 devices used for acquisition, from 5 different subjects from the database. A variety of hand poses, as well as scales and lighting conditions were captured.

4.2.7 Database Statistics

Once the database was collected, a general overview of its participants could be provided in the form of graphs. As can be noted from Fig. 4.5, participants' age ranges from below 20 (university freshmen) to over 40, with most of the participants being between 20 and 30 years old. Unsurprisingly, few participants preferred acquiring images of their right hand, the number of right-handed individuals being 3 times greater. Also, the ratio of men to women present in this set of images is 2:1.

4.2.8 Labeling Ground-Truth Landmarks

A generic processing pipeline requires palmprints to be extracted from hand images in a consistent manner (i.e. using specific landmarks).

The landmarks and the extracted ROIs are provided as potential benchmarks in future tests related to ROI extraction. Several other palmprint images with their corresponding ROIs are provided in Fig. 4.6.

Fig. 4.4 Inter and intra variation from the acquired database, 'NUIG_Palm1'. The order of devices is maintained as in Table 4.4.

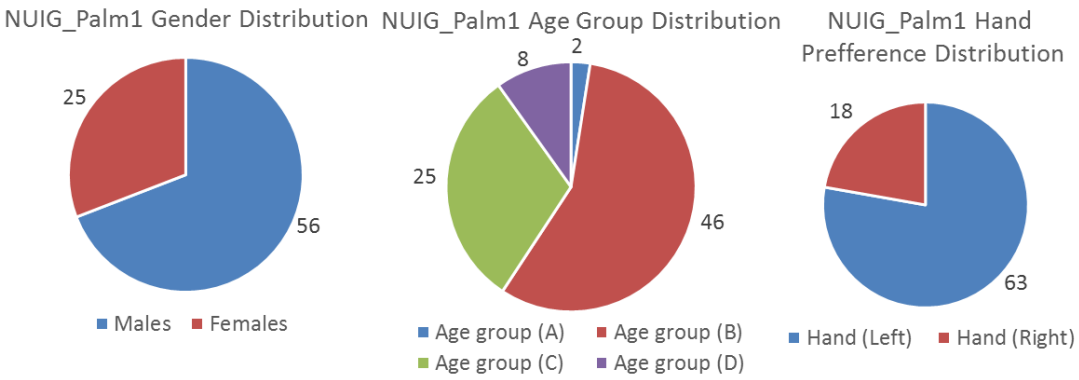


Fig. 4.5 Participants' distributions found in NUIG_Palm1: (left) Gender, (center) Age Group, (right) Hand Preference.

Fig. 4.6 Samples of hand images and their extracted ROI.

4.2.9 Conclusions

The current Section presented an overview of the database of unconstrained palmprints publicly released in August 2017 with the article "Unconstrained palmprint as a smartphone biometric", published in *IEEE Transactions on Consumer Electronics*, under the name 'NUIG_Palm1'.
A detailed description of the design process, as well as the collection stages are provided.

4.3 Database for Automatic ROI Extraction - NUIG_Palm2

This section introduces the collection of a large database of images that supports the training of unconstrained palmprint ROI extraction.

4.3.1 Motivation and Context

In order to address the lack of labeled hand images aligned with the acquisition protocol of palmprints, a novel setup composed of a webcam and a Leap Motion device was constructed. This setup allows for the collection of a large number of images that are already labeled at the moment of acquisition (by the Leap Motion Software) with the 3D coordinates of all finger-joints. The main idea behind this database was to use the keypoints provided by the Leap Motion and extract a palmprint ROI from the hand image.

4.3.2 Leap Motion - Previous Usage

Initially the Leap Motion (LM) device received interest thanks to its novel human-computer interaction, but struggled to find a role other than entertainment. An initial attempt at classifying Auslan number symbols was made by Potter *et al.* [149], but revealed many shortcomings, especially from the hand tracking software, primarily because of its early stages of development. Dimartino [150] implemented a system that can register new static and dynamic gestures, then recognize them using a predefined distance based on the number of standard deviations between the test gesture and all of the stored gesture classes. Marin *et al.* [151] use an SVM to classify gestures based on features outputted from the LM, as well as a Microsoft Kinect device. The outputted feature vectors differ because the sensing strategy is different. Distances, elevations and angles were derived from LM, whereas Kinect provided various information regarding the curvature of fingers.

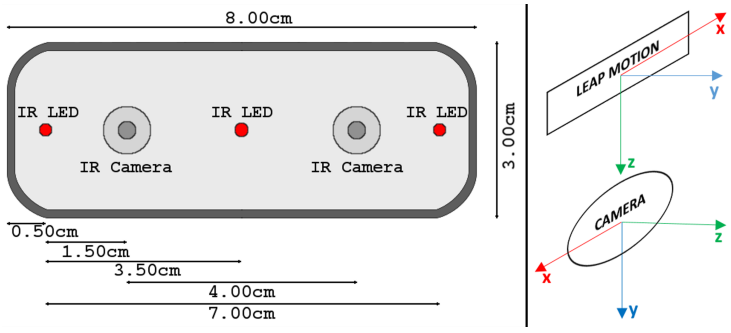


Fig. 4.7 *Left*: Diagram and position of the IR LEDs and cameras, shown on the Leap Motion. *Right*: Axes of the Leap Motion device vs. Axes corresponding to pinhole camera.

4.3.3 Acquisition Setup Based on Leap Motion

Fig. 4.8 *Left:* Webcam-Leap Motion setup used throughout the acquisition phase. *Center:* Example acquisition scenario using a dummy hand. *Right:* Coordinates outputted by the Leap Motion API, projected using the approach described in Section 4.3.4

Ever since the launch of the Leap Motion (LM) device in 2012, its Application Program Interface (API) has been made available to many programming languages and frameworks. At the moment there is a clear movement towards integration with Virtual Reality (VR) systems. In fact, the LM's use in conjunction with VR headsets has been so successful under its project *Orion*, that newer VR or Augmented Reality (AR) headsets are looking to integrate the LM hardware by default [152].

The LM is composed of two Infrared (IR) cameras that are positioned 2 cm from the sides and are flanked by 3 IR LEDs which flood the scene with light. A diagram indicating the placement of these elements and the dimensions of the device is provided in Fig. 4.7.

As mentioned in [153], the exact way LM works is kept a secret, but we are aware that it uses the images acquired from the two infrared (IR) cameras to determine the finger joints (without using stereo vision for depth-map estimation) and track their positions throughout the frames.

The device's accuracy was evaluated by two independent studies [153],[154] under laboratory conditions with the help of robotic arms/rigs. They have concluded that for static setups, the reading was accurate within +/- 0.2 mm, whereas in dynamic setups the reading was accurate within +/- 1.2 mm.

As part of the contribution of this chapter, a novel setup was made (Leap Motion + webcam), allowing the acquisition of a large number of hand images and the necessary labels to extract a hand's palmprint region. The LM was mounted on top of a Logitech Pro9000 webcam, with a distance between the 2 devices' centers of 40 mm, as noted in Fig. 4.8. The LM's Software Development Kit (SDK) for Python2.7 was used to read the finger joints' position relative to the LM's coordinate system. Because the API is referring to bones instead of joints, the start of a bone was considered to be the corresponding joint. Furthermore, as part

Table 4.6 Mapping of bones used by the Leap Motion API, as well as the points used in the experiments of the current paper.

Joint number	Joint index	Bone name	Finger	Used for ROI extraction
1	11	*Metacarpal*		✓
2	12	*Proximal phalange*		✓
3	13	*Intermediate phalange*	*Thumb*	✗
4	14	*Distal phalange*		✗
5	21	*Metacarpal*		✓
6	22	*Proximal phalange*		✗
7	23	*Intermediate phalange*	*Index*	✗
8	24	*Distal phalange*		✗
9	31	*Metacarpal*		✓
10	32	*Proximal phalange*		✗
11	33	*Intermediate phalange*	*Middle*	✗
12	34	*Distal phalange*		✗
13	41	*Metacarpal*		✓
14	42	*Proximal phalange*		✗
15	43	*Intermediate phalange*	*Ring*	✗
16	44	*Distal phalange*		✗
17	51	*Metacarpal*		✓
18	52	*Proximal phalange*		✗
19	53	*Intermediate phalange*	*Pinky*	✗
20	54	*Distalphalange*		✗

of a convention within the API, the thumb is listed as having 4 bones, although it actually has 3 bones, the 4th being considered of length 0. The bones, their corresponding fingers and indexes are listed in Table 4.6.

4.3.4 Projection of 3D Points onto Webcam Image Plane

The projection of points with real-world coordinates (3D) onto a camera's image plane (2D) is a well-known problem, with the solution in Equation (4.1), as described by Zhang [155]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A_{3,3} \begin{bmatrix} R_{3,3} & t_{3,1} \end{bmatrix} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{bmatrix}, \tag{4.1}$$

where (u,v) represent the projected 2D coordinates, (X,Y,Z) are the point's world 3D coordinates, $R_{3,3}$ is the rotation matrix for the 3 axes (x,y,z), $t_{3,1}$ is the translation vector for

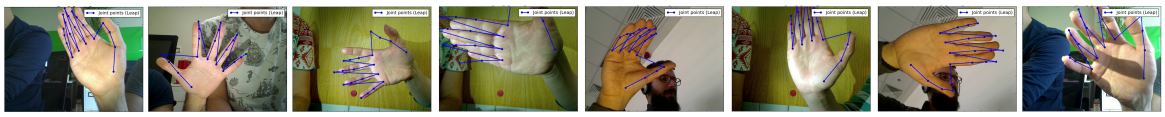


Fig. 4.9 Images and point projections from the database NUIG_Palm2, which has been made available to the research community.

the 3 axes and $A_{3,3}$ is also known as the intrinsic camera matrix, defined by Equation 4.2:

$$A = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4.2}$$

where f_x, f_y are the focal lengths in pixels on the Ox and Oy axes, and v_0, u_0 are the optical center of the image. The camera's intrinsic matrix is obtained by calibrating it with a checkerboard using the OpenCV library for Python [156].
The camera's extrinsic matrix, also known as the camera's pose in the field of Structure from motion, can be determined from the essential matrix. However, in order to obtain the essential matrix, correspondences between 2 different images have to be made. For this reason, the LM database includes a calibration stage for every acquisition session. This stage requires the capturing of 2 images that are shifted by a small amount along the camera's x axis in the range of 2-5 cm. Then specific descriptors are extracted and compared, e.g. Scale Invariant Feature Transform (SIFT) [115] or Speeded-Up Robust Features (SURF) [157]. For the recovery of the camera's pose, the OpenCV library [156] for Python was used.
Finally, it is important to note the underlying discrepancy of the two devices' coordinate systems, represented in Fig. 4.8. Taking this difference into account, as well as the distance from the camera's optical center to the LM's coordinates center, the necessary offsets were added to the values read using the LM (i.e. 40 mm were added on the z axis, as well as 5 mm added on the y axis). For consistency, the 3D coordinates that are saved are relative to the camera's axes, not the original LM coordinates. In an effort to improve the reproducibility of experiments, the Python script used to acquire the images was released on Github [1]

4.3.5 Acquisition Protocol

An acquisition protocol was developed, where every participant was asked to hold their hand in front of the acquisition setup at a distance of around 30cm. The main variations considered

[1]The Python 2.7 script can be accessed at "https://github.com/AdrianUng/Leap-Motion-project-points-onto-image"

were the presence of strong shadows, glare on the hand's surface, as well as varied hand pose. A smartphone's flashlight was used to direct strong light from multiple angles to the surface of the participant's hand, thus enabling all regions of the hand to be either affected by glare or by shadows of varying intensity.
The poses that were considered correspond to those scenarios associated to palmprints in 'the wild': straight fingers but either the thumb or a side finger (index or pinky) touching one of the central fingers. Furthermore, as many orientations as possible were considered, but this number depended on the limits of LM device or the participant's range of motion.
Furthermore, several backgrounds were used, but this was not the main focus of the collection. However, most of the acquisitions included 2 types of backgrounds: 'cluttered office' and 'skin-color background'.

The acquisition session was estimated to last around 15-20 minutes, with intervals of around 20-30 seconds where one of the above mentioned factors was varied. The span of each interval was used to acquire image frames from the webcam, together with the 3D coordinates (according to the webcam's coordinate system) as well as the 2D coordinates. This process was monitored by the investigator by looking at the real-time projection of 3D coordinates onto the webcam's image plane.
The resolution at which the images were saved was the maximum possible output possible for the webcam model that was used - 1200x1600 pixels.

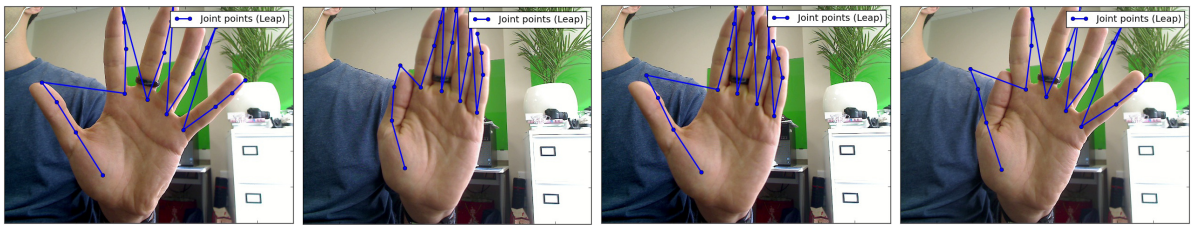
4.3.6 Landmarks

The landmarks considered for ROI extraction are the metacarpal joints from the 4 main fingers, together with the proximal and intermediate phalanges of the thumb. These joints best characterize the surface of the palm, and represent the most accurately read values by the LM because the points are found on the same plane.
Since the projection of LM 3D points lacks accuracy, as can be noted in Fig. 4.10 and Fig. 4.9, the 2D points can be considered an approximation of the ground truth (GT).

4.3.7 Privacy and Consent forms

All participants that took part in this database were informed of the intention to make the database public prior to their acquisition session. To address current GDPR regulations a Consent Form was signed, following the design used for supporting documents of NUIG_Palm1 database.

Because of the proximity to the participants, as well as the angle of the acquisition device, some of the images contain exposed regions of the participants' face. Great care was taken to



(a) (b) (c) (d)

Fig. 4.10 Projected points' spatial variation depending on the hand's posture:*a*) all fingers spread, *b*) all fingers closed, *c*) closed fingers and parted thumb, *d*) parted fingers, closed thumb

mask any facial distinctive traits by employing specific blurring filters in order to maintain the anonymity of all participants.

4.3.8 Collection

The collection phase was initiated in April 2018, when 10 participants were acquired from within our department. This corresponds to session number 4 in Fig. 4.11.
A 2nd phase took place in May 2018 at the FotoNation office in Bucharest, Romania where 6 participants were enrolled. This corresponds to sessions number 2 and 3 in 4.11.
A 3rd phase took place in June 2018 at the FotoNation office in Galway, Ireland. This session added images from 7 more subjects. A 4th and final phase took place in February 2019, which increased the database to its final number of 26 participants. This phase corresponds to session 6 and a part of session 4.

4.3.9 Database Statistics

The final database contains a total of 24,631 images acquired from 26 participants. The distribution of images according to background (/session location), participant age group and participant gender are presented in Fig. 4.11.

4.3.10 Conclusion

The images collected with the setup described in Section 4.3.3 will be made available to the research community under the name 'NUIG_Palm2' through a Licensing Agreement. The 3D coordinates, 2D point projections and the source code used for acquisition and testing will also be provided, to guarantee that experiments are reproducible. It is the authors' hope that this database will aid the development of unconstrained palmprint ROI extraction algorithms.

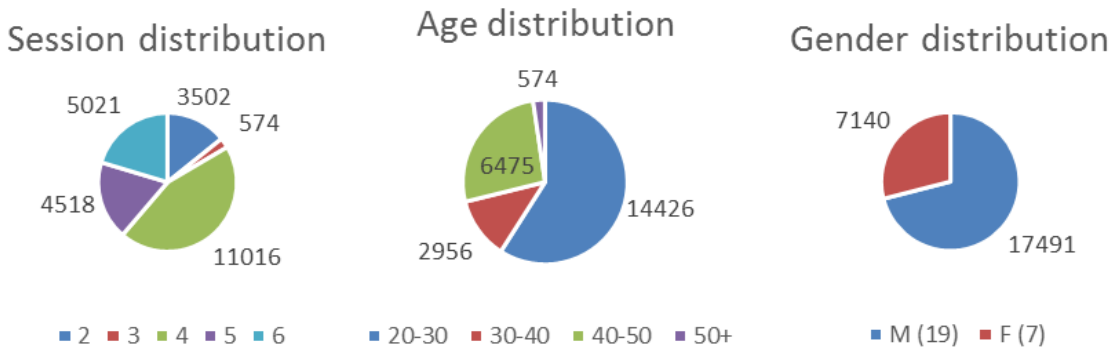


Fig. 4.11 Distribution of meta-information regarding the contents of NUIG_Palm2 (26 participants). *(left)*: number of images for each session: (2) Office white background, (3) home cluttered background, (4) Office green background, (5) Wooden background, (6) office and wooden background. *(center)*: number of images for each age category, where A:20-30, B:30-40, C:40-50 and D:50+ *(right)*: number of images attribute to male/female gender.

4.4 Chapter Conclusions

This chapter introduces one of the primary contributions of this thesis - the database of hands collected to address the challenges associated with palmprint recognition in "the wild". They follow the guidelines of standard palmprint recognition systems, but with the challenges associated with an unconstrained environment.

Firstly, a Proof of Concept (PoC) database was designed to determine the main challenges of palmprint recognition. Secondly, a large scale database is collected using a variety of devices under various lighting conditions and with backgrounds describing complex scenes. Thirdly, a set of images with labeled points allow for training automatic ROI extraction algorithms in unconstrained conditions.

Chapter 5

Palmprints in the Wild - Preliminary and Exploratory Experiments

This chapter represents a collection of initial experiments that are based on the Proof of Concept database (PoCDB) of palmprints described in Section 4.1 of Chapter 4. The palmprint recognition pipeline is composed of several stages, namely: acquisition and segmentation from input image, palmprint Region of Interest (ROI) extraction, followed by feature extraction and sample comparison. The experiments presented in this chapter explored the pre-processing stages, as presented in Fig. 5.1. The findings led to several peer-reviewed conference publications.

An initial set of experiments revolved around skin segmentation, in order to process the hand pixels instead of the background pixels. This is normally a straightforward task in conventional and constrained palmprint databases (such as the ones presented in Section 3.1 of Chapter 3), but represents a considerable challenge for unconstrained palmprint databases (such as the ones presented in Section 3.3 of Chapter 3).

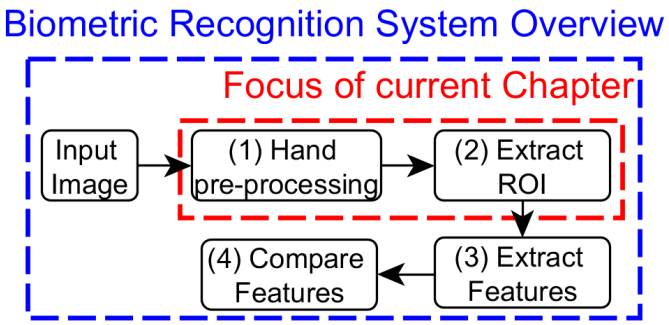


Fig. 5.1 Generic overview of a hand-based biometric recognition system using optical sensors, with the focus of the current chapter.

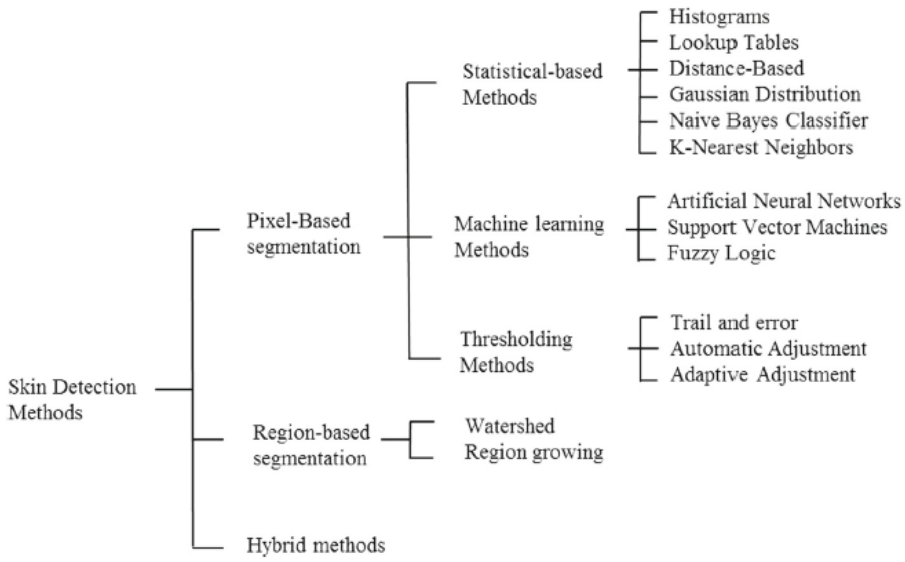


Fig. 5.2 Taxonomy of skin segmentation techniques, as defined by Naji *et al.* [1]. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Artificial Intelligence Review, Naji, S., Jalab, H.A. & Kareem, S.A., "A survey on skin detection in colored images", ©2019.

The use of various color spaces for skin segmentation is presented in Section 5.1. Section 5.2 describes an approach for Region of Interest (ROI) extraction based on Harris corners. 5.3 presents conclusions for the chapter.

5.1 Skin-color Pixel Segmentation

5.1.1 Related Work

Naji *et al* [1] provide an overview of approaches for skin segmentation in color images. According to this taxonomy, the most successful approaches are the ones based on machine learning methods. Kim *et al* [158] train two popular CNN architectures: VGG [123] and a network-in-network architecture based on 20 Inception Modules [159]. The results reported for patch-wise training (slicing the input image into several windows) were better than whole-image training (the input image being downsampled). Based on the literature,
However, statistical-based and thresholding methods are generally faster than methods based on stacked convolutions, making them more suitable for deployment in a resource-constrained environment [160].

5.1.2 Definition and Rationale

Segmentation is defined as the process of extracting a region from an image based on certain properties which strongly describe that region. In the case of skin segmentation, the main objective is to extract from an image all pixels which are identified as belonging to the class of 'skin' [161].
The general model is formulated as the probability of a pixel belonging to the foreground (skin class) as being $P(m(x_i,y_i)|\ m(x_i,y_i)\ in\ C_{skin})$ where (x_i,y_i) represent the value a pixel has in the position defined by the x and y coordinates. Similarly, the probability of a pixel to belong to the background can be formulated as $P(m(x_i,y_i)|\ m(x_i,y_i)\ in\ C_{non-skin})$. It is important to notice that a pixel cannot belong to both classes, which turns this into a binary classification task.

A model is required to determine the distribution of values which best characterize the class C_{skin} such that it has little to no overlap with class $C_{non-skin}$, describing the background.

5.1.3 Color Spaces for Skin Segmentation

There are a variety of ways to represent images, many color spaces having been developed to answer to specific technological requirements. Either attempting to mimic the human visual system or serving specific applications such as skin detection or perception uniformity across devices, there are many ways of encoding color information. This subsection provides an overview of several well-known and several that are less known color spaces, with the ultimate aim of having them implemented for the task of skin-color detection.

A. RGB and nRGB

The most convenient way of looking at an image is to consider its colors as being the sum of primary colors – red, green and blue. This represents the basis of the RGB color model, represented as a three-dimensional system of coordinates where any color is characterized using for each axis values ranging from 0 to 255.The normalized version of RGB uses a similar representation, but considers normalized values of its R, G and B components relative to their sum. For every pixel location the equation 5.1 is used to compute the new R', G' and B' levels such that equation 5.2 is true for all all pixels.

$$\begin{cases} R' = R/R+G+B \\ G' = G/R+G+B \\ B' = B/R+G+B \end{cases} \tag{5.1}$$

$$R'+G'+B' = 1 \tag{5.2}$$

B. YCbCr

Historically, YCbCr was developed in the early stages of television (tv) development, from the need to send a more compact representation of information sent through communication channels. Once the information would arrive at the local tv box, it would decode the signal and represent the images with RGB. The novelty is represented by the separation of colors from the grayscale information into tow difference signals. The Y, Cb, and Cr components represent a linear combination of RGB values, obtained with the equation 5.3:

$$\begin{cases} Y' = 0+(0.299\cdot R'_D)+(0.587\cdot G'_D)+(0.114\cdot) \\ C_B = 128-(0.168736\cdot R'_D)-(0.331264\cdot G'_D)+(0.5\cdot B'_D) \\ C_R = 128+(0.5\cdot R'_D)-(0.418688\cdot G'_D)-(0.081312\cdot B'_D) \end{cases} \tag{5.3}$$

where R', G' and B' correspond to RGB values defining a specific color.

C. HSI

The HSI color model was developed to provide color encoding to existing monochrome broadcasts without having to change the luminance (black and white). Therefore any color represents a non-linear combinations of RGB values into three components:

- Hue, corresponding to the attribute of a visual sensation appearing to be closer to red, green or blue
- Saturation, which indicates the 'purity' of the color relative to its brightness. At low levels of saturation the colors tend to fade away into a generic gray.
- Intensity (or Value) which provides the grayscale (black and white) information. HSI tends to be perceptually closer to the way the human visual system tends to represent colors.

The HSI is computed using normalized RGB components. Considering R, G and B to be the un-normalized values, the Intensity component is computed using 5.4:

$$I = \frac{1}{3}(R+G+B) \tag{5.4}$$

A color's Hue (θ) is defined as the angle between the location of a color within the HIS triangle to the line from White to Red. Equation 5.5 defines a color's hue in terms of its normalized color components as:

$$\begin{cases} A = \left[(r-\frac{1}{3})^2 + (b-\frac{1}{3})^2 + (g-\frac{1}{3})^2\right]^{1/2} \\ B = \frac{2}{3}(r-\frac{1}{3}) - (b-\frac{1}{3}) - (g-\frac{1}{3}) \\ \theta = \arccos\left[\frac{B}{A(\frac{2}{3})^{1/2}}\right] \end{cases} \tag{5.5}$$

A color's saturation is defined as how far the color is located from the center of the HSI cylinder. Colors located at the outer edge of the triangle are fully saturated while pastel colors are located near the center of the cylinder. The Saturation (S) is defined as in equation 5.6:

$$S = 1 - 3 \times min(r,g,b) \tag{5.6}$$

D. CIE L*a*b

The CIE L*a*b color space was designed by the International Commision on Illumination in 1976, to approximate human vision and aspires to be perceptually uniform in its representation of color. It expresses color as three values:

- L* for lightness from black (0) to white (100)
- a* from green to red
- b* from blue to yellow.

The non-linear transformation into L*a*b color space is performed using equations 5.7, 5.8 and 5.9, having already converted RGB values into XYZ:

$$\begin{cases} X = R \cdot 0.4124 + G \cdot 0.3576 + B \cdot 0.1805 \\ Y = R \cdot 0.2126 + G \cdot 0.7152 + B \cdot 0.0722 \\ Z = R \cdot 0.0193 + G \cdot 0.1192 + B \cdot 0.9505 \end{cases} \tag{5.7}$$

$$
\begin{aligned}
L^* &= 116\, f\left(\frac{Y}{Y_n}\right) - 16 \\
a^* &= 500\left(f\left(\frac{X}{X_n}\right) - \left(\frac{Y}{Y_n}\right)\right) \\
b^* &= 200\left(f\left(\frac{X}{X_n}\right) - \left(\frac{Z}{Z_n}\right)\right),
\end{aligned}
\tag{5.8}
$$

where

$$
f(t) = \begin{cases} \sqrt[3]{t}, & if\ t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29}, & otherwise \end{cases}
\tag{5.9}
$$

$$
\delta = \frac{6}{29}
$$

The X_n, Y_n, Z_n in Equation 5.8 correspond to the CIE XYZ tristimulus values of the reference white point. Under Illuminant D65 with normalization $Y = 100$, they are considered to have the values: $X_n = 95.0489$, $Y_n = 100$, $Z_n = 108.8840$.

E. Ohta (I1I2I3)

A lesser known color space is the Ohta color space [162], also known as I1I2I3 is specifically designed for color segmentation. The main idea is to represent colors using a coordinate system that is uncorrelated (has a basis close to the principal components of the colors in the image, making them maximally discriminant). It can be estimated from RGB coordinates using the equation 5.10:

$$
I_1 = \frac{R+G+B}{3}; I_2 = \frac{R-B}{2}; I_3 = \frac{2G-R-B}{4}
\tag{5.10}
$$

F. TSL

TSL is a perceptual color space defining color as Tint (degree to which a stimulus can be described as similar to or different from another stimuli that are described as red, green, blue, yellow and white), Saturation (colorfulness of a stimulus relative to its own brightness) and Lightness (brightness of a stimulus relative that appears white in similar viewing conditions). It was initially developed as a solution to face detection through skin segmentation [163]. The conversion from RGB values is based on equation 5.11:

$$\begin{cases} \frac{1}{2\pi}\arctan\left(\frac{r'}{g'}+\frac{1}{4}\right) & if g' > 0 \\ \frac{1}{2\pi}\arctan\left(\frac{r'}{g'}+\frac{3}{4}\right) & if g' < 0 \\ 0 & if g' = 0 \end{cases} \tag{5.11}$$

$$S = \sqrt{\frac{9}{5}(r'^2 + g'^2)}$$

$$L = 0.299R + 0.587G + 0.114B$$

where r' and g' are defined based on the normalized r and g components:

$$r' = r - \frac{1}{3}\; g' = g - \frac{1}{3} \tag{5.12}$$

5.1.4 Modeling Skin Color based on a Distribution of Pixels

Considering a set of images displaying skin in a variety of scenes, from several devices and under various lighting conditions, the probability distribution of the skin pixels can be modeled, such that similar pixels in other images (scenes) can be classified as being 'skin'. This estimation can be modeled as a Gaussian distribution based on the histogram of all pixels considered as 'skin' in the training set [160]. This process is called back-projection and can be visualized in Fig. 5.3. Given the fact that the experiments rely on skin regions extracted from hand images, it is important to remove the line/edge information prior to computing the skin-pixel distribution. Inspired by Yogarajah *et al* [164], a level of confidence of 95% is chosen based on the histogram of skin-pixels from the training set.

For every color space, a threshold is decided for every color channel separately, the final skin mask being the result of an AND operator applied to the input image. For instance, in the case of RGB, the distribution of Red, Green and Blue values are determined separately, the final threshold region bein a cube found within the RGB cube. The same methodology applies to all color channels of nRGB, I1I2I3. In the case of L*a*b and YCbCr, only the chromatic channels are modeled. In the case of HSI or TSL, only the chromatic channel was used.
Hereon this model will be named histogram-based back-projection (HBP)

5.1.5 Clustering of Pixels with K-means

Clustering techniques are used for unsupervised classification and are used to determine the separation of an input pattern set into a number of clusters (or classes) based on a given metric relative to the centers of those clusters . One of the most popular clustering algorithms

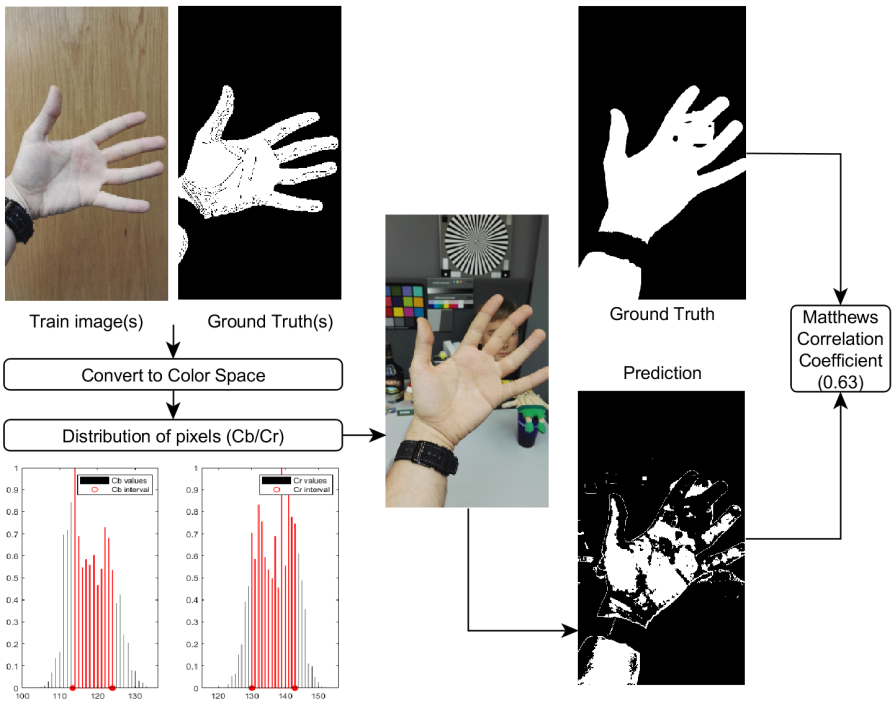
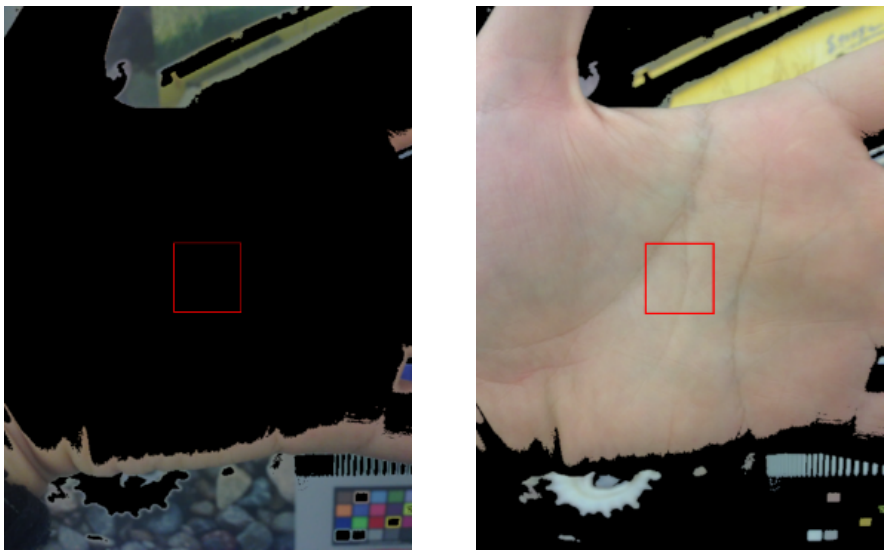


Fig. 5.3 Visualization of the HBP Model estimation applied to Cb/Cr distributions of training pixels [164].



(a) Background cluster (b) Skin cluster

Fig. 5.4 Approach for determining which cluster contains the skin: a) 1st cluster, containing mostly black pixels in the sampling region in the center, b) 2nd cluster, having no black pixels in the sampling region (red square). The black pixels correspond to pixels which are not part of the respective cluster.

is the k-means [165], due to its low complexity $O(n)$. The squared error for a clustering L of a pattern set $\mathscr{X}$ (containing K clusters) is defined as in equation 5.13:

$$e^2(\mathscr{X},\mathscr{L}) = \sum_{j=1}^{K}\sum_{i=1}^{n_j}\left|\left|x_i^j - c_j)\right|\right|^2 \tag{5.13}$$

where x_i^j is the i-th pattern belonging to the j-th cluster and c_j is the centroid of the j-th cluster.

For the application of skin-segmentation two clusters were considered, separating the pixels into the two considered classes – background and skin. There are, however, a few challenges in adapting the use of k-means to the use case of skin segmentation in hand images.

Firstly, the clusters' pixel distribution is not determined a priori. This meant that an approach had to be developed where the cluster containing skin pixels is determined. Based on the assumption that the central part of the image always overlaps a portion of the palmprint, a square region is defined with the task of sampling pixel values. Counting the black pixels contained in this region allowed the identification of the 'background' cluster. This process is visualized in Fig. 5.4, where a palmprint occupies a great part of the image, including the center. The red square is determined as being 1/6th of the input image's width or height (depending on which one is smaller).
Secondly, in order to reduce the influence of the random initialization step of the clusters' center, the clustering was repeated 10 times for each image.

Because clustering approaches use distances for classifying pixels, each pixel could be described by several channels of a color space. For instance, in the case of RGB or I1I2I3, the distances were computed with respect to centroids defined by (R,G,B) and (I1,I2,I3) coordinates. This approach is different from the adaptive thresholding approach defined in Section 5.1.4, where an interval is computed for every channel separately, the final mask of skin pixels being the result of an AND operator between the three thresholds.

5.1.6 Experimental Methodology

This section describes the experiments performed and the evaluation of the obtained results. The approaches presented in Section 5.1.4 and 5.1.5 were trained on a distribution of skin pixels extracted from a series of images. These models are then tested on a different distribution of hand images.

A. Images used for Train/Testing

A subset of the Proof of Concept (PoC) database of palmprints was considered for this experiment. There are 4 cases being investigated:

- Skin tone belonging to person1, of Middle-eastern ethnicity (108 images)
- Skin tone belonging to person2, of Caucasian ethnicity (108 images)
- Skin tone belonging to person3, of South-Asian ethnicity, with darker skin color (108 images)
- All of the above mentioned images (324 images)

All images (both training and testing) and their corresponding ground truth masks were initially resized to 20% of their initial resolution, by using bilinear-interpolation. This step was included to increase the speed of experiments. No other pre-processing stages were performed on the images after resizing.

B. Ground-truth Labels

For the considered images, binary masks denoting skin regions were manually annotated using the freely available GNU Image Manipulation Program (GIMP) [1]. They were made the same resolution as the color images.

C. K-fold Cross-Validation

Usually cross-validation is intended for situations where small-scale databases are available, to remove the bias of training scenarios [166]. Initially the images are shuffled, followed by a split into 8 parts, where each part is used for training and the remaining 7 are used for testing. This applies only to the approach described in Section 5.1.4, as it requires a training phase. For a fair comparison, the same order of the shuffled images was maintained throughout all setups (for each color space).

D. Metrics for evaluation

For evaluation purposes the Matthews Correlation Coefficient (MCC) [167] was used. It is defined using the rate of classifying pixels as either True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), with the equation:

[1] Available for download from 'https://www.gimp.org/'.

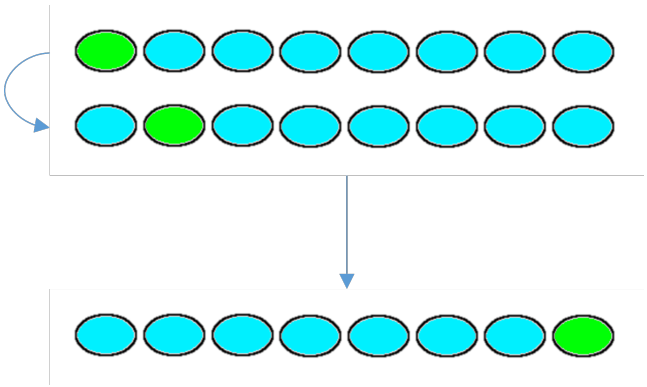
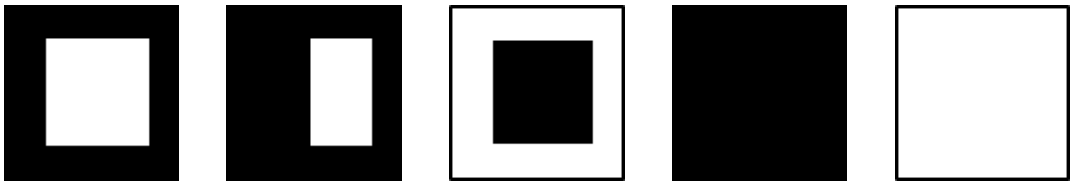


Fig. 5.5 Visualization of the K-fold cross-validation process [166], where the green parts are used for training and the rest are used for testing. After obtaining results for each scenario, the average performance is computed.



(a) Ground truth mask (b) Prediction mask (c) Reversed segmentation (d) Failed segmentation 1 (e) Failed segmentation 2

Fig. 5.6 Visualization of the Matthews Correlation Coefficient (MCC) [167]: if (a) corresponds to the ground truth mask (white is the region that should be segmented and black the region left out), (b) represents the predicted segmentation mask (MCC value between -1 and 1), then (c) corresponds to a perfectly reversed segmentation mask (MCC value of -1), (d) and (e) are cases of failed segmentation.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5.14}$$

The MCC ranges from -1 to 1, having 0 as a neutral position. If MCC = 1, the segmentation is considered to be perfect. Conversely, if MCC = -1, the segmentation result is perfectly reversed (TP and TN are both equal to 0). If MCC = 0, then all predictions are one class or the either (in our case 'background' or 'skin'). Several cases are presented in Fig. 5.6, with their corresponding MCC score.

Table 5.1 Results of skin-pixel segmentation using the histogram-based back-projection approach.

MCC Score	RGB	nRGB	I1I2I3	L*a*b	YCbCr	HSI(H)	TSL(T)
User1	0.4784	0.4619	0.5418	0.4765	0.4765	0.4043	0.4189
User2	0.631	0.5402	0.4639	0.4532	0.4141	0.4353	0.4202
User3	0.471	0.3819	0.4669	0.4766	0.4304	0.4437	0.5103
All users	0.518	0.467	0.469	0.454	0.400	0.420	0.438

Table 5.2 Results of skin-pixel segmentation using the clustering-based approach.

MCC Score	RGB	nRGB	I1I2I3	L*a*b	YCbCr	HSI (H)	HSI (HS)	TSL (T)	TSL (TS)
User1	0.447	0.447	0.431	0.821	0.816	0.570	0.667	0.51	0.506
User2	0.599	0.599	0.589	0.773	0.813	0.526	0.668	0.425	0.425
User3	0.412	0.412	0.399	0.837	0.813	0.524	0.580	0.467	0.474
All users	0.486	0.486	0.473	0.810	0.814	0.540	0.638	0.467	0.468

5.1.7 Results

The two approaches described in Sections 5.1.4 and 5.1.5 are used to predict the skin pixels using the methodology presented in Section 5.1.6. This sub-section presents an evaluation of their performance.

Results obtained using the histogram-based back-projection approach are presented in Table 5.1. They are generally very poor, with the average MCC score being around 0.45 across most color spaces. The best performing color space for the adaptive histogram thresholding method is RGB, with 0.518 for the case training and testing on all 3 users. All other color spaces reported MCC scores bellow 0.5, the closest being I1I2I3 with 0.463. When only the images for user2 were used for training/testing, the obtained MCC score was 0.631, reported by the RGB color space.

Results using the clustering approach are presented in Table 5.2. The highest MCC score for the case using images from all three users was reported for YCbCr and L*a*b, which reported MCC scores of 0.814 and 0.810. The other color spaces reported MCC scores of around 0.5. It is interesting to note that L*a*b color space reported higher MCC scores than YCbCr did for user1 and user3. YCbCr consistently reported MCC scores greater than 0.8 across all users.

It is also worth mentioning how including the Saturation channel in HSI improved the clustering of skin pixels, the MCC score increasing by 0.060 for user3 and by 0.142 for user2.

5.1.8 Conclusion

Skin segmentation is a considerable challenge in an unconstrained environment, especially when the background contains colors that are similar to skin. Furthermore, it is difficult to accurately classify a pixel based solely on its chromaticity, as it is not a strong enough identifier. Changing the color space can help with this task, as well as including other channels of information such as saturation (as observed in Table 5.2), but results are limited [161].
The results in this Section were also presented as part of a conference paper at the IEEE International Conference on Consumer Electronics (ICCE) 2016 [168].

5.2 Proof of Concept Palmprint Recognition Pipeline - a Case Study

This Section presents a Proof of Concept implementation of a palmprint recognition pipeline using images from the PoC palmprint database described in Section 4.1 of Chapter 4. The challenges associated with each stage of the pipeline were instrumental in shaping the general direction of my PhD research.

Following the segmentation results obtained in Section 5.1, an approach that extracts the palmprint Region of Interest (ROI) was proposed for hand images acquired in an unconstrained environment. This approach relies on an iterative selection of Harris corners based on their strength, to determine the central location of the hand region. This process is explained in Section 5.2.2. For the feature extraction and comparison phase, SIFT descriptors [115] were considered. A proof of concept scenario is described in Section 5.2.6.

5.2.1 Initial Hand-Background Separation

The skin pixel segmentation strategy described in Section 5.1.2 would lead to many false positives detections when deployed in an unconstrained environment. If one would take into account all possible skin tones in all lighting conditions (with shadows and various temperature of light), the resulting skin-tone interval would be too broad, affecting its robustness. The better performing K-means based segmentation (described in Section 5.1.5) was used to provide an initial separation of the hand from the background. Taking as an example the image in Fig. 5.7, two clusters of pixels are obtained. Using the same approach described in Fig. 5.4, the cluster containing skin-pixels was determined.

(a) Hand image (b) Cluster 1 (c) Cluster 2

Fig. 5.7 K-means segmentation of hand pixels, where (a) displays the input hand image, (b) outlines the background pixels, contained in cluster 1, and (c) outlines the skin pixels, contained in cluster 2. The black regions in (a) and (b) represent pixels that are not part of those clusters.

5.2.2 Iterative Selection of Corners and ROI Extraction

Introduced by Chris Harris and Mike Stephens [60], the Harris corner detection has been established in the field of Computer Vision for many applications, such as: image alignment and stitching, image retrieval, object recognition, etc.

After having selected an initial cluster candidate for the hand region (segmentation need not be perfect), all Harris corners are detected. A filtering stage is considered for the detected, to remove outliers corners, as described in Algorithm 1 and depicted in Fig. 5.8a-5.8c

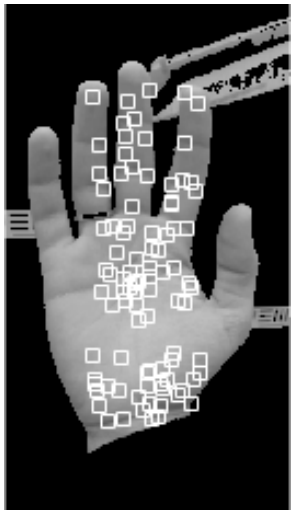
Once the cloud of corners determines a central position of the palm, a square region is defined. The square is centered on the coordinates obtained as being the center of the point cloud containing the filtered Harris Corners. The size of the square is considered to be 1/6th of the size of the image's larger side:

$$ROI_{side} = max(image\_height, image\_width) \tag{5.15}$$

5.2.3 Inspecting the ROIs

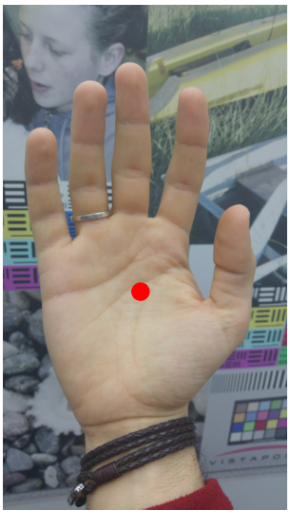
The resulting ROIs have been inspected visually and classified as being:

(a) Selected features, level1



(b) Selected features, level2

(c) Selected features, level3



(d) Detected center of palm

(e) ROI

Fig. 5.8 Stages undertaken for ROI extraction: a) Iterative selection of Harris corners corresponding to k-means based hand segmentation from Fig. 5.7; b) Center of the palm, computed as the arithmetic mean of remaining corners in (a) and its corresponding ROI.

Result: Corners satisfying the conditions
for *i=1* $\rightarrow$ *3 (i = number of stages)* **do**
  Determine average corner strength (*average_strength*) ;
  for *j=1* $\rightarrow$ *N (N = number of remaining corners)* **do**
    if $corner\_strength^{(j)} < average\_strength^{(i)}$ **then**
      Discard corner ;
    else
      Keep corner ;
    end
  end
end
$hand\_center(x,y) = \left(\frac{1}{N}\sum_{k=1}^{N} x_k, \frac{1}{N}\sum_{k=1}^{N} y_k\right)$

Algorithm 1: Iterative selection of Harris corners in hand region.

- 'successful', where most if not all of the palm region is extracted, with few or no background pixels being included
- 'ambiguous', where the extracted ROI contains half or more of the palm area but also includes a considerable number of background pixels
- 'clear failures', where either the ROI contains a small region of the palm area (less than half) and/or too many background pixels.

Two cases that result in successes are displayed in Fig. 5.9b and 5.9d. Two clear failures are included in Fig. 5.9f and 5.9h, where the ROI contains little of the hand's palmprint.

Fig. 5.10 displays several 'ambiguous' cases, where even though much of the palmprint surface is contained in the ROI, too much of the background is being included (Fig. 5.10a and 5.10b). The case where the scale of the hand within the image does not correspond to the predetermined ROI size (Equation 5.2.2), is displayed in Fig. 5.10c. This leads to the ROI containing the entire hand. A case where even though no background pixels are included, the ROI fails because of the reduced palmprint surface, as presented in Fig. 5.10c.

5.2.4 Limitations of Proposed ROI Extraction

A number of aspects regarding this approach should be mentioned before proceeding:

- The extraction algorithm does not include a rotation normalization step. However, this could be done after the extraction, based on a vertically aligned ROIs. Such an alignment was suggested by Zhao *et al* [169], who used SIFT key-points to apply an afine-transformation to the test palmprint ROI.

(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 5.9 (above) Successfully extracted palmprint ROIs, (below) Failed palmprint ROI extraction. a), c), e) and g) present the entire image with b), d), f) and h) their corresponding ROI.

- The extracted ROI is scaled relative to the image's size, not the hand's size. This can be a problem as it can include fewer palm pixels than background (as in Fig. 5.10d, or Fig. 5.10c).

- The approach is based on corners, which are assumed to be found more in the palm area. As a consequence the approach becomes sensitive to the segmentation map, or how centered the hand is within the image. If the segmentation is noisy (includes background regions with strong corners), there is a strong tendency to center the ROI to various parts of the background, like in Fig. 5.9h.

- The ROI extraction approach is difficult to evaluate. There are no ROI masks that can be used to evaluate the extraction, which forces a system based on visual estimation of the extraction.

(a) (b) (c) (d)

Fig. 5.10 ROIs that can be considered as being 'ambiguous' due to: a), b) too much background being included in the ROI, c) the ROI includes most of the input image, d) not enough surface of the palmprint being covered.

5.2.5 Considered Feature Extraction and Comparison

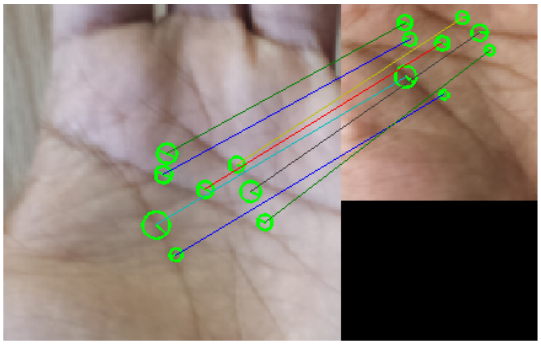
In order to address most of the shortcomings of the extraction algorithm, such as the absence of rotation and scale normalization, the features that were considered had to be robust to rotation and scale variation. Starting with the introduction of Scale Invariant Feature Transform (SIFT) [115], several such features invariant to rotation and scale have been developed, of which we mention Speeded Up Robust Features (SURF) [157] and Oriented fast and Rotated BRIEF (ORB) [117]. This category of feature extraction approaches was described in Section 3.3.1 of Chapter 3.

For the purpose of this Proof of Concept pipeline, SIFT descriptors were used to compare a query with several stored samples. The first step is to extract SIFT descriptors from two images. These feature descriptors are then compared based on a k-d tree algorithm [170] which determines the nearest neighbor. Sometimes the second-nearest match may be very close to the first, due to noise or other distortions. In this case, Lowe's rule is applied, where the ratio of closest-distance to second-closest-distance is taken. A match is accepted only if this ratio is less than 0.7. This helps reduce the number of false matches up to 90% while discarding 5% of the false non-matches [115]. Such a successful match is represented in Fig. 5.11 between two hand images from the same subject, but with different hand (and therefore ROI) orientation.

5.2.6 Proof of Concept - Comparison Trial

In this scenario three ROIs are considered as being stored during a 'training phase', where their extracted SIFT descriptors are stored locally for matching. A test ROI's extracted SIFT descriptors are matched with the set from the training phase.

(a) (b) (c) (d)



(e)

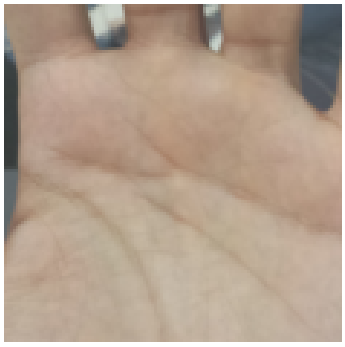
Fig. 5.11 a)-d) Hand images and their corresponding ROIs, e) ROIs matched using SIFT descriptors

The corresponding SIFT descriptors extracted from three ROIs are saved locally (Fig. 5.12c-5.12e) and matched with the descriptors extracted from two test ROIs (Fig. 5.12a and 5.12b). After comparison with an ROI sample from a different ID (Fig. 5.12a), samples Fig. 5.12c-5.12e resulted in 5, 5 and 6 inliers. When the samples in Fig. 5.12c-5.12e are compared with the ROI in Fig. 5.12a, the comparison results in 10, 24 and 20 inliers. By defining a threshold for the number of inliers, one is able to determine if a test ROI belongs to the same class or not. Furthermore, if there are several images stored for the same class, a voting scheme can be developed, where the majority vote is considered as being the final decision.

5.2.7 Conclusion

A PoC solution for cross-device authentication is presented, with initial results based on a limited database acquired in unconstrained conditions. A novel palmprint ROI extraction is proposed, based on Harris corner detection. The contents of this Section were part of a magazine article published in the *IEEE Consumer Electronics Magazine* with the title "Palm

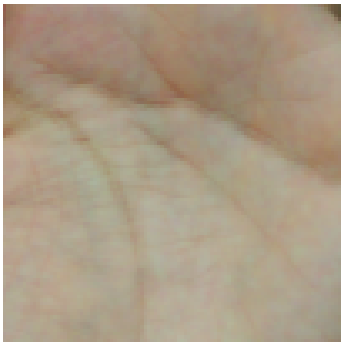
(a) 92 SIFT descriptors detected



(b) 66 SIFT descriptors detected

(c) 38 SIFT descriptors detected

(d) 53 SIFT descriptors detected



(e) 45 SIFT descriptors detected

Fig. 5.12 ROI used in a Proof of Concept comparison scenario. The ROIs in b), c), d), and e) correspond to the same ID, whereas the user in a) belongs to a different ID.

print as a smartphone biometric: another option for digital privacy and security", vol. 5, no. 3, pp. 71-78, 2016.

5.3 Chapter Conclusions

The work presented in this chapter focused on the pre-processing stages of a biometric recognition pipeline using palmprints acquired in unconstrained conditions. As such, the presented results show that the color-based segmentation of the hand in unconstrained conditions is not sufficient on its own. Texture and shape should be both taken into consideration, either by using an AAM or CNN such as autoencoders. However, these solutions introduce their own limitations, especially the need for large training databases.
A third approach would use hand pose regression strategies, thus by-passing the requirement for accurate hand segmentation. Such a solution is proposed in Chapter7 of this thesis, based

on a large database of images using an acquisition setup that allowed the automatic labeling of hand images. This database is introduced in Section 4.3 of Chapter 4.

Chapter 6

Unconstrained Palmprints - Baseline Experiments

This chapter focuses on the feature extraction and feature comparison/classification stages of the biometric recognition pipeline, as presented in Fig. 6.1.

Section 6.1 gives a general context for the experiments carried out based on the database of images NUIG_Palm1. Section 6.2 presents an overview of the experiments carried out and the methodology developed for each scenario that is investigated (e.g. information regarding the ROI extraction, pre-processing, classification approaches and feature extraction techniques). Section 6.3 reports the obtained results in the considered training/testing scenarios. Section 6.4 presents conclusions and directions for future work.

The material in this chapter was originally published in the *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp.334-342, August 2017 as " Unconstrained palmprint as a smartphone biometric". It has been modified to integrate with the rest of the thesis.

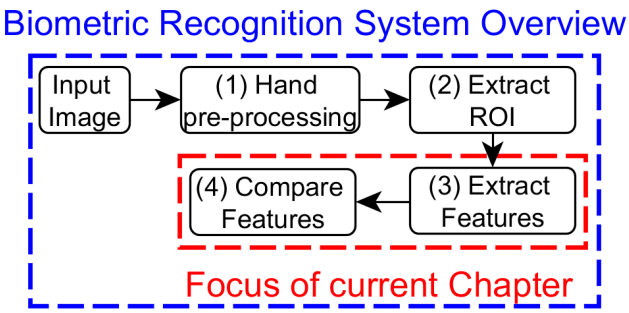


Fig. 6.1 Generic overview of a hand-based biometric recognition system using optical sensors, with the focus of the current chapter.

6.1 General Context

Palmprint recognition is a well established field, an overview of existing literature being presented in Chapter 3. With the introduction of NUIG_Palm1 database (for a detailed description, the reader is referred to Section 4.2 of Chapter 4), the problem of cross-device palmprint recognition in unconstrained conditions was considered for the first time. Furthermore, it is important to underline the restrictions imposed to any algorithm that is expected to run on an embedded platform: (i) computation load, (ii) potential to be run on an embedded chip and (iii) fast inference time. In the year 2017 (time of publication of this work) the tools for embedding CNNs onto embedded chip-sets [171] was in its infancy. This is the main reason why no solutions based on CNNs were included in the following experiments.

The main objective of the experiments presented in this chapter is a Proof of Concept (PoC) regarding the use of unconstrained palmprints for cross-device authentication. These experiments therefore represent a set of 'baseline recognition results', to be used for reference in future papers. Several state of the art feature extraction techniques are implemented and evaluated with various classifiers which are easily embeddable on a chip.

A series of training/testing strategies are considered for experimentation, making use of the properties of NUIG_Palm1 database. Every hand in the database was acquired with 5 devices in 4 scenarios (corresponding to lighting conditions: 'indoor normal' and 'indoor dark', as well as background: 'complex scene' and 'wooden texture'), leading to a total of 20 images per hand class. Based on this distribution of images, we define two distinct training strategies (also represented in Fig 6.2b):

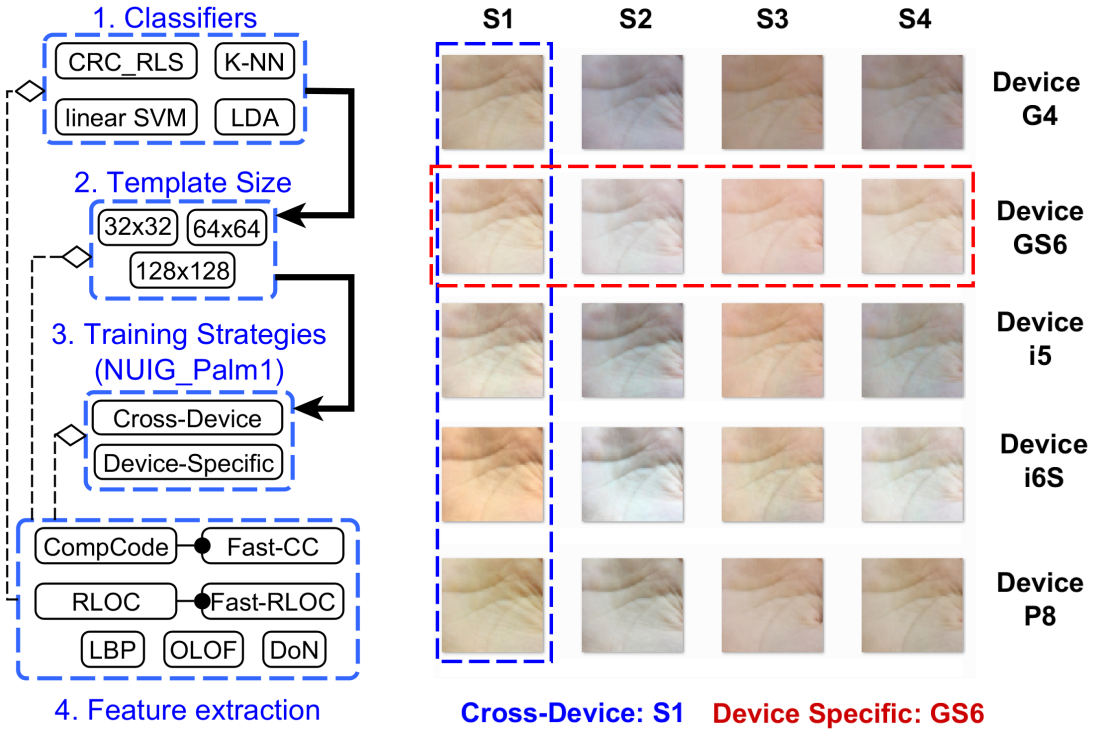
- Cross-device (CD), where the training set contains ROIs from several devices acquired in one scenario (from four possible lighting-background combinations)
- Device-specific (DS), where the training set contains ROIs from a single device in several lighting-background conditions.

6.2 Experimental Methodology

6.2.1 Overview of Experiments

This section provides an outline of the considered experiments. They are organized in four groups, as represented in graphical form in Fig. 6.2a:

1. What is the preferred classifier, from a collection of classifiers, based on Cross-Device training strategy (Scenario 1) (Section 6.2.1)



(a) Experimental workflow

(b) Training strategies

Fig. 6.2 (a) Succession of items being investigated: classifiers, palmprint sample size and training strategies. The performance of feature extraction approaches is included throughout the entire experimental setup; (b) a visual representation of the Cross-Device and Device-Specific training scenarios for a hand class from NUIG_Palm1 database.

2. What is the optimum size for the extracted palmprint ROI palmprint samples, where the size of the ROI is varied throughout experiments, from (32x32) to (64x64) and (128x128) pixels. For sub-sampling the bicubic interpolation is being used.

3. What is the preferred strategy for training:

   (a) Cross-device (Section 6.2.1)

   (b) Device-specific (Section 6.2.1)

4. Considering several feature extraction approaches (found in the literature), which one provides the best results. These are used throughout all experimental combinations (Section 6.2.1).

A. Classification Strategy Evaluation

This set of experiments is designed to compare several classifiers, such as Collaborative Representation Classifier with Regularized Least Squares (CRC_RLS), Support Vector Machines (linear kernel), K-Nearest Neighborhood (number of neighbors being equal to the number of training samples for each class) classifier and Fisher Discriminant Analysis (LDA) [172]. The scikit-learn [13] Python library was used to perform the recognition experiments for linear SVM, LDA and KNN, with scripts released on Github [1]. The CRC_RLS classifier was implemented in Matlab, with scripts being available in the same Github repository. As mentioned in Section 6.2.1 the training strategy used for these experiments corresponds to Cross-Device. Setup 1 (S1) was used, with 'indoor normal' lighting conditions, having as background the poster with complex scenes.

In an effort to reduce the space occupied by reporting results (in Section 6.3), only the optimum sample size was mentioned. Otherwise, the total number of experiments considered would have been (11 feature $\times$ 4 classifiers $\times$ 3 sample sizes) which equals to 132. In this way only 44 results are reported.

B. Cross-device Training (CD_Train)

The classifier reporting the best results during the evaluation of classifiers was used for this set of experiments. All the images that are part of a lighting/background setup are used as the training set one at a time (k-fold cross validation). For instance, images captured in cluttered background under 'indoor normal' lighting condition (setup S1) with all five devices make up the training set of images (total of 405 images), while the rest make up the testing set (total of 1,215 images).

The training set is then changed to the other lighting/background setups (finishing with Setup S4) and results are averaged, giving a better perspective of the performances and challenges of the database.

C. Device-specific Training (DS_Train)

The training set for each class is using all the images from one device at a time. This results in training with 4 palmprint images for each class, thus covering both lighting conditions (total of 324 images). The remaining 16 images from the other devices for each class are used as query (probe) images (total of 1296 images). The sample sizes used are the ones which yielded the best results in the CD_Train experiment.

[1] Scripts used for experiments are available at "github.com/AdrianUng/unconstrained-palmprint-recognition"

The results from this scenario should be compared with expectations made for every device depending on their imaging capabilities, provided in Table 4.4 (in Section 4.2 of Chapter 4). Based on these parameters we expect the order of performance (the first being the best performing and the last being the least performing) for the classifier trained with images originating from one device at a time to be the following: GS6, G4, i6S, P8, and finally i5. While devices G4, GS6, i6S, and P8 benefit from optical stabilization, i5 does not. Further, the image stabilization technology differs, G4 and GS6 having stabilization on 3 axis, while i6S and P8 have stabilization on 1 axis.

D. Feature Extraction Investigation

Several feature extraction methods are employed in the baseline experiments to obtain a diversified set of discriminative features.

Methods encoding the orientation at pixel level were considered (OLOF [92], CompCode [89], RLOC [58] and their Fast-implementation [99]) as well as two texture descriptors (LBP [91] and DoN [84]) were included in the experimental phase.

Fast-CompCode, Fast-RLOC and DoN correspond to current state of the art results reported on other palmprint databases (HKPU [17] and IITD [18]). CompCode, RLOC and OLOF are included thanks to their status as fundamental feature extraction techniques for palmprint recognition. They have withstood the test of time, reflected in their usage as benchmark techniques for any novel feature extraction (e.g. [71], [84], [99], [125]). Following is a formal definition of these feature extraction techniques.

6.2.2 ROI Extraction and Images' Pre-processing

Along with the images, a set of manually placed key-points were included in the NUIG_Palm1 database. These 5 key-points mark the finger bases for the outer fingers (index and pinky) and the middle of the finger-valley for the two central fingers (middle and ring fingers).
If we denote the first two points as X_1, X_2 and the last two points as X_4 and X_5, then the middle of these segments are represented by X_{12} and X_{45}. These two landmarks are then used to create a reference withing the image, allowing the hand's normalization of in-plane rotation, as demonstrated in Fig. 6.3, where the extracted ROI is contained within the black square.

In order to improve the contrast of the extracted ROI samples, the Contrast-limited Adaptive Histogram Equalization (CLAHE) [173] algorithm was applied to every palmprint sample before the feature extraction phase. This process (visualized in Fig. 6.4) increased the contrast and emphasized the palm lines, thus improving the recognition rates.

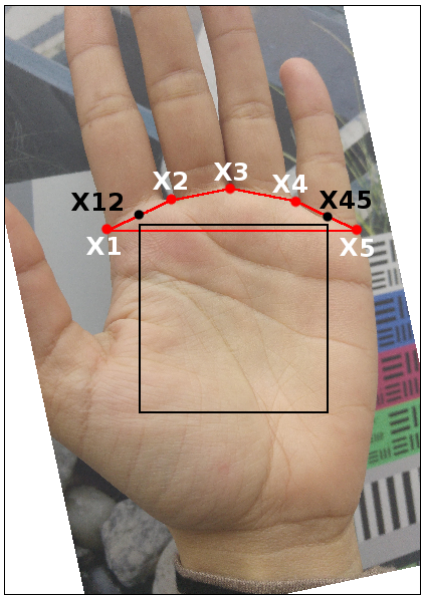


Fig. 6.3 Visualization of the palmprint ROI extraction using the provided key-points in NUIG_Palm1. The hand image's rotation is normalized using the X_{12} and X_{45} key-points, after which a square is defined, containing the ROI.

6.2.3 Feature Extraction Approaches used in Experiments

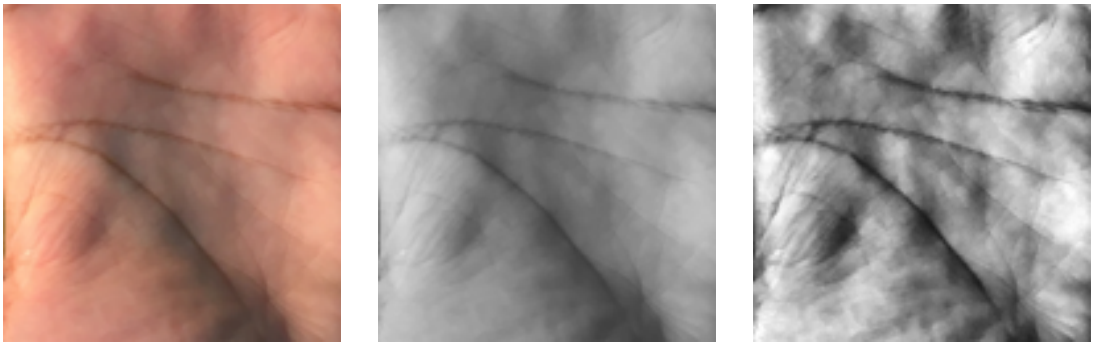
A. Competitive Code (CompCode)

Competitive Code (CompCode) is one of the fundamental algorithms used for palmprint feature extraction. Kong *et al.* [89] make the assumption that every pixel in the palmprint ROI sample is found on a line. A bank of Gabor filters is defined, as in Equation 6.1 (represented in Fig. 6.5 (a)-(f)).

$$\psi(x,y,\omega,\theta) = \frac{\omega}{\sqrt{2\pi}\kappa} e^{-\frac{\omega^2}{8\kappa^2}(4x'^2+y'^2)} \left(e^{i\omega x'} - e^{-\frac{\kappa^2}{2}} \right) \tag{6.1}$$

where $x' = (x-x_0)\cos\theta + (y-y_0)\sin\theta$, $y' = (x-x_0)\sin\theta + (y-y_0)\cos\theta$; (x_0,y_0) is the center of the function; θ is the radial frequency in radians per unit length and θ is the orientation of the Gabor functions in radians. The κ is defined by $\kappa = \sqrt{2\ln 2}\left(\frac{2^\delta+1}{2^\delta-1}\right)$, where δ is the half-amplitude bandwidth of the frequency response, which is between 1 and 1.5. When σ and δ are fixed, ω can be derived from $\omega = \kappa/\sigma$.
The input ROI sample *I(x,y)* is then convolved with the bank of Gabor filters as in Equation 6.2:

$$C_i = I(x,y) * \psi_R(x,y,\omega,\theta_i), \quad i = 1,\ldots,6 \tag{6.2}$$



(a) RGB image (b) Grayscale image (c) CLAHE improved

Fig. 6.4 Improving the contrast in ROI images: (a) ROI represented in RGB, (b) ROI converted to Grayscale, (c) Contrast-improved ROI using CLAHE [173].

where *(x,y)* are the image's pixels, ψ_R is the real part of 2D-Gabor filter response with radial frequency ω and orientation θ_i. The dominant line orientation is determined with a competitive rule, in this case being arg_{min} (as in Equation 6.3) because the palm lines are associated to dark pixels:

$$C = arg\ min_i(C_i), \tag{6.3}$$

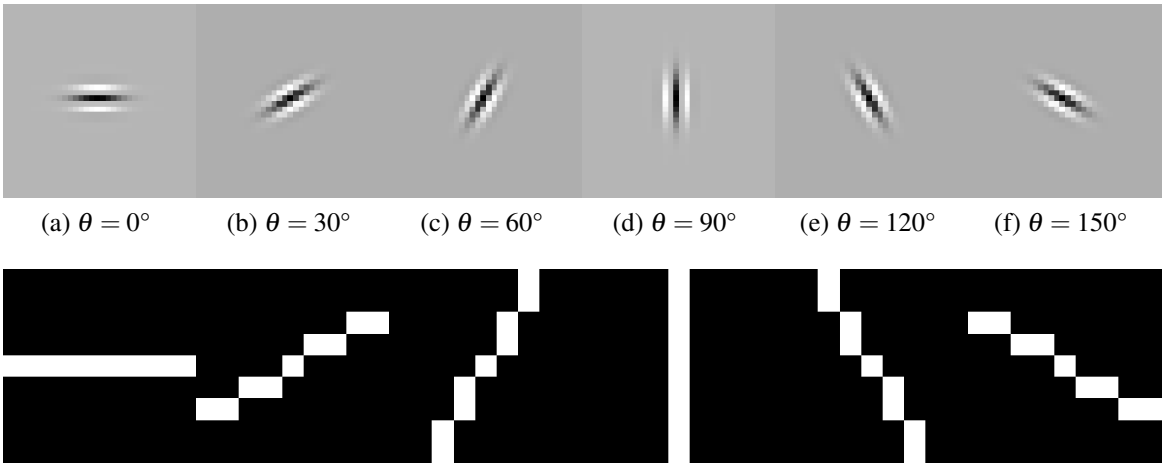
At every pixel's location is stored the class of the lowest recorded filter response, thereby obtaining a feature map as presented in Fig. 6.8b.

B. Robust Line Orientation Code (RLOC)

Robust Line Orientation Code (RLOC) [58] is defined as a modified finite Radon transform, interpreted as a summation of image pixels describing a line in a 9x9. These filters are represented in Fig. 6.5 (g)-(l). To extract RLOC, filters with 6 orientations (as in CompCode) are convolved with the input ROI image using Equation 6.2, followed by a pixel-wise competitive rule defined in Equation 6.3 applied to the filter responses. An example feature map is presented in Fig. 6.8c.

C. Fast-CompC, Fast-RLOC

A fast implementations of CompCode (Fast-CompC) and RLOC (Fast-RLOC) was defined in [99] and reported better overall results than their original implementations. By reducing the number of filter responses being used, from 6 to 2, a more discriminative feature is obtained. The selected two orientations need to be orthogonal , therefore there are 3 pairs of orientation (when considering 6 orientations to define the bank of filters) for CompCode and RLOC



(a) $\theta = 0°$ (b) $\theta = 30°$ (c) $\theta = 60°$ (d) $\theta = 90°$ (e) $\theta = 120°$ (f) $\theta = 150°$

(g) $\theta = 0°$ (h) $\theta = 30°$ (i) $\theta = 60°$ (j) $\theta = 90°$ (k) $\theta = 120°$ (l) $\theta = 150°$

Fig. 6.5 Filters used for convolution to obtain Compcode [89] and RLOC [58]: (a)-(f) Gabor filters, (g)-(l) MFRAT filters. White positions in the 9x9 MFRAT filters correspond to the pixels included in the summation process.

defined as:

$$C_{F,i} = arg\ min_i(C_i, C_{i+3}),\ i = 1, ..., 3, \tag{6.4}$$

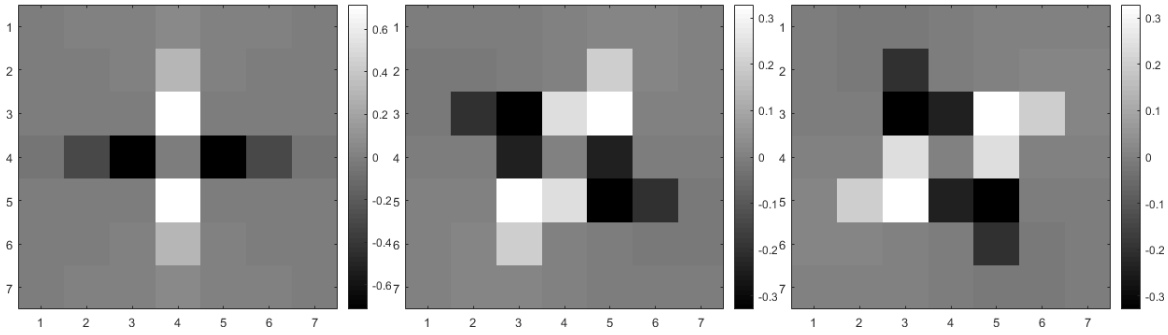
where $C_{F,i}$ represents the extracted RLOC and CompCode filter responses. According to the authors, the pair of orthogonal orientations chosen for the fast implementation does not matter, but some differences were noted in the experimental part. As a notation, every fast implementation of CompCode and RLOC contains the pair number used for that feature extraction, as mentioned in Equation 6.4. Corresponding pairs of Fast-CompC and Fast-RLOC feature maps are presented in Fig. 6.8d-6.8f and Fig. 6.8g-6.8i respectively.

D. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) [91] is a simple yet powerful texture operator labeling the pixels of an image by thresholding the neighborhood of each pixel, considering the result as a binary number:

$$LBP_P = \sum_{p=0}^{P-1} (g_p - g_c)2^p,\ s(x) = \begin{cases} 1, if x \geq 0 \\ 0, otherwise \end{cases}, \tag{6.5}$$

where P is the size of the kernel used (3x3) and p is the index of the neighboring pixels in the region, whereas c represents the value of the central pixel in the kernel. An example LBP feature map is presented in Fig. 6.8j.



(a) OLOF pair 1 (b) OLOF pair 2 (c) OLOF pair 3

Fig. 6.6 OLOF filters used to convolve with the palmprint ROI.

E. Othogonal Line Oriented Features (OLOF)

Orthogonal Line Orthogonal Features (OLOF) [92] is based on 2D Gaussian filters to obtain a weighted average of line-like regions:

$$f(x,y,\theta) = exp\left[\left(\frac{xcos\theta + ysin\theta}{\delta_x}\right)^2 - \left(\frac{-xsin\theta + ycos\theta}{\delta_y}\right)^2\right], \tag{6.6}$$

where θ are the orientation of the filter, δ_x and δ_y denote the horizontal and vertical scale parameters. Throughout experiments δ_x and δ_y were set as 1.8 and 0.5. An orthogonal filter is defined with:

$$OF(\theta) = f(x,y,\theta) - f(x,y,\theta + \frac{\pi}{2}) \tag{6.7}$$

Three ordinal filters ($OF(0)$, $OF(\frac{\pi}{6})$ and $OF(\frac{\pi}{3})$) are convolved with the ROI sample. Three bit codes are then determined based on the sign of filter responses. They are represented in Fig. 6.6. An example OLOF feature map is presented in 6.8k.

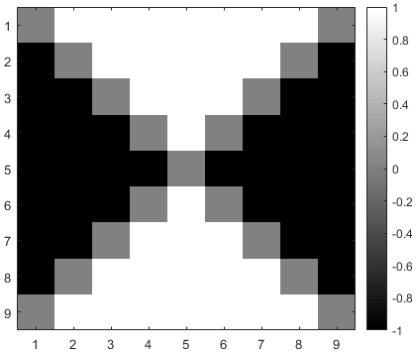


Fig. 6.7 DoN filter used to recover the 3D shape information from the palmprint ROI.

F. Difference of Vertex Normal Vectors (DoN)

Difference of Vertex Normal Vectors (DoN) [84] represents a 3D feature descriptor recovered from a 2D image. Each point/pixel on the image plane is corresponding to a vertex on the palmprint surface. For every point having two neighboring regions R_i^1 and R_i^2, its DoN feature is computed with:

$$DoN(i) = \tau\left(\sum_{j\in R_i^1} z_j - \sum_{j\in R_i^2} z_j\right),\ \tau(\alpha) = \begin{cases} 0, & \alpha < 0 \\ 1, & \alpha \geq 0 \end{cases} \tag{6.8}$$

Practically, in order to construct the feature extractor a filter needs to be constructed using:

$$f_{i,j} = \begin{cases} 1, & if|i| > |j| \\ -1, & if|i| < |j| \\ 0, & otherwise \end{cases} \tag{6.9}$$

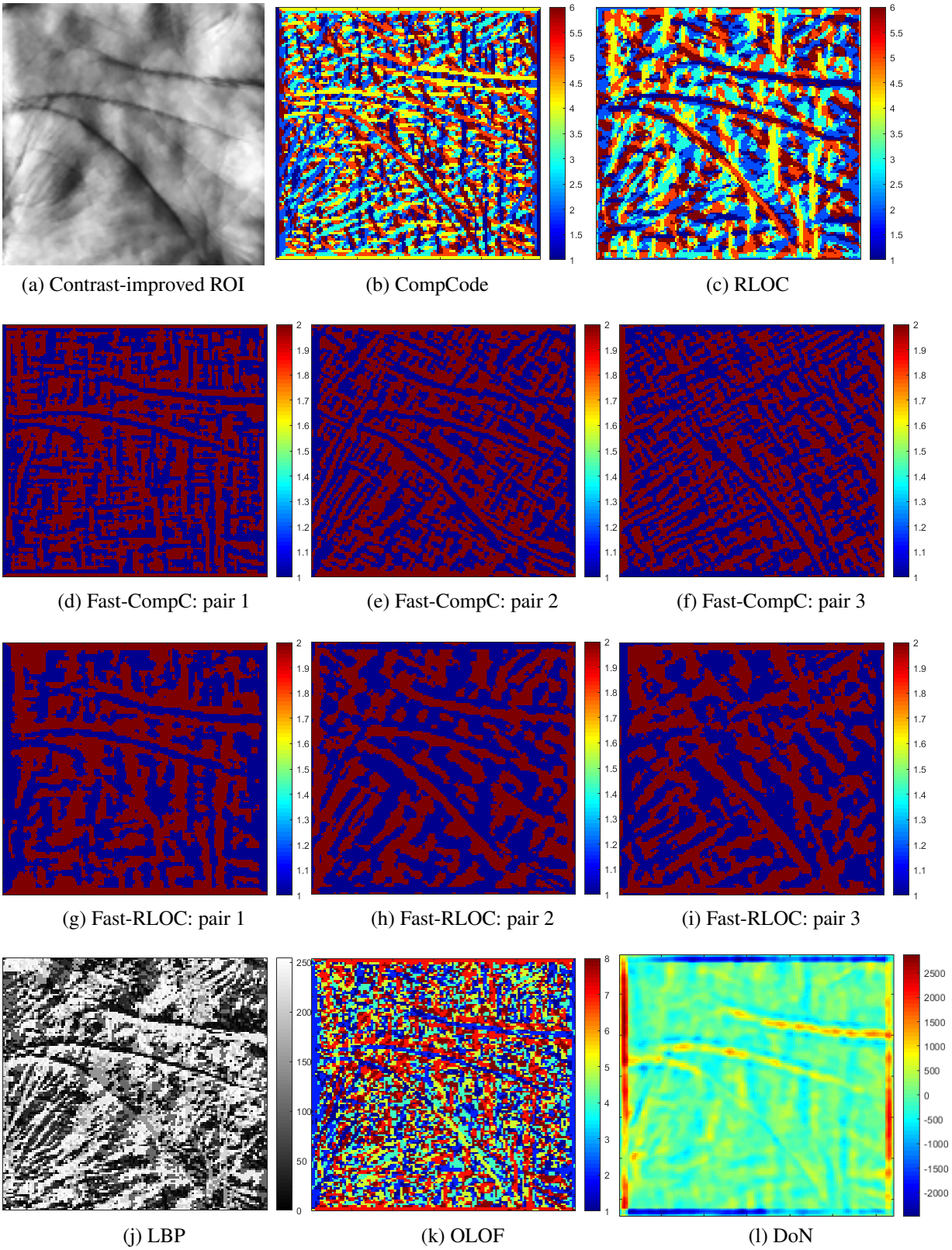
where i, j are the indexes $i, j \in [-B, B]$. The filter size is $(2B+1,\ 2B+1)$. In order to obtain the final feature image, the ROI sample is convolved with the filter defined in Equation 6.9. The filter is also represented in Fig. 6.7. An example feature map of the DoN feature maps used during experiments is presented in Fig. 6.8l.

6.2.4 Classification Techniques

A. Support Vector Machine with Linear Kernel

The Support Vector Machine (SVM) was first developed by Cortes and Vapnik [174] for pattern recognition and function regression. Given an identically and independently distributed training example set $\{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$, where $\boldsymbol{x} \in R^N$, y $\in \{-1, 1\}$. The kernel function can map the training examples in input space into a feature space such that the mapped training examples are linearly separable. In order to have a better classification result, we maximize the margin C of separation between patterns. The problem can be converted to maximize the following dual optimization problem:

$$W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{6.10}$$



(a) Contrast-improved ROI (b) CompCode (c) RLOC

(d) Fast-CompC: pair 1 (e) Fast-CompC: pair 2 (f) Fast-CompC: pair 3

(g) Fast-RLOC: pair 1 (h) Fast-RLOC: pair 2 (i) Fast-RLOC: pair 3

(j) LBP (k) OLOF (l) DoN

Fig. 6.8 Visual representations of feature extraction techniques considered for classification.

subject to:

$$\begin{gathered} \sum_{i=1}^{n} \alpha_i y_i = 0, \\ for \;\; \alpha_i \in [0,C] \;\; and \;\; i \in [0,n] \end{gathered} \tag{6.11}$$

The decision function becomes:

$$\begin{gathered} f(x) = sign\left(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + \mathbf{b}\right) \\ b = y_r - \sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_r, y_r) \end{gathered} \tag{6.12}$$

where $(\mathbf{x}_r, y_r)$ is any training example and K is the linear kernel being used.
The two class approach is extended to multi-class by implementing 'one-vs-the-rest' multi-class strategy, thus training k models, according to the total number of classes.

B. Fisher's Discriminant Analysis

Fisher's Linear Discriminant Analysis (LDA) ([175] p.106) searches for the projection of a database such that the ratio between the inter-class variance and the intra-class variance is minimized.

The LDA classifier [176] is based on the Bayes classification rule, where for a given vector x, it is assigned to the class c_k when the following inequality is satisfied:

$$p(c_k|x) > p(c_j|x) \;\; for\, all \;\; k \neq j \tag{6.13}$$

These posterior probabilities can not be directly measured, but can be derived from estimates of the priori probabilities and the class distribution according to the Bayes formula:

$$p(c_k|x) = \frac{p(c_k)p(x|c_k)}{p(x)} \tag{6.14}$$

where $p(x|c_k)$ is the probability density function for the vector within k class, $p(c_k)$ is the prior probability for class k and usually assumed to be equal for all classes, $p(x)$ is the probability density function of the input space and is also a constant over all the classes. Then the decision rule referred in Equation 6.13 is simplified to:

$$p(x|c_k) > p(x|c_j) \;\; for\, all \;\; k \neq j \tag{6.15}$$

The probability density functions for all the classes are assumed to follow a multivariate Gaussian distribution:

$$p(x|c_k) = \frac{1}{\sqrt{(2\pi)^f det(C)}} exp(-\frac{1}{2}(x-\mu_k)^T C^{-1}(x-\mu_k)) \tag{6.16}$$

where x is the vector to be classified, f is the dimension of the vector, C is the common covariance matrix of all the classes, and μ_k is the mean value of class k.

LDA is included as a reference for classical pattern recognition solutions to classification.

C. K-Nearest Neighborhood

The Nearest Neighborhood (K-NN) classifier represents a classic approach to classification [177]. Given a function $\rho : \chi \times \chi \to \mathbb{R}$, we determine the distance between two elements of the domain χ. For instance, if $\chi = \mathbb{R}^d$, then ρ can be the Euclidean (L2) distance:

$$\rho(x,x') = \|x-x'\| = \sqrt{\sum_{i=1}^{d}(x_i - x'_i)^2} \tag{6.17}$$

Let $S = \{(x_1,y_1),\ ...\ ,(x_m,y_m)\}$ be a sequence of training examples. For each $x \in \chi$, let $\pi_1(x),\ ...\ ,\pi_m(x)$ be a reordering of $\{1\ ,\ ...\ ,\ m\}$ according to their distance to x, $\rho(x,x_i)$. That is, for all $i < m$:

$$\rho(x,x_{\pi_i(x)}) \leq \rho(x,x_{\pi_{i+1}}(x)). \tag{6.18}$$

When $K > 1$, the label y with the most votes is considered:

$$y_{\pi_i}(x) :\ i \leq K \tag{6.19}$$

The K-NN classifier continues to be used thanks to its low complexity [127].

D. CRC_RLS Classification Strategy

As palmprint recognition is a small sample size classification problem, a collaborative representation classifier with regularized least squares (CRC-RLS) is used to determine the identity of the query image [178], [128]. Let $D = [D_1,\ D_2,...,\ D_k]$ be the dictionary which denotes the training palmprint images of the k subjects (available in the database). $D_i = [v_{i,n_1},\ v_{i,n_2},...,\ v_{i,n_i}]$ are the training palm images of the i^{th} person and the total number of training palm images of the same person is denoted as n_i. A query palmprint image y can

be collaboratively coded over the dictionary D [178] by using:

$$a = Xy, \tag{6.20}$$

where

$$X = (D^T D + \gamma \cdot I_{mat})^{-1} D^T, \tag{6.21}$$

where γ is the regularization parameter and I_{mat} is the identity matrix. The regularized residual ε_i for each subject i in the dictionary D over this coding scheme is calculated using:

$$\varepsilon_i = \frac{\|y - D_i \cdot a_i\|_2}{\|a_i\|_2} \tag{6.22}$$

From the regularized residuals, the identity of the query image y can be calculated with:

$$Identitiy(y) = arg\ min(\varepsilon_i) \tag{6.23}$$

This specific choice of classification scheme is employed based on not only its success rate in face recognition, gender classification, palmprint recognition etc., but also because it is up to 1,600 times faster than the state of the art sparse representation based classifiers [178]. Such a high performing, computationally light classifier may be widely adopted in resource constrained consumer devices such as smartphones.

6.2.5 Resources Used

The tools used to carry out the experiments in this chapter include:

- Intel Core i7-2600 CPU @ 3.40 GHz
- 16 Gb of RAM
- Matlab 2013a
- scikit-learn 0.20.1 for Python 2.7
- CompCode and OLOF implementation (Matlab) used by Khan *et al.* [54] [1]
- RLOC, Fast-CompCode, Fast-RLOC and DoN implementation (Matlab), along with the Collaborative Representation Classifier, available as a Github repository [2].

[1] Implementation of CompCode, ContCode and OLOF available to download at Zohaib Khan's website "https://sites.google.com/site/zohaibnet/Home/codes". The archive in question is named **PP_ContCode.zip**.

[2] Codes available at the Github repository "https://github.com/AdrianUng/unconstrained-palmprint-recognition".

6.3 Results

6.3.1 Classification strategy evaluation (Cross-Device)

Several traditional classifiers were evaluated under the same conditions as the cross-device training setup described in CD_Train using for training the images in Scenario 1 ('indoor lighting', from all devices). This leads to 5 images per hand class being used for training (405 in total) and 15 for testing (1215 in total). The sample size used was the one which reported the best results for each feature extraction technique in Table 6.2.
The performance of classifier-feature extraction combinations are presented in Table 6.1 in the form of Equal Error Rate (EER). The EER is commonly used to compare biometric recognition systems, as it denotes the point of intersection between the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). The lower the EER, the better a system is considered.

Table 6.1 Equal Error Rates (EER) (%) for CD_Train training using various classifiers. The overall best result is colored in green, with best results for each classifier colored with yellow. The lowest obtained EER for every feature extraction technique is bolded.

Feature extraction \ Classifier	Sample size	Linear SVM	K-NN (N=4)	LDA	CRC_RLS
CompCode	32x32	**13.66**	21.74	18.60	14.09
Fast-CompC1		14.57	17.18	17.67	**14.38**
Fast-CompC2		12.23	14.42	12.59	**12.06**
Fast-CompC3		**12.51**	14.68	13.90	12.65
RLOC	64x64	10.86	13.23	13.79	**10.82**
Fast-RLOC1		13.20	13.23	17.26	**13.03**
Fast-RLOC2		11.07	27.07	14.62	**10.98**
Fast-RLOC3		11.19	12.52	15.23	**10.90**
LBP	32x32	13.17	**12.98**	16.15	13.13
OLOF	32x32	13.09	14.23	14.90	**12.67**
DoN	64x64	12.31	20.82	14.22	**10.08**
Average	-	12.53	16.55	15.36	**12.25**

The best performance of the SVM with linear kernel is reported when using RLOC (10.86%). The behavior of this classifier is very similar to that of CRC_RLS. The best performance of the KNN with N=4 is reported when using Fast-RLOC3 (12.52%). Similarly, the LDA classifier obtains its best result (12.59%) when using a fast implementation (Fast-CompC2).

The best EER is reported for the proposed CRC_RLS classifier, with 10.08% being obtained when using the DoN feature. The next best result is reported by the SVM with linear kernel when using RLOC as feature extraction. Overall the CRC_RLS classifier reports the

lowest EERs regardless of the feature extraction technique used, from the 11 considered approaches obtaining the minimum (feature-wise) in 8 cases. Considering the performance reported for these classifiers, the CRC_RLS is selected for the rest of the experiments.

6.3.2 Cross-device Training (CD_Train)

The CD_Train results are presented in Table 6.2, indicating that the CRC_RLS classifier is fairly robust to alignment errors, considering the comparison is done at pixel level. Please note the fast implementations of CompCode and RLOC contain 'Fast' in their name and an indication of the pair used for their computation, as detailed in Equation 6.4.

Table 6.2 CD_Train averaged Recognition Rates (RR) and averaged Equal Error Rates (EER). The two best results are colored in green and yellow, for every feature extraction technique the best result is indicated by the bolded numbers.

Feature Extraction \ Sample size	32x32		64x64		128x128	
	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)
CompCode	**67.03**	**14.03**	69.81	15.69	56.68	21.98
Fast-CompC1	**66.64**	**14.60**	69.15	16.63	54.59	25.23
Fast-CompC2	**73.12**	**12.31**	71.31	15.68	55.72	25.62
Fast-CompC3	**70.74**	**13.11**	69.07	16.94	51.37	27.90
RLOC	70.57	12.20	77.06	10.88	71.46	15.37
Fast-RLOC1	67.32	13.50	**74.54**	**13.10**	67.73	18.29
Fast-RLOC2	69.44	12.31	**75.61**	**10.99**	60.47	15.82
Fast-RLOC3	70.06	12.58	**76.81**	**11.67**	69.79	17.04
LBP	**72.71**	**13.61**	72.59	14.89	68.37	15.96
OLOF	**73.02**	**12.94**	70.47	16.34	56.58	24.37
DoN	77.67	9.49	**79.87**	**10.06**	71.77	15.36

The best result is achieved by the DoN feature with 79.87% RR and 10.06% EER, closely followed by Fast-RLOC3 and Fast-RLOC2, at around 77% RR and 11% EER. Both scenarios used ROI samples of size 64x64.
When using ROI samples of size 32x32 the best performance is reported by Fast-CompC2 with 73.12% RR and 12.31% EER. Similar performance is obtained with OLOF (73.02% RR and 12.94% EER) and LBP (72.71% RR and 13.61% EER).

It is interesting to note how the recognition performance decreases as the size of the ROI sample increases, a powerful sub-sampling (from the initial size of the ROI) compensates slight misalignment. The lowest recognition results are obtained when using ROI sample of size 128x128, with Fast-CompC1 (51.37% RR and 27.9% EER). In this case the best results are reported by DoN (71.77% RR and 15.36% EER) and RLOC (71.46% RR and 15.37% EER).

The results reflect the appropriate size of the ROI sample for each feature extraction technique. It is important to note that features like CompCode, LBP or OLOF are strongly affected by ROI misalignment, which explains why the best results are reported for ROI of size (32x32) pixels.

6.3.3 Device-specific Training (DS_Train)

Receiver Operating Characteristic (ROC) curves for the DS_Train recognition experiments are provided in Fig. 6.9, whereas the RR and EER are presented in Table 6.3.

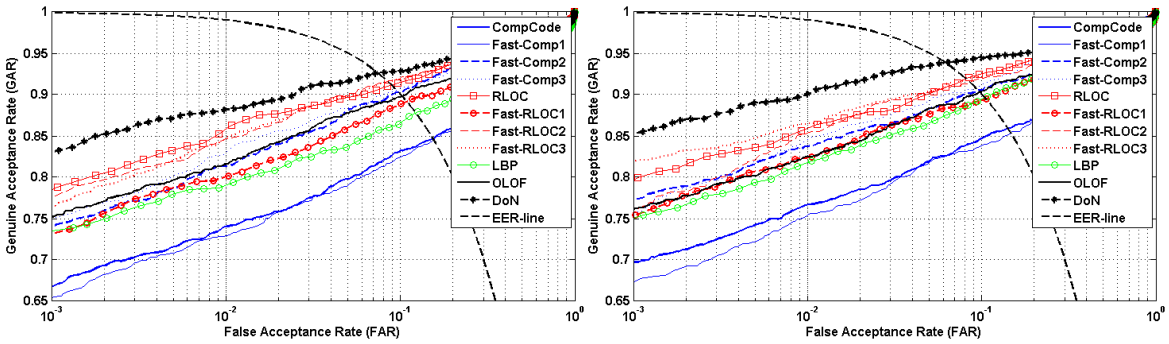
Table 6.3 DS_Train RR (%) and EER (%) results. Overall best results are marked with green, with best results for every device being colored with yellow. Best results reported for every feature are bolded.

Feature extraction	Sample size	G4		GS6		i5		i6S		P8		Average	
		RR	EER	RR	EER	RR	EER	RR	EER	RR	EER	RR	EER
Compcode	64x64	71.06	15.24	**74.31**	**14.15**	72.45	14.54	69.98	16.68	72.83	14.83	72.12	15.08
Fast-Comp1	64x64	72.07	15.56	72.76	14.56	**72.38**	**14.53**	70.83	16.22	71.6	15.41	71.92	15.25
Fast-Comp2	32x32	**79.86**	**9.44**	80.86	9.51	80.09	9.86	77.85	11.26	79.86	9.36	79.70	9.88
Fast-Comp3	32x32	**79.32**	**9.89**	81.02	9.21	77.93	9.77	78.39	10.09	79.93	9.39	79.31	9.66
RLOC	64x64	82.56	8.87	**83.02**	**7.98**	81.94	8.65	81.17	9.88	81.79	8.74	82.09	8.82
Fast-RLOC1	64x64	77.62	10.99	79.94	10.63	**79.32**	**10.28**	76.69	11.94	78.62	10.54	78.43	10.87
Fast-RLOC2	64x64	81.64	8.38	81.48	8.63	**82.95**	**7.91**	80.01	9.41	80.86	8.74	81.38	8.61
Fast-RLOC3	64x64	81.79	8.52	**84.03**	**8.14**	81.02	9.42	80.86	10.00	81.79	9.09	81.89	9.03
LBP	32x32	76.77	12.29	**79.55**	**10.21**	77.62	11.23	77.31	12.86	78.47	11.27	77.94	11.56
OLOF	32x32	79.32	10.14	**80.09**	**9.73**	78.86	10.14	76.31	11.05	79.47	10.08	78.80	10.22
DoN	64x64	86.34	7.40	87.81	6.10	85.96	7.63	84.41	8.41	85.8	7.46	**86.06**	**7.40**

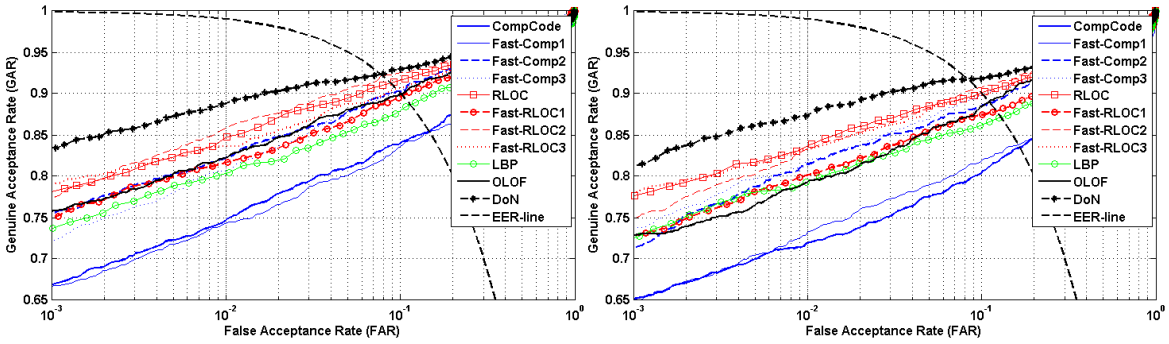
The overall best result is reported for DoN when trained with images from GS6, with 87.81% RR and 6.10% EER. The next best result corresponds to training with images from G4, obtaining 86.34% RR and 7.40% EER.
The DoN provides the overall best results, regardless of the training device (the average RR and EER being 86.06% and 7.40%.) The next best results are obtained using Fast-RLOC2 in the i5 scenario, with 82.95% RR and 7.91% EER.

The lowest result is reported for Compcode training with images from i6S, with 69.98% RR and 16.68% EER. Similar performance is reported for Fast-CompC1, also with training images from i6S, with 70.83% RR and 16.22% EER. This observation is also backed by the ROC plots in Fig. 6.9, where the curves corresponding to CompCode and Fast-CompC1 report the lowest values.

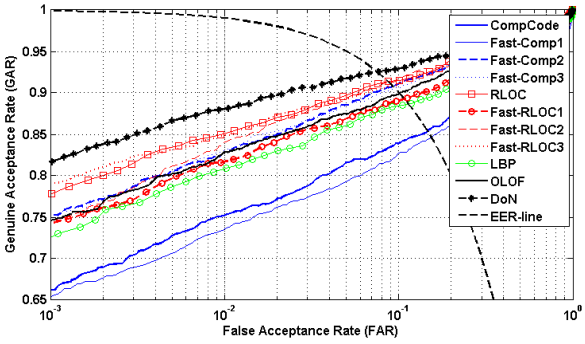
The lowest results are consistently provided by i6S, with the highest EER reported for every feature extraction technique (Table 6.3). This is also reflected in the ROC plots in Fig. 6.9d. This behavior is not correlated with the sensor's low resolution, but with the



(a) Train set:G4 (b) Train set:GS6



(c) Train set:i5 (d) Train set:i6S



(e) Train set:P8

Fig. 6.9 ROCs from DS_Train scenario, using for training only images from one device at a time.

behavior of the **Auto mode** in terms of exposure time and ISO (used during acquisition). In the choice between motion blur and lower Signal to Noise Ratio (SNR), i6S overestimates its capabilities and increases the exposure time to unacceptably long values for real use (e.g. 0.25s). As it can be noted from Fig. 6.9c, i5 (which represents an older model of the same family of devices) follows a similar trend. On the other hand, the key traits which favor GS6's performance are the bigger sensor, better resolution and the overall optical design.

If we are to consider the value of 10^{-3} for False Acceptance Rate (equivalent to having 1 falsely accepted image for every 1,000 comparisons) as benchmark for a biometric recognition system, then DoN gives the best result in DS_Train, with a Genuine Acceptance Rate (GAR) of 85.50% when training with images acquired with GS6. The GAR This is an acceptable performance for authentication on smartphones for device unlocking, or in conjunction with other forms of authentication (used as a soft biometric).

6.4 Conclusions

There is an ongoing effort to perform recognition of biometric characteristics across several devices and in unconstrained conditions, an especially difficult scenario, but with high applicability in day-to-day commercial (online) activities. Biometric characteristics in this category include face [179], iris [180] or fingerprint [181] recognition. A challenging aspect of these approaches lies in training with images from one device and comparing with biometric samples acquired with a different device. Previously, Jia *et al.* [10] performed such recognition scenarios across three consumer devices using palmprint recognition, but in very constrained conditions.
This work adds palmprint recognition to the literature on unconstrained biometric recognition across consumer devices.

This chapter demonstrated the feasibility of user authentication on smartphones based on palmprint biometrics. The unconstrained nature of images from the palmprint database NUIG_Palm1 (introduced in Section 4.2 of Chapter 4) has proven to be a considerable challenge. Recognition scenarios were designed according to the attributes of NUIG_Palm1 (cross-device and device-specific).

With an average Recognition Rate (RR) of 86.06% and average Equal Error Rate (EER) of 7.4%, the Collaborative Representation Classifier via Regularized Least Squares (CRC_RLS) classifier using Difference of Vertex Normal Vectors (DoN) features provided the best results (device-specific case), making it a candidate for a solution using the palmprint as a biometric characteristic for authentication on/using consumer devices.

The fact that results are better in the device-specific case than in the cross-device one suggests that training the classifier with images from several lighting scenarios enhances the performance as compared to training with images from several devices. In real life case, this can be managed by routinely updating the training set stored in the consumer device or central database, thus reducing the probability for the system to encounter an image acquired under considerably different conditions when compared to the stored samples.
Secondly, the camera sensor size influences the recognition rate in noisy conditions, devices having a larger sensor (1.26" vs. 1.32") performed better than smaller sensors. Having a larger sensor increases the amount of light passing through the lens, therefore reducing the overall noise levels. We can expect a similar result with other biometric recognition scenarios that rely on a camera for acquisition.
Thirdly, the palmprint ROI extraction stage revealed the absence of an algorithm that could be robust to background, as well as hand pose.

Future work aims to develop an approach for unconstrained ROI extraction of the hand images. Furthermore, misalignment mitigation techniques should be explored, to compensate for inaccurate ROI extraction.

Chapter 7

Unconstrained Palmprint ROI Extraction

This chapter addresses one of the key stages required for robust and reliable palmprint authentication, specifically the ROI extraction. This step is important as it will influence the comparison trial stage.
The chapter is organized as follows: Section 7.1 presents an overview of the context and a short literature review, Section 7.2 describes the proposed architecture, Section 7.3 introduces resources used throughout experiments - databases, neural networks and the tools used. Section 7.4 outlines the adopted experimental methodology for the tasks evaluated. An overview of experiments is presented at the end of Section 7.4. Section 7.5 introduces the obtained results, with Sections 7.6 and 7.7 discussing them.

7.1 Related Work

The related work for this chapter spans a number of different research fields, including:

- palmprint region of interest extraction, where the main objective is to extract the palmprint region and use it to discriminate between individuals' identity (discussed in Section 3.2 of Chapter 3).
- applications using the Leap Motion sensor which has been used for static or dynamic hand gesture recognition (the reader is referred to Section 4.3.2 of Chapter 4.)
- hand pose estimation, where the location of the hands' joints are determined. In the most recent literature CNN based techniques have shown the most promise and our review will focus primarily on CNN based techniques.

7.1.1 Hand Pose Estimation - a Literature Review

Hand pose estimation is defined as the effort of estimating the positions of a hand's joints using either 2D or 3D coordinates (relative to the camera). Considerable efforts have been made using depth-based systems [182], but more recently the vision-based approaches (visible range) have made considerable progress, especially thanks to advancements in body pose estimation. The two tasks are different in very few ways, mainly in the variability of the labeled joints. Bodies tend to be upright, with arm and leg movements taking place in specific patterns. However, strong articulations, small or barely visible joints and occlusions are the main challenges that are faced. Comparatively, hands face more potential orientations and fingers easily become obstructed from view, although the context does provide some cues regarding the potential location of fingers/joints.
Both hand and body pose are specific tasks of the general problem of pose detection, which means that advancements in one area will improve the other as well. It is common to have an algorithm performing well on one pose detection task to be deployed on other pose detection tasks as well. Typical examples include the OpenPose project [183], which is a library that supports body, hand and also head pose detection on still images.

Toshev *et al.* introduce DeepPose, a solution to human pose detection using CNNs. Their approach involves several stages of refinement for every (x, y) prediction of joints (elbow, wrist, etc.) using a sliding window.

Tomson *et al.* [184] combine a CNN architecture with conditional probabilities derived from pairs of joints to create a unified model. They use the predictions provided by the CNN (local model) for body parts and make a *higher-level spatial model* based on the probability of correctly predicted pairs of joints. The spatial model is not designed to improve the predictions that are already correct, but to penalize potential false positives. Tomson *et al.* report better results on body pose databases when compared to DeepPose.

Wei *et al.* [185] extend the work on Pose Machines [186] and introduce Convolutional Pose Machines (CPM) to progressively determine the probability maps for body pose landmarks in RGB images. They initialize predictions with a deep neural network and further refine them with a 2nd input which is connected to the main architecture. Several residual stages of convolutions lead to progressively lower losses, Wei *et al.* recommending 6 stages. The probability maps (outputs) of the network represent masks that are 8 times smaller than the input image, in the original paper being of (368x368) pixels.

Newell *et al.* [187] connect residual information from several scales, obtaining a network they entitle Stacked Hourglass Network (SHGN), after the successive up-scaling and down-scaling of layers. Compared to the unified model described by Tomson [184], SHGN does not use probabilities, sufficient contextual information being derived from this change in

scale. Predictions of joints are iteratively improved using intermediate supervision after each hourglass module. 2 modules are recommended for an optimum balance between accuracy and resources used. For practical reasons Newell *et al.* used probability maps (output) that are 4 times smaller than the input image, in the original paper being (256x256) pixels. Furthermore, *L2* is used as the loss function during training. Reported results surpass the ones provided by CPM when using 6 stages.

Mueller *et al.* [147] introduce a database of hands to help with 2D hand pose estimation. The images were generated using a Generative Adversarial Network (GAN) entitled *GeoConGAN*. Mueller *et al.* train a Resnet [188] architecture with both 2D and 3D joint positions, making the hand pose estimator especially robust in the context of hands that are partially covered. Their GANerated Hands database provides both sets of images.

Zimmerman *et al.* [189] estimate the 3D coordinates of a hand's joints from a single RGB image. The pipeline is made up of 3 different CNNs. They suggest a CNN which segments and crops the hand from the image called *HandSegNet*, a 2nd that estimates the pose (joints) called *PoseNet*, followed by a 3rd network called *PosePrior* that outputs the 3D structure of the hand in relative, normalized 3D coordinates in a canonical frame.The architecture for PoseNet takes from CPM the 1st stage and its weights (any common layers are initialized with weights learned from body pose regression [185]), followed by 2 stages containing five Convolutional layers. Similar to CPM, the 3 prediction stages output heatmaps for corresponding joint position are 8 times smaller than the input image of (256x256) pixels.

Gomez-Donoso *et al.* [190], [191] use a CNN-based approach to detect/segment hands from a frame and perform regression for every joint. Faster R-CNN [70] was used for segmentation and ResNet50 [188] was used for landmark regression. The used images were initially acquired with a setup of 4 webcams located at different angles, from which the finger-joint labels were read with a Leap Motion device. The authors claim an average error of 10 pixels per hand image. However, Gomez-Donoso *et al.* do not provide a comparison of their approach with other methods for hand-pose regression, nor scale the distances between predicted and ground truth data.

7.2 Proposed CNN Architecture

The previously used approaches to the ROI extraction/segmentation in the context of palmprint recognition have already been discussed in Chapter **3.2**. The proposed CNN for unconstrained ROI extraction attempts to combine elements from two established CNNs - InceptionV3 [192] and Segnet basic [193]. The hypothesis is that the Segnet encoder is able to provide contextually rich information regarding the hand's location/pose, which when

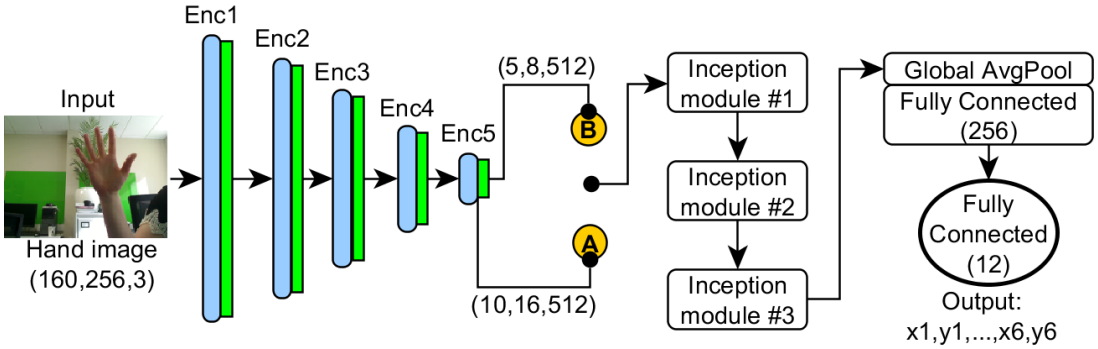


Fig. 7.1 Diagram with the architecture of the proposed network Segnet-Inception network. Enc1, Enc2, etc. stand for the encoding stages of Segnet. Output is the set of 6 points (x_i,y_i) describing the palmprint key-points with (x,y) coordinates. Points **A** and **B** represent the input to the Inception modules that are used, depending on the selected network architecture - **A** for Proposed network or **B** for altV7. A detailed outline of this architecture is given in Table 7.1

coupled with a chain of Inception modules form an improved architecture for point regression tasks.

Combining two architectures attempts to bring together two or more networks that either have the same function (e.g. classification, segmentation, etc.) or perform different functions that have some overlap in their learning (e.g. classification and segmentation). Therefore, examples from the former category are Inception-Resnet [194], which brings to the Inception modules residual information from previous layers, or SPDNN [195], which reinforces networks' weak points with other networks' strong points in a parallel manner, with their outputs being merged at the end. Bazrafkan *et al.* [196] considered both cases of segmentation as well as classification, having provided an implementation for the task of low-quality iris segmentation with 3 network architectures. In the second category there is also HyperFace [197] which combines the outputs from layers found at several depths in AlexNet [198]. The outputs are then concatenated and used as a starting point for multiple output branches of the network, each specialized for a given task.

The proposed architecture couples elements from two networks designed for different tasks: Inception for image classification (object detection), and SegNet for image segmentation. Starting with the Segnet architecture [193], the output of the last Convolutional layer (before MaxPooling) from the last Encoding stage (point **A**) is directed towards the chain of Inception Modules (IM), as represented in Fig. 7.1. The exact point of forking is better described in Table 7.1, detailing the succession of layers. The reason for choosing the output of this particular layer is the compactness of information it has encoded. The structure of IMs

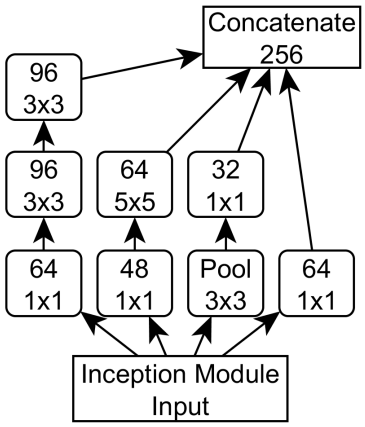


Fig. 7.2 Inception module architecture, as implemented in Keras platform [199] according to Szegedy *et al.* [192]. Every block (except Average Pool) represents a 2D convolution. The number in the upper half of the block is the number of filters, whereas the number in the lower half is the kernel size. The input is of shape (10,16,512) or (5,8,512)

follows the implementation described in [192] which uses (1x1) convolutions with 'ReLu' activation, as well as batch normalization, represented in Fig. 7.2.

The scenario in which the chain of IMs is connected to point **B** in Fig. 7.1 corresponds to the alternative implementation of the proposed network, having as input to the chain of IMs the output of the MaxPooling layer. This represents an even more compact representation of the input image compared to the fork in point **A**.
It is interesting to observe the behavior of these two networks and determine if the image representation in fork **A** aids the regression task, compared to the fork in **B**.

The loss function used for the regression of the selected points is the average Euclidean Distance (ED) for (x,y) coordinates, defined in Equation (7.1):

$$Loss = \frac{2}{N}\sum_{i=1}^{N/2}\sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2}\,, \tag{7.1}$$

where $i = 1,..,N/2$ and $N = 12$ total number of x or y coordinates, (x',y') representing the predicted coordinates.

7.3 Resources Used

This section contains information about the data used for training and testing the networks. It also contains information about the models evaluated in this paper and the methods used to evaluate them.

Table 7.1 Architecture of the proposed networks. For simplicity positions 20*, 21* and 22* were not fully reported. Please refer to Fig. 7.2 for the structure of the used Inception Modules.

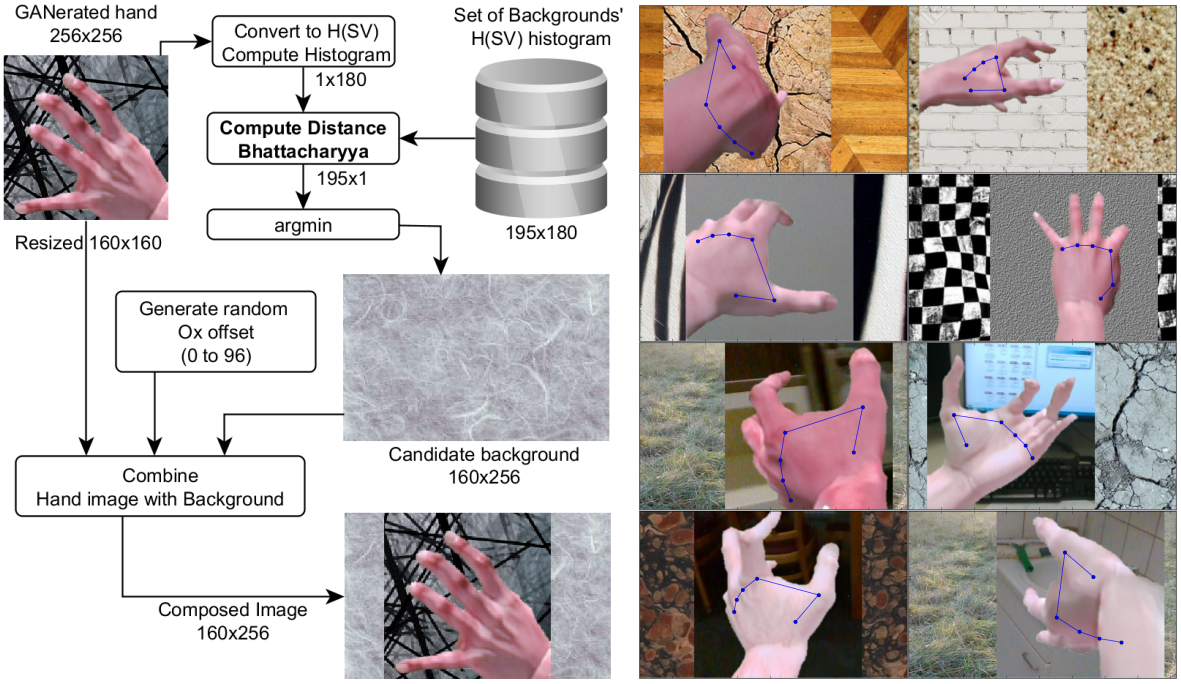
No.	Name	Kernel	Dimensionality
1	Input	-	160x256x3
2	Conv. + ReLU + BN	3x3	160x256x64
3	Conv. + ReLU + BN	3x3	160x256x64
4	MaxPool	2x2	80x128x64
5	Conv. + ReLu + BN	3x3	80x128x128
6	Conv. + ReLu + BN	3x3	80x128x128
7	MaxPool	2x2	80x128x64
8	Conv. + ReLu + BN	3x3	40x64x256
9	Conv. + ReLu + BN	3x3	40x64x256
10	Conv. + ReLu + BN	3x3	40x64x256
11	MaxPool	2x2	20x32x256
12	Conv. + ReLu + BN	3x3	20x32x512
13	Conv. + ReLu + BN	3x3	20x32x512
14	Conv. + ReLu + BN	3x3	20x32x512
15	MaxPool	2x2	10x16x512
16	Conv. + ReLu + BN	3x3	10x16x512
17	Conv. + ReLu + BN	3x3	10x16x512
18	Conv. + ReLu + BN *(fork point **A**)*	3x3	10x16x512
19	MaxPool *(fork point **B**)*	2x2	5x8x512
*	Inception Module x 3 (Fig. 7.2)	-	10x16x256 or 5x8x256
20	Global Average Pool	-	256
21	FC + ReLu	-	256
22	FC + ReLu	-	12x1

7.3.1 Databases Used

A. NUIG_Palm2

The first database used for training is the database collected with the Leap Motion device, entitled 'NUIG_Palm2' (NUIGP2), previously introduced in Section 4.3 of Chapter 4. A total of 18,303 labeled images were used for the training set and the remaining 1,103 were kept for testing. Because of the limitations of the 2D projection, 763 images were selected for evaluation purposes (for NUIGP2). The following augmentations were applied to the training set:

- vertical flip: images are being flipped along the width
- Gaussian blur: images are blurred with a kernel whose size is randomly chosen from 3 options - (8x8, 10x10, 14x14)



(a) Adapting GAN_H database (b) Sample images from

Fig. 7.3 (a) Algorithm for adapting the hand images from the GANerated Hands (GAN_H) database to our use-case. 195 texture images were selected from the database in [200], (b) Sample images from the adapted database. Even though all hand images are successfully paired with compatible backgrounds based on their color information, the semantic information of the backgrounds' contents is more relevant in the top two rows than the two bottom rows.

- motion blur: images are blurred with a linear motion blur whose kernel size and orientation angle are chosen randomly from (3x3, 5x5, 7x7) pixels and (0, 45,90, 135) degrees respectively. For this step the pyblur version 0.2.3 library was used.

- image stretching: the images are stretched along either the width or the length, with a random value - (0.03%, 0.05% or 0.075%). After stretching, the images are cropped so they have the initial shape.

- rotation: the images are rotated with a randomly chosen number of degrees selected from the set (7,-5,-3,-2,2,3,5,7).

B. GANerated Hands

The 2nd database is made up of synthetic hands introduced by Mueller *et al.* [147], entitled GANerated Hands (GAN_H). This database contains 2 folders: one with synthetic hands

that are holding objects, and another with synthetic hands only. For our setup only the folder without objects was considered. It contains 143,449 hand images which do not require the augmentation applied to NUIG_Palm2.
However, because the networks' input image size is (160,256,3), the images from GAN_H, of size(256,256,3) needed to be re-scaled to (160,160,3) and placed in a (160,256,3) frame. However, instead of leaving this difference (96 pixels) with either black or white pixels, we chose to use images from the database of images released by Cimpoi *et al.* entitled 'Textures in the Wild' [200] and combine the two sets to pair with the backgrounds of the hand images having similarly colored backgrounds. After converting the hand image to the HSV color space, the histogram of the Hue layer was computed using 180 bins. Then the lowest Battacharya distance between that histogram and the list of 195 histograms corresponding to all potential backgrounds was determined. The lowest distance was considered as an appropriate candidate, leading to composed images that appeared natural. A better illustration of this process is presented in Fig. 7.3.

$$D(H_{im}, H_{bgr}) = \sqrt{1 - \frac{1}{\sqrt{\bar{H_{im}}\bar{H_{bgr}}N_{bins}^2}} \sum^{N_{bins}} \sqrt{H_{im}H_{bgr}}}, \tag{7.2}$$

where N_{bins} is the number of bins used (180), H_{im} and H_{bgr} are the computed histograms of the hand image and the background respectively, with $\bar{H_{im}}$ and $\bar{H_{im}}$ being defined by Equation (7.3):

$$\bar{H_{im}} = \frac{1}{N_{bins}} \sum^{N_{bins}} H_{im} \tag{7.3}$$

C. NUIG_Palm1 and MOHI

NUIG_Palm1 (NUIGP1) is a recently introduced database of palmprints [19] acquired with smartphones in an unconstrained context. The hand pose was chosen by the subjects, against two types of backgrounds - one with real-world scenes and another with wooden background similar to skin color. This is currently the only database of palmprint images acquired in such conditions, as mentioned previously (in Chapter3). As described previously in Section 4.2 of Chapter 4, the acquired hands have many orientations in a variety of lighting conditions. The database currently contains 81 subjects, providing a total of 1,616 images.

MOHI [201] is a database of hands acquired for hand contour recognition in the context of smartphone imaging, but fit the use case defined for the problem of palmprint ROI extraction. Like palmprint databases, the RGB images contain hands with stretched fingers in a variety of

(a) NUIGP1 (b) MOHI

Fig. 7.4 Sample images from NUIGP1 and MOHI.

orientations, against a uniform/white background. The main reason for using this database is the means of acquisition which relied on a smartphone camera in several lighting conditions, as well as the large number of classes (200) spread across 3 sessions (total of 2,500 images).

D. BIOmix

To prove the usefulness of the proposed CNN described in Section 7.2, as well as the data in NUIG_Palm2, another database was made by combining images from all previous ly mentioned databases. The training set is composed of 50% of the NUIGP1 [19] (820 images from 41 subjects) and 50% of the MOHI [201] (1,500 images from 100 subjects) database, which along with 8,145 images from NUIG_Palm2 (from 7 subjects) are augmented with the process detailed earlier in Section 7.3.1. The training set is finally complemented by 50,000 images from the GANerated_Hands database, leading to a total number of 112,790 images. The remaining images from NUIGP1 (795 images from 40 subjects) and from MOHI (1,000 from 100 subjects) are used for testing the proposed network's performance for 2D point regression, as well as ROI extraction success rate.
This mixed database will hereon be mentioned in the current paper as BIOmix, to emphasize the combination of the two separate palmprint databases with the two hand pose databases.

Furthermore, the performance of both NUIG_Palm2 and GAN_H will be evaluated on the previously described test set (NUIGP1 and MOHI). For this comparison to be fair, the same (NUIGP1 and MOHI) images that are part of BIOmix are included in these cases as well. In the case of NUIGP2 all training images required the augmentation described in Section 7.3.1 whereas in the case of GAN_H only non-synthetic hand images required augmentation.

7.3.2 Tested Convolutional Neural Networks

The CNNs being compared are detailed in Table 7.2 and can be broadly classified into 5 groups:

1. Proposed network (forking point **A**), as described in Section 7.2

2. Alternative implementation (altV7) of proposed network (forking point **B**), as described in Section 7.2

3. Classification CNNs - InceptionV3, ResNet50 (previously used for keypoint regression in [202], [191] and [147])

4. CNNs for Palmprint RoI Extraction, previously used for this purpose by Bao [66] and Chatfield [68]

5. Pose regression CNNs using probability heatmaps for joints - Stacked Hourglass Network (SHGN) [187], Convolutional Pose Machines (CPM) [185], PoseNet [189].

There is a further branching in the aforementioned algorithms, considering that some include a pre-processing stage of segmenting and cropping the hand prior to pose regression. This approach was not stressed, because palms/hands are positioned close to the center of the image (as can be seen in Fig. 7.4). Furthermore, there is only one hand in every image that requires processing. Because the proposed network used the entire image as input, also SHGN, CPM, and Posenet were implemented in the same way, not using a sliding window.

Table 7.2 Architecture details of the CNNs being compared throughout the experimental phase.

Network name	Strategy	Output size	Total parameters	Number of float operations	Convolutional layers
Proposed	(x,y) coordinates	12x1	15,662,748	31,347,752	34
altV7			15,662,748	31,347,752	34
InceptionV3			22,858,028	45,836,556	95
Resnet50			36,177,292	72,433,801	53
Bao			18,354,108	36,703,500	4
Chatfield			60,981,964	121,947,614	5
SHGN	probability heatmaps	40x64x6	6,594,960	13,236,981	98
CPM		20x32x6	40,958,564	81,908,010	41
PoseNet			25,908,818	51,807,130	23

7.3.3 Tools Used

The following equipment and software was used to obtain the reported results:

- GPU: TitanX-Pascal 1080
- CUDA version: 9.0
- cuDNN version: 7.1.4
- tensorflow version: 1.8.0
- Keras version: 2.2.2
- OpenCV version: 3.4.2.16
- Github repositories used: SHGN source code[1] from Github.
- Scripting codes used for training/evaluation are released in a Github repository [1]

7.4 Experimental Methodology

A first analysis of the algorithms is provided regarding the 2D point regression problem, followed by a series of experiments centered on the ROI overlap when using ground-truth landmarks versus predicted ones. The images in NUIGP1 and MOHI were labeled using an in-house developed Python script[2] which was released on Github. These points are defined by 6 pairs of (x,y) coordinates representing the location of palmprint landmarks corresponding to the finger joint locations described in Section 4.3 of Chapter 4, where NUIG_Palm2 (NUIGP2) is introduced.
The trained networks are able to generalize the position of joints from the hands in the test set, allowing the ROI (Region of Interest) to be extracted.

7.4.1 2D-point Regression

The proposed network was trained on the databases detailed in Section 7.3.1, having its performance compared with existing approaches. The Euclidean Distance between the

[1] A fork of the used repository can be found at "https://github.com/AdrianUng/Stacked_Hourglass_Network_Keras". Forked from yuanyuanli85.

[1] Please visit "https://github.com/AdrianUng" for a full list of repositories used.

[2] The Python script used for labeling hand images can be accessed at: "https://github.com/AdrianUng/Label-image-with-points".

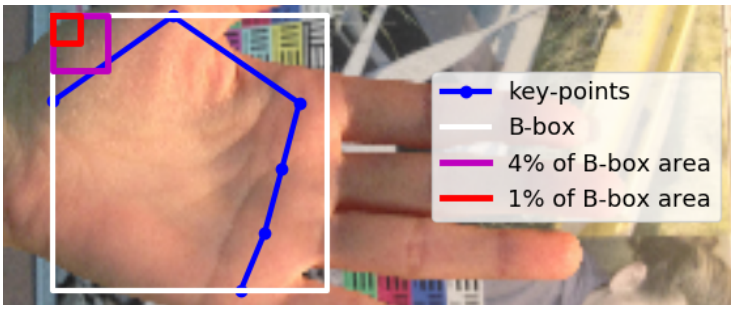


Fig. 7.5 Visualization of the PCK reference used during evaluation. The white square represents the area defined by the palmprint landmarks' bounding box (B-box). The magenta and red squares represent the area obtained using 20% and 10% of B-box's side. These correspond to 4% and 1% of B-box's surface.

predicted and ground truth pixels was used in order to evaluate the pixel error:

$$Error(P^{gt}_{(x,y)}, P^{pred}_{(x,y)}) = \frac{1}{2}\sqrt{(x^{gt} - x^{pred})^2 + (y^{gt} - y^{pred})^2} \tag{7.4}$$

where $P^{gt}_{(x,y)}, P^{pred}_{(x,y)}$ represent the instances of ground truth-prediction pairs being evaluated for a given point.

For error visualization, the Percentage of Correct Keypoint (PCK) is adopted. This metric is commonly used in pose estimation [185], [187], [184], [189] and is especially useful for the normalization of the prediction error relative to object's scale. PCK is defined as the normalization of prediction error relative to a reference size, which in our case is 10% (or 20%) of the palmprint bounding box side. This reference is defined as the side of the bounding box containing all 6 palmprint landmarks P1-P6. It is then possible to formally define PCK error and the PCK rate as in Equation 7.5:

$$\begin{aligned} PCK_{error}(P^{gt}_{(x,y)}, P^{pred}_{(x,y)}) &= \frac{Error(P^{gt}_{(x,y)}, P^{pred}_{(x,y)})}{reference} \\ PCK_{rate}(PCK^{Point\ j}_{error}) &= |PCK^{Point\ j}_{error} < thresh|, \end{aligned} \tag{7.5}$$

where the variable *thresh* ranges between 0 and 1 and is used to count the number of images for a given PCK error. PCK rates should be as high as possible (100%) and correspond to PCK errors as close to 0 as possible. Fig. 7.5 provides a visual aid regarding the thresholds used to compute the PCK rates throughout experiments.

Throughout this paper the word average refers to the arithmetic mean, either computed image-wise (across all points P1-P6) or globally (across images, for specific points).

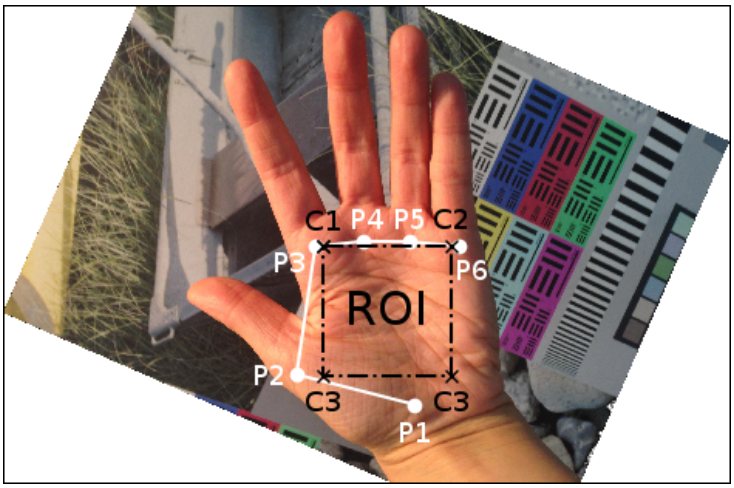


Fig. 7.6 Visualization of the ROI extraction procedure using an image from NUIGP1. Based on the predicted landmark points $P1,...,P6$ the hand's rotation is normalized and the ROI's side and center are computed according to the algorithm described in Section 7.4.2.

In the case of networks that rely on heatmaps to determine the position of the finger joints (SHGN, PCM and PoseNet), first the predicted heatmaps are converted to (x,y) coordinates and then compared to the ground truth positions. For every joint the maximum value from its probability distribution is used for the (x,y) positions.

In order to keep the Results section as compact and as practical as possible, we will include the PCK error corresponding to two representative points (P3 and P6). Although points P4 and P5 are used during the ROI extraction stage, their PCK error is part of an interval of error defined by points P3 and P6. For an image-wise overview of the PCK error, the average PCK is also included.

7.4.2 Palmprint ROI Extraction

As mentioned previously in Section 4.3 of Chapter 4, there are 6 landmarks $(P1,..,P6)$ available to describe the palmprint area. Initially the intermediate points $P34$ and $P56$ are determined and used to normalize the hand's rotation. If the difference of the x coordinates belonging to $P34$ and $P56$ is negative, an extra 180°is added to the rotation angle. Point P1 is used to further determine if the ROI is following the orientation of the hand in the image. Having rotated the image and palmprint landmarks as displayed in Fig. 7.6, the side of the ROI square is computed as being 90% of the distance between $P3$ and $P6$. Adding 1/2 of the ROI side to the y coordinate of the equivalent point $P36$ determines the ROI's center.
In order to validate the extraction of the ROI using predicted points for $P1,...,P6$ (ROI_{pred}), an overlap factor (Intersection over Union metric) with the ROI extracted with ground truth

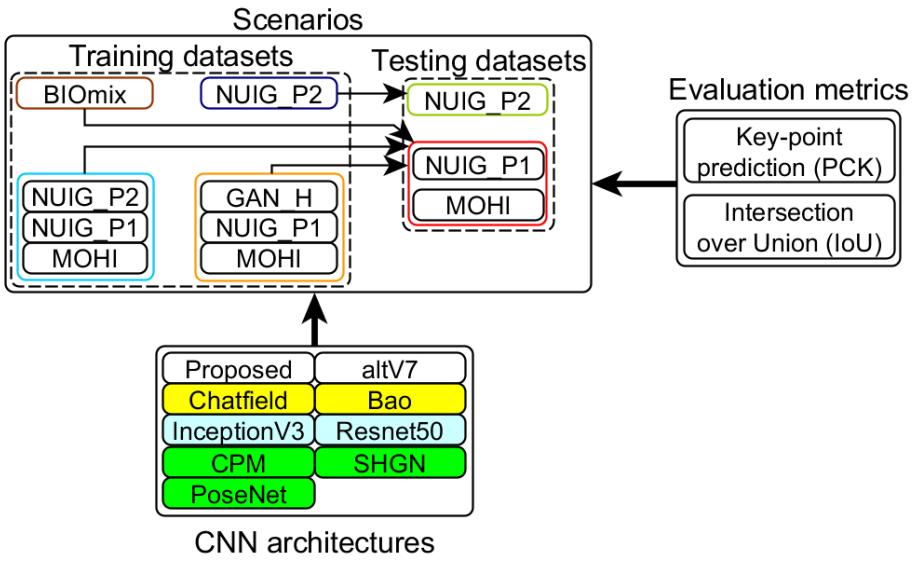


Fig. 7.7 Overview of experiments in a diagram form. The list of CNN architectures, along with the training/testing scenarios and the evaluation metrics are represented in this way to better describe the structure of experiments.

points (ROI_{gt}) is defined in Equation (7.6):

$$ROI_{IoU} = \frac{|ROI_{gt} \bigcap ROI_{pred}|}{|ROI_{gt} \bigcup ROI_{pred}|} \tag{7.6}$$

If the ROI_{IoU} factor (defined as an Intersection over Union (IoU)) is greater than 0.5, the ROI_{pred} is considered as being successful (correct). Otherwise it is counted as being a failed (incorrect) ROI extraction. The IoU is a commonly used metric in the field of object detection, where bounding boxes are traced around objects. The 0.5 threshold for of the ROI_{IoU} factor is empirically determined and is considered as the limit to which landmarks define ROIs sufficiently centered on the palm.

7.4.3 Overview of Experiments

In order to better keep track of the databases that are used for training and testing, depending on the evaluation task, a summary is provided in Fig. 7.7 and Table 7.3.

The first set of experiments deals with the main task of point regression and includes 4 distinct training/testing scenarios.

Scenario 1 trains and tests the networks in Table 7.2 on NUIGP2 database. Predicted keypoints from this scenario are presented and discussed based on specific visual examples,

in Section 7.5.1. Scenarios 2, 3.1 and 4.1 use for training images from NUIGP2, GAN_H and BIOmix respectively and are tested on NUIGP1 and MOHI.

The second set of experiments is centered around the ROI extraction and evaluates the ROI_{IoU} factor between the ROIs extracted using predicted points relative to the ROIs extracted using ground truth points. The networks in Table 7.2 are trained with BIOmix (scenario 3.2) and GAN_H (scenario 4.2) and tested on NUIGP1 and MOHI images.
Please note that networks in 3.1, 4.1 and 3.2, 4.2 are the same but the evaluation is focused on a different task (key-point regression for the former and ROI_{IoU} factor for the latter).

A discussion on the relationship between the training databases (BIOmix vs. GAN_H) and the two tasks (keypoint prediction and ROI extraction) is given in Section 7.6, where visual samples from Scenarios 3 and 4 are taken into consideration.

Table 7.3 Training/testing sets used in the experimental phase, alongside PCK reference used (10% or 20% of palmprint bounding box side).

No.	Task	Train database	Test database(s)	PCK reference
1	2D point regression	NUIGP2	NUIGP2 (same database)	20% of palm print b_{box}
2		NUIGP2	NUIGP1, MOHI (cross-database)	10% of palm print b_{box}
3.1		GAN_H		
4.1		BIOmix		
3.2	ROI extraction	BIOmix		
4.2		GAN_H		

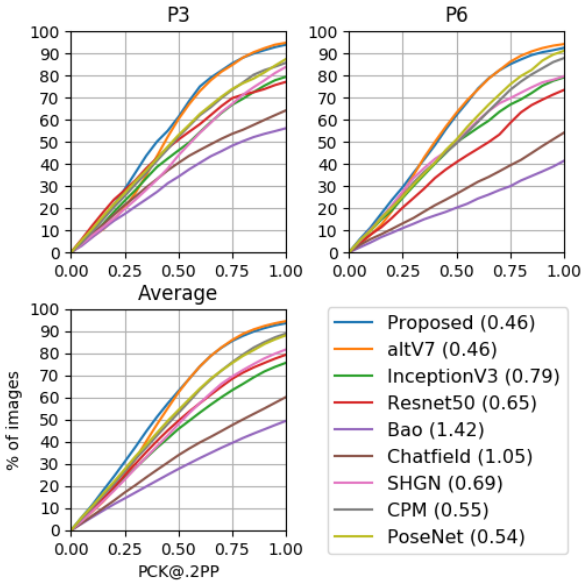
7.5 Results

7.5.1 2D Point Regression

Scenario 1. Train: NUIGP2, Test: NUIGP2

In this scenario all of the training images were sourced from NUIGP2 and networks were tested on 5 subjects that were not part of the training distribution. The test set contained 5,796 images in total.
PCK errors and rates (defined in Equation 7.5) for all scenarios in Table 7.3, for the networks summarized in Table 7.2, are presented in Fig. 7.8. It is important to mention that in order to compensate for labeling errors found in the training/testing datasets, a PCK reference of 20% of the bounding box containing all ground truth keypoints (as in Fig. 7.5) was considered, instead of 10%.

One can note that in general, the Proposed network, along with its alternative implementation (altV7) provide the best results. The PCK rates corresponding to a PCK error of 1.0 (the



(a) Test: NUIGP2

Fig. 7.8 Evaluation of palmprint point regression trained on the NUIGP2 database. The last plot represents the image-wise average PCK error (P1-P6). The average PCK (points P3-P6) is shown next to the network name. Best seen in color.

prediction error being scale-normalized as in Section 7.4.1) place the Proposed network and altV7 on the first position, with 94.50% of testing images, followed by PoseNet (87.95%), CPM (85.80%) and SHGN(84%). The worst performing network is Bao, successfully predicting the P3 keypoint in about 56% of images. Generally key-point P3 is considered to be the most accurately labeled during training phase.

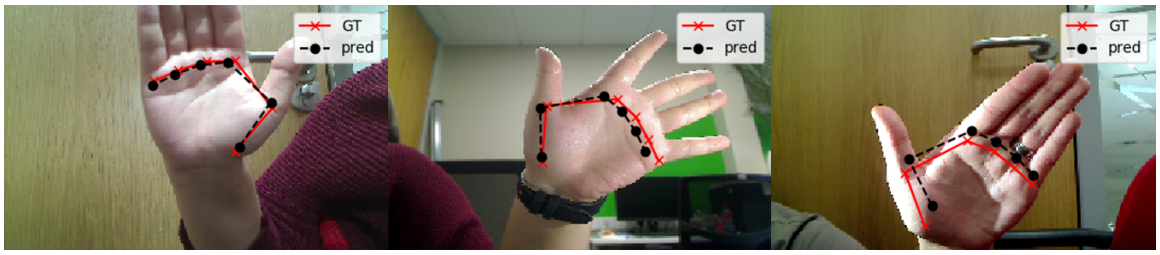
The situation is similar when analyzing the results for keypoint P6, with altV7 providing the highest PCK rate (94.25%) followed by the Proposed network (92%), PoseNet (91%) and CPM (88%). As in the case of keypoint P3, Bao_CNN had the lowest PCK rate, of about 41.5%.

Network performance is assessed in the 3rd graph of Fig. 7.8, displaying the PCK error averaged between all points P1-P6. For a scale-normalized prediction error of 1.0, the Proposed network and altV7 provide the highest PCK (about 94%), followed by the networks designed for pose detection: CPM and PoseNet (about 89%), and SHGN (82%), followed by ResNet50 (80%) and InceptionV3 (75%). The lowest PCK rates are reported for Chatfield (60%) and Bao_CNN (50%).

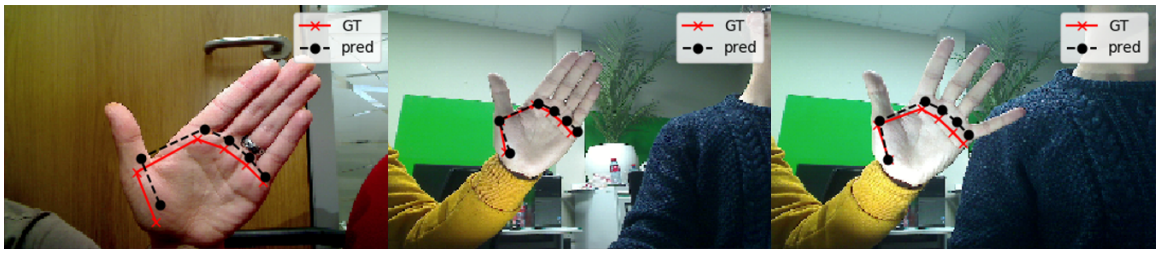
This separation of networks leads to a grouping into 3 categories, based on results: best performing (Propsed and altV7, CPM, and PoseNet), average performing (InceptionV3 and ResNet), and least performing (Chatfield and Bao_CNN).

Visual analysis of Scenario 1: NUIGP2

This subsection contains a discussion about the keypoint predictions for the proposed network, listing examples of both successful and unsuccessful cases.



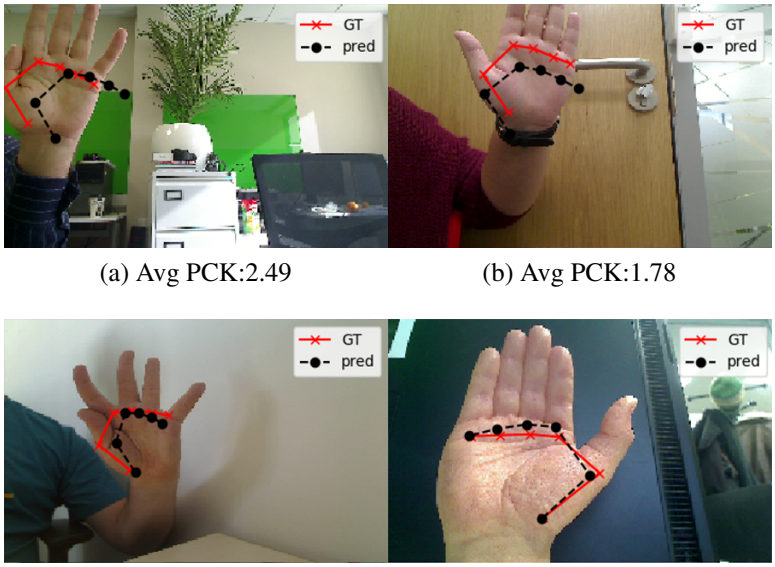
(a) Avg PCK:0.16 (b) Avg PCK:0.68 (c) Avg PCK:0.40



(d) Avg PCK:0.33 (e) Avg PCK:0.21 (f) Avg PCK:0.30

Fig. 7.9 Several samples from NUIGP2 test set representing successful cases.

An especially interesting case is displayed in Fig. 7.9a, where the hand is presented against a wooden background, the fingers are closed and partially cropped. In this case the network predicted the keypoints correctly. Fig. 7.9b displays a similar case where the network failed to predict the keypoints correctly as they have a strong translation. Please note the position of ground truth keypoints, acquired by the Leap Motion, which tend to lean slightly towards the edge of the hand.
Fig. 7.9d and 7.9c present the same hand pose but in two distinct lighting conditions. It is interesting to note the positions of predicted keypoints in both sub-figures, as they retain their spacial coherence regardless of the lighting.
Fig. 7.9e and 7.9f contain the same hand but with different hand posture. The positions of predicted points are maintained throughout the change of posture, unlike the ground truth points (acquired with Leap Motion).
It is worthwhile observing that in most cases mentioned above, the predicted points are positioned very close to the base of the main fingers, above what are considered Ground Truth (GT) points. This is explained by the bias introduced by the Leap Motion, used to label the images for training/testing.



(a) Avg PCK:2.49 (b) Avg PCK:1.78

(c) Avg PCK:0.66 (d) Avg PCK:0.33

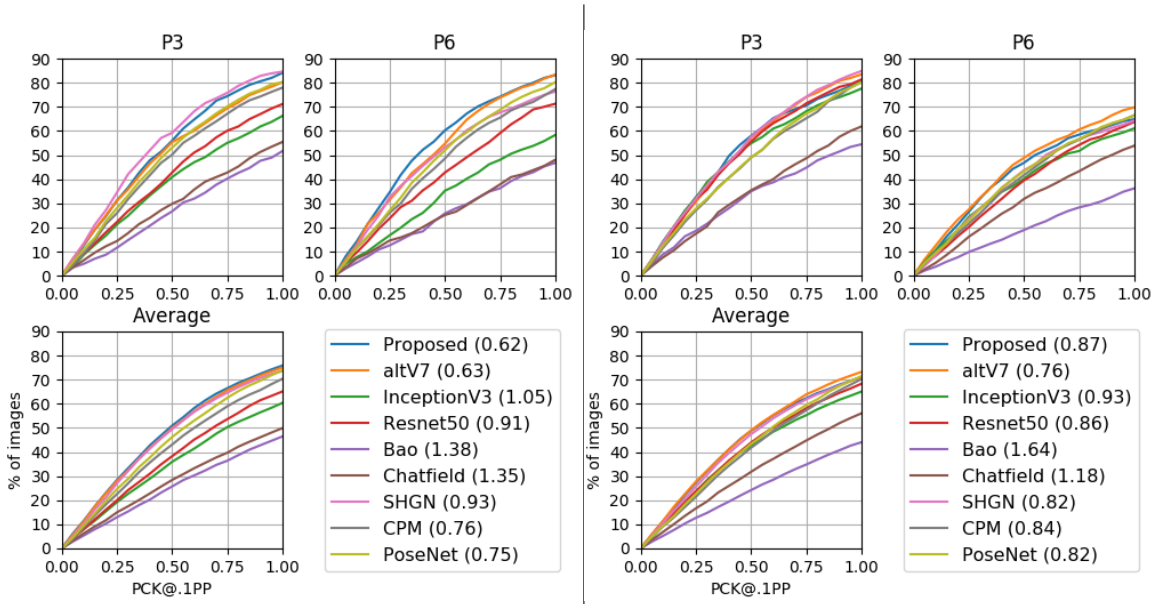
Fig. 7.10 Several samples from NUIGP2 test set representing failure cases.

Cases such as the one in Fig. 7.10a reflect the effect of the hand's strong translation to the edge of the image.
The image in Fig. 7.10c represents a previously unseen case, where the thumb is partially covering the index finger. Although it can be considered a failed case, the surface described by the predicted points overlaps most of the palm area. Similarly, in Fig. 7.10d the points P3-P5 describe a slight in-plane rotation away from the thumb.

The failed cases can be loosely grouped into two categories: hard failures, such as the ones presented in Fig. 7.10a and 7.10b, and soft failures represented by Fig. 7.10c and 7.10d. In order to better differentiate between soft and hard failures, the ROI_{IoU} factor (defined in Section 7.4.2) using the two extracted ROIs (using predicted versus ground truth points) is taken into consideration for training Scenarios 3 and 4.

Scenario 2. Train: NUIGP2, Test: NUIGP1, MOHI

In this scenario the bulk of training images was sourced from NUIGP2 (alongside 50% of the images from NUIGP1 and MOHI) and tested on NUIGP1 and MOHI (subjects that were not included in the training set). Note the main difference between the NUIGP2 set the and NUIGP1 set and MOHI is the way labels were provided for the palmprint landmarks. While the training images in NUIGP2 were generated using the setup described in Section 3.3 of Chapter 4, NUIGP1 and MOHI were hand labeled.
PCK rates are presented in Fig. 7.11 for two landmark points (P3 and P6), as well as for the



(a) Test: NUIGP1 (b) Test: MOHI

Fig. 7.11 Networks trained on NUIGP2, tested on NUIGP1 and MOHI. Global average PCK error (P3-P6) is shown next to the network name. Best seen in color.

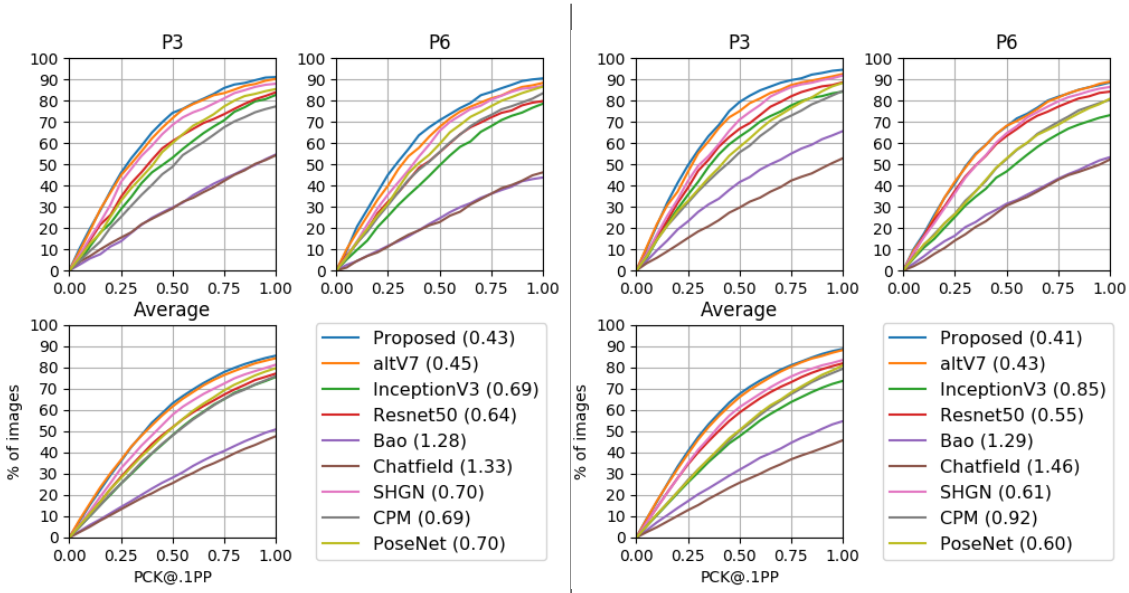
image-wise PCK error. P3 is considered to be the most accurately (and consistently) labeled point in NUIGP2. Labels for landmark point P6 from the leap motion device contained the highest levels of measurement error.

In the case of NUIGP1 images, SHGN provides better PCK rates (higher) when compared to the proposed network in the case of P3, but overall the two networks, alongside altV7 achieve the best performance. The proposed network also achieves the lowest global PCK error, of 0.62. For a PCK error of 1.0, the best performing networks are SHGN and the proposed network with a PCK rate of 85% of keypoints predicted, followed by altV7 and PoseNet (80%), and CPM (78%). The worst performing network is Bao, with a PCK rate of 52% of keypoints predicted.

For a PCK error of 1.0, it is noticed that for point P6, the proposed network and its alternative (altV7) achieve a PCK rate 83% correctly predicted keypoints, followed by Posenet (80.60%) and CPM and SHGN (77%). As with P3, the worst performing network is Bao with a PCK rate of 48%. From the image-wise PCK error distribution, one can notice the separation of 3 groups of networks, based on their performance: better performing ones (Proposed, altV7, SHGN, CPM and PoseNet), average performing (ResNet and InceptionV3) and networks that perform poorly (Bao, Chatfield). The separation of these groups tends to be between 10-15% in terms of PCK rate.

All networks were observed to exhibit a slight improvement in performance when predicting the position of P3 for images in the MOHI test set. Considering the PCK error

reference of 1.0, the best performing networks are SHGN (85%) and altV7 (83.50%), followed by ResNet50 (81.40%) and Proposed network (80.80%). The least performing network is Bao with a PCK rate pf 58%. However, this represents a 6% increase in PCK rate, compared to NUIGP1 dataset. In the case of P6 the predictions are worse and can be justified by the extreme hand positions found in MOHI, but also the labeling errors in the training distribution of NUIGP2. For a PCK error of 1.0 the best performing network is altV7 with a PCK rate of 70%, followed by PoseNet (66.60%) and Proposed, SHGN and ResNet (64.50%).



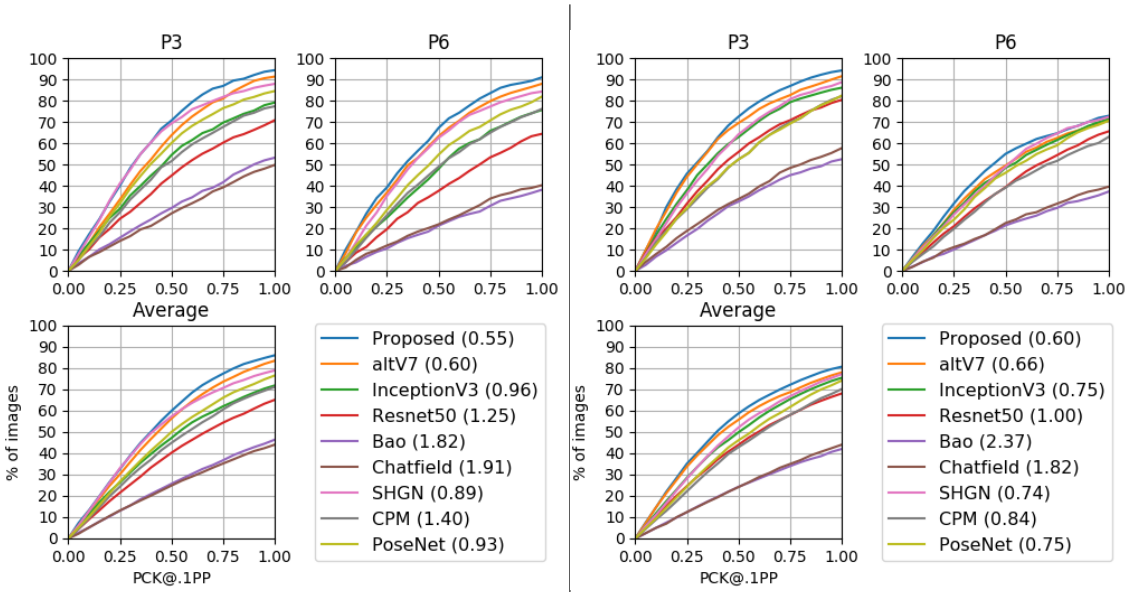
(a) Test: NUIGP1 (b) Test: MOHI

Fig. 7.12 Networks trained on GAN_H, tested on NUIGP1 and MOHI. Global average PCK error is shown next to the network name. Best seen in color.

Scenario 3.1. Train: GAN_H, Test: NUIGP1, MOHI

This scenario sourced most of its training images from GAN_H (130,000) and 50% of the images from NUIGP1 and MOHI. Compared to Scenario 2, GAN_H provides considerably lower error rates, as presented in Fig. 7.12 regarding points P3, P6 and image-wise average PCK.

In the case of test images from NUIGP1, most networks had PCK rates of over 80% (corresponding to a PCK error of 1.0) for both P3 and P6. The PCK error distribution between P3 and P6 is very similar, which is expected considering the lack of labeling errors in the training set from GAN_H. For a PCK error of 1.0, the proposed network achieves a PCK rate of 91% for point P3, followed by altV7 (90%), SHGN (88%) and PoseNet (86%).



(a) Test: NUIGP1 (b) Test: MOHI

Fig. 7.13 Networks trained on BIOmix, PCK error is presented for P3, P6 and image-wise Average (P1-P6). Global average PCK is shown next to the network name. Best seen in color.

The proposed network similarly achieves a PCK rate 90.50% for point P6 and is the most accurate. AltV7 (88%), SHGN and PoseNet (86.50%) and CPM (83.50%) provide the next best results.

The results reported for the MOHI images are either similar or better than those obtained on NUIGP1. For point P3 at a PCK error of 1.0, the best PCK rates are obtained with the Proposed network (94.50%), followed by altV7 and SHGN (92.30%), ResNet and PoseNet (88.60%). In the case of point P6, the best PCK rates are obtained by the Proposed net and altV7 (89%), SHGN (86%) and ResNet (84.20%).

Fig. 7.12 shows that the average PCK error is under 1.0 for all networks except Bao and Chatfield (both NUIGP1 and MOHI). Furthermore, in the case of the proposed network and altV7, the average PCK error is below 0.5 for both test sets.

Scenario 4.1. Train: BIOmix, Test: NUIGP1, MOHI

As described in Section 7.3.1, this training scenario combines the two previous databases, where about 50% of images are taken from GAN_H and the other half are taken from NUIGP2, but also including images from NUIGP1 and MOHI.

When testing on images from NUIGP1, the Proposed network provides the best results in all cases presented in Fig. 7.13. For a PCK error of 1.0, the highest PCK rate corresponding to point P3 is provided by the proposed net (94.30%), followed by altV7 (91.40%), SHGN

(87.40%) and PoseNet (84.80%), InceptionV3 (79%) and CPM (71%). The lowest PCK rate is provided by Chatfield, with 50%. In the case of P6 the best PCK rate is reported by the proposed net (91%), followed by altV7 (88%), SHGN (84.30%) and PoseNet (82%). All other networks report PCK rates that are below 80%.

Predictions for MOHI test images are similar or better for point P3, the Proposed network having the highest PCK rate (94%) and altV7 (91.60%), followed by SHGN (88.80%), InceptionV3 (86.20%), PoseNet (82.20%) and ResNet (80.50%). However, in the case of point P6 all PCK rates are lower than those obtained for NUIGP1, the highest rate being given by the proposed network (73%), followed by SHGN (72.40%), InceptionV3 (71.40%), PoseNet and altV7 (70.80%). All other networks achieve PCK rates that are lower than 70%.

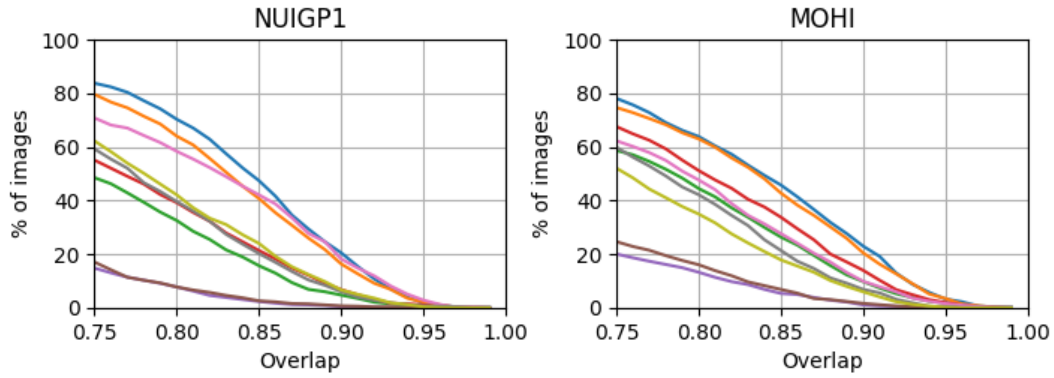
It is also interesting to note in Fig. 7.13 that InceptionV3 performs well on the task of point regression, obtaining performances similar to specialized networks like CPM or PoseNet, which rely on probability heatmaps.

7.5.2 ROI Extraction: BIOmix vs. GAN_H

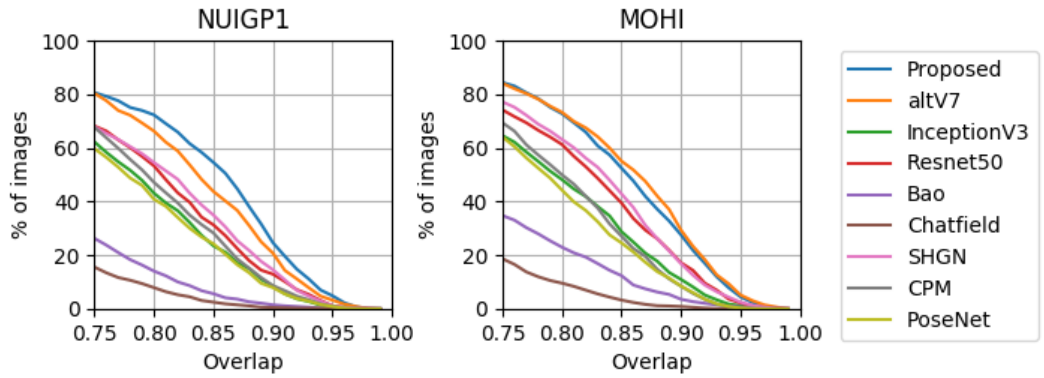
An evaluation of how the Regions of Interest (ROIs) are extracted using predicted points (ROI_{pred}) versus ground truth points (ROI_{gt}) is presented in this sub-section.

The ROI_{IoU} rates are reported in Fig. 7.14, where two training scenarios are considered: GAN_H and BIOmix (as summarized in Table 7.3). When comparing ROI_{pred} versusROI_{gt}, if the ROI_{IoU} factor is greater than 0.5, then that ROI is considered to be a correct detection because it covers a significant part of the palmprint.

When comparing the results obtained for NUIGP1 when training with BIOmix (Fig. 7.14a) versus GAN_H (Fig. 7.14b), one can notice a very similar trend in the case of the proposed network and altV7. As expected, having a greater proportion of images from GAN_H improves the successful ROI extraction rates from the lesser performing networks in Fig. 7.14a, such as Bao, Chatfield, but also InceptionV3 and Resnet50. The observed increase in performance can be explained by the labeling accuracy used during training brought by the image from GAN_H. A main cause for failures, as reported in the original paper of GAN_H [147], is brought by the skin-colored backgrounds, of which NUIGP1 has approximately 50% (both training and testing). However, it is important to note that the NUIGP1 and MOHI images used during training improved this behavior, in the case of the Proposed network limiting the failure cases to 27 out of 43 (GAN_H) versus 8 out of 16 failed ROIs (as can be noted in Table 7.14). Other cases include scenes that were not covered by the training set, such as the hand occupying a much larger or much smaller area of the image. In both scenarios the best results were provided by the Proposed network.



(a) Train database: BIOmix



(b) Train database: GAN_H

Fig. 7.14 Percentage of images having a specific overlap factor when training networks with (a) BIOmix and (b) GAN_H. The graph on the *left* corresponds to images from NUIGP1 (795), whereas the graph on the *right* corresponds to images from MOHI (1,000).

When testing on MOHI, Fig. 7.14 shows a similar jump in performance when training with GAN_H, significantly pronounced in the case of Bao, InceptionV3 and Resnet, but also altV7. The proposed network is marginally outperformed by altV7 based on the ROI_{IoU} rate. However, the rate of incorrectly extracted ROIs is still higher than when training with BIOmix (24 versus 16).

The proposed network reaches the lowest rate of incorrectly extracted ROIs when trained on BIOmix, having 3 failures for MOHI and 16 for the NUIGP1 test sets. Close performance is obtained with altV7 (7 images), followed by PoseNet (8 images) and SHGN (27 images). The least performing networks are, as in the other experiments, Chatfield and Bao.
It is also worth noting that when training with GAN_H, the lowest rate of failed ROIs is reached by PoseNet (21 instances), followed by AltV7 (23 images), CPM (26 images) and the proposed network (27 images).

Table 7.4 ROI extraction success and failure rates, based on the predictions provided by the neural networks used in the testing phase. A total of 1,795 palmprint images were used: 795 from NUIGP1 and 1,000 from MOHI. IoU corresponds to the allowed ROI_{IoU} factor (used as a threshold).

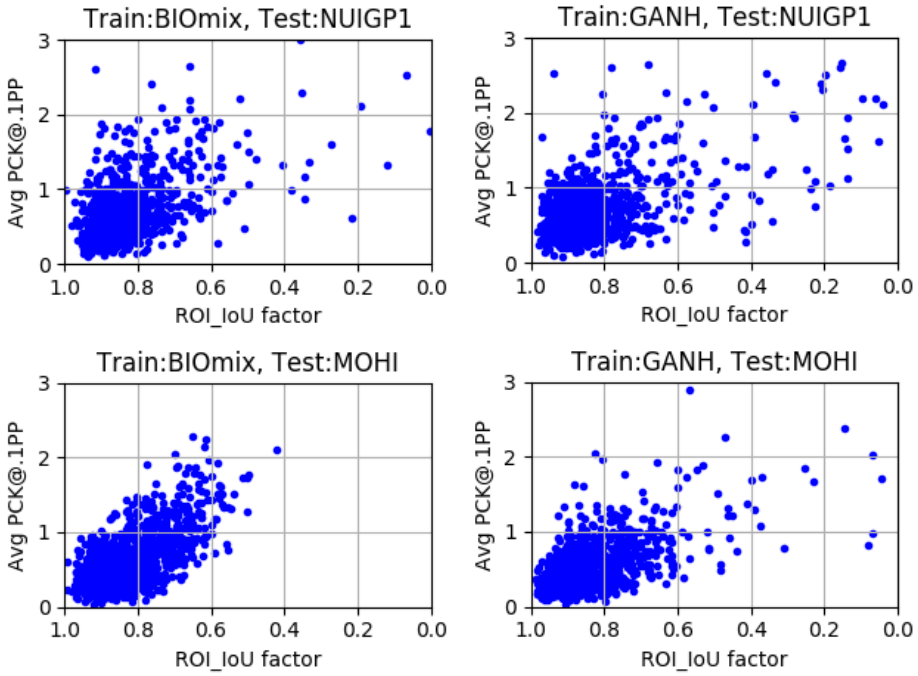
Test set	Network	Train: **BIOmix** IoU>0.5	Train: **BIOmix** IoU<0.5	Train: **GAN_H** IoU>0.5	Train: **GAN_H** IoU<0.5
NUIGP1	*Proposed*	779	16	752	43
	altV7	771	24	771	24
	InceptionV3	686	109	730	65
	ResNet50	702	93	737	58
	Bao_CNN	467	328	561	234
	Chatfield_CNN	452	343	508	287
	SHGN	707	88	684	111
	CPM	689	106	759	36
	PoseNet	730	65	732	63
MOHI	*Proposed*	997	3	973	27
	altV7	993	7	977	23
	InceptionV3	926	74	887	113
	ResNet50	955	45	889	111
	Bao_CNN	475	525	707	293
	Chatfield_CNN	555	445	584	416
	SHGN	973	27	912	88
	CPM	974	26	974	26
	PoseNet	992	8	979	21

7.5.3 Keypoint Error vs. ROI Extraction

Based on Fig. 7.15, where the ROI_{IoU} factor for every image is plotted against the average PCK error (reference used is 10% of the bounding box side). Fig. 7.15a and Fig. 7.15b follow the training procedure corresponding to Scenarios 3 and 4 (detailed in Table 7.3). The ROI_{IoU} values range between 1.0 and 0.0 (a low value indicating poorer performance). The average PCK (computed image-wise) ranges from 0.0 to 3.0 (a high value indicates a low score). The majority of instances are clustered in an area corresponding to (1.0, 0.6) for ROI_{IoU} and (0.0, 2.0) for average PCK.

As the avg PCK values increase, we can identify the images that have a high ROI_{IoU} factor although the average PCK is high(er), like in the case of NUIGP1 in Fig. 7.15a. Compared to BIOmix, training with GANH leads to generally lower average PCK image-wise, but also introduces more sparse image instances corresponding to a low ROI_{IoU}, as is the case in Fig. 7.15b.

Based on this clustering we can conclude that the combination of images composing the training database BIOmix leads to a better performing network, when compared to GAN_H (described in Section 7.3.1).



(a) (b)

Fig. 7.15 Plotting keypoint prediction error against ROI_{IoU} for Scenarios 3 and 4. The PCK is computed using a reference of 10% of the ground truth keypoints' bounding box's side.

7.5.4 Run-time of Networks

A random set of 200 images selected from the test set (100 from NUIGP1, 100 from MOHI) were passed to the networks tested on BIOmix. A batch size of 1 was used for the task of inference, the final time recorded being divided by 200. These results are presented in Table 7.5.

Using a sampling frequency of 24 frames per second (FPS), we determine the required time to perform inference on the images sampled from 1 second. We conclude that PoseNet, along with the Proposed network and alt-V7 are the fastest networks. All networks considered for experiments are capable of achieving inference at real-time speed (video-rate).

7.5.5 IoU relative to Palmprint Bounding Box Surface

We show the versatility of the Proposed network to the hand's scale in Fig. 7.16, where the IoU factor is represented as a function to the size of the palmprint bounding box (as represented in Fig. 7.5) relative to the entire image.

Table 7.5 Time required for networks to perform inference on 1 image.

Model name	Inference (ms)
Proposed	11
alt-V7	11
InceptionV3	20
ResNet50	14
Bao	0.004
Chatfield	5
SHGN	12
CPM	18
PoseNet	10

The Proposed network is trained on BIOmix and test images correspond to NUIGP1 and MOHI. The average size of palm bounding boxes are centered around the value of 16%. In contrast, the hand images in NUIGP1 tend to vary more in size, the mean value of palm bounding box centering around the value of 20% (of the image's size), with some instances being larger than 30%.

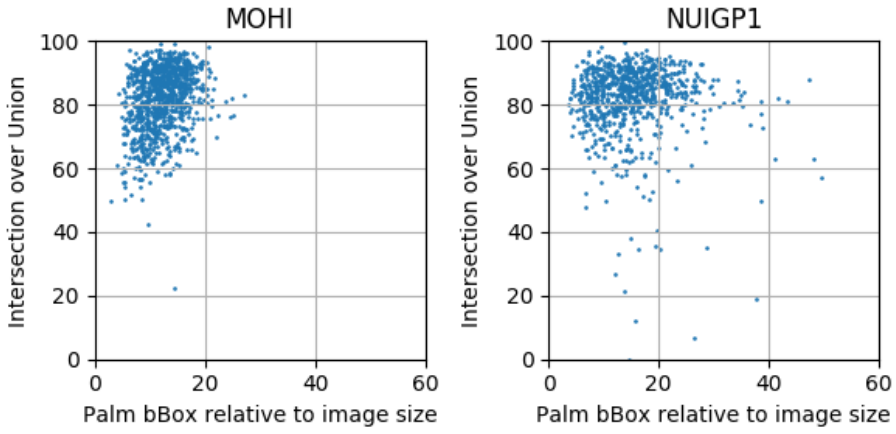
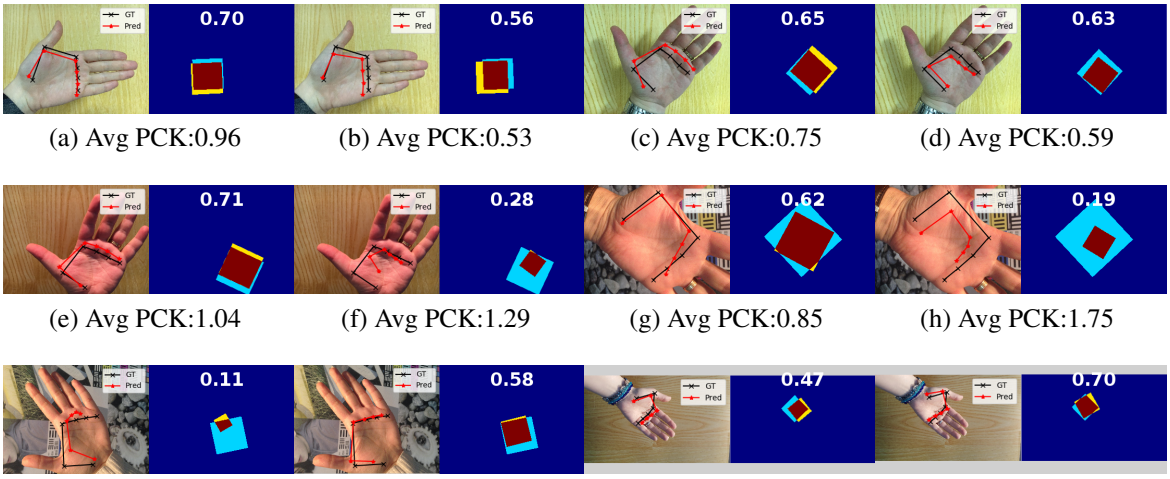


Fig. 7.16 Intersection over Union, represented as a function of the surface of the bounding box containing the palm landmarks (relative to the size of the image), as represented in Fig. 7.5.

7.6 Visual Analysis: Impact of Training Database

This section provides a visual evaluation of how the training databases influence the two tasks - keypoint prediction and consequently the palmprint ROI extraction. For this purpose,



(a) Avg PCK:0.96 (b) Avg PCK:0.53 (c) Avg PCK:0.75 (d) Avg PCK:0.59

(e) Avg PCK:1.04 (f) Avg PCK:1.29 (g) Avg PCK:0.85 (h) Avg PCK:1.75

(i) Avg PCK:2.00 (j) Avg PCK:1.25 (k) Avg PCK:1.05 (l) Avg PCK:1.17

Fig. 7.17 Visual examination of samples from NUIGP1 samples, using the proposed architecture. a)-d) BIOmix and GAN_H correct ROI, e)-h) BIOmix correct ROI, GAN_H failed ROI, i)-l) BIOmix failed ROI, GAN_H correct ROI. Next to a hand image is the overlap of ROI_{gt} and ROI_{pred}.

a series of NUIGP1 and MOHI test images from the two training scenarios (scenario 3.1 and 4.1 in Table 7.3) are used as reference.
We have considered only the results provided by the Proposed network because it achieves the best results overall (PCK rates and ROI_{IoU} rates).

7.6.1 Testing on NUIGP1

Three cases are considered for discussion, of which two samples are presented on each row in Fig. 7.17.

Although the average PCK error in Fig. 19b is lower than the one in Fig. 19a, the ROI_{IoU} factor is comparatively lower. The predicted keypoints in Fig. 7.17c reflect the influence of images from NUIGP2, with points P3-P6 being placed closer to the base of the main fingers, which introduce some distance between them and the ground truth (GT) points. The keypoints in Fig. 7.17d are placed below the P3-P6 GT points which slightly improves the ROI_{IoU} factor. The placement of points is only lacking in scale, most likely caused by the size of the hand relative to the image.

The 2nd row displays successful cases of Proposed net when trained with BIOmix and failures when trained with GAN_H. Similar to Fig. 7.17a, Fig. 7.17e positions the predicted key-points above the GT points, while correctly mapping the structure of the hand. The most likely factors explaining the failure in Fig. 7.17f are the background color, as well as

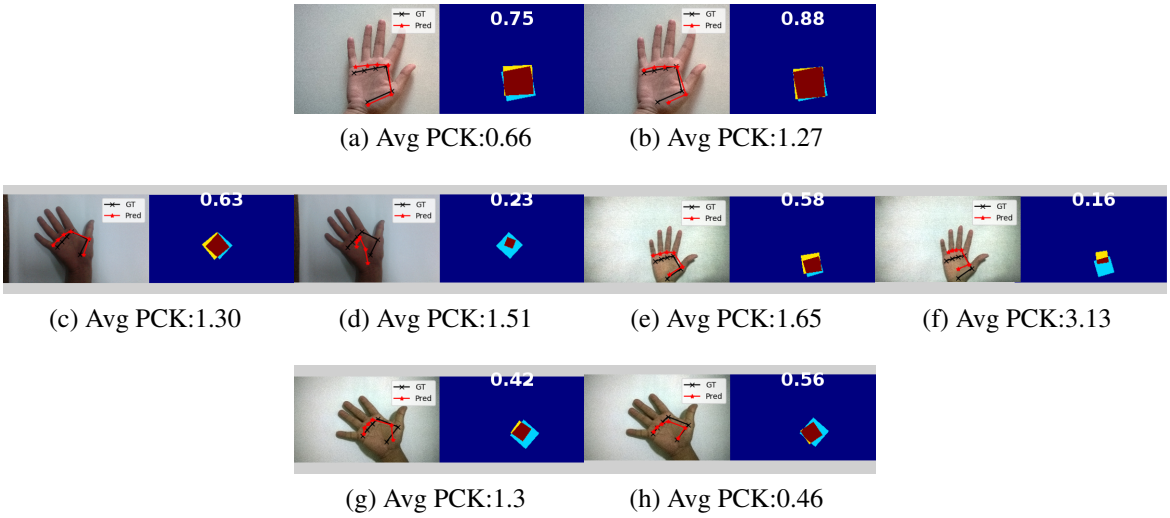
the absence of the hand's wrist. The images in GAN_H tend to always include some wrist information.
A 2nd faiulre case is displayed in Fig. 7.17g. Although the network did not place the main keypoints (P3-P6) correctly, P1 and P2 are very close to the GT points, keeping the average PCK error below 1.0. As in Fig. 7.17d, the predictions in Fig. 7.17h cover a smaller part of ROI_{gt}, but in this case it is considered a failed extraction. The scale of the hand, along with cropped fingers are the most likely causes for failure.

The 3rd row presents the only two cases where the proposed architecture trained with BIOmix failed the ROI extraction but succeeded when trained with GAN_H. The most accurately placed keypoint in Fig. 7.17j is P3. Even though these predictions are not entirely accurate (average PCK error is 1.25), the achieved ROI_{IoU} is 0.58, which covers the most relevant part of the palmprint.
The case presented in Fig. 7.17k is very close to being a successful prediction considering the overlap factor of 0.47. As in Fig. 7.17a and Fig. 7.17e, the predicted points in Fig. 7.17k are very close to the base of the main fingers, while the keypoints in Fig. 7.17l lead to a larger ROI_{IoU} factor, although the average PCK error is slightly larger.

7.6.2 Testing on MOHI

The 1st row in Fig. 7.18 presents a vertically placed hand in the center of the image, where both training scenarios lead to an ROI_{IoU} above 0.50, but Fig. 7.18a has a lower average PCK error compared to Fig. 7.18b. The main difference between the two cases is a slight rotation when training with BIOmix. Whereas in Fig. 7.18a points P4-P6 are slightly rotated upward, Fig. 7.18b maintains a better alignment overall with the GT points.

The 2nd row exemplifies one of the cases where training with BIOmix provided an ROI_{IoU} greater than 0.5 but training with GAN_H resulted in an ROI_{IoU} factor smaller than 0.5. Similar to the case in Fig. 7.18a, the predicted points in Fig. 7.18c lead to the ROI having a slight rotation towards the thumb, with the best predictions being for points P2 and P3. Even though in Fig. 7.18d the ROI extraction is failed, all points are found on the palm, the network having correctly determined the position of the right hand present in the image. The main cause for this failure is most likely the dark conditions of acquisition, considering that lighting tends to be uniform throughout the synthetic images of GAN_H. BIOmix contains many instances of hands exposed to strong light or shadows, making the network more robust to such scenarios.
The case displayed in Fig. 7.18f represents another failed ROI extraction when training with GAN_H, this time caused by the scale of the hand with respect to the image. Even if the predicted keypoints in Fig. 7.18e are found at the base of the main fingers, with a



(a) Avg PCK:0.66 (b) Avg PCK:1.27

(c) Avg PCK:1.30 (d) Avg PCK:1.51 (e) Avg PCK:1.65 (f) Avg PCK:3.13

(g) Avg PCK:1.3 (h) Avg PCK:0.46

Fig. 7.18 Visual examination of samples from MOHI samples, using the proposed architecture. a)-b) BIOmix and GAN_H correct ROI, c)-f) BIOmix correct ROI, GAN_H failed ROI, g)-h) BIOmix failed ROI, GAN_H correct ROI. Next to a hand image is the overlap of ROI_{gt} and ROI_{pred}.

considerable distance from the GT points, the IoU factor for the ROIs (0.58) is greater than 0.5. It is important to note that, similar to Fig. 7.18d, all predicted keypoints from the failed ROI extraction are still found on the hand.

The 3rd row represents the only case where training with BIOmix resulted in a failed ROI extraction (out of the total 3 failed cases), but GAN_H was successful. Fig. 7.18g shows a failed extraction, with the overall rotation of ROI turning away from the thumb. This is caused by having less accurate predictions for points P3-P5 than for P6. Furthermore, points P3-P5 are are closer to the base of the fingers. In a similar manner, in Fig. 7.18h the best predicted keypoints are P1, P5 and P6. However, keypoints P3-P5 are placed closer to the center of the palm, giving the ROI a rotation towards the thumb, thus providing a greater ROI_{IoU} than in Fig. 7.18g.

7.7 Conclusions and Future Work

The chapter explored the potential of using CNNs to accurately predict landmarks that can be used to extract the Region of Interest from a hand contained in an image. A comparison and analysis were made with previously used CNN architectures, as well as keypoint regression within the context of hand pose detection.

The first two points (P1 and P2) were the most difficult to mark (both as a human operator,

as well as the predictive algorithm), as indicated by the networks as well (lowest reported accuracy). In the absence of other cues from the hand, the task of labeling is very difficult. This is not an issue when specific physical pointers (i.e. other hardware elements) are employed, or if the training/testing data is obtained in a synthetic manner with 3D models, where the labeling error is effectively equal to zero. The issue of correctly labeled data is important, as it limits the evaluation of models trained with that data.
It is important to note that points closer to the edge of the hand (or other elements of the hand), with more structural references, would help in reducing the labeling error of landmarks, thus providing a more consistent landmark database. This applies to points P3-P6, which report the smallest errors across all experimental scenarios. The use of other points, that can be more accurately labeled in the dataset is something that has to be considered for future work.

Adapting hand pose regression techniques to the problem of palmprint ROI extraction was validated by the results presented, especially for those difficult cases where no constraints are imposed on the user.

Using 2D key-point prediction to determine the landmarks required for ROI extraction provided some good results, but they are limited. In future experiments 3D information regarding the hand's pose will be included, to aid the regression task. Furthermore, a fully convolutional approach for palmprint ROI will be developed, by adapting powerful object detection architectures (such as Faster-RCNN [70] or Mask-RCNN [203]) to include an 'angle' parameter that would align a bounding box to the hand's orientation. Alternatively, after accurately detecting the palmprint ROI and landmarks, one would then be able to determine and apply affine transformations that would further correct distortions caused by the hand's pose. This has become possible with the introduction of the database of palmprints NUIGP2, as well as other large-scale databases collected in unconstrained conditions.

The proposed network (containing several Inception modules) is an efficient architecture for key-point regression, requiring further investigation of its inference capabilities on other hand databases and with deeper configurations.

Finally, combining images from several databases aids the robustness of the network, as was demonstrated with the better results obtained when training with BIOmix.

Chapter 8

Unconstrained Region of Interest - Performance Evaluation

This chapter outlines the difference in recognition performance between Regions of Interest (ROI) extracted using ground truth (GT) versus predicted landmarks, as obtained with the extraction approach described in Chapter 7. The images used throughout the experiments of this chapter belong to the palmprint database NUIG_Palm1 (NUIGP1), introduced in Section 4.2 of Chapter 4.

The generic biometric recognition pipeline is presented in Fig. 8.1, with the stages of recognition being investigated in this chapter: ROI extraction and feature recognition (comparison).

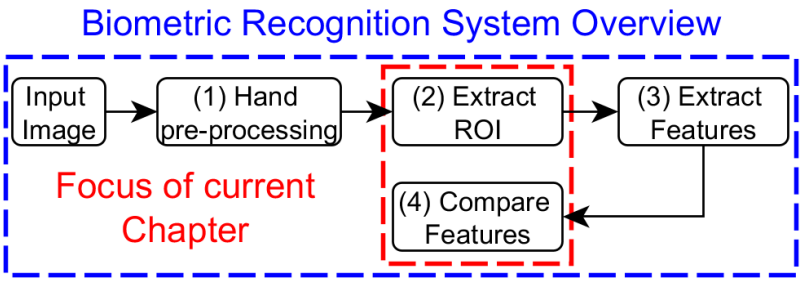


Fig. 8.1 Biometric recognition pipeline, with focus of current chapter.

The chapter is organized as follows: Section 8.1 introduces an overview of ROI extraction and its evaluation in the literature, Section 8.2 provides an overview of the experiments and formulates the main assumptions being investigated, Section 8.3 describes the experimental methodology, whereas Section 8.4 presents results obtained and interprets them. Section 8.5 presents the chapter's conclusions and future directions of research.

8.1 Introduction

Over the years there have been introduced many algorithms for Region of Interest (ROI) extraction (overview presented in Section 3.2 of Chapter 3). However, the literature regarding ROI extraction evaluation and comparisons is scarce, the first study of the impact of ROI extraction strategy on the Recognition Rate being provided by Ito *et al* [44]. They compare with three approaches [38], [37], [55] and manually placed landmarks (ground truth), using hand images with open and closed fingers. They use the Equal Error Rate (EER) in recognition experiments to highlight the robustness of their ROI extraction algorithm to various hand poses.

Khan *et al* [54] introduce one of the earliest analyses of key-points used for palmprint ROI extraction. They consider a subset of 300 hand images and evaluate the point prediction error (Chebyshev distance) from the manually placed ground truth landmarks. They also consider the extracted sample's absolute rotation error, as well as the scale error. Aykut *et al* [64] determine the prediction error to their Active Appearance Model (AAM) for every point, as well as specific regions of their hand model. Scale normalization of the error is not considered, as the scale remains consistent throughout the images.

Bao *et al* [66] introduce a shallow Convolutional Neural Network (CNN) that predicts the key-points used for ROI extraction. This approach is compared to Zhang's method [38], which is the most commonly used. The tangent line between the two finger-valleys is determined, then using these two points to normalize the hand rotation. Root Mean Square Error (RMSE) is used to evaluate the point predictions, followed by an evaluation inspecting the Correct Recognition Rate (RR) of extracted palmprint samples. The two approaches have similar performance, but the CNN architecture fares better than Zhang's [38] approach especially when noise is added to the images.

8.2 Objectives of Chapter

In this work we have a number of top-level goals:

Experiment 1. Investigate how the Recognition Rate (RR) and Equal Error Rate (EER) are affected by the misalignment of palmprint ROIs by using the Intersection over Union (IoU) metric applied to palmprint ROI extraction. Assumptions are made regarding how these results are affected based on the distribution of ROIs (extracted using Ground Truth or Predicted labels) and evaluated in the Results section. Several combinations of these distributions for training and testing are taken into consideration and evaluated.

Experiment 2. Take into account the attributes of the NUIG_Palm1 database, and design specific training and testing procedures. This represents an extension of Section 6.3 of Chapter 6, which provided baseline recognition results. It is expected to obtain lower performance when training in the 'cross-device' scenario (more details in Section 8.3).

Experiment 3. Include a misalignment mitigation strategy that improves the recognition accuracy. Translations and rotations of the extracted ROIs are used in several scenarios that evaluate this impact.

Experiment 4. As the main focus is placed on the influence of ROI misalignment, a generic feature extraction technique (Local Binary Pattern) and generic classifier (K-Nearest Neighbor) are used throughout experiments.

8.3 Overview of Experiments

This section provides an outline of the selected recognition experiments. They can be broken down into four categories, based on the aspect they are investigating (also represented in diagram form in Fig. 8.2):

1. How the recognition performance varies in the 'cross-device' and 'device-specific' training strategies using ground truth (GT) and predicted (Pred) ROIs. The experiment is outlined in Section 8.3.1 of this chapter.

2. How the recognition results are affected by the misalignment between ROI_{gt} and ROI_{pred}, investigated through the split between train and test images. The experiment is outlined in Section 8.3.2 of this chapter.

3. How much the misalignment can be addressed if one considers extracting translated and rotated ROIs from each hand image, investigated by including translation and rotation to the key-points used for ROI extraction. The experiment is outlined in Section 8.3.3 of this chapter.

4. How the recognition performance is affected by the ROI size, as it varies between (32x32), (64x64) and (128x128) pixels. The extracted ROIs are resized by using bilinear interpolation. Based on the results obtained in Section 5.3 of Chapter 5, the size of (64x64) pixels report the highest recognition values.

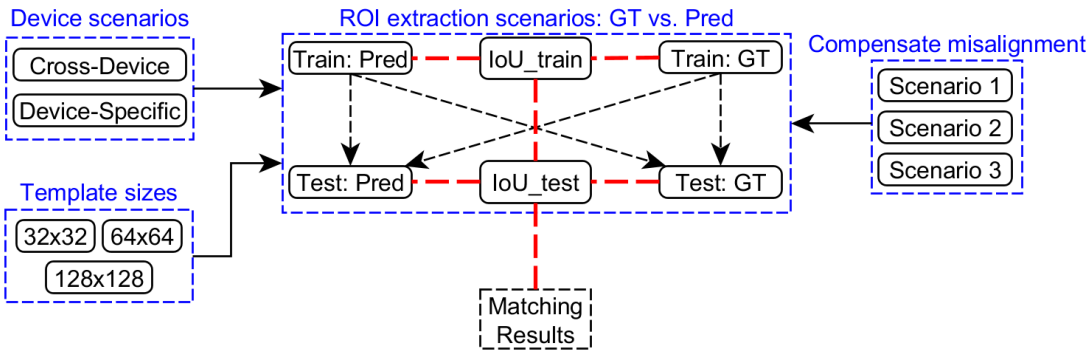


Fig. 8.2 Overview of experiments, in graphical form.

8.3.1 NUIG_Palm1: Recognition Scenarios

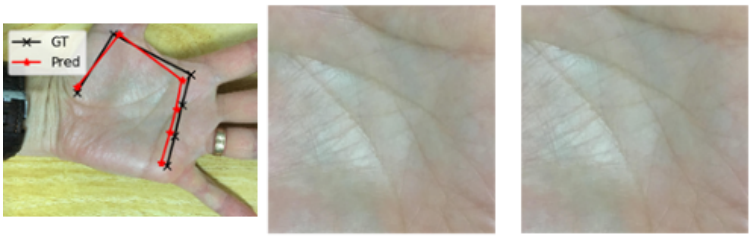
The database NUIG_Palm1 (NUIGP1) contains in total 81 subjects (classes). Almost half of these (41 classes) were used to train the ROI extraction algorithm described in Chapter 7, leaving 40 subjects for recognition experiments. Each class contains about 20 images, representing palmprint images acquired with the camera of 5 smartphones. The labels '_S1', '_S2', '_S3' and '_S4' correspond to the lighting levels and backgrounds used during acquisition (described in Section 4.2 of Chapter 4).

Similar to Section 6.2.1 of Chapter 6, two distinct training/testing strategies are defined, based on the devices that are used for image acquisition:

1. Cross-device: where '_S1' images from all devices are used for training, the rest being used for testing. Several ROI sizes, classifiers and feature extraction techniques are tested in order to determine the best combinations.

2. Device-specific: where '_S1' to '_S4' images from a single device are used for training, with the rest (i.e. from other devices) are kept for testing. The best classifier-feature extraction combination from the cross-device scenario is used.

In Sections 6.3.2 and 6.3.3 of Chapter 6 we have concluded that the device-specific scenario was outperforming cross-device recognition scenario. The classifier-feature extraction combination reporting the best results was CRC_RLS with DoN when training with device-specific images from GS6 (87.81% Recognition Rate and 6.10% Equal Error Rate).
This behavior is expected to manifest in the experiments outlined in this chapter as well. However, it is unclear how the misalignment will affect the recognition procedure, and if the best results will still belong to the scenario training with images acquired with GS6.



(a) Palmprint prediction with overlap $IoU = 82.92\%$

(b) Palmprint prediction with overlap $IoU = 52.21\%$

Fig. 8.3 Palmprint landmarks used for ROI extraction - visual comparison of misalignment for (a) high IoU factor and (b) low IoU factor.

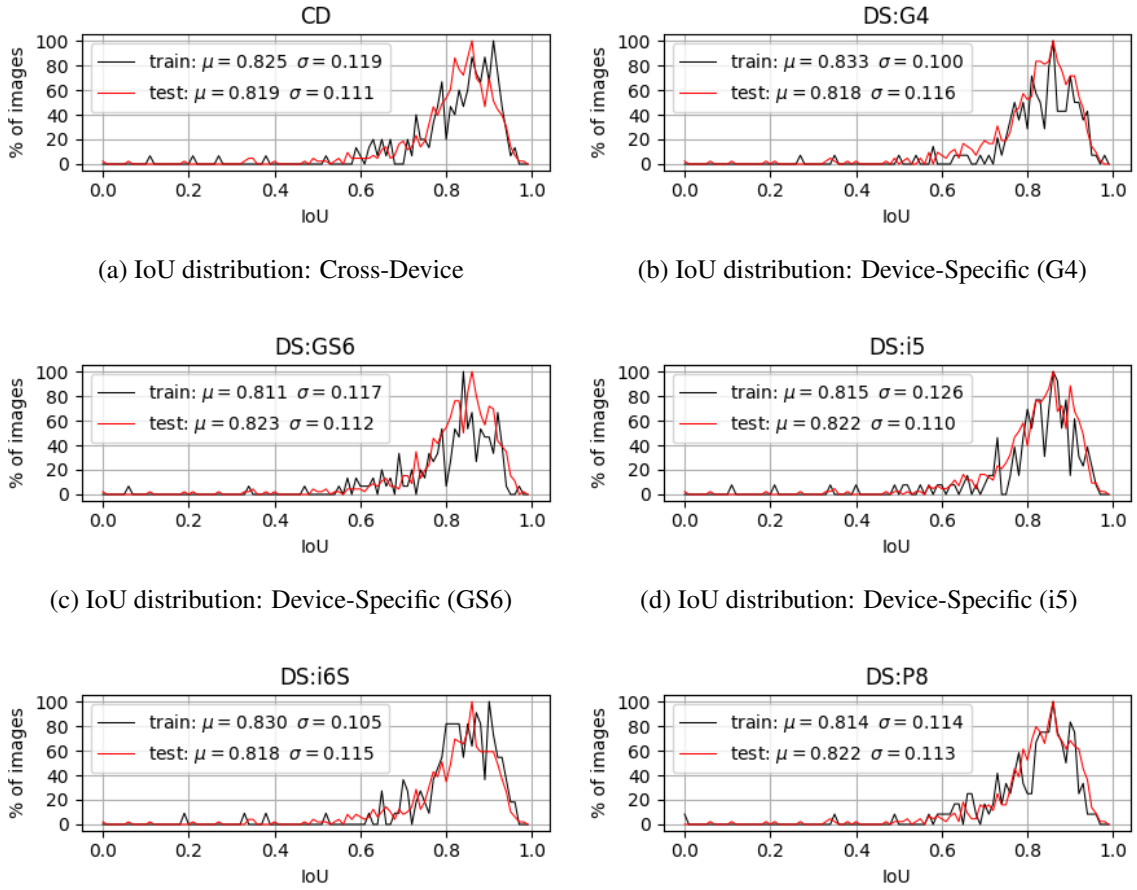
8.3.2 Evaluating the Misalignment

Any misalignment between the stored and query samples affects the recognition task (comparison trial). For this reason the Intersection over Union (IoU) factor was used, in order to investigate the correlation between high IoU and high Recognition Rate. Fig. 8.3 provides two samples of extracted ROIs with low and high IoU factor.

The overlap between ROIs extracted with Ground Truth landmarks (ROI_{GT}) versus Predicted landmarks (ROI_{Pred}) was determined for each train/test scenario, as presented in Fig. 8.4. Depending on the training scenario (either cross-device or device-specific), we associate an IoU rate to both training (S_{train}) and testing (S_{test}) sets. In other words, the IoU is computed for every image in S_{test} and S_{train} (the distribution of these values being presented in Fig. 8.4).

In all scenarios (cross-device and device-specific) in Fig. 8.4, the distribution of IoU factor for S_{test} is very similar to the distribution of S_{train}. The mean (μ) is generally centered at 0.82 in most cases, with standard deviation (σ) being around 0.11 for all cases.

It is interesting to notice that the IoU distribution in S_{train} for GS6 has the lowest mean (0.811) of all cases, with a high standard deviation (0.117). It is expected that training with ROI_{Pred} from GS6 will affect the recognition rate considerably, compared to training with ROI_{GT}.



(a) IoU distribution: Cross-Device

(b) IoU distribution: Device-Specific (G4)

(c) IoU distribution: Device-Specific (GS6)

(d) IoU distribution: Device-Specific (i5)

(e) IoU distribution: Device-Specific (i6S)

(f) IoU distribution: Device-Specific (P8)

Fig. 8.4 Distribution of Intersection over Union (IoU) in the train/test scenarios explored in (a) Cross-Device (CD) and (b)-(f) Device-Specific (DS) scenarios. For each distribution (train and test) the mean (μ) and standard deviation (σ) are defined.

Training with ROI_{Pred} samples from G4 and i6S is expected to be the least affected by the misalignment, as the mean of S_{train} is high (0.833 and 0.83) and the standard deviation is low (0.10 and 0.105).

We formulate several assumptions regarding the impact of ROI misalignment on the recognition task, as outlined in Table 8.1. As point of reference we use the segmentation map defined using the ground truth landmarks:

- **Assumption 1:** High IoU rate in both S_{train} and S_{test} leads to high RR. This translates to training with ROI_{GT} and testing with ROI_{GT} (maximum IoU for both training and testing sets).

Table 8.1 ROI misalignment evaluation scenarios, based on what key-points were used to extract the palmprint ROI samples. Ground truth key-points (GT) as well as Predicted key-points (Pred) are considered, allowing for 4 possible combinations. These scenarios are defined to investigate the assumptions defined in Section 8.3.2.

#	Scenario label	Train	Test	Expected performance
1	**GT-exclusive**	ROI_{GT}	ROI_{GT}	Highest (no failed ROIs in train/test sets)
2	**Pred-exclusive**	ROI_{Pred}	ROI_{Pred}	Low (failed ROIs in train/test sets)
3.1	**GT-Pred**	ROI_{GT}	ROI_{Pred}	Low (failed ROIs in test set)
3.2	**Pred-GT**	ROI_{Pred}	ROI_{GT}	Low (failed ROIs in train set)

- **Assumption 2:** Low IoU in both S_{train} and S_{test} leads to low RR. This translates to training with ROI_{Pred} and testing with ROI_{Pred}.

- **Assumption 3:** Lower IoU rate in S_{train} compared to S_{test} (or the other way around) leads to low Recognition Rate (RR). This corresponds to two scenarios where ROI_{GT} or ROI_{Pred} are used for training and the other is used for testing.

8.3.3 Mitigating ROI Extraction Misalignment

Several training strategies are considered, aiming to reduce the misalignment (translational and rotational) between the stored image samples and query, as presented in Fig. 8.5. This introduces 8 translation cases and 2 rotation cases. Three scenarios are defined, based on the number of palmprint ROIs extracted from one hand image:

- **Scenario 1**: only one ROI extracted from a hand image (the central ROI).

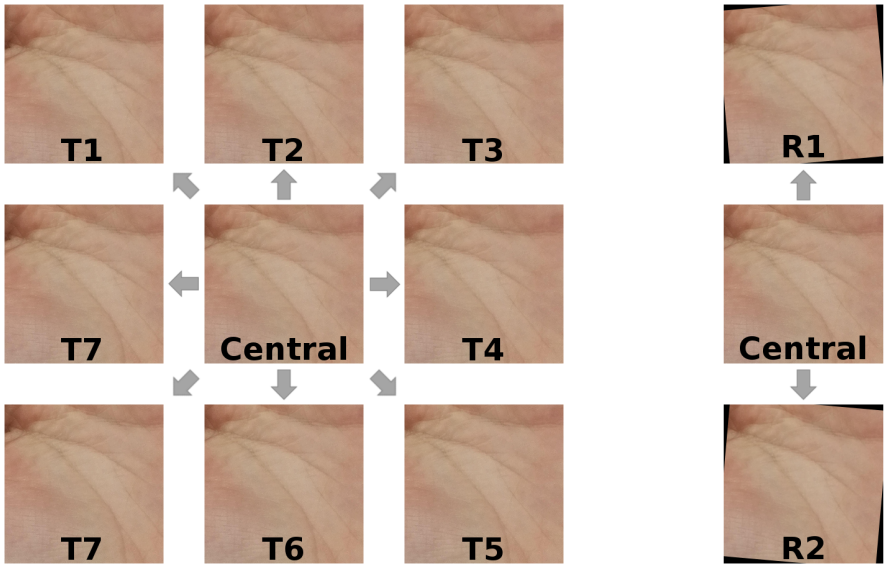
- **Scenario 2**: central ROI and translated ROIs by 50 pixels into the 8 possible directions.

- **Scenario 3**: central ROI, translated ROIs and rotated ($\pm$5 degrees) ROIs (central and translated ones). This results in extracting 27 ROIs from a single hand image.

Training is carried out with each of the three aforementioned scenarios and comparison trial (testing) is performed only with ROIs extracted using scenario 1.

8.3.4 NUIGP1: ROI Extraction Approach

The landmarks used for ROI extraction follow the model of hand joints, described in Section 3 of Chapter 4. Each category of landmarks (ground truth or predicted) generate their specific distribution of training/testing ROIs.



(a) Translation of ROI

(b) Rotation of ROI

Fig. 8.5 ROI extraction strategy considered for training stage. (a) displays the translation of extraction coordinates into 8 directions with 50 pixels and (b) displays the rotation of ROIs (central and translated) by ± 5 degrees.

In general, the ROIs extracted using predicted landmarks represent translated and/or slightly rotated versions of their ground-truth counterparts. This is expected to be compensated by including translated and rotated ROIs during training.
In order to be consistent with the comparison, the extracted ROIs that were considered as having failed (with less than 0.5 IoU factor, as in Section 2.4 of Chapter 7) were included in their corresponding training or testing sets. Considering that the devices G4, i5 and i6S have either one or several failed ROIs in their corresponding '_S1' acquisition scenario (as displayed in Table 8.2), it is expected that they report lower recognition rates than GS6 and P8. This is explained by the fact that '_S1' images are used during training in both cross-device and device-specific strategies. The device class containing most of the 'failed' ROI extractions is i5. It is, therefore expected that recognition results for those scenarios be lower than the scenarios training with other devices.

Based on the number of failed ROI extractions in Table 8.2 and the train/test distributions listed in Table 8.1, several assumptions can be made:

1. as the most failed ROI extractions correspond to i5, it is expected that when only i5 images are used for training (device-specific, in the Pred-GT and Pred-exclusive), the results will be the lowest.

Table 8.2 The number of images that failed ROI extraction using the approach described in Chapter 7, organized depending on the device or acquisition setup.

Device name	S1	S2	S3	S4
G4	1	1		
GS6		1	1	1
i5	2	3		
i6S	1	1	1	
P8		2		1

2. similarly, when training with i5 images in the GT-Pred case, the results are likely to be the best (as the i5 images are the most significant source of error).

8.3.5 Feature Extraction

As the chapter's focus is not placed on feature extraction performance, but on the evaluation of misalignment, any texture descriptor can be used. We have therefore chosen the Local Binary Pattern (LBP) because of its texture descriptive properties. Two LBP kernel sizes were considered (3x3 and 5x5). The best results were delivered when using a filter of radius $R = 3$ and $P = 24$ (kernel of size 5x5), as opposed to the standard LBP using $R = 1$ (kernel size of 3x3). This changes the potential values of the pattern to range from 1 to 2^{24} instead of from 1 to 2^8. Recognition results when using LBP with $R = 1$ (kernel size of 3x3) are not reported, as they were too low.

Similar to the experiments in Chapter 6, all images had their contrast improved using the Contrast Limited Adaptive Histogram Equalization (CLAHE) [173] algorithm prior to the feature extraction stage.

8.3.6 Classification: Nearest Neighborhood

For the recognition phase only one classifier was considered, as the focus of the chapter is placed on the effects of ROI misalignment, not on the performance of the classifier itself.

The K-Nearest Neighborhood classifier (KNN) continues to be used ([74], [67], [127]) thanks to its reduced complexity, as it is non-parametric. The classifier uses $K = 3$ votes (chosen empirically) and the Euclidean (L2) distance for determining the class (from training set) which is closest to the query sample.

8.3.7 Resources Used

The tools used to run the experiments include:

- For ROI extraction:
  - GPU: TitanX-Pascal 1080
  - CUDA version: 9.0
  - cuDNN version: 7.1.4
  - tensorflow version: 1.9.0
  - Keras version: 2.2.2
- scikit-learn version: 0.19
- scikit-image version: 0.14.2
- OpenCV version: 3.4.2.16
- github repositories with the implementation of experiments[1].

8.4 Results

8.4.1 Cross-device (CD) Training

Results from the scenario where images from scenario '_S1' from all devices were used for training, are provided in Table 8.3. For these experiments all ROI_{GT}/ROI_{Pred} combinations were explored. The green colored cells correspond to the best results in that specific category. In all cases the extended training scenario (translation + rotation) report the best performance across all categories. The difference in terms of EER between results obtained with an extended training scenario (III) versus none (I), are 8.06% for GT-exclusive, 12.32% for Pred-exclusive and around 15% for GT-Pred and Pred-GT scenarios.

The cross-device recognition results are similar to the ones described in Section 6.3.2 of Chapter 6 (reaching 80% RR and 10% EER), with the highest Recognition Rate (RR) of 83.89% (Equal Error Rate of 9.54%) being achieved in the GT-exclusive case. However, this result corresponds to Training scenario III, whereas in training scenario I the RR is 70.97% and EER is 17.60%. This validates the inclusion of an extended stage of ROI extraction that considerably increases the chances of a correct match. Similar performance was reported for the Pred-exclusive and Pred-GT cases, of around 75% RR and 16% EER. This is explained by the number of failed ROIs contained in the training and testing distributions in these

[1]Implementation of the approaches used can be found at "github.com/AdrianUng/Palmprint-recognition-using-misaligned-ROIs".

Table 8.3 Recognition Rates (RR) and Equal Error Rates (EER) corresponding to the Cross-Device training strategy. Best results overall are highlighted with green, best results in each training scenario are highlighted with yellow.

Training strategy	GT-exclusive (Train: ROI_{GT}, Test: ROI_{GT})						Pred-exclusive (Train: ROI_{Pred}, Test: ROI_{Pred})					
	128x128		64x64		32x32		128x128		64x64		32x32	
	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)
I	65.10	20.40	70.97	17.60	64.93	18.86	46.48	31.59	51.17	28.42	49.66	28.20
II	73.83	14.04	81.38	11.24	76.01	14.76	64.26	21.98	67.62	18.88	65.77	21.12
III	78.36	13.20	83.89	9.54	76.17	14.47	69.63	18.82	74.50	16.10	68.12	19.28
	GT-Pred (Train: ROI_{GT}, Test: ROI_{Pred})						**Pred-GT (Train: ROI_{Pred}, Test: ROI_{GT})**					
	128x128		**64x64**		**32x32**		**128x128**		**64x64**		**32x32**	
I	33.72	37.40	37.92	35.54	37.42	34.51	40.60	33.41	44.80	31.10	47.99	29.96
II	54.53	27.80	60.57	24.79	55.03	25.70	57.71	24.02	66.44	20.95	60.74	23.68
III	59.90	24.40	67.28	21.39	60.91	24.43	67.11	19.99	75.17	16.01	67.45	20.40

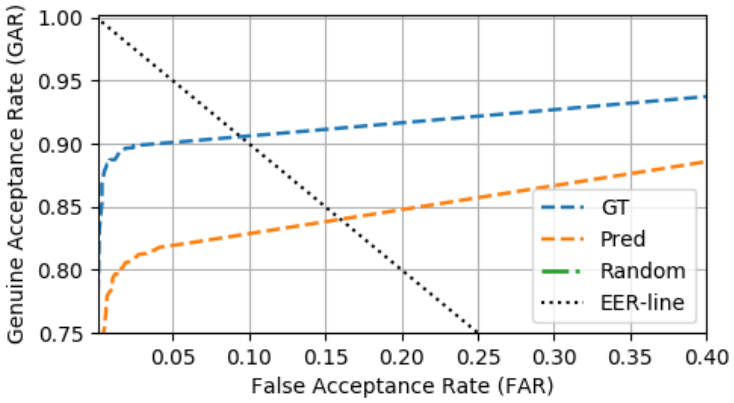


Fig. 8.6 Receiver Operating Characteristic (ROC) in the cross-device case. For training the GT-exclusive (GT) and Pred-exclusive (Pred) scenarios were used.

two cases. Whereas Pred-exclusive supports 4 failed ROIs in the training set and 12 failed ROIs in the testing set, Pred-GT reports slightly better results because its testing set has no instances of failed ROI extraction.

The lowest overall performance is reported for GT-Pred (67.28% RR and 21.39% EER in the extended training scenario). This attempts to match images that essentially belong to two distinct distributions (in terms of how the ROI is extracted). This result strongly suggests that the training and test sets should be obtained using the same (or a similar) ROI extraction technique.

Receiver Operating Characteristic (ROC) curves corresponding to Pred-exclusive and GT-exclusive scenarios from Table 8.3 are displayed in Fig. 8.6. For a reference value of 10^{-3} for False Acceptance Rate (FAR), corresponding to the probability of a test sample of being mistakenly accepted as a different class, the Genuine Acceptance Rate (GAR) is reported as the probability of a test sample to be correctly classified for every 1,000 attempts.

The KNN with LBP reports a GAR of 79.15% in the GT-exclusive case. Conversely, at the same value of FAR the Pred-exclusive scenario reports a GAR value of 62.60%.

8.4.2 Device-specific (DS) Training

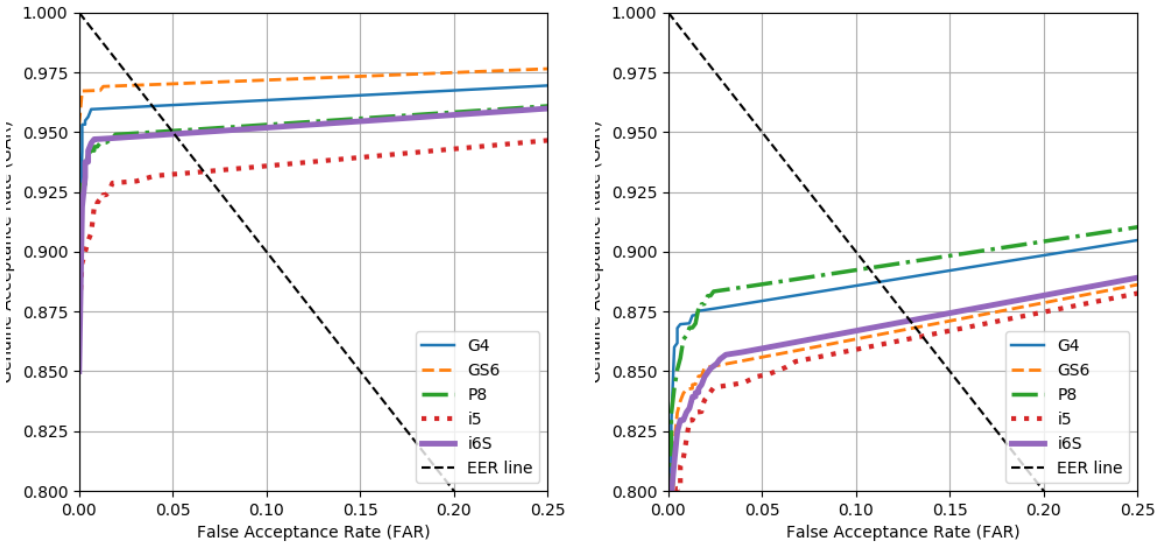
Recognition results for DS training scenarios are listed in Tables 8.4 (GT and Pred-exclusive cases) and 8.5 (GT-Pred and Pred-GT cases).

The overall best result corresponds to the GT-exclusive scenario training with images from GS6, of 94.66% RR and 3.06% EER. This corresponds to the extended training scenario III, which represents an improvement of 12% (RR) and 13.44% (EER) from training scenario I. The device scenario with the lowest performance is i5, as expected (as was also reported in Section 4.3 of Chapter 6).
It is worth noting the impact of including an extended training scenario, as the EER difference between scenario I and scenario III is around 13% for all device cases, the exception being i5 (whose difference is 9.96%).

It is interesting to note that in the Pred-exclusive scenario the ranking of devices (in terms of recognition performance) is changed, training with images from P8 reporting the best result of 83.18% RR and 10.71% EER. P8 corresponds to the device-training scenario reporting the lowest difference in terms of EER between the GT and Pred-exclusive scenarios, of 5.71% (all other device cases report differences greater than 6.50%).
As expected, training with images from i5 in Pred-exclusive scenario leads to the lowest performance, especially considering it is the device-case having the most ROI extraction failures (5 in total). However, it is interesting to note that the EER difference between this case and GT-exclusive is only 6.79%, which is the 2nd lowest variation between training scenarios.
Similar to the GT-exclusive case, the EER drops significantly when including an extended training scenario. The lowest differences correspond to device cases GS6 (14.96%) and i6S (14.93%), with P8 having the largest difference, of 17.27%.

The ROCs associated to the best scenarios in GT and Pred-exclusive cases are represented in Fig. 8.7. For an analysis of the GAR at a determined value of 10^{-3} for FAR, the device-scenario GS6 reports the value of 96.19% (GT-exclusive) and 71.47% (Pred-exclusive). The scenario using images from P8 is on the 3rd best position in terms of GAR, with 90.49% in GT-exclusive and on the 1st position in the Pred-exclusive scenario (79.43%).

All training-testing scenarios up to this point correspond to the first two evaluation scenarios in Table 8.1 (training and testing exclusively with ROI_{GT} or ROI_{Pred}). Following is an analysis of GT-Pred and Pred-GT scenarios, with recognition results introduced in Table 8.5.



(a) Device-specific training: ground truth ROIs

(b) Device-specific training: predicted ROIs

Fig. 8.7 Receiver Operating Characteristic in the device-specific setup. For simplicity, the experiments using ROIs extracted with ground truth labels (8.7a) were separated from the ones using predicted labels (8.7b).

Table 8.4 Recognition Rates (RR) and Equal Error Rates (EER) corresponding to the Device-Specific training scenario, GT-exclusive and Pred-exclusive recognition scenarios.

Device	*Scenario*	**GT-exclusive (Train: ROI_{GT}, Test: ROI_{GT})**						**Pred-exclusive(Train: ROI_{Pred}, Test: ROI_{Pred})**					
		128x128		**64x64**		**32x32**		**128x128**		**64x64**		**32x32**	
		RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)
G4	**I**	62.64	22.50	70.64	17.83	71.59	16.83	46.78	30.76	54.32	27.26	55.26	24.79
	II	83.52	9.30	92.15	4.99	90.42	6.27	75.35	15.98	80.22	12.85	80.85	12.35
	III	87.91	6.67	93.88	3.95	91.84	5.95	79.28	14.01	84.93	11.29	84.46	10.66
GS6	**I**	64.21	20.17	72.84	16.50	73.94	14.85	45.68	32.19	53.53	28.23	60.13	25.15
	II	84.62	7.27	91.68	3.83	89.80	3.82	68.60	19.71	76.14	15.01	77.71	15.13
	III	90.89	4.83	94.66	3.06	93.88	3.95	73.00	17.02	80.85	13.27	80.38	13.28
i5	**I**	64.72	21.47	70.87	16.65	72.44	15.35	45.35	32.67	49.29	30.58	53.54	27.51
	II	75.75	13.22	85.20	8.13	84.25	8.27	63.46	21.25	72.91	17.55	71.97	17.09
	III	80.16	10.10	89.45	6.69	88.82	6.83	71.81	16.92	79.37	13.48	77.80	14.07
i6S	**I**	63.94	20.96	68.03	18.77	71.02	17.45	44.72	31.08	51.65	27.77	56.38	26.50
	II	82.36	10.40	88.66	6.86	86.93	7.81	69.61	18.72	77.01	15.19	76.85	16.20
	III	87.09	7.39	92.44	5.12	89.92	6.81	73.86	16.39	81.73	12.84	79.84	14.62
P8	**I**	60.85	23.04	70.13	18.63	72.17	14.51	45.91	30.61	54.87	27.98	57.08	25.84
	II	75.47	12.85	85.85	6.88	87.11	6.31	65.57	18.68	78.14	13.55	78.77	12.18
	III	83.18	9.30	90.41	4.99	89.47	5.98	74.84	15.05	83.18	10.71	83.02	10.05

Table 8.5 Recognition Rates (RR) and Equal Error Rates (EER) corresponding to device-specific training scenario, GT-Pred and Pred-GT recognition scenarios.

Device	Scenario	GT-Pred (Train: ROI_{GT}, Test: ROI_{Pred})						Pred-GT (Train: ROI_{Pred}, Test: ROI_{GT})					
		128x128		64x64		32x32		128x128		64x64		32x32	
		RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)	RR(%)	EER(%)
G4	I	34.22	37.42	38.15	34.79	43.17	33.08	41.44	33.71	48.19	30.50	52.28	29.33
	II	59.65	23.79	70.80	19.21	69.07	19.09	66.41	21.30	73.16	16.54	75.20	15.04
	III	66.09	20.38	75.20	16.68	74.25	16.91	75.20	16.25	80.53	12.58	80.85	12.69
GS6	I	33.91	36.18	40.97	33.59	43.17	31.99	39.09	34.18	48.04	30.21	52.59	29.04
	II	60.13	22.64	70.64	18.87	71.90	17.90	58.71	23.81	69.86	19.32	71.90	17.77
	III	71.27	17.84	77.55	15.43	78.81	15.06	69.07	18.07	77.71	15.78	78.34	14.38
i5	I	37.48	35.63	44.57	32.04	46.61	31.08	36.22	35.84	44.25	32.51	47.24	30.51
	II	56.54	23.35	67.72	18.90	68.66	18.22	54.33	24.55	69.61	19.13	73.23	17.01
	III	66.14	19.55	77.01	15.63	76.54	13.72	69.29	17.64	79.69	13.13	80.79	12.99
i6S	I	34.96	36.47	41.26	34.08	45.04	32.56	37.95	35.22	45.20	32.50	45.98	30.90
	II	57.01	24.91	67.09	20.32	67.87	20.08	63.31	19.23	78.11	14.41	75.12	14.28
	III	65.98	21.08	74.80	17.31	74.80	17.73	74.33	14.00	85.04	9.42	81.57	11.10
P8	I	32.55	37.31	40.57	33.31	47.64	31.00	37.89	34.41	45.44	31.82	48.58	30.04
	II	50.31	27.26	68.71	17.90	69.34	16.07	59.59	22.74	71.23	17.45	73.11	15.95
	III	65.09	21.49	76.57	14.55	77.36	14.65	72.17	17.26	80.82	12.00	80.82	11.87

The best performing device-training scenario in the GT-Pred case corresponds to i5, with 76.54% RR and 13.72% EER. This outcome is not surprising, as i5 contains the most failed ROI extractions among the five device cases of the palmprint database. As the GT-Pred case uses ROI_{GT} for training and ROI_{Pred} for testing, the problem associated to the i5 failed ROIs is removed.

It is interesting to note that unlike the results from GT-exclusive, Pred-exclusive in the device-specific as well as cross-device scenarios, in GT-Pred case there are 3 (out of 5) cases where the optimum result (for that category) corresponds to using ROIs of size 32x32 instead of 64x64. For the GT-Pred case the number of best results corresponding to ROIs of size 32x32 is 4 out of 5. The intuition behind this phenomenon is that a strong down-sampling attenuates the impact noise, or in our case the misalignment, has when the distance is being computed between two images.

The best result for the Pred-GT case is reported by i6S, with 85.04% RR and 9.42%. This result reflects how the failed ROI extraction was defined, as having an IoU of less than 50%. The fact that i6S has the best result in this case means that its failed ROIs are still relevant for recognition (as compared to the ROI_{GT} images from the other device-cases). In fact this result outperforms the best result obtained in the Pred-exclusive case (83.81% RR and 10.71% EER).

The least performing device scenario in the Pred-GT case was expected to be i5, but instead it is GS6 (78.34% RR and 14.38% EER), which suggest a considerable difference between the ROI_{Pred} samples from the GS6 device case relative to the ROI_{GT} samples from the other device cases.

As it was observed in Tables 8.3 and 8.4, the extended training scenario III also corresponds to the best results in GT-Pred, as well as Pred-GT cases. The EER difference between scenario I and scenario III is comparable to what was described previously for the cross-device case, of around 13-17%.

8.4.3 Using IoU Distribution to Interpret Recognition Rates

It is difficult to quantify the influence of the ROI IoU distribution on the recognition task, regardless of the device case (CD or DS). The distributions of ROI_{IoU} for the training and testing sets are so similar that one would be unable to accurately estimate the effects of misalignment in a Pred-exclusive scenario.

We can conclude that ROI_{IoU} alone is not sufficient to quantitatively estimate the impact of misalignment on the task of recognition.

8.5 Conclusion and Future Work

The current chapter introduced a series of experiments using palmprint ROIs extracted with either ground-truth (ROI_{GT}) or predicted (ROI_{Pred}) landmarks. Several scenarios are explored, of which (as expected) the scenario using only ROIs extracted with ground truth points reported the best result of 3.06% EER when using the GS6 images for training. This corresponds to the conclusion obtained in Section 4 of Chapter as well.

A surprising result was obtained when training with ROI_{Pred} and testing with ROI_{GT} in the device-specific case, which reported better results (9.42% EER) than when training and testing with ROI_{Pred} (10.71% EER). This is attributed to the specific cases of ROI extraction failures and how similar the images used for training are to the test images. This behavior is not observed in the cross-device case. We can conclude that it is recommended to maintain the same ROI extraction strategy for both training and testing distributions of images.

We propose several ideas for future work that can aid the comparison trial stage, based on a predictor that would take into account the following:

1. the IoU between ROI_{GT} and ROI_{Pred} (as explored in this chapter). This is partially accounted for by including an extending training scenario, as described in Section 8.3.3 of this Chapter. Furthermore, to better account for scale variation of palmprint ROIs, the translation operation should be relative to a specific distance (i.e. using the landmarks used for ROI extraction) instead of having a fixed value of 50 pixels.

2. the relative rotation difference between ROI_{GT} and ROI_{Pred}. This is partially mitigated by the extended training process, but should include more rotations.

3. image quality assessment tools, to evaluate how much information is contained in the image. Especially in smartphones, when faced with low light conditions the on-board image processing techniques resort to heavy smoothing (to reduce noise), ultimately removing many of the low-level features found on palmprints.

Future work also includes exploring ROI alignment approaches in the recognition pipeline, as well as classification strategies that are robust to rotation misalignment.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

In this dissertation, smartphone user authentication using palmprint recognition was studied.

The palmprint is a powerful and valid biometric characteristic, its large scale adoption on consumer devices being 'one camera' away. This was the main hypothesis behind the work contained in this thesis, formulated in Chapter 2. The migration and deployment of palmprint recognition onto consumer devices are not just feasible, but likely to happen in the foreseeable future. Chapter 3 presented a detailed literature review of palmprint recognition, focusing on the pipeline of a biometric recognition system: acquisition (databases), palmprint ROI extraction, feature extraction techniques and classification strategies. As hand databases become more unconstrained, the solutions for ROI extraction need to become more robust.

The challenging environment of operation, where acquisition is affected by many factors raises significant challenges for any biometric recognition system relying on the device's camera. The fact that hands change shape in such a dynamic way (as opposed to faces or irises, where it is unusual for them to not be vertically aligned) has imposed certain constraints onto the hand's posture in conventional palmprint recognition systems. However, a constraint-free environment imposes the requirement to normalize the palmprints, which are affected by the hand's pose. All these factors limit the use of conventional pipelines for palmprint pre-processing, as they rely on background removal. Initial work on hand segmentation using skin color thresholding was presented in Section 5.1 of Chapter 5. The best results were reported when using the YCbCr colorspace, with an approach based on K-means clustering of pixels. Furthermore, this clustering approach was part of a Proof of Concept system relying on the iterative detection and removal of Harris corners, with the main aim of extracting a fixed-size ROI centered on the palm. Considering this solution was tested on images where hands had several orientations, a feature that was robust to both scale

and rotation was required, which is why SIFT descriptors were used. However, this solution for ROI extraction was not suitable as it did not take into account the size/scale of the hand. This limited its use to either a local region of the palmprint (if most of the image was covered by the palm) or included too many background pixels (if the hand occupied a small region of the image).

One of the key contributions of this thesis was the release of two hand databases acquired in unconstrained or very challenging conditions. The hand database NUIG_Palm1, described in Section 4.2 of Chapter 4, was the first of its kind and has already made an impact in the research community.

At the time of writing of this thesis, there were few solutions using Deep Learning to extract the palmprint ROI. This is partly explained by the lack of hand databases that could be used for training such data-hungry solutions. NUIG_Palm2, introduced in Section 4.3 of Chapter 4, addressed this and is suitable for training Deep Learning solutions for palmprint ROI extraction.

A proof of concept was introduced in Chapter 6, where baseline results (to be used for future reference) were obtained with a Collaborative Representation Classifier with Regularized Least Squares (CRC_RLS). The CRC_RLS was compared with several conventional classifiers (Fisher's Discriminant Analysis, K-Nearest Neighborhood and Support Vector Machine with linear kernel) and outperformed them. The CRC_RLS is suitable for a smartphone-based biometric recognition system as it has a light computational load, is very fast and can be embedded. The lowest Equal Error Rate was achieved when training with images from several lighting conditions, in the device-specific case (average EER of 7.4%). A series of state-of-the-art feature extraction techniques for extracting palmprint-specific features were implemented and their performance compared. The Difference of Vertex Normal Vectors (DoN) was determined to perform the best in conjunction with the CRC_RLS classifier.

Chapter 7 investigated the potential of landmark regression to be used for palmprint ROI extraction. This strategy was especially promising when dealing with unconstrained hand pose. Using a Leap Motion device placed next to a webcam, a large (labeled) database of hands was collected in a short time (NUIG_Palm2) and used to train several Deep Learning networks for the task of 2D key-point regression. A novel architecture was suggested, which was shown to outperform other approaches on several hand databases. An evaluation strategy that normalizes the prediction error was introduced and applied to several training/testing scenarios. Besides the key-point prediction error, which was normalized to the palmprint's scale, the Intersection over Union (IoU) was another metric used to evaluate the performance of the 9 Convolutional Neural Networks (CNNs) considered for experiments.

Chapter 8 provided an analysis of the previously extracted palmprint ROIs using the proposed network described in Chapter 7. Several recognition scenarios were defined, according to the attributes of the palmprint database NUIG_Palm1. The impact of misalignment was investigated by attempting to correlate the rate of IoU for palmprint ROIs based on their distribution in the training/testing sets. Unfortunately, the IoU alone proved to be insufficient to estimate the effect of ROI misalignment on the Recognition Rate. To mitigate the impact of inaccurate ROI extraction, an extended training strategy was shown to increase the EER from 28% to 16% in the cross-device scenario, and from 28% to 11% in the device-specific scenario. Based on several ROI extraction scenarios (using either ground truth or predicted landmarks for the train/test sets) the obtained results recommended to maintain the same ROI extraction strategy in both training and testing sets.

9.2 Future Work

One potentially significant future work would be to improve the landmark prediction inference time (as in YOLO [204] or Faster R-CNN [72]) and also include a rotation parameter associated to the box prediction of bounding boxes. This would adapt the bounding box prediction strategy to a palmprint ROI prediction approach, which would take into account the hand's orientation.

Furthermore, a stage including the affine transformation of the ROI could be considered, as the ROIs of hands having an out-of-plane rotation will appear distorted. This can be achieved by recovering the 3D hand pose structure from a 2D hand image by using a CNN [189]. Alternatively, a post ROI extraction step for alignment should be considered, similar to [205] where SIFT features from two samples are compared and used to align them. This alignment stage can act as a preliminary stage of authentication, as the number of SIFT matches indicates whether or not the query sample belongs to that class.

To support the deployment of palmprint recognition solutions onto consumer electronics devices, specific strategies could be investigated for reducing the size (and number of computations) that a neural network is expected to have, without affecting the overall performance. Such strategies include Knowledge Distillation [140], which considerably reduces the number of computations in a network while lowering the classification accuracy by several points. As reported by Shao *et al.* [34], the run-time was improved fourfold. The development of new approaches for network distillation represents a viable solution to situations where a compromise has to be reached between run-time and performance.

Finally, the multiple-device acquisition of palmprints could also be taken into consideration, using palmprint databases that were inspired by NUIG_Palm1 (the first database of its kind) but larger in size (e.g. XJTU-UP [34] or MPD [35]).

References

[1] S. Naji, H. A. Jalab, and S. A. Kareem, "A survey on skin detection in colored images," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 1041–1087, 2019.

[2] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, 2016.

[3] P. Corcoran and C. Costache, "Biometric technology and smartphones: A consideration of the practicalities of a broad adoption of biometrics and the likely impacts," in *2015 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 1–7, IEEE, 2015.

[4] W. Jansen, "Authenticating users on handheld devices," in *Proceedings of the Canadian Information Technology Security Symposium*, pp. 1–12, 2003.

[5] "iphone 5s specifications (gsmarena)." URL="gsmarena.com/res.php3?sSearch=iphone+5s" [accessed on 24.10.2019].

[6] T. AA, 25th October 2017. URL:"androidauthority.com/new-palmid-tech-lets-unlock-devices-wave-hand-809956/" [accessed on 5.11.2019].

[7] "First biometric authentication solution for consumer augmented reality headsets." URL:"helpnetsecurity.com/2018/05/09/augmented-reality-authentication/".

[8] S. Agarwal, "What is kaios and why is it the 3rd most popular mobile os?," 11th April 2019. URL="https://www.kaiostech.com/" accessed on [24.10.2019].

[9] J. Davis, "Kaios shows why it is critical to africa's digital ambition," 15th November 2019. URL="https://telecoms.com/500934/kaios-shows-why-it-is-critical-to-africas-digital-ambition/" accessed on 20.04.2020.

[10] W. Jia, R. X. Hu, J. Gui, Y. Zhao, and X. M. Ren, "Palmprint Recognition Across Different Devices," *Sensors (Switzerland)*, vol. 12, no. 6, pp. 7938–7964, 2012.

[11] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to biometrics*. Springer Science & Business Media, 2011.

[12] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[14] S. Z. Li, *Encyclopedia of Biometrics: A.*, vol. 2. Springer Science & Business Media, 2009.

[15] A. K. Jain and J. Feng, "Latent palmprint matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 31, no. 6, pp. 1032–1047, 2009.

[16] A. W.-K. Kong, D. Zhang, and G. Lu, "A study of identical twins' palmprints for personal verification," *Pattern Recognition*, vol. 39, no. 11, pp. 2149–2156, 2006.

[17] The Hong Kong Polytechnic University, "PolyU Palmprint Palmprint Database, "http://www.comp.polyu.edu.hk/˜biometrics/"."

[18] Indian Institute of Technology Delhi, "IIT Delhi Touchless Palmprint Database (Version 1.0)," 2014. URL:"http://web.iitd.ac.in/ajaykr/Database_Palm.htm".

[19] A.-S. Ungureanu, S. Thavalengal, T. E. Cognard, C. Costache, and P. Corcoran, "Unconstrained palmprint as a smartphone biometric," *IEEE Transactions on Consumer Electronics*, vol. 63, pp. 334–342, aug 2017.

[20] Bogazici University, Istanbul, Turkey, "Bosphorus Hand Database." http://bosphorus.ee.boun.edu.tr/hand/Home.aspx.

[21] The Chinese Academy of Sciences, Automation Institute , "CASIA Palmprint Database." http://biometrics.idealtest.org/.

[22] College of Engineering, Pune-411005(An Autonomous Institute of Government of Maharashtra), "Palmprint Dataset." URL:http://www.coep.org.in/resources/coeppalmprintdatabase.

[23] M. A. Ferrer, F. Vargas, and A. Morales, "Bispectral contactless hand based biometric system," in *CONATEL 2011*, pp. 1–6, IEEE, 2011.

[24] L. Zhang, L. Li, A. Yang, Y. Shen, and M. Yang, "Towards contactless palmprint recognition: A novel device, a new benchmark, and a collaborative representation based identification approach," *Pattern Recognition*, vol. 69, pp. 199 – 212, 2017.

[25] A. Kumar, "Toward more accurate matching of contactless palmprint images under less constrained environments," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 1, pp. 34–47, 2018.

[26] Q. Xiao, J. Lu, W. Jia, and X. Liu, "Extracting palmprint roi from whole hand image using straight line clusters," *IEEE Access*, 2019.

[27] E. Yoruk, H. Dutaugaci, and B. Sankur, "Hand biometrics," *Image and vision computing*, vol. 24, no. 5, pp. 483–497, 2006.

[28] S. Aoyama, K. Ito, T. Aoki, and H. Ota, "A Contactless Palmprint Recognition Algorithm for Mobile Phones," in *Proceedings of International Workshop on Advanced Image Technology 2013*, (Nagoya), 2013.

[29] J. S. Kim, G. Li, B. Son, and J. Kim, "An empirical study of palmprint recognition for mobile phones," *IEEE Transactions on Consumer Electronics*, vol. 61, pp. 311–319, aug 2015.

[30] K. Tiwari, C. J. Hwang, and P. Gupta, "A palmprint based recognition system for smartphone," in *2016 Future Technologies Conference (FTC)*, pp. 577–586, IEEE, 2016.

[31] M. Izadpanahkakhk, A. Uncini, S. H. Zahiri, M. T. Gorjikolaie, and S. M. Razavi, "Novel mobile palmprint databases for biometric authentication," *International Journal of Grid and Utility Computing*, 2019.

[32] M. Choraś and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Analysis and Applications*, vol. 15, no. 1, pp. 73–85, 2012.

[33] M. Afifi, "11k hands: gender recognition and biometric identification using a large dataset of hand images," *Multimedia Tools and Applications*, pp. 1–20, 2017.

[34] H. Shao, D. Zhong, and X. Du, "Deep palmprint recognition via distilled hashing coding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.

[35] Y. Zhang, L. Zhang, X. Liu, S. Zhao, Y. Shen, and Y. Yang, "Pay by showing your palm: A study of palmprint verification on mobile platforms," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 862–867, IEEE, 2019.

[36] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[37] L. Leng, G. Liu, M. Li, M. K. Khan, and A. M. Al-Khouri, "Logical conjunction of triple-perpendicular-directional translation residual for contactless palmprint preprocessing," in *2014 11th International Conference on Information Technology: New Generations*, pp. 523–528, IEEE, 2014.

[38] D. D. Zhang, W. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Transactions on pattern analysis and machine intelligence*, 2003.

[39] Y. Zhou and A. Kumar, "Human identification using palm-vein images," *IEEE transactions on information forensics and security*, vol. 6, no. 4, pp. 1259–1274, 2011.

[40] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *2008 15th IEEE International Conference on Image Processing*, pp. 281–284, IEEE, 2008.

[41] G. Badrinath and P. Gupta, "Palmprint based recognition system using phase-difference information," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 287–305, 2012.

[42] K. Tiwari, D. K. Arya, G. Badrinath, and P. Gupta, "Designing palmprint based recognition system using local structure tensor and force field transformation for human identification," *Neurocomputing*, vol. 116, pp. 222–230, 2013.

[43] M. Hammami, S. B. Jemaa, and H. Ben-Abdallah, "Selection of discriminative sub-regions for palmprint recognition," *Multimedia tools and applications*, vol. 68, no. 3, pp. 1023–1050, 2014.

[44] K. Ito, T. Sato, S. Aoyama, S. Sakai, S. Yusa, and T. Aoki, "Palm region extraction for contactless palmprint recognition," in *2015 International Conference on Biometrics (ICB)*, pp. 334–340, IEEE, 2015.

[45] A. Poinsot, F. Yang, and M. Paindavoine, "Small sample biometric recognition based on palmprint and face fusion," in *2009 Fourth International Multi-Conference on Computing in the Global Information Technology*, pp. 118–122, Aug 2009.

[46] W. Chen, Y. Chiang, and Y. Chiu, "Biometric verification by fusing hand geometry and palmprint," in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*, vol. 2, pp. 403–406, Nov 2007.

[47] N. Charfi, H. Trichili, A. M. Alimi, and B. Solaiman, "Local invariant representation for multi-instance toucheless palmprint identification," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 003522–003527, IEEE, 2016.

[48] M. K. Balwant, A. Agarwal, and C. Rao, "Online touchless palmprint registration system in a dynamic environment," *Procedia Computer Science*, vol. 54, pp. 799–808, 2015.

[49] G. K. O. Michael, T. Connie, and A. T. B. Jin, "Touch-less palm print biometric system.," in *VISAPP (2)*, pp. 423–430, 2008.

[50] M. Franzgrote, C. Borg, B. J. T. Ries, S. Bussemaker, X. Jiang, M. Fieleser, and L. Zhang, "Palmprint verification on mobile phones using accelerated competitive code," in *2011 International Conference on Hand-Based Biometrics*, pp. 1–6, IEEE, 2011.

[51] A. Morales, M. A. Ferrer, C. M. Travieso, and J. B. Alonso, "Multisampling approach applied to contactless hand biometrics," in *2012 IEEE International Carnahan Conference on Security Technology (ICCST)*, pp. 224–229, Oct 2012.

[52] T. Chai, S. Wang, and D. Sun, "A palmprint roi extraction method for mobile devices in complex environment," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 1342–1346, IEEE, 2016.

[53] X. Sun, Q. Xu, C. Wang, W. Dong, and Z. Zhu, "Roi extraction for online touchless palm vein based on concavity analysis," in *2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 1123–1126, May 2017.

[54] Z. Khan, A. Mian, and Y. Hu, "Contour code: Robust and efficient multispectral palmprint encoding for human recognition," in *2011 International Conference on Computer Vision*, pp. 1935–1942, IEEE, 2011.

[55] Y. Han, Z. Sun, F. Wang, and T. Tan, "Palmprint recognition under unconstrained scenes," in *Asian Conference on Computer Vision*, pp. 1–11, Springer, 2007.

[56] X. Liang, D. Zhang, G. Lu, Z. Guo, and N. Luo, "A novel multicamera system for high-speed touchless palm recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[57] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Transactions on Image Processing*, vol. 15, pp. 1803–1815, July 2006.

[58] W. Jia, D.-S. Huang, and D. Zhang, "Palmprint verification based on robust line orientation code," *Pattern Recognition*, vol. 41, no. 5, pp. 1504–1513, 2008.

[59] L. Shang, J. Chen, P.-G. Su, and Y. Zhou, "Roi extraction of palmprint images using modified harris corner point detection algorithm," in *International Conference on Intelligent Computing*, pp. 479–486, Springer, 2012.

[60] C. G. Harris, M. Stephens, *et al.*, "A combined corner and edge detector.," in *Alvey vision conference*, vol. 15:50, pp. 10–5244, Citeseer, 1988.

[61] H. Javidnia, A. Ungureanu, and P. Corcoran, "Palm-print recognition for authentication on smartphones," in *2015 IEEE International Symposium on Technology and Society (ISTAS)*, pp. 1–5, IEEE, 2015.

[62] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 1, pp. 122–124, IEEE, 2001.

[63] J. Doublet, O. Lepetit, and M. Revenu, "Contact less hand recognition using shape and texture features," in *Signal Processing, 2006 8th International Conference on*, vol. 3, IEEE, 2006.

[64] M. Aykut and M. Ekinci, "Developing a contactless palmprint authentication system by introducing a novel ROI extraction method," *Image and Vision Computing*, vol. 40, pp. 65–74, 2015.

[65] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874, 2014.

[66] X. Bao and Z. Guo, "Extracting region of interest for palmprint by convolutional neural networks," *2016 6th International Conference on Image Processing Theory, Tools and Applications, IPTA 2016*, no. iii, 2017.

[67] M. Izadpanahkakhk, S. Razavi, M. Taghipour-Gorjikolaie, S. Zahiri, and A. Uncini, "Deep Region of Interest and Feature Extraction Models for Palmprint Verification Using Convolutional Neural Networks Transfer Learning," *Applied Sciences*, vol. 8, p. 1210, jul 2018.

[68] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

[69] G. Jaswal, A. Kaul, R. Nath, and A. Nigam, "Deeppalm-a unified framework for personal human authentication," in *2018 International Conference on Signal Processing and Communications (SPCOM)*, pp. 322–326, IEEE, 2018.

[70] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[71] Y. Liu and A. Kumar, "A deep learning based framework to detect and recognize humans using contactless palmprints in the wild," *arXiv preprint arXiv:1812.11319*, 2018.

[72] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[73] L. Leng, F. Gao, Q. Chen, and C. Kim, "Palmprint recognition system on mobile devices with double-line-single-point assistance," *Personal and Ubiquitous Computing*, vol. 22, no. 1, pp. 93–104, 2018.

[74] S. Minaee and Y. Wang, "Palmprint recognition using deep scattering convolutional network," *arXiv preprint arXiv:1603.09027*, 2016.

[75] A. Meraoumia, F. Kadri, H. Bendjenna, S. Chitroub, and A. Bouridane, "Improving biometric identification performance using pcanet deep learning and multispectral palmprint," in *Biometric Security and Privacy*, pp. 51–69, Springer, 2016.

[76] D. Zhang, W. Zuo, and F. Yue, "A comparative study of palmprint recognition algorithms," *ACM computing surveys (CSUR)*, vol. 44, no. 1, p. 2, 2012.

[77] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *pattern recognition*, vol. 42, no. 7, pp. 1408–1418, 2009.

[78] D. P. Dewangan and A. Pandey, "A survey on security in palmprint recognitio: a biometric trait," *Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET)*, vol. 1, p. 347, 2012.

[79] W. Jia, B. Zhang, J. Lu, Y. Zhu, Y. Zhao, W. Zuo, and H. Ling, "Palmprint recognition based on complete direction representation," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4483–4498, 2017.

[80] J. Chen, Y.-S. Moon, M.-F. Wong, and G. Su, "Palmprint authentication using a symbolic representation of images," *Image and Vision Computing*, vol. 28, no. 3, pp. 343–351, 2010.

[81] R. Raghavendra and C. Busch, "Robust palmprint verification using sparse representation of binarized statistical features: a comprehensive study," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*, pp. 181–185, ACM, 2014.

[82] W. Jia, R.-X. Hu, Y.-K. Lei, Y. Zhao, and J. Gui, "Histogram of oriented lines for palmprint recognition," *IEEE Transactions on systems, man, and cybernetics: systems*, vol. 44, no. 3, pp. 385–395, 2013.

[83] Y.-T. Luo, L.-Y. Zhao, B. Zhang, W. Jia, F. Xue, J.-T. Lu, Y.-H. Zhu, and B.-Q. Xu, "Local line directional pattern for palmprint recognition," *Pattern Recognition*, vol. 50, pp. 26–44, 2016.

[84] Q. Zheng, A. Kumar, and G. Pan, "A 3d feature descriptor recovered from a single 2d palmprint image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 6, pp. 1272–1279, 2016.

[85] G. Li and J. Kim, "Palmprint recognition with local micro-structure tetra pattern," *Pattern Recognition*, vol. 61, pp. 29–46, 2017.

[86] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893 vol. 1, June 2005.

[87] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[88] S. Murala, R. Maheshwari, and R. Balasubramanian, "Local tetra patterns: a new feature descriptor for content-based image retrieval," *IEEE transactions on image processing*, vol. 21, no. 5, pp. 2874–2886, 2012.

[89] A.-K. Kong and D. Zhang, "Competitive coding scheme for palmprint verification," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1, pp. 520–523, IEEE, 2004.

[90] X. Wang, H. Gong, H. Zhang, B. Li, and Z. Zhuang, "Palmprint identification using boosting local binary pattern," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, pp. 503–506, IEEE, 2006.

[91] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 7, pp. 971–987, 2002.

[92] Z. Sun, T. Tan, Y. Wang, and S. Z. Li, "Ordinal palmprint represention for personal identification [represention read representation]," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 279–284, IEEE, 2005.

[93] X. Wu, K. Wang, and D. Zhang, "Palmprint texture analysis using derivative of gaussian filters," in *2006 International Conference on Computational Intelligence and Security*, vol. 1, pp. 751–754, IEEE, 2006.

[94] Z. Guo, D. Zhang, L. Zhang, and W. Zuo, "Palmprint verification using binary orientation co-occurrence vector," *Pattern Recognition Letters*, vol. 30, no. 13, pp. 1219–1227, 2009.

[95] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE transactions on instrumentation and measurement*, vol. 59, no. 2, pp. 480–490, 2009.

[96] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *2008 15th IEEE International Conference on Image Processing*, pp. 281–284, IEEE, 2008.

[97] L. Zhang, H. Li, and J. Niu, "Fragile bits in palmprint recognition," *IEEE Signal processing letters*, vol. 19, no. 10, pp. 663–666, 2012.

[98] L. Fei, Y. Xu, W. Tang, and D. Zhang, "Double-orientation code and nonlinear matching scheme for palmprint recognition," *Pattern Recognition*, vol. 49, pp. 89–101, 2016.

[99] Q. Zheng, A. Kumar, and G. Pan, "Suspecting less and doing better: New insights on palmprint identification for faster and more accurate matching," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 633–641, 2016.

[100] L. Fei, Y. Xu, and D. Zhang, "Half-orientation extraction of palmprint features," *Pattern Recognition Letters*, vol. 69, pp. 35–41, 2016.

[101] M. Tabejamaat and A. Mousavi, "Concavity-orientation coding for palmprint recognition," *Multimedia Tools and Applications*, vol. 76, no. 7, pp. 9387–9403, 2017.

[102] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964–973, 2008.

[103] G. Peters, N. Krüger, and C. Von Der Malsburg, "Learning object representations by clustering banana wavelet responses," *Proceedings of the 1st STIPR*, pp. 113–118, 1997.

[104] L. Fei, B. Zhang, Y. Xu, and L. Yan, "Palmprint recognition using neighboring direction indicator," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 6, pp. 787–798, 2016.

[105] L. Fei, J. Wen, Z. Zhang, K. Yan, and Z. Zhong, "Local multiple directional pattern of palmprint image," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 3013–3018, IEEE, 2016.

[106] Y. Xu, L. Fei, J. Wen, and D. Zhang, "Discriminative and robust competitive code for palmprint recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 2, pp. 232–241, 2016.

[107] F. Zhong and J. Zhang, "Face recognition with enhanced local directional patterns," *Neurocomputing*, vol. 119, pp. 375–384, 2013.

[108] A. R. Rivera, J. R. Castillo, and O. O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE transactions on image processing*, vol. 22, no. 5, pp. 1740–1752, 2012.

[109] L. Fei, G. Lu, W. Jia, J. Wen, and D. Zhang, "Complete binary representation for 3-d palmprint recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 12, pp. 2761–2771, 2018.

[110] S. Iitsuka, K. Ito, and T. Aoki, "A practical palmprint recognition algorithm using phase information," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, IEEE, 2008.

[111] J. Chen and Y.-S. Moon, "Using sift features in palmprint authentication," in *2008 19th International Conference on Pattern Recognition*, pp. 1–4, IEEE, 2008.

[112] A. Morales, M. Ferrer, and A. Kumar, "Towards contactless palmprint authentication," *IET Computer Vision*, vol. 5, no. 6, p. 407, 2011.

[113] Q. Zhao, X. Wu, and W. Bu, "Contactless palmprint verification based on SIFT and iterative RANSAC," in *2013 IEEE International Conference on Image Processing*, pp. 4186–4189, IEEE, sep 2013.

[114] W. Kang, Y. Liu, Q. Wu, and X. Yue, "Contact-free palm-vein recognition based on local invariant features," *PloS one*, vol. 9, no. 5, p. e97548, 2014.

[115] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, nov 2004.

[116] N. Charfi, H. Trichili, A. M. Alimi, and B. Solaiman, "Novel hand biometric system using invariant descriptors," in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 261–266, IEEE, 2014.

[117] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "Orb: An efficient alternative to sift or surf.," in *ICCV*, vol. 11:1, p. 2, Citeseer, 2011.

[118] B. G. Srinivas and P. Gupta, "Palmprint based verification system using surf features," in *International Conference on Contemporary Computing*, pp. 250–262, Springer, 2009.

[119] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, jun 2008.

[120] L. Dian and S. Dongmei, "Contactless palmprint recognition based on convolutional neural network," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 1363–1367, IEEE, 2016.

[121] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[122] A. S. Tarawneh, D. Chetverikov, and A. B. Hassanat, "Pilot comparative study of different deep features for palmprint identification in low-quality images," *arXiv preprint arXiv:1804.04602*, 2018.

[123] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[124] A. Hassanat, M. Al-Awadi, E. Btoush, A. Al-Btoush, G. Altarawneh, *et al.*, "New mobile phone and webcam hand images databases for personal authentication and identification," *Procedia Manufacturing*, vol. 3, pp. 4060–4067, 2015.

[125] R. Ramachandra, K. B. Raja, S. Venkatesh, S. Hegde, S. D. Dandappanavar, and C. Busch, "Verifying the newborns without infection risks using contactless palmprints," in *2018 International Conference on Biometrics (ICB)*, pp. 209–216, IEEE, 2018.

[126] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[127] A. Genovese, V. Piuri, K. N. Plataniotis, and F. Scotti, "Palmnet: Gabor-pca convolutional networks for touchless palmprint recognition," *IEEE Transactions on Information Forensics and Security*, 2019.

[128] L. Zhang, Y. Shen, H. Li, and J. Lu, "3d palmprint identification using block-wise features and collaborative representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1730–1736, 2015.

[129] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?," *IEEE transactions on image processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[130] A. Jalali, R. Mallipeddi, and M. Lee, "Deformation invariant and contactless palmprint recognition using convolutional neural network," in *Proceedings of the 3rd International Conference on Human-Agent Interaction*, pp. 209–212, ACM, 2015.

[131] Z. D. P. Xin, P. Xin, L. Xiaoling, and G. Xiaojing, "Palmprint recognition based on deep learning," in *6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)*, IET, 2015.

[132] L. Fei, G. Lu, W. Jia, S. Teng, and D. Zhang, "Feature extraction methods for palmprint recognition: A survey and evaluation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 2, pp. 346–363, 2018.

[133] S. Zhao, B. Zhang, and C. P. Chen, "Joint deep convolutional feature representation for hyperspectral palmprint recognition," *Information Sciences*, vol. 489, pp. 167–181, 2019.

[134] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1735–1742, IEEE, 2006.

[135] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European conference on computer vision*, pp. 499–515, Springer, 2016.

[136] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[137] J. Svoboda, J. Masci, and M. M. Bronstein, "Palmprint recognition via discriminative index learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 4232–4237, IEEE, 2016.

[138] L. Zhang, Z. Cheng, Y. Shen, and D. Wang, "Palmprint and palmvein recognition based on dcnn and a new large-scale contactless palmvein dataset," *Symmetry*, vol. 10, no. 4, p. 78, 2018.

[139] D. Zhong, Y. Yang, and X. Du, "Palmprint recognition using siamese network," in *Chinese Conference on Biometric Recognition*, pp. 48–55, Springer, 2018.

[140] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[141] H. Shao, D. Zhong, and Y. Li, "Palmgan for cross-domain palmprint recognition," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1390–1395, IEEE, 2019.

[142] X. Du, D. Zhong, and P. Li, "Low-shot palmprint recognition based on meta-siamese network," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 79–84, IEEE, 2019.

[143] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[144] H. Shao and D. Zhong, "Few-shot palmprint recognition via graph neural networks," *Electronics Letters*, 2019.

[145] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[146] V. Kanhangad, A. Kumar, and D. Zhang, "Contactless and pose invariant biometric identification using hand surface," *IEEE transactions on image processing*, vol. 20, no. 5, pp. 1415–1424, 2010.

[147] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, "Ganerated hands for real-time 3d hand tracking from monocular rgb," in *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[148] Article 29 Working Party, "Opinion 3/2012 on developments in biometric technologies," April 2012. URL:"https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/".

[149] L. E. Potter, J. Araullo, and L. Carter, "The leap motion controller: a view on sign language," in *Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, pp. 175–178, ACM, 2013.

[150] M. DiMartino, "A statistical hand gesture recognition system using the leap motion controller," Master's thesis, CALIFORNIA STATE UNIVERSITY, NORTHRIDGE, 2016.

[151] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," *2014 IEEE International Conference on Image Processing, ICIP 2014*, pp. 1565–1569, 2014.

[152] "VRgineers and Leap Motion Team Up to Create the World's First Professional VR Headset with Integrated Hand Tracking." https://vrgineers.com/vrgineers-leap-motion-team-create-worlds-first-professional-vr-headset-integrated-hand-tracking/. Accessed: 2018-08-09.

[153] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the Leap Motion Controller," *Sensors (Switzerland)*, vol. 13, no. 5, pp. 6380–6393, 2013.

[154] J. Guna, G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik, "An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking," *Sensors (Switzerland)*, vol. 14, no. 2, pp. 3702–3720, 2014.

[155] Z. Zhang, "A Flexible New Technique for Camera Calibration (Technical Report)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2002.

[156] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[157] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[158] Y. Kim, I. Hwang, and N. I. Cho, "Convolutional neural networks and training strategies for skin detection," in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3919–3923, IEEE, 2017.

[159] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, pp. 1–9, 2015.

[160] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, vol. 3, pp. 85–92, Moscow, Russia, 2003.

[161] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 1, pp. 122–124, IEEE, 2001.

[162] Y.-I. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Computer graphics and image processing*, vol. 13, no. 3, pp. 222–241, 1980.

[163] J.-C. Terrillon, M. David, and S. Akamatsu, "Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 112–117, IEEE, 1998.

[164] P. Yogarajah, J. Condell, K. Curran, A. Cheddad, and P. McKevitt, "A dynamic threshold approach for skin segmentation in color images," in *2010 IEEE International Conference on Image Processing*, pp. 2225–2228, IEEE, 2010.

[165] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.

[166] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14:2, pp. 1137–1145, Montreal, Canada, 1995.

[167] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000.

[168] A.-s. Ungureanu, H. Javidnia, C. Costache, and P. Corcoran, "A review and comparative study of skin segmentation techniques for handheld imaging devices," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, (Las Vegas), pp. 530–531, IEEE, jan 2016.

[169] Q. Zhao, W. Bu, and X. Wu, "Sift-based image alignment for contactless palmprint verification," in *2013 International Conference on Biometrics (ICB)*, pp. 1–6, IEEE, 2013.

[170] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *cvpr*, vol. 97, p. 1000, Citeseer, 1997.

[171] Y. Wang, L. Xia, T. Tang, B. Li, S. Yao, M. Cheng, and H. Yang, "Low power convolutional neural networks on a chip," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 129–132, IEEE, 2016.

[172] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[173] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, pp. 474–485, Academic Press Professional, Inc., 1994.

[174] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[175] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1:10. Springer series in statistics New York, 2001.

[176] H. Zhang, Y. Zhao, F. Yao, L. Xu, P. Shang, and G. Li, "An adaptation strategy of using lda classifier for emg pattern recognition," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 4267–4270, IEEE, 2013.

[177] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[178] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *2011 International conference on computer vision*, pp. 471–478, IEEE, 2011.

[179] S. Gupta, N. Gupta, S. Ghosh, M. Singh, S. Nagpal, M. Vatsa, and R. Singh, "Facesurv: A benchmark video dataset for face detection and recognition across spectra and resolutions," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pp. 1–7, IEEE, 2019.

[180] S. Barra, A. Casanova, F. Narducci, and S. Ricciardi, "Ubiquitous iris recognition by means of mobile devices," *Pattern Recognition Letters*, vol. 57, pp. 66–73, 2015.

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

[182] H. Cheng, L. Yang, and Z. Liu, "A Survey on 3D Hand Gesture Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, p. 1, 2015.

[183] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, 2017.

[184] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Advances in neural information processing systems*, pp. 1799–1807, 2014.

[185] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 4724–4732, 2016.

[186] V. Ramakrishna, D. Munoz, M. Hebert, J. A. Bagnell, and Y. Sheikh, "Pose machines: Articulated pose estimation via inference machines," in *European Conference on Computer Vision*, pp. 33–47, Springer, 2014.

[187] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*, pp. 483–499, Springer, 2016.

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

[189] C. Zimmermann and T. Brox, "Learning to estimate 3d hand pose from single rgb images," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4903–4911, 2017.

[190] F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Large-scale multiview 3D hand pose dataset," *Image and Vision Computing*, vol. 81, pp. 25–33, 2019.

[191] F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Robust hand pose regression using convolutional neural networks," in *Iberian Robotics conference*, pp. 591–602, Springer, 2017.

[192] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, IEEE, jun 2016.

[193] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *Cvpr 2015*, p. 5, nov 2015.

[194] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

[195] S. Bazrafkan and P. Corcoran, "Semi-parallel deep neural networks (spdnn), convergence and generalization," *arXiv preprint arXiv:1711.01963*, 2017.

[196] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An end to end deep neural network for iris segmentation in unconstrained scenarios," *Neural Networks*, vol. 106, pp. 79–95, 2018.

[197] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. XX, no. Xx, pp. 1–16, 2017.

[198] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[199] F. Chollet, J. Allaire, *et al.*, "R interface to keras." https://github.com/rstudio/keras, 2017.

[200] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing Textures in the Wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3606–3613, IEEE, jun 2014.

[201] A. Hassanat, M. Al-Awadi, E. Btoush, A. Al-Btoush, G. Altarawneh, *et al.*, "New mobile phone and webcam hand images databases for personal authentication and identification," *Procedia Manufacturing*, vol. 3, pp. 4060–4067, 2015.

[202] F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Large-scale Multiview 3D Hand Pose Dataset," pp. 1–23, jul 2017.

[203] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[204] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, 2017.

[205] Q. Zhao, W. Bu, and X. Wu, "Sift-based image alignment for contactless palmprint verification," in *2013 International Conference on Biometrics (ICB)*, pp. 1–6, IEEE, 2013.

Appendix A

Application: Ethics Research Committee

OÉ Gaillimh
NUI Galway

RESEARCH ETHICS COMMITTEE APPLICATION FORM

For Applicant to complete:

Applicants' Name: Dr. Peter Corcoran

Title of Project: Publishing Database of Palmprint Biometric Data for use by research community

For Ethics Committee use only:

Reference Number: | *Date received:*

Review Date: | *Outcome:* ☐ Approval ☐ Provisional Approval ☐ Deferral ☐ Approval Declined

Applicant informed (Date):

Please complete form and select YES / NO options as appropriate. An electronic version of this form is also available on the NUI Galway website (*http: / / www.nuigalway.ie / research / vp_research / ethics.htm*).

An application will only be accepted for review by the NUI Galway Research Ethics Committee (REC) if it is completed fully and the relevant enclosures are received. Refer to the accompanying Guidance Notes when completing the form and complete the checklist on the next page before submitting the form. Where you have received permission to do this, or similar research in another institution, please provide evidence of permission with this application.

Please submit your completed application: application form; protocol; participant consent form(s); patient information sheet(s); Questionnaire(s); as one single PDF document.

Address to send application: (Hard copy with signatures)
NUI Galway Research Ethics Committee
Research Support Services
Unit 8, Business Innovation Centre
NUI Galway

Email address: (pdf) **eithne.oconnell@nuigalway.ie**

SUBMISSION CHECKLIST

Please indicate if the following have been enclosed by selecting YES/NO/Not applicable options below.
Please forward copies of the form and relevant enclosures required as outlined below.

		YES	NO	
1	Electronic Copy of **complete** application. (single PDF document – with **all** relevant attachments)	X		
1	Hard Copy of **complete** application form (with **all** attachments)	X		

		Electronic Copy YES	Hard Copy YES	NO	Not applicable
1	Participant consent form(s)	X	X		
1	Participant information sheet(s)	X	X		
1	Copy of Confirmation of Garda vetting				X
1	Sample letters (GP, Recruitment etc)				X
1	Copy of Risk Assessment Form**				X
1	Copy of Principal Applicant CV **(2A4 pages max)** (plus that of primary supervisor if principal applicant is a PhD student)	X	X		
1	Annex 1**				X
1	Annex 2***				X
1	Annex 3**** (1 copy per procedure for which risk identified)				X

* *Please indicate if not yet finalised.*
** *If the study involves the use of a new medicinal product or medical device, or the use of an existing product outside the terms of its product licence*
*** *If the study includes the use of ionizing or non-ionising radiation, radioactive substances or X rays*
**** *Please complete for each hazardous procedure*

STUDY DESCRIPTORS

Select all descriptors that apply to this study:

Competent volunteer	X	Cross-over		Biological material	
Healthy volunteer		Case-study		Foetal material	
Patient volunteer		Longitudinal		Hazardous materials	
'Incompetent' patients		Cross-sectional		Invasive procedures	
Children (under 18 yrs)		Placebo		Devices (in licence)	
Observational		Therapeutic		Medicinal products	
Interview		Controlled		(in licence)	
Questionnaire		Double-blind		Devices	
Record-based	X	Single-blind		(outside licence)	
Randomised		Prospective		Medicinal products	
Non-randomised		Retrospective		(outside licence)	

SECTION 1 Applicant(s) Details

1. Title of project:

Developing a PalmPrint Biometric Database for Internal NUIG Research Studies

2. Principal Applicant: *(All correspondence will be sent to this address unless indicated otherwise.)*

Family Name: Corcoran **Forename**: Peter **Title**: Dr.

Contact address *(for correspondence regarding application)*:
College of Engineering & Informatics, NUIG

Tel: 087-2705964 **Fax**: N/A **Email**: peter.corcoran@nuigalway.ie
dr.peter.corcoran@ieee.org

Mobile Number / Other Contact Number: 087-2705964

Present appointment of PA: Statutory Lecturer & SFI PI

Qualifications of PA: PhD (1987)

3. Other Investigator(s):

Family Name: Costache **Forename**: Claudia **Title**: Dr.

Department: College of Engineering & Informatics

Institution: NUIG

Tel: 086-1255674 **Fax**: N/A **Email**: Claudia.iancucostache@nuigalway.ie

Mobile Number / Other Contact Number: 086-1255674

Present appointment: PostDoctoral Researcher SFI

Qualifications: PhD (2011)

Family Name: Ungureanu **Forename**:Adrian-Stefan **Title**: Mr.

Department: College of Engineering & Informatics

Institution: NUIG

Tel: 083 432 4965 **Fax**: N/A **Email**: a.ungureanu1@nuigalway.ie

Present appointment: PhD Student

Qualifications: Ms.E (2014)

Family Name: Javidnia **Forename**: Hossein **Title**: Mr.

Department: College of Engineering & Informatics

Institution: NUIG

Tel: 083 460 9753 **Fax**: N/A **Email**: h.javidnia1@nuigalway.ie

Present appointment: PhD Student

Qualifications: Ms.E (2014)

4. Principal Applicant: *(All correspondence will be sent to this address unless indicated otherwise.)*

Name	Department / Institute	Appointment
Peter Corcoran	College of Engineering & Informatics, NUIG	Statutory Lecturer & SFI PI

5. Funding Sources:

(i) Has any funding been obtained / sought by the investigator in respect of this study?

Funding applied for:	YES		NO		Not applicable	
Funding secured:	YES	X	NO		Not applicable	

(ii) Name of sponsoring organisation from which funding has been obtained / sought?

The study is part of a research programme funded by Science Foundation of Ireland (Project title: Next Generation Imaging for Smartphone and Embedded Platforms; Project ID 13/SPP/I2868).

(iii) Does the Investigator(s) have any direct involvement in the sponsoring organization?

e.g. financial, share-holding etc:	YES		NO	X	Not applicable	

If YES, give details:

NOTE: Where the research programme has already received funding approval, please attach the letter of offer to this application.
Confirmation Letter held by NUIG Research Office; Note that this project is part of a larger proposal and seed-funding award to establish a new research center.

6. Proposed start date and duration of study:

Proposed Start date: 01 Sept 2015

Duration (months): 6 -12 months for initial data collection and refinement; follow-on research activity up to 5 years.

7. Signature of relevant personnel:

Principal Applicant declaration

The information in this application form is accurate to the best of my knowledge and belief and I take full responsibility for it.

I understand that it is my responsibility to obtain institutional approval where appropriate before the project takes place.

I agree to supply interim and final reports to the Research Ethics Committee from which approval was granted for this project.

I agree to advise the Research Ethics Committee from which approval was granted for this project and any local researchers taking part in the proposal of any material changes to the proposal or any adverse or unexpected events that may occur during this project.

I agree to advise the Research Ethics Committee in the event of premature termination, suspension or deferral of this project and to provide a report outlining the circumstances for such termination, suspension or deferral.

Signature of Principal Applicant: _____ **Date**: _____

Co-Signed by Supervisor where the P.A. is a Student: _____ **Date:** _____

Head of Department/Supervisor

I am fully aware of the details of this project and agree for it to continue as outlined here. I can confirm that the necessary facilities and resources are available to the researcher.

Name: _Dr. Peter Corcoran
Department: _Lead PI, Center for Cognitive, Connected & Computational Imaging_

Signature: _____ **Date**: 4/6/2015

SECTION 2 Study Details

*This section must be completed. A copy of the protocol should be enclosed with the application form but it is **not** sufficient to complete questions by referring to the protocol.*

8. Aims and objectives of study (i.e. what is the intention of the study, key research questions?)

The aim of this study is to build a database of palmprints acquired with smartphone cameras. This will enable a study of the detection/recognition rates of algorithms developed for palmprint based authentication on smartphone devices. In turn this will encourage and facilitate the further development of improved algorithms and advance the reliability of biometrics on smartphones. The end goal is improved cyber-security for mobile devices & services.

9. Scientific/theoretical background[1] to study (Approx. 250 words)

There is increasing interest in the use of smartphones to authenticate users. Today most authentication uses PIN codes and passwords but the use of fingerprints have been established in recent models of the iPhone. More recently the smartphone industry has become interested to use iris biometrics as a more robust method of authentication, but there are major challenges in acquiring iris image with smartphone and a custom optical system is needed. Palmprint has been used for biometrics in previous research and does not require special optics or additional equipment to capture the image with a modern smartphone. However there are challenges to automate the detection of the main region of interest (ROI) and to reliably extract the main features of the palm from images acquired in unconstrained conditions.
In order to test and develop suitable palmprint biometric techniques for smartphones there is a need for a public database of such images so that researchers can compare and validate new pattern recognition techniques and acceptance and rejection rates for these biometrics.

10. Description of Research (i.e. what do you intend to do?)

Our study aims to validate some current recognition algorithms by using the data acquisition capabilities of the latest smartphone models. Enhanced modes of data acquisition will also be developed and tested. The database will allow for testing and comparison of the effectiveness of different techniques.

11. List procedures or investigations involving risks to participants' well-being or safety (what, when, how often and risks associated with all procedures)

No direct risks to health, safety or well-being.
Only risk is the loss or theft of a participant's biometric data. As this will be anonymized the risk of identity theft or equivalent misuse of data is low. As palmprints are not in general use as a personal biometric the consequence of loss or theft is also low.

12. Study design (tick as appropriate)

Survey/Questionnaire		Interviews	
Case Study		• individual	
Observational		• group	
Action research		• person-to-person	
Record based	X	• telephone	
Cohort		• electronic	
Case control			
Other		Forms of Recording	
(please specify)		• Video	
		• Audio	
		• Photography	X
		• Notes	
		• Electronic recording	X

[1] A succinct background to be provided and to include reference to published work

13. Size of the study (including controls): c. 200 persons

(i) How was the size of the study determined?

Data from c. 200 persons will be gathered within a 6-12 month timeframe. This is based on similar preliminary studies in the biometric literature.

(ii) Was there formal statistical input into the overall study design? YES NO **X**

(iii) What method of analysis will be used?

A variety of image processing and pattern matching algorithms will be employed to verify quality of the data and to perform studies to determine ROC curves (ROC curve is a graphical plot that illustrates the performance of a classification system), etc. These are well-know and established techniques used in the biometric literature.

14. Where[2] will the study take place and in what setting?

Data will be gathered at NUIG in Engineering and Information Technology buildings to facilitate the access of university participants. The acquired data will be processed and stored on a computer server in a secure facility. Access to this data will be restricted to authorized SFI researchers.

15. Does the study involve:

(i) distribution of a questionnaire? YES: ☐ NO: ☒ **X**

If YES, please append a copy of the questionnaire to this application. Please indicate whether the appended questionnaire is: Non-validated: ☐ Validated: ☐

(ii) the use of a existing medicinal product or medical device? YES ☐ NO ☒ **X**

If YES, is this medical product or device being used within the terms of its current product licence?

YES ☐ NO ☐

If NO, please complete **Annex 1** of this application.

(ii) the use of a new medicinal product or medical device? YES ☐ NO ☒ **X**

If YES, please complete **Annex 1** of this application.

(iii) the use of ionising or non-ionising radiation, radioactive substances or X rays?

YES ☐ NO ☒ **X** If YES, please complete **Annex 2** of this application.

16. Peer Review / Critique[3]

Has the protocol been subject to peer review? YES ☐ NO ☒ **X**

If the review formed part of the process of obtaining funding, please give the name and address of the funding organisation:

N/A

If the review took place as part of an internal process, please give brief details:

N/A

If no review has taken place, please explain why and offer justification for this:

Non-medical study.

17. Does the study fall into any of the following categories?

Pilot:	YES [X]	NO [ ]	Not applicable [ ]
Multi-centre study	YES [ ]	NO [ ]	Not applicable [X]

If this is a multi-centre study, please complete the following details, otherwise go to question 17.

(i) Which centres are involved?

Contact Name	Department/Centre
Dr. Peter Corcoran	Center for Cognitive, Connected & Computational Imaging

(ii) Which ethics committees have been approached, and what is the outcome to date?

N/A

(iii) Who will have overall responsibility for the study?

Dr. Peter Corcoran

(iv) Who has control of the data generated?

Center for Cognitive, Connected and Computational Imaging [College of Engineering & Informatics]

SECTION 3 Recruitment of participants

18. Who is being studied?

If non-competent persons are being studied, please give details of reasons for non-competence

All persons will be competent. They are expected to operate the imaging function on a smartphone.

19. How will be the participants in the study be:

(i) Selected?

General members of the population.

(ii) Recruited? (Please append advertisement materials to application)

External advertising will not be used.
It is expected that sufficient volunteers can be recruited within the College of Engineering and Informatics and within the Industry Partner supporting this research.

20. What criteria will be used for inclusion and exclusion of participants?

(i) Inclusion criteria:

N/A

(ii) Exclusion criteria:

N/A

21. How many participants will be recruited and of what age groups?

Data from c. 200 persons will be gathered within a 6-12 month timeframe.
Ages from 18 up. We do not envisage any upper age limit as long as volunteers are competent to operate the camera function of the smartphone.

22. If applicable, how will the control group in the study be:

(i) Selected?

N/A

(ii) Recruited? (please append advertisement materials to application)

N/A

23. What criteria will be used for inclusion and exclusion of the control group?

(i) Inclusion criteria:

N/A

(ii) Exclusion criteria:

N/A

24. If applicable, how many controls will be recruited and of what age group?

N/A

25. Are the participants/controls included in this study involved in any other research investigation at the present time?

YES: ☐ NO: ☒

If YES, please give details

26. Will participants receive any payment or other incentive to participate?

YES: ☒ NO: ☐

(i) If YES, give details of incentive per participant?

A small stipend (< 50 euro) may be paid to some participants who agree to making their data part of a public dataset. This is to offset time taken to acquire the relevant biometric data.

If YES, what is the source of the incentive?

Stipend will be paid by the industry partner.

SECTION 4 Consent

27. Is written consent for participation in the study to be obtained?

YES: [X] NO: [ ]

If YES, please attach a copy of the consent form to be used *(Guidance on consent is given in the Guidance Notes)*

If NO written consent is to be obtained, please explain why

Attached.

28. How long will the subject have to decide whether to take part in the study?

(If less than 24 hours, please justify)

As the study does not involve medical data, is voluntary and as the data collection process is without any direct risk to the participant, a minimum period is not envisaged; where volunteers are well-informed and comfortable with the nature of the study data may be volunteered on the spot. However participants who are not comfortable will be encouraged to take the time to reflect overnight prior to participation.

29. Does the study include participants whom are not competent English speakers and/or do not comprehend spoken or written English?

YES: [ ] NO: [X]

If YES, give details of special arrangements made to assist these participants

30. Please attach a copy of the written participant information sheet

If NO information sheet is to be given to participants, please justify

Attached.

31. If you are recruiting from vulnerable groups (Children under 18 years of age; People with learning difficulties; Unconscious or severely ill participants; Other vulnerable groups e.g. dementia, psychological disorders, etc.), please specify and justify

N/A

(ii) What special arrangements have been made to deal with the issues of consent and assent for vulnerable participants e.g. is parental or guardian agreement to be obtained, and if so in what form?

N/A

(iii) In what way, if any, can the proposed study be expected to benefit the individual who participates?

N/A

32. *Answer this question only where invasive or other interventions are planned which could be a risk to a pregnancy*
Are women of childbearing potential included in this study?

YES: [ ] NO: [ ]

If YES, does the protocol/participant information sheet address the following:

- scientific justification
- negative teratogenic studies
- warning participants that foetus may be damaged
- requirement for initial negative pregnancy test
- forms of contraception defined

- duration of use to exceed drug metabolism
- exclude those unlikely to follow contraceptive advice
- notify investigator if pregnancy suspected.

If NO, please explain

SECTION 5 Details of interventions

33. Does the study involve the use of a new medicinal product or medical device, or the use of an existing product outside the terms of its product licence?

YES: [ ] NO: [X]

If YES, please complete Question 33 and Annex 1 of the Application Form.

34. Does the study involve investigations and/or interventions on either participants or controls?

(Please tick YES/NO as appropriate. If YES, details should be available in the protocol)

Investigation/Intervention		YES		NO
Self completion questionnaires		YES	X	NO
Interviews/interview administered questionnaires		YES	X	NO
Video/audio tape recording		YES	X	NO
Physical examination		YES	X	NO
Internal physical examination		YES	X	NO
Venepuncture*		YES	X	NO
Arterial puncture*		YES	X	NO
Biopsy material*		YES	X	NO
Other tissue/body sample*		YES	X	NO
Imaging investigation (not radiation)	X	YES		NO
Other investigations not part of normal care		YES	X	NO
Additional out patient attendance		YES	X	NO
Longer inpatient stays		YES	X	NO
Local anesthesia		YES	X	NO
General anesthesia		YES	X	NO

Other – please detail

N/A

Please indicate and justify where treatment is withheld as a result of taking part in the project.

N/A

35. Will any ionising or non-ionising radiation, or radioactive substances or X-Rays be administered to a participant?

YES: [ ] NO: [X]

If YES, please compete Annex 2 of the Application Form.

36. Where research conducted in a general practice setting, will all GPs whose patients will be involved, be required to sign to indicate that they are aware of and in agreement with the planned project?

YES: [ ] NO: [ ] Not applicable: [X]

If NO, please explain why not

N/A

* *Please see Guidance Notes*

SECTION 6 Risks and ethical problems

37. Are there any potential risks to participants?

YES: [ ] NO: [X]

If YES, please complete **Annex 3** for each procedure for which a potential risk occurs.

38. Could this study cause any discomfort or distress, either physical or mental?

YES: [ ] NO: [X]

If YES, estimate the degree and likelihood of discomfort or distress entailed and the precautions to be taken to minimise them.

N / A

Please include other potential embarrassments to the subject that should be explained prior to obtaining consent (e.g. state of undress etc)

39. What particular ethical problems or issues do you consider to be important or difficult with the proposed study?

Study involves creation of a database of personal biometric palmprint data. As personal data is involved, every measure must be taken to ensure the data cannot be used for other purposes than what is specified in the project and in information sheet. Secure storage and data anonymity should be ensured and subjects should be adequately informed.
In order to ensure all potential ethical issues regarding data protection are addressed in our project, we have consulted the following:
(i) Article 29 Data Protection Working Party - Opinion 3/2012 on developments in biometric technologies (1); attached with this application.
(ii) EU state of the art research in biometrics field and best practices for privacy and data protection, project BEAT (Biometrics Evaluation and Testing) (2). The deliverables **D9.2: Guidelines for Privacy and Data Protection** document output from BEAT project is attached to this application.
(iii) We have contacted the Data Protection Commissioner and awaiting a response on Irish public policy/guidelines. If different from European guidelines, this will be taken into account once we have a response.

(i) Will treatments provided during the study be available if needed at the end of the study?

YES: [ ] NO: [ ] Not applicable: [X]

(ii) If NO, is this made clear in the participant information sheet?

YES: [X] NO: [ ]

If NO, please give reasons

N / A

SECTION 7 Indemnity

Product liability and consumer protection legislation make the supplier and producer (manufacturer) or any person changing the nature of a substance, e.g. by dilution, strictly liable for any harm resulting from a consumer's use of a product.

(Please refer to Page 8 of the '*Guidance Notes on Completing the Application Form*' for information on indemnity.)

40. Arrangements for indemnification[4] / compensation

(i) What arrangements have been made to provide indemnification and/or compensation in the event of a claim by, or on behalf of, a participant for negligent harm?

Considered low risk in our study. Biometric data is not widely used at present and a live biometric is required. In addition authentication procedures with biometrics are invariably supervised. Future use of biometrics in unsupervised contexts (e.g. via smartphone) may pose a future risk. However the stored data will be only available in anonymous form so this is considered a hypothetical scenario with very low risk as unsupervised authentication workflows always incorporate liveness detection.

(ii) What arrangements have been made to provide indemnification and/or compensation in the event of a claim by, or on behalf of, a participant for non-negligent harm?

The only potential for harm to occur to a participant would require a criminal act on the part of a 3rd party. The terms & conditions of the consent form will explicitly preclude harm arising from such criminal acts.

(iii) Will an undergraduate student be involved directly in conducting the project?

YES: ☐ NO: ☒

41. In cases of equipment or medical devices, have appropriate arrangements been made with the manufacturer to provide indemnity?

YES: ☐ NO: ☐ Not applicable: ☒

If YES, please give details and enclose a copy of the relevant correspondence with this application

42. In cases of medicinal products, have appropriate arrangements been made with the manufacturer to provide indemnity?

YES: ☐ NO: ☐ Not applicable: ☒

If YES, please give details and enclose a copy of the relevant correspondence with this application

[4] *Where there is more than one institution /organisation involved in the study, each institution /organization is responsible for its own indemnity cover, and confirmation of such cover must be appended to the application.*

SECTION 8 Confidentiality

43. Will the study include the use of any of the following?

Audio/Video recordings YES: [X] NO: [ ]

Observation of participants: YES: [ ] NO: [ ]

If YES to either:

(i) How are confidentiality and anonymity to be ensured?

Digital data will be identified by "subject ID", rather than name. There will be no digital records linking Subject ID with person identity. Stored data will be limited to the relevant biometric. Some metadata might also be retained, limited to: age range, gender, race. As the study does NOT involve face data, no facial records will be retained in the database. Only extracted hand and palmprint data will be retained. Data will be stored according to subject ID.

(ii) What arrangements have been made to obtain consent for these procedures?

A participant Consent fomr is to be signed by volunteers.

(iii) What will happen to the tapes at the end of the study?

The non-biometric elements (Consent forms in paper format) will be deleted.
Specifically, biometric elements such as palm image will be retained as elements of the database.

44. Will the study data be held on computer?

YES: [X] NO: [ ]

If YES, will the data be held so that participants cannot be identified from computer files (i.e. no name, address, medical chart number or other potential identifier such as GMS or RSI number?

YES: [X] NO: [ ]

If NO, please give reasons

45. Will records (preferably paper records) linking study participant ID with identifying features be stored confidentially? (Please refer to the REC policy on Data Retention: http://www.nuigalway.ie/research/vp_research/documents/ethics_committee_docs/datapolicy.pdf)

YES: [X] NO: [ ]

Please give details of arrangements for confidential storage

The Consent Forms which are the only records linking Subject ID to person identity will be kept in a secure location on NUIG premises by the principal investigator. The digital data will be kept in a separate physical location on a computer server.

For how long will records be retained prior to destruction?

The records will be retained for internal use for the duration of the project, up to 5 years. Once the data collection phase of the project is completed, all paper records linking Subject IDs to names will be destroyed.

46. Will the participants' medical records be examined by investigators in the study?

YES: ☐ NO: ☒

If YES, will information relevant **only** to this study be extracted: YES: ☐ NO: ☐ Not applicable: ☒

(i) If extra information is extracted, please justify

N/A

(ii) What, if any, additional steps have been taken to safeguard the confidentiality of personal medical records?

N/A

47. Will research workers outside the employment of NUI Galway examine medical or other personal records?

YES: ☒ NO: ☐

If YES, it is the responsibility of the Principal Applicant to ensure that research workers understand that: Information obtained about and from research participants is confidential to the study and must not be divulged except in legitimate methods of study data presentation or exceptional circumstances as discussed and agreed with the principal investigator.

Please ensure that you complete the checklist on the front cover of this application form and include all relevant enclosures.

THANK YOU.

ANNEX 1

This form is to be used if the study involves the use of a new medical product or medical device, or the use of an existing product outside the terms of its product licence.

(i) Does this project have Irish Medicines Board approval or has an application been made?

YES: ☐ NO: ☐ Not applicable: ☐ Application is at present with IMB: ☐

If approval applied for, state date of application: ☐

(ii) Is a pharmaceutical or commercial company arranging this trial?

YES: ☐ NO: ☐

If YES, attach indemnification.
If NO, has the licensing authority been notified? YES: ☐ NO: ☐

(iii) Does the drug(s) or medical device have a product license(s) for the purpose for which it is to be used?

YES: ☐ NO: ☐

If YES, please give details

(iv) Is any drug or medical device being supplied by a company with a Clinical Trial Exemption Certificate or in response to an investigator with a Clinical Trial Exemption, or Doctors' Exemption?

YES: ☐ NO: ☐

If YES, give details of:

Clinical Trial Certificate Number: ☐

Clinical Trial Exemption Number: ☐

Doctors' Exemption Number: ☐

(v) Details of drug use or medical device *(please complete the table below)*

Approved name: ☐

Generic name: ☐

Trade name: ☐

Strength	Dosage	Frequency	Route	Duration of course

(vii) Who will administer the drug or fit the medical device?

(viii) If a medical device, has the device been through acceptance and safety testing?

YES ☐ NO ☐

Please give details

(ix) Who is supplying the drug(s)/medical device? (If imported, name country)

(x) Who will dispense the drug(s)/medical device?

What is their qualification to dispense the drug(s)/medical device?

(xi) Does the organisation and performance of this trial conform to European Directives on Good Clinical Practice?

YES ☐ NO ☐

If no, please detail and explain

ANNEX 2

This form is to be used if the study involves the use of ionizing or non-ionising radiation, radioactive substances or X-Rays. A competent Radiation Protection Advisor must be involved in implementing this section.

A. RADIOACTIVE SUBSTANCES

(i) Details of substances to be administered *(please complete the table below)*

Investigation	Radionucleide	Chemical form	Quantity of radioactivity to be administered (MBq)	Route	Frequency

(ii) Estimated Effective Dose (Effective Dose Equivalent) (mSv)
(Please supply source of reference or attach calculation)

(iii) Absorbed dose to organ or tissues concentrating radioactivity (mGy) (Specify dose and organ)
(Please supply source of reference or attach calculation)

(iv) Administration of Radioactive Substances Advisory Committee certificate holder to oversee / administer substance

Name of Person:	
Position:	
Certificate No.:	

I have assisted in and approve the protocol and arrangements that have been made in this project for the administration of the radioactive substance(s).

Signature: _____ Date: _____

B. X-RAYS

(i) Details of radiographic procedures *(please complete the table below)*

Investigation	Organs	Frequency

(ii) Estimated Effective Dose (Effective Dose Equivalent) (mSv)
(Please supply source of reference or attach calculation)

C. NON IONISING RADIATION

(i) Details of procedures *(please complete the table below)*

Investigation	Organs	Frequency

(iv) Who has given safety advice?

Name of Person:	
Position:	
Qualification to advise:	

I have assisted in and approve the safety of the protocol and arrangements that have been made in this project

Signature: _____ Date: _____

ANNEX 3

Risk Assessment Form – Procedures Involving Human Subjects

Procedure no.:

Title of Procedure:

Name of Assessor(s):

Assessment Date:

Does this procedure already have ethical approval?YES ☐ NO ☐

If YES, enter Approval No. and Expiry Date: Approval No:

Expiry Date:

1. Please provide a brief description of the procedure;

2. Location in which the Procedure will take place

(e.g. Research Laboratory – Room No. , Teaching Laboratory – Room No., Hospital clinic – specify, etc)

3. Subject(s) to be used

How many human participants?	

(tick as appropriate)

Undergraduate student(s)	☐
Postgraduate student(s)	☐
University staff or campus personnel	☐
Members of the general public	☐

4. What is the level of any potential risks for participants? [To be explained BEFORE obtaining consent]

None	☐
Minimal only	☐
Moderate	☐
Significant	☐

(ii) If the risk is other than minimal, please give details and likelihood of risk occurrence

(ii) If the risk is other than minimal, please give details of precautions taken to minimise the risk

5. Actions to be taken in the event of adverse response or medical emergency

Please provide details of arrangements to deal with adverse events, including reporting to the relevant authorities and follow-up

6. Appropriate level of supervision required for procedure (please tick as appropriate)

Post-graduate researcher	
Research/ lecturing Staff	
Paramedical personnel	
Medical personnel – Nurse	
Medical personnel – Doctor	
Medical personnel – Other	

If other personnel, please specify title and/or required qualification

7. Other documentation required for this assessment

Pre-test subject questionnaire	
Detailed protocol	
Other	

If other documentation is required, please describe

8. Signature

Signed: _____ **Date:** _____

Signature of Principal Applicant

FOR COMPLETION BY HEAD OF DEPARTMENT

Risk Assessment Form – procedures involving human subjects

In the Department / Institute / Center of:

Procedure no.:

Title of Procedure:

Name of Assessor(s):

Assessment Date:

9. Approval of Procedure

- ☐ Granted
- ☐ Subject to conditions (see below)
- ☐ Refer to Hospital Ethics Committee

Other, please specify

10. Comments and / or conditions

11. Signature

Signed: _____ **Date:** _____

Signature of Head of Department / Centre

(Please copy this Annex as necessary)

Reference list

1. Article 29 Data Protection Working Party - Opinion 3/2012 on developments in biometric technologies http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf
2. Biometrics Evaluation and Testing (BEAT) https://www.beat-eu.org/

Appendix B

Consent Form

OÉ Gaillimh
NUI Galway

Dept. of Electrical & Electronic Engineering
College of Engineering & Informatics
National University of Ireland, Galway
University Road, Galway Ireland
Phone: +353 91 524411
Fax:

Consent Form

Participant Identification Number: ……

Name of Researcher: ……

I, (name of subject) ……

Agree to take part in the research project:

Developing a PalmPrint Biometric Database for Internal NUIG Research Studies

I confirm that I have read the Information Sheet dated …… for the above study and have had the opportunity to ask questions. I confirm that the nature, demands and possible risks of the research have been explained to me and I understand and accept them. I understand that my consent is entirely voluntary and that I may withdraw at any time from the research project without explanation or penalty.

YES ☐ NO ☐

I further consent to have my biometrical data collected made publicly available for further research, subject to terms and conditions specified in the License Agreement. I understand no link between identity and biometrical data will be shared, and thus identification based on this data is not possible.

YES ☐ NO ☐

Note that under this agreement NUIG Galway accepts no liability for the consequences of malicious or criminal acts by 3rd parties and the consenter agrees that they will not pursue future claims arising from misuse of the consenter's biometric data where said misuse originates through such acts.

Name:…….. Date: ……. Signature: …….

Declaration by the Investigator

I confirm that I have provided an Information Sheet and explained the nature and effect of the procedures to the participant and that his/her consent has been given freely and voluntarily.

Name:…….. Date: ……. Signature: …….

Appendix C

Information Sheet

Information Sheet on PalmPrint Biometrics

OÉ Gaillimh
NUI Galway

You are being invited to take part in this research study. Before you decide, it is important for you to understand why the research is being done and what it involves. This Information Sheet will tell you about the purpose, risks and benefits of this research study.

If you agree to take part, we will ask you to sign a Consent Form. Scientific language can be difficult to understand, if there is anything that you are not clear about, please ask questions. Please take as much time as you need to read it.

You should only consent to participate in this research when you feel that you understand what is being asked of you, and you have had enough time to think about your decision. We would like to remind you that taking part in this research is completely voluntary. If you decide to take part, please be aware that you can still withdraw at any time and without giving a reason. Your rights will not be affected in any way.

Purpose of the research?

The aim of this research is to create and publish a database of palm-prints' images that reflect scenarios of actual use and have only been acquired using smart phones to verify their users' identity. At the end of the project, the database will be made publicly available to allow researchers to compare and validate their results.

By taking part in the study there is no direct medical benefit to you. It is hoped that information obtained during the study will be useful scientifically and may benefit understanding of biometric science and the development of new personal authentication technologies for use in smartphones.

What is your part in the research?

You will be required to come to the engineering building at NUIG at an agreed suitable time on a single occasion for a maximum of 45 minutes. On the day, you will be asked to sign a consent form and then the devices will be shown to you and there will be a demonstration of what you are required to do during acquisition. The acquisition is expected to last about 45 minutes.

What are the risks?

There are no risks to health, safety or well-being. Also your identity will be kept anonymous. You are providing personal biometric data, but this will be kept in a secure environment and cannot be associated with you as the data is stored anonymously. The Consent form will be the only link between your identity and the Subject ID number used to mark pictures in digital storage. The Consent forms will be destroyed at the end of the project, which will last up to 5 years.

Note that palmprints are not currently in any widespread use for biometric authentication and we do not envisage that they will come into significant usage in the near future. When palmprints are adopted more widely it is most likely that they will be used as a *secondary biometric* – that is, they will be used to cross-check fingerprint or iris (eye) data providing an additional (*secondary*) authentication. To falsify your identity a cyber-criminal would need both your iris/fingerprint scan and your palmprint. Thus, as your data will be annonymized, we consider the risk involved in making your palmprint available for public research through this database to be negligible.

To reflect this all of the individual researchers working on this project are providing their own biometric data as part of the the final public release database.

What if I do not want to take part?

You can discontinue your participation in the research study at any time. Should you feel at any stage that you want to discontinue being a participant in the study, then this is a dealt with in an unhesitating and confidential manner where you have the option of withdrawing form the study without the risk of information being disclosed.

What happens to the information?

Your personal biometric data will not be associated with you during or after this project. Biometric data will be stored in digital form using an identifier such as "subject XX". Some basic data such as age-range, gender and ethnic background might be saved with the data, but that is all. The image and associated data

will be stored in a secure lab facility and handled in confidence whereby results of the participants as well as their confidentiality are the first priority of the researchers carrying out the study.

Your identity is contained in the Consent form which will be kept in a secure location, separately from your biometric data, and only the lead PI will have access to your personal data.

What if someone goes wrong?

There are at least 3 experts taking part at this project and all acquired images will be checked by them. Also, your digital data will be kept separately from the paper data. If we suspect someone outside the project has gained access to the digital data, we will destroy the paper data linking your identity to the "subject ID". If we suspect someone outside the project has gained access to the paper data we will rename the digital images with different "subject ID". In both cases you will be informed.

What happens at the end of the study?

At the end of the data collection phase the biometric data will be used to present results in scientific papers; this data, which is retained for internal research at NUIG will be kept for the duration of the project, up to 5 years. At the end of the project the records that link you, as an individual, with your biometric data will be destroyed. Prior to this they will be stored separately and only the lead PI will have access to your personal data.

A random subset of the data contained in the main database will later be made publicly available. Only data from those who have agreed on the consent form will be made public and only 50% of that data, selected randomly, will be contained in the final database that is released publicly.

It is important to note that at all stages your information will be anonymised. All digital data gathered from the research will be held by the principle investigator for internal use no longer than 5 years. Data that is released publicly may remain available for public research for a longer period, but there will be no links to the individuals who originally contributed their data.

Further information

Please do not hesitate to ask us to clarify on any points which seem unclear.

Principal Investigators:

Dr. Peter Corcoran, College of Engineering & Informatics, NUIG.

Email: *Peter.Corcoran@nuigalway.ie*

Other Investigators:

Dr. Claudia Costache, College of Engineering & Informatics, NUIG.

Email: *Claudia.iancucostache@nuigalway.ie*

Hossein Javidnia, College of Engineering & Informatics, NUIG.
Email: *h.javidnia1@nuigalway.ie*

Adrian-Stefan Ungureanu, College of Engineering & Informatics, NUIG.
Email: *a.ungureanu1@nuigalway.ie*

We would like to remind you that taking part in this research is completely voluntary. If you decide to take part, please be aware that you can still withdraw at any time and without giving a reason. Your rights will not be affected in any way. If you decide to take part, please sign the Consent Form. This Information Sheet is for you to keep.

Appendix D

Acquisition Protocol

Experimental Protocol

PalmPrint Data Collection

The following is a study protocol to follow for each subject.
After advertisement, the volunteers who contact us to be part of the study will be invited to attend an information presentation & consent meeting.

Information Presentation
Volunteers will be asked to carefully read the Information Sheet. After reading the Information sheet volunteers will be invited to ask the investigator questions or request additional explanations.

A demonstration of the devices in the laboratory will next take place. The investigator will explain all conditions to volunteers.

If a volunteer decides to take part in the study, he/she will then be asked to sign the Consent Form. The length of the acquisitioning (approximately 20 minutes) is clearly stated in the document and the investigator will draw the volunteer's attention to this before proposing a suitable appointment timetable according to the volunteer's availability.

Biometric Data Session
Volunteers will be asked to choose one hand to capture images. A typical session will involve capturing images using multiple devices and in various lighting conditions and against different background.

One group of smartphone devices used to capture palm-print images is now given. This grouping uses 4 different, well known smartphone devices:

A) Samsung S6 Plus
B) Huawei P8
C) LG 4G
D) iPhone 6S
E) iPhone 5

Images will be acquired using each device with variations in 2 main parameters. These represent the most challenging variations for processing palmprint images.

A) Light
B) Background

For each parameter we defined 2 main variations.

A) Light (Indoor Dark, Indoor Normal)
B) Background (Artificial background with pictures, Wooden Surface)

Based on the number of parameters and their variations, every subject will contribute with 20 images to the database.
This database is unique because all images are acquired by the users themselves. After a short explanation, the user places his/her hand at a comfortable distance in front of them and follows the parameters listed previously.

In order to provide a useful database, the smartphones must be consistent with the market coverage today. For this reason the database should include popular brands like Apple, Samsung devices. Future plans include increasing the number of devices used.

Automatic smartphone settings are to be used in all scenarios.

Appendix E

License Agreement

LICENSE AGREEMENT FOR NON-COMERCIAL RESEARCH USE OF NUIG_Palmprint1: Palmprint Biometric Database

Introduction: The goal of our biometrics research is to develop new techniques, technology, and algorithms for the automatic authentication of humans using smartphones. As part of this research, NUI Galway are involved in an ongoing effort to collect a database of biometric imagery for palmprints. The database is meant to aid research efforts in the general area of developing, testing and evaluating authentication algorithms. The Center for Cognitive, Connected and Computational Imaging, College of Engineering and Informatics, having offices at National University of Ireland Galway, University Road, Galway, Ireland, owns copyright of the collection of biometric images and serves as the source for the database known as Palmprint Biometric Database.

Release of the Database: To advance the state-of-the-art in human authentication, the Palmprint Biometric Database obtained at NUI Galway will be made available to researchers on a case-by-case basis. All requests for the database must be submitted to the NUI Galway **LICENSOR** by the researcher's institution on behalf of the individual researcher or research unit (henceforth the "Licensee").

Effective as of _____ [EFFECTIVE DATE]

Center for Cognitive, Connected and Computational Imaging, College of Engineering and Informatics, having offices at National University of Ireland Galway, University Road, Galway, Ireland, (hereinafter "LICENSOR") and

_____ [Affiliation of LICENSEE]

having offices at

_____ [Address of LICENSEE]

(hereinafter "LICENSEE"), in consideration of the mutual covenants contained herein, the parties, intending to be legally bound hereby, hereto agree as follows:

1. LICENSOR developed certain valuable intellectual property, known as NUIG_Palmprint1 (Palmprint Biometric Database, hereinafter "Database") containing biometric data dissociated from personal data.

2. LICENSOR desires to grant a license to LICENSEE for the use of the Database in the scope of non-commercial research. This license will in no case be considered a transfer of the Database and the LICENSOR owns and continues to own all intellectual property rights in the Database.

3. LICENSEE shall have and shall obtain no rights with respect to the Database or any portion thereof and shall not use the Database except as expressly set forth in this Agreement.

4. Subject to the terms and conditions of this Agreement, LICENSOR hereby grants to LICENSEE for research use only, for a period of 2 years starting at the effective date above mentioned, renewable upon the discretion of LICENSOR, a royalty-free, nonexclusive, nontransferable, license subject to the following conditions:

4.1 The Database is only for the referred use of LICENSEE and, in a need-to-know basis, of those direct research colleagues who belong to the same research institution as LICENSEE and have been made aware of and have agreed in writing to adhere to the terms of this license.

4.2 The Database will not be copied nor distributed in any form other than for backup of LICENSEE.

4.3 In the case of the Database being modified, the LICENSOR will contact the LICENSEE so that all copies of the Database are permanently destroyed. Both parties (LICENSOR and LICENSEE) will sign a new license for the remaining time to complete the original 2 year period and the LICENSEE will receive the updated Database version.

4.4 The Database will only be used for research purposes and will not be used nor included in commercial applications in any form (e.g., original files, encrypted files, files containing extracted features, etc).

4.5 In any work made public, whatever the form, based directly or indirectly on any part of the Database, our efforts in constructing the database should be acknowledged as: “Portions of the research in this paper use the Palmprint Biometric Database collected in Center for Cognitive, Connected and Computational Imaging, College of Engineering and Informatics, National University of Ireland Galway.

4.6 Those seeking to include renderings of more than 10 images from the Palmprint Biometric Database in reports, papers, and other documents to be published or released must first obtain approval in writing from the **LICENSOR**.

4.7 Two years after the EFFECTIVE DATE, in case the license agreement is not renewed in writing by both parties (LICENSOR and LICENSEE), the license will expire and LICENSEE shall permanently destroy all copies of the Database.

5. LICENSOR shall comply with the Irish and EU legislation in force.

6. Data Protection: It is LICENSEE sole responsibility, to comply with all applicable data protection laws including Irish, European, and the laws of the jurisdiction within which the Licensee resides or operates.

7. Indemnification: LICENSEE agrees to indemnify, defend, and hold harmless the NUI Galway Center for Cognitive, Connected and Computational Imaging, College of Engineering and Informatics employees and agents, individually and collectively, from any and all losses, expenses, damages, demands and/or claims based upon any such injury or damage (real or alleged) and shall pay all damages, claims, judgments or expenses resulting from LICENSEES and its Researcher's use of the Palmprint Biometric Database.

8. **Keep Confidential:** Each party agrees that it shall take all reasonable precautions to keep confidential the other party's Confidential Information, and shall not, without the prior written consent of the other, use, disclose, copy or modify the other party's Confidential Information other than as necessary for the exercise of its rights, and performance of its obligations, under this Agreement.

9. This Agreement is subject to Irish law and the jurisdiction of the Irish Courts.

IN WITNESS WHEREOF, the parties hereto have executed this Agreement in duplicate originals by their duly authorized officers or representative

[Representative, Affiliation, Address of LICENSEE]

LICENSOR	**LICENSEE**
Prof. Peter Corcoran	_____
Director of Center for Cognitive, Connected	_____
and Computational Imaging (C3Imaging),	_____
College of Engineering and Informatics, National	_____
University of Ireland Galway,	_____
University Road, Galway, Ireland	

Signature LICENSOR: Signature LICENSEE: