Criminal and victim identification based on soft biometrics

Chan, Frodo Kin-Sun

2017

Chan, F. K.-S. (2017). Criminal and victim identification based on soft biometrics. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/70576>

<https://doi.org/10.32657/10356/70576>

NANYANG
TECHNOLOGICAL
UNIVERSITY

Criminal and Victim Identification based on Soft Biometrics

Frodo Kin Sun Chan

School of Computer Science and Engineering

A Thesis Submitted to the Nanyang Technological University in
Fulfillment of the Requirement for the Degree of

Doctor of Philosophy

2017

Abstract

Criminal and victim identification is always important in forensic investigation. However, it can be a very challenging problem for identifying criminals and victims in digital media when only their non-facial body sites are available in evidence images. These criminals and victims can be masked gunmen, paedophiles, and victims in child pornographic and voyeur images.

To solve the above problem, several novel alignment and identification approaches are proposed in this thesis. Firstly, lower leg geometry is proposed as a soft biometric trait for criminal and victim identification. This study provides a foundation for further research based on body geometry. Secondly, leg geometry and hair follicles are proposed to align the androgenic hair patterns in consideration of viewpoint and pose variations, which were ignored by a recent paper suggesting androgenic hair patterns for identification. Experiments on 1,138 high and low resolution images from 283 different legs show that the proposed alignment algorithms provide improvements of 5%-10% on different experimental settings.

Thirdly, a new approach is developed to improve the identification of androgenic hair patterns significantly. In the past, it was believed that androgenic hair patterns in low resolution images are not a distinctive biometric trait because of the previous result. A new algorithm, which makes use of leg geometry to align lower leg images, large feature sets (about 60,000 features) extracted through multi-directional gridding systems to increase discriminative power and robustness, the partial least squares (PLS) method to handle imbalanced training data and to perform the multi-grid feature fusion, and scheme

generating more positive samples to increase robustness against viewpoint and pose variation, is proposed. Experimental results on 1,493 low resolution leg images with large viewpoint and pose variations from 412 legs demonstrate the proposed multi-grid feature fusion algorithm provides a significant improvement.

Fourthly, estimation of soft biometric traits (e.g. height, BMI, age) is proposed. Recent studies demonstrated that soft biometric traits can be effectively estimated through comparative verbal descriptions (e.g. taller and shorter) given by witnesses. The usage of descriptions is extended to non-verbal comparative features extracted from legs for soft biometrics estimation. The proposed algorithm is examined on more than 2000 front and side leg images. The experimental results demonstrate that the proposed algorithm offers comparable performance with the existing methods.

Fifthly, the distinctiveness of human skin texture is evaluated by comparing with blood vessels. Recent studies show that blood vessel patterns can be visualized in different body parts. However, their visualization methods have limitations on images with low quality and subjects with high concentration of body fat, high concentration of melanin, and dense hair. In this study, more than 6000 inner forearm and thigh images were collected from a laboratory environment and the Internet with large pose, viewpoint, resolution, and illumination variations. The experimental results indicate the potential use of skin texture for criminal and victim identification.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Adams Wai-Kin Kong, for his continuous support of my PhD study. With his patience, motivation, immense knowledge, and guidance, it helped me a lot for the research and writing of this thesis.

I would like to thank reviewers, who spend time on reading my papers and thesis and provide me insightful and constructive comments.

I thank my fellow labmates in Biometrics and Forensics Lab – Arfika, Dr. Su, Abhik, Chaoying, Xiaojie, Yanzhu, Peicong, Quang, Xingpeng, Huang Yi, Yong, Hengyi, Shakeela, Rubha, Hamid, and Rabi for having meals and chatting with me.

Last but not least, I would like to thank Edwin, Helen, Roger, Mengchye, Peh Ping, Chris, Seline, Kin Siong, Peh Erh, Beng Lay, Wong Mui, Alvin, Alicia, Peggy, Pastor Anthony, Edmond, and Lok Man for their spiritual supports and countless prayers to me.

To my Sovereign Lord and my parents,

for their unconditional love and support

Table of Contents

Abstract	i
Acknowledgement	iii
Table of Contents	v
List of Figures	ix
List of Tables	xii
List of Publications and Awards	xiii
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	5
1.3 Contributions	6
1.4 Organization	7
1.5 Datasets	8
Chapter 2 Background and Related Work	10
2.1 Biometric Traits.....	10
2.1.1 Hand Geometry.....	10
2.1.2 Soft Biometric Traits	11
2.2 Criminal and Victim Identification Methods	14
2.2.1 Some Computational Methods	14
2.2.2 Androgenic Hair Patterns Matching	18
2.2.3 Soft Biometrics and Witness Descriptions	20
2.2.4 Blood Vessel Patterns Matching.....	22
2.2.5 Skin Texture Matching	24
2.3 Image Alignment Methods.....	26
2.3.1 Point Registration	26
2.3.2 Procrustes Analysis.....	28
2.4 Feature Extraction Methods	31
2.4.1 Shape-Based Feature Extraction.....	31
2.4.2 Edge-Based Feature Extraction	32
2.4.3 Texture-Based Feature Extraction	33

2.5 Feature Matching Methods.....	34
2.5.1 Distance Based Approach.....	34
2.5.2 Learning Approaches.....	35
2.5.2.1 Partial Least Squares for Large Feature Set	35
2.5.2.2 Learning to Rank	38
2.6 Summary	43
Chapter 3 Proposed Feature Extraction Schemes for Soft Biometric Traits	44
3.1 Overview	44
3.2 Evaluation Datasets	45
3.2.1 Data Collection	45
3.2.2 Preprocessing.....	48
3.3 An Algorithm Used in the Evaluation.....	50
3.4 Experimental Results.....	57
3.5 Summary	60
Chapter 4 Proposed Alignment Algorithms on Soft Biometric Traits for Boosting Matching Performance.....	61
4.1 Using Leg Geometry to Align Androgenic Hair Patterns	61
4.1.1 Overview	61
4.1.2 Proposed Alignment Algorithm.....	63
4.1.2.1 Edge Sampling.....	63
4.1.2.2 Image Registration.....	67
4.1.3 Feature Matching.....	69
4.1.4 Experimental Results.....	70
4.1.5 Summary.....	77
4.2 Using Hair Follicles with Leg Geometry to Align Androgenic Hair Patterns	78
4.2.1 Overview	78
4.2.2 Proposed Alignment Algorithm.....	79
4.2.2.1 Androgenic Hair Follicle Extraction	81
4.2.2.2 Image Registration.....	83
4.2.3 Experimental Results.....	84
4.2.4 Summary.....	86
Chapter 5 Proposed Identification Algorithm for Enhancing the Robustness and Accuracy of Soft Biometric Traits	87

5.1 Overview	87
5.2 Datasets and Preprocessing	88
5.3 Proposed Identification Algorithm.....	93
5.3.1 A Positive Sample Generation Scheme	95
5.3.2 A Low Resolution Alignment Method	96
5.3.3 Feature Extraction.....	98
5.3.4 Two Grid Systems	99
5.3.5 PLS Regression.....	102
5.4 Experimental Results.....	104
5.5 Discussion	117
5.6 Summary	121
Chapter 6 Proposed Estimation Algorithm for Deducing Soft Biometric Traits.....	123
6.1 Overview	123
6.2 Proposed Approach	126
6.2.1 Template Generation and Image Alignment Scheme.....	127
6.2.2 Feature Extraction.....	128
6.2.3 Comparative Feature and Rank Pattern Generation Scheme.....	133
6.2.4 OpRank Algorithm	135
6.3 Experimental Results.....	138
6.4 Summary	153
Chapter 7 Proposed Identification Algorithm for Comparison with Soft and Hard Biometric Traits	155
7.1 Overview	155
7.2 Testing database	160
7.3 A Skin Texture Matching Algorithm	165
7.3.1 Alignment and Positive Sample Generation.....	167
7.3.2 Feature Extraction Based on Two Gridding Systems and Classification.....	168
7.4 Experimental Results.....	170
7.5 Discussion	181
7.6 Summary	186
Chapter 8 Conclusion.....	187
8.1 Summary of Main Contributions.....	187
8.2 Future Work	190

References	193
Appendices.....	207
Appendix A: References of the Internet Images in Figure 5.2.....	207
Appendix B: References of the Internet Images in Figure 7.1	209
Appendix C: Some Results of the Skin Texture Algorithm.....	211
Appendix D: Some Failure Cases	216

List of Figures

Figure 1.1 The photos of riots.....	3
Figure 2.1 An example of simple mean operation and generalized Procrustes analysis for calculating a common shape template.	30
Figure 3.1 Examples of lower leg shapes for this evaluation.	46
Figure 3.2 High resolution images in the dataset.....	47
Figure 3.3 18 original images of different legs in the dataset.....	48
Figure 3.4 The preprocessing step: (a) the input image, (b) the region of interest defined by six edge points A, A', M, M', B and B'	51
Figure 3.5 Some examples of 142×298 preprocessed images.	52
Figure 3.6 Alignment of the same legs.	53
Figure 3.7 Alignment of different legs.	53
Figure 3.8 Intersection points on the aligned leg shapes	56
Figure 3.9 Cumulative match characteristic curves	59
Figure 4.1 The schematic diagram of the overall low resolution leg identification system.	62
Figure 4.2 Illustration of the weakness of the regular sampling method.....	64
Figure 4.3 Illustration of the centerline and the center point by Do.	66
Figure 4.4 The proposed sampling scheme applied to the leg boundaries in Figure 4.2.	67
Figure 4.5 Three examples of aligned images from different pairs of legs.	68
Figure 4.6 Illustration of the dynamic grid system in [4].	69
Figure 4.7 The CMC curves obtained from different alignment methods.....	71
Figure 4.8 The alignment results given by different alignment methods.	74
Figure 4.9 The CMC curves obtained from the proposed alignment algorithm with Su et al.'s method [4] on different channels and grayscale.	75
Figure 4.10 Different channels of images.....	76
Figure 4.11 A leg is captured at different viewing angles: (a) -45°, (b) -22.5°, (c) 0°, and (d) 22.5° with respect to the subject.	79
Figure 4.12 A schematic diagram of the identification process using the high resolution alignment algorithm.	80
Figure 4.13 The process of androgenic hair follicle extraction	83
Figure 4.14 CMC curves obtained from the high resolution images.	85
Figure 5.1 Some samples of left and right legs in the front pose datasets. The relative scale is kept for showing the sample variations.....	89
Figure 5.2 A comparison between the testing images and criminal images in real cases.	91
Figure 5.3 The region of interest is defined by six points A, A', M, M', B, and B'	92
Figure 5.4 The proposed preprocessing scheme.	93
Figure 5.5 A schematic diagram of the leg identification algorithm.	94
Figure 5.6 Illustration of the centerline and the center point defined by Do.	97
Figure 5.7 A leg with four orange dots is captured at different viewpoints.	102

Figure 5.8 (a)-(l) are the directional grid systems and (m) is the dynamic grid system.	102
Figure 5.9 CMC curves from (a) the first fold, (b) the second fold, and (c) the probe set is the Low Quality Set and the gallery set is the Image Set 2.	106
Figure 5.10 Comparison between the proposed algorithm and the preliminary method using the large feature sets for the probe sets of (a) the Image Set 1 and (b) the Low Quality Set.	109
Figure 5.11 CMC curves were obtained from different algorithmic settings to analyze different computational components in the proposed algorithm.	113
Figure 5.12 CMC curves for different image resolutions of (a) the Image Set 1 and (b) the Low Quality Set.	116
Figure 6.1 A simulation of an evidence photo obtained by enforcement agency for child sexual abuse case.	125
Figure 6.2 Examples of voyeur images collected randomly from the Internet.	125
Figure 6.3 A schematic diagram of the proposed approach.	127
Figure 6.4 Illustration of Gabor feature extraction process.	131
Figure 6.5 (a)-(l) are the directional grid systems.	133
Figure 6.6 The concept of the rank pattern.	135
Figure 6.7 Some examples in the (a) front leg and (b) side leg database.	139
Figure 6.8 The first column and the second column are height, BMI, and age distributions of the front and side leg databases. (a-b) Height, (c-d) BMI, and (e-f) age.	140
Figure 6.9 CS curves and MAEs of the proposed OpRank and different methods from the front leg database.	144
Figure 6.10 CS curves and MAEs of the proposed OpRank and different methods from the side leg database.	145
Figure 6.11 CMC curves of the estimated soft biometric features with (a) Islam et al.'s method [131] and (b) Su et al.'s method [4] on the front leg database.	149
Figure 6.12 CMC curves of the estimated soft biometric features with (a) Islam et al.'s method [131] and (b) Su et al.'s method [4] on the side leg database.	149
Figure 6.13 CS curves and MAEs of the proposed OpRank with rank patterns and other methods for (a-b) height, (c-d) BMI, and (e-f) age estimations from the front leg database.	151
Figure 6.14 CS curves and MAEs of proposed OpRank with rank patterns and other methods for (a-b) height, (c-d) BMI, and (e-f) age estimations from side leg database.	152
Figure 7.1 Examples of terrorist and rioter images.	156
Figure 7.2 Blood vessels visualized from images collected from the Internet [136].	157
Figure 7.3 Images without clear blood vessels by using the method [136].	157
Figure 7.4 Images skin texture from different forearms and thighs.	158
Figure 7.5 Skin texture sample images from different studies.	160
Figure 7.6 Sample images in the four NTU datasets.	163
Figure 7.7 Examples from the Internet forearm dataset.	164
Figure 7.8 Examples from the Internet thigh dataset.	165

Figure 7.9 Skin texture from different locations of the same (a) forearms and (b) thighs.	166
Figure 7.10 A schematic diagram of the proposed skin texture matching algorithm. ...	167
Figure 7.11 Illustration of the gridding systems.	169
Figure 7.12 Skin texture samples: (a) original images, (b) red channels, (c) green channels, and (d) blue channels	169
Figure 7.13 The CMC curves from (a) NTU color forearm dataset and (b) NTU color thigh dataset.	173
Figure 7.14 The CMC curves from (a) Internet forearm dataset and (b) Internet thigh dataset.	175
Figure 7.15 Top ten retrievals in a search of (a) inner forearms and (b) inner thighs given by the skin texture matching algorithm.	176
Figure 7.16 Comparison of skin texture from color images and blood vessels from NIR images.	178
Figure 7.17 The comparisons between two image resolutions (142×298 pixels and 35×74 pixels).....	179
Figure 7.18 Comparison of the proposed algorithm and the state-of-the-art texture recognition methods.....	181

List of Tables

Table 2.1	Characteristics of soft biometric traits. [20]	13
Table 3.1	The details of dataset.	47
Table 3.2	A summary of the selected features.	58
Table 3.3	A summary of the experimental results.	59
Table 5.1	The details of the datasets.	89
Table 5.2	A summary of the experimental results.	107
Table 5.3	A summary of the experimental results of the proposed algorithm and the one without color information	110
Table 5.4	A summary of the experimental results with different settings of the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.	114
Table 5.5	A comparison between PLS and linear SVM on the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.	115
Table 6.1	Gabor parameters used in the two layers.	131
Table 6.2	Gender of two databases.	139
Table 6.3	The MAEs between one leg and two legs of front leg database.	146
Table 7.1	Summary of the difference between the previous studies and this study.	159
Table 7.2	The details of the Internet datasets.	163
Table 7.3	Experimental results on (a) NTU color forearm dataset and (b) NTU color thigh dataset.	173
Table 7.4	A summary of the experimental results. (a) Internet forearm dataset and (b) Internet thigh dataset.	175

List of Publications and Awards

Publications

1. **F.K.S. Chan** and A.W.K. Kong, “Using leg geometry to align androgenic hair patterns in low resolution images for criminal and victim identification,” in *International Conference on Pattern Recognition*, 2014, pp. 495-500.
2. M.R. Islam, **F.K.S. Chan**, and A.W.K. Kong, “A preliminary study of lower leg geometry as a soft biometric trait for forensic investigation,” in *International Conference on Pattern Recognition*, 2014, pp. 427-431.
3. **F.K.S. Chan** and A.W.K. Kong, “Using hair follicles with leg geometry to align androgenic hair patterns,” in *European Intelligence and Security Informatics Conference*, 2015, pp. 137-140.
4. **F.K.S. Chan** and A.W.K. Kong, “A further study of low resolution androgenic hair patterns as a soft biometric trait,” *Image and Vision Computing*, under the major revision.
5. **F.K.S. Chan**, X. Li, and A.W.K. Kong, “A study of distinctiveness of skin texture for forensic applications through comparison with blood vessels,” *IEEE Transactions on Information Forensics and Security*, accepted.
6. **F.K.S. Chan** and A.W.K. Kong, “Towards using comparative features to estimate soft biometric trait from legs,” to be submitted.

Awards

1. Granted an IAPR Travel Award of US\$1,000 for attending the *International Conference on Pattern Recognition*, 2014 (50 out of 247).
2. Received a subsidized rate for attending *European Intelligence and Security Informatics Conference*, 2015.

Chapter 1

Introduction

1.1 Motivation

Nowadays, criminal and victim identification, including two processes, identification (one-to-many matching), and verification (one-to-one matching), is an integral part of forensic investigation procedures. The identification process means the existing records in a database are compared with the biometric traits found at a crime scene. But the verification process means the biometric traits found at a crime scene are compared with those taken from an arrested suspect.

Biometric traits refer to human physiological or behavioral characteristics, e.g. DNA sample, latent fingerprint, handwriting, and face image. They have two types, hard biometric traits and soft biometric traits. Hard biometrics are the traits that are expected to be universal, permanent, and unique to each individual. Soft biometrics e.g. height, BMI, race, skin color, and eye color are the traits that may not be universal, permanent, or unique but useful for forensic agencies in narrowing down suspects and speeding up the database search.

Many biometric traits, such as DNA, fingerprint, face, and palmprint, are useless to legal cases where only digital images with non-facial body sites of criminals or victims are left for identification. These criminals or victims can be related to child sexual abuse, riots, and terror attacks. In addition, the frequency of the related crimes is increasing. Statistics

Canada reported that there were over 5 million child sexual abuse images on the internet and 15,662 websites hosting child sexual abuse images [1]. In riots, many countries keep experiencing, e.g. U.S [96], U.K. [139], Rome [140], and Brazil [141]. Figure 1.1 shows several photos of different riots. In 2011, the England riot caused approximately 2,500 shops and businesses looted [94]. In 2011, the Stanley Cup riot in Vancouver [95] was about 100,000 people in the streets. The severity of terrorist activities is well-known. Paris, Jakarta, and Barrels attacks are a few well-known examples. Islamic State of Iraq and the Levant (ISIS) and other terrorist groups always use images and videos to communicate with their followers and threaten governments. Normally, the criminals always wear face masks to avoid identification, but may expose non-facial body sites. Obviously, the current biometric traits adopted by law enforcement agencies are not sufficient. Thus, the need of new biometric traits is clear for the cases where the faces of criminals or victims are not observed in the digital images.

To resolve this challenging identification problem, skin marks [2], tattoos [3], androgenic hair patterns [4], soft biometrics [5-7], and vein patterns hidden in color images [8] have been proposed but they have their own limitations. Skin marks and tattoos are not universal. And skin marks are only good for high resolution images. Rioters are usually locals with casual dressing style, e.g. T-shirts, short sleeve shirts, and shorts. Their body parts are frequently exposed in front of cameras. Photos of masked rioters showing non-facial skin can be easily found. Riot images can be captured by reporters, who always use high quality DSLR cameras. Child sexual abuse images are usually close-up images captured by smartphone cameras or consumer digital cameras. These images are not necessarily captured by surveillance cameras, because by 2016 and by 2018, there will be

2.16 and 2.56 billion smartphone users worldwide, respectively [97]. Images captured by surveillance cameras are not within the scope of this thesis.

Figure 1.1 The photos of riots. (a) London riots 2011 [11], (b)-(c) anti-government protesters in Bahrain, Manama [12-13], (d)-(e) Athens riots 2011 [14], and (f) Brazil riots 2013 [15].

Androgenic hair patterns [4] were suggested to extract from low resolution leg images for personal identification. However, the previous study neglected the importance of alignment before matching two patterns and its matching method was neither discriminative nor robust to pose and viewpoint variations. Because of the previous result, it was believed that androgenic hair patterns in low resolution images are not a distinctive biometric trait. In this thesis, several alignment and identification algorithms are proposed to overcome its weaknesses.

Soft biometric traits, which were effectively estimated through comparative verbal descriptions (e.g. taller and shorter) given by witnesses, were combined for identification [5-7]. The studies concentrated on estimating soft biometrics based on verbal descriptions and used full body of a target subject (e.g. suspect) to compare with reference subjects. Legs often appear in sexual offense images, e.g. child sexual abuse images and voyeur images, but estimating soft biometrics, such as BMI and height of legs is not a straightforward task for human beings. In this thesis, a new algorithm extends the usage of comparative descriptions to non-verbal features extracted from legs for soft biometrics estimation.

Blood vessel patterns or vein patterns [8], which are extracted by vein visualization method in color images, were suggested for identification in digital media. Although blood vessel patterns can be visualized in different body parts, the visualization method has limitations on the skin images with low quality, high concentration of body fat, high concentration of melanin, and dense hair. In this thesis, human skin texture is proposed to overcome the limitations because blood vessels in color images are in fact derived from

skin images and human skin contains distinctive features, like wrinkles, freckles, spots, scars etc.

1.2 Research Objectives

The aim of the thesis is to propose new soft biometric traits and efficient algorithms for criminal and victim identification. Although their faces are masked, other body parts can be exposed. In this thesis, the leg geometry is proposed as one of soft biometric traits.

To overcome the weaknesses of androgenic hair patterns mentioned in the previous section, several alignment and identification algorithms are proposed. For the alignment, leg geometry and hair follicles are proposed as the features for aligning high and low resolution leg images. For the identification, a new algorithm, which makes use of leg geometry to align lower leg images, large feature sets (about 60,000 features) extracted through multi-directional gridding systems to increase the discriminative power and robustness and the partial least squares (PLS) method to handle imbalanced training data and to perform the multi-grid feature fusion, is proposed. To further enhance the performance of the PLS models trained on very limited positive samples, one to three images per model in the experiments, and to further increase robustness against viewpoint and pose variations, a scheme is designed to generate more positive samples from a single image.

In order to extend the usage of comparative descriptions to non-verbal features extracted from legs for soft biometrics estimation, the shapes of legs in a reference database are first grouped according to their ranks of soft biometrics e.g. $20 < \text{BMI}$ and $\text{BMI} < 23$. The

shapes of legs in the same rank are further clustered by the k-means algorithm and the means are used as references for comparing the shapes of input legs to extract comparative features. The comparative features are applied to train and test the proposed OpRank algorithm, which is a novel probabilistic model optimizing the correlation between the ranks of soft biometrics and the comparative features with the regularization of the ranking information. The proposed algorithm is examined on more than 2000 front and side leg images.

To overcome the limitations of vein patterns in color images, a study is performed to compare the state-of-the-art vein visualization methods and a skin texture matching algorithm. They are evaluated on a skin dataset, which contains 1312 inner arm images and 1048 inner thigh image, and an Internet dataset, which contains 1509 inner arm images and 1257 inner thigh images. In addition, one experiment is used to compare the performance of color skin images and NIR vein images on the same skin texture algorithm. To evaluate the robustness of the skin texture algorithm, an experiment is designed to evaluate on low resolution images.

1.3 Contributions

The experiments and analysis confirm the feasibility of the proposed biometric features and methods for criminal and victim identification. Firstly, lower leg geometry is shown to be an effective soft biometric trait and it provides a foundation for further research on criminal and victim identification based on body geometry. Secondly, the alignment problem of androgenic hair patterns is handled by using hair follicles and leg geometry for

high and low resolution images. Thirdly, the androgenic hair patterns are shown to be distinctive by using our proposed algorithm. Fourthly, the usage of comparative descriptions is extended to non-verbal features extracted from legs for soft biometrics estimation. Fifthly, the skin texture is shown to be useful for criminal and victim identification.

Based on the above approaches, three conference papers are published and mentioned in the list of publications (page xiii). Two journal papers are under major review and another will be submitted soon. Two travel awards were given to the author for these conference papers (page xiv).

1.4 Organization

The details of the research are shown in the following chapters. In Chapter 2, some related works are reviewed. Then, a study of lower leg geometry as a soft biometric trait is presented in Chapter 3. And then, Chapter 4 describes the alignment algorithms with androgenic hair patterns to improve high and low resolution leg identification. Furthermore, Chapter 5 reports the new learning approach for improving the identification of androgenic hair patterns. Chapter 6 mentions an algorithm to extend the usage of comparative descriptions to non-verbal features extracted from legs for soft biometrics estimation. Chapter 7 evaluates the vein visualization methods and a skin texture matching algorithm. Finally, conclusion and future work are given in Chapter 8.

1.5 Datasets

Since different algorithms are proposed in different chapters, 11 datasets are applied to evaluate the algorithms in the thesis. In Chapter 3 and 4, the high resolution and low resolution leg datasets are formed by segmenting and resizing the data collected from 1138 images of 283 legs. In these two datasets, the images are captured from two occasions with an interval of around 11 days. The poses of the subjects, who are Arab, Bengali, Caucasian, Chinese, Eurasian, Indian, Javanese, Malay, Punjabi, and Vietnamese, are not strictly fixed.

In Chapter 5, a front pose leg dataset is collected from two occasions with an interval of at least one week. There are three sets of image from different subjects, who are mainly Chinese, Indian, Vietnamese, other Asians, Iranians, Europeans, Arabs, and South Americans. In the first two sets, they are collected from two different occasions and 536 images from 412 legs for each set are used to simulate good quality images collected at crime scene. In the third set, 423 low quality images from 281 legs are used to simulate bad quality images collected at crime scene. They have problems of blur, out of focus, viewpoints, shadows, scaling, and illumination. In addition, the left and right legs of the same subject are considered as the legs of two independent subjects in the experiments. To simulate different resolutions, the images with four different resolutions (25 dpi, 18.75dpi, 12.5dpi, and 6.25 dpi) are also generated from the front pose leg dataset for experiments.

In Chapter 6, a front pose leg dataset and a side pose leg dataset are used for evaluation. The front pose leg dataset is originally from the first two image sets of the front pose leg dataset in Chapter 5, but the images with missing demographic information (age, BMI, and height) are removed. In total, 1070 leg images are included. From the side pose

leg dataset, 934 leg images, which are mainly from Chinese, Indian, Vietnamese, Caucasian, Indonesian, Eurasian, European, and Javanese, are collected with side pose. Also, the left and right legs of the same subjects are considered as the legs of two independent subjects in the experiments.

In Chapter 7, 6 image datasets, NTU color inner forearm datasets, NTU near infrared (NIR) inner forearm dataset, NTU color inner thigh dataset, NTU NIR inner thigh dataset, an Internet inner forearm dataset, and an Internet inner thigh dataset are used. For each NTU dataset, the images are captured from two occasions with an interval of around 14.53 days. In NTU color forearm and thigh datasets, 656 forearm and 524 thigh images from 344 and 272 subjects, who are mainly Chinese, Malay, Indian, Vietnamese, and Caucasian in Singapore, are collected. Images in the NTU color forearm (thigh) datasets have corresponding images in the NTU NIR forearm (thigh) dataset, which is collected by using a near infrared camera. These four datasets contain totally 4720 images. In the Internet datasets, 213 forearm and 229 thigh images are collected as probe images for first occasion and 640 forearm and 504 thigh images are collected as gallery images for the second occasion. The original sizes of these images range from 35×50 pixels to 903×1150 pixels and they can have problems of blur, shadows, out of focus, compression, viewpoint, pose, illumination, and scaling variations. For increasing the gallery size, they are combined with the images in the second occasion of the NTU color forearm and thigh datasets to be 1296 forearm and 1028 thigh images.

Chapter 2

Background and Related Work

In this chapter, some background information related to the proposed alignment and identification algorithm is provided. In Section 2.1, two related biometric traits, which are hand geometry and soft biometric traits, are reviewed. In Section 2.2, different criminal and victim identification methods, which include some common methods, androgenic hair patterns [4], comparative descriptions from soft biometrics [5-7], and blood vessel patterns [8], are explained. In Section 2.3, two alignment methods are introduced because they are used for aligning the images and generating the reference image in the proposed algorithms. In Section 2.4, three different type of feature extraction methods are mentioned. Section 2.5 describes the feature matching methods including distance based approach and learning approaches. Section 2.6 gives a summary of this chapter.

2.1 Biometric Traits

2.1.1 Hand Geometry

Leg geometry is a new concept, but the hand geometry [9-10] is definitely not a new term in the biometric community. Based on the successful implementation of hand geometry in biometric applications, the leg geometry is expected to be useful for forensic investigation and personal identification.

Hand geometry recognition is not a new technique and has over 30-year history. It has been widely used for physical access, attendance tracking, and personal verification. The most famous application is the use in the 1996 Olympic game for controlling and protecting physical access to the Olympic Village [16]. Hand geometry considers not only the hand shape, but also the palm and fingers structures. In 2002, Bulatov et al. [17] suggested 30 global features of hand geometry, e.g. finger length, finger width, and finger perimeter.

To extract those features from the hand, there are two traditional ways. The first one is to restrict degrees of freedom of hand by using a flat platform and pegs or pins and the second one is to place a hand on a flat platform without pegs or pins (e.g. scanner) [18]. The most recent one is unconstrained and contact-free. It means that no specific platform, pegs or pins is required for hand image acquisition [19]. However, hand geometry is only suitable for security applications because it is usually obtained with the cooperation with user. Since the users are free to show any gestures or hold something in the real-life photos of criminals and victims, hand geometry is difficult to capture. On the other hand, the leg geometry, which is still able to be captured at a distance and is not easily changed in comparison with hand geometry, is preferred to hand geometry for criminal and victim identification.

2.1.2 Soft Biometric Traits

In recent years, soft biometric traits draw much attention [5, 20-21] simply because of the low computational cost and simplicity in retrieval. From a report published in New

Scientist [22], the US government's Intelligence Advanced Research Projects Activity (IARPA) was seeking a new way for faster and more accurate search and retrieval of the biometric features. The agent announced a contest for challenging teams of the country's top researchers to recognize people from the discarded data, which possibly included lots of soft biometric traits.

Soft biometric traits are different from traditional biometric traits, like face, fingerprint, iris, and palmprint. Normally, the traits are neither permanent nor very distinctive; they can be height, BMI, size, gait, scars, tattoos or other features. Since they lack of distinctiveness and permanence, they cannot be used alone. However, soft biometric traits can be very useful to narrow the searching range. Besides, they can be fused with the classical biometric traits to increase the system reliability. Reid et al. [5] also showed that there is possibility to identify individuals by using a larger set of soft biometric traits. Dantcheva et al. [20] suggested a list of soft biometric traits, including age, height, BMI, and body shapes. In their suggestions, accessories, like clothes color, glasses are classified as the soft biometrics also.

Dantcheva et al. also summarized the characteristics of soft biometric traits, which can be shown in Table 2.1. Accordingly, Reid et al. showed similar descriptions in the paper [5]. They mentioned that the soft biometric traits require less user cooperation and less computation compared to hard biometrics.

Table 2.1 Characteristics of soft biometric traits. [20]

Characteristics of soft biometric traits	Description
Human compliant	Traits are conformed to natural human description labels.
Computationally efficient	Sensor and computational requirements are marginal.
Enrolment free	Training of the system is performed off-line and without prior knowledge of the inspected individuals.
Deducible from classical biometrics	Traits can be partly derived from images captured for primary (classical) biometric identifier (e.g. eye color from iris images).
Non-intrusive	Data acquisition is user friendly or can be fully imperceptible.
Identifiable from a distance	Data acquisition is achievable at long range.
Not requiring the individual's cooperation	Consent and contribution of the subject are not needed.
Preserving human privacy	The stored signatures are visually available to everyone and serve in this sense privacy.

Although soft biometric traits have so many advantages, labeling of the soft biometric traits can be a headache problem. Previously, categorical labels have been used for describing the soft biometric traits. However, human can make mistakes and their labeling can be quite subjective, which means it may not be consistent. In order to deal with this problem, Reid et al. [5] proposed to use relative measurements in the labeling. Such concept of relative measurements is extended to use for estimation of soft biometrics in Chapter 6.

Furthermore, shapes of different body parts were used as soft biometric traits in Reid et al. [5] and Dantcheva et al. [20]. They indicated that the body shape could be a

common and useful soft biometric trait. In this study, leg geometry is proposed as a soft biometric trait for image alignment and feature matching.

2.2 Criminal and Victim Identification Methods

Identifying criminals and victims is always a challenging task for forensic investigation. There were many methods suggested to solve this problem in the past. In the following five sub-sections, some computational methods are explained in Section 2.2.1. Then, three recent methods, including androgenic hair patterns, comparative descriptions from soft biometrics, and blood vessel patterns are described in Sections 2.2.2 – 2.2.4. The last sub-section reviews the possibility of skin texture on forensic cases.

2.2.1 Some Computational Methods

In order to cover a wide range of forensic demands, biometric researchers have put their efforts on developing effective computational methods for criminal and victim identification. Some enhance the current manual or semi-automatic recognition approaches used by law enforcement agencies and some tackle new demands from sexual offenses, riots, digital crimes, and terrorism. They include, but are not limited to, latent print enhancement [23-24], tattoo retrieval [25-27], face sketch recognition [28-31], shoe print recognition [32], dental identification [33-34], authorship verification [35], body skin mark identification and related studies [36-38], androgenic hair patterns matching [4], soft biometrics and witness descriptions [5-7], and vein patterns matching [8].

Latent prints, which are one type of fingerprints, are not visible to the naked eye but can be found at the scene of a crime. Normally, their quality is poor and their fingerprint surface area is small when they are compared with plain or rolled prints. To enhance its quality, Yoon et al. [23] suggested to capture its orientation information by short-time Fourier transform and generate the hypothesized orientation fields by randomized RANSAC [39]. Since fingerprints may overlap on top of each other, Chen et al. [24] applied relaxation labeling technique and fingerprint singularity information to separate the fingerprints.

Tattoos are one of the important traits of criminal identification because some hidden information related to a suspect's criminal history, which can be gang membership, previous convictions, and years spent in jail, may be revealed by tattoos. Manger [25] implemented content-based image retrieval techniques to solve a large scale tattoos problem. In order to detect and classify tattoos in unconstrained conditions, Heflin et al. [26] introduced an automatic tattoo segmentation and used LBP-like [40] feature descriptor to represent tattoos features. On the other hand, when a suspect's tattoo is not available, a sketch of a tattoo drawn based on the eyewitness's description is used to find out the suspect. Han et al. [27] made use of Canny edge detector and SIFT to extract sketch and tattoo image. Then, sparse representation based classification (SRC) [41-42] was employed for matching.

Face sketch recognition is a common method to assist police to identify suspects. Since a sketch is constructed with the subjective description of eyewitnesses, Yuen et al. [28] applied a searching method using relevance feedback to narrow down the searching space by human input. To improve the matching performance between a face photo and its

sketch, Wang et al. [29] suggested two approaches for making face photos and a query sketch in the modality. 1) All of the face photos in the gallery are first transformed to sketches. 2) The query sketch is first transformed to a photo. They found that the second one is better in the experimental results. However, they tested on artistic sketches, which are very different from real implementation. In order to handle the sketches from non-artistic drawing and facial composite software, Han et al. [30] implemented component-based representation approach to search the sketch and the corresponding mugshot photo. Nejati et al. [31] introduced a recognition method to estimate and reduce distortions due to personal sketching bias before attempting to recognize the sketch.

Footwear marks, which are usually found at crime scenes, can be identified by searching the most similar footwear print available in the large database. Different types of footwear prints were mentioned in Tang et al. [32]. To speed up large database and cope with partial prints, color prints, and crime scene marks, Tang et al. [32] suggested that the footwear print pattern is clustered as the combination of shape features, which are like lines, circles, and ellipses, and represented by an Attributed Relational Graph (ARG) [43]. The Footwear Print Distance (FPD), which was derived from Earth Mover's distance [44], was introduced to compute the similarity between prints.

Dental information, which is stored in dental radiographs, is one of few biometrics for postmortem (PM) identification, especially under the severe circumstances in mass disasters (e.g. airplane crashes). Automatic process for handling dental information is still demanded. Nassar et al. [33] introduced an automatic way to construct the dental charts, which is a useful guide for tooth-tooth matching, by using appearance based feature and

string matching technique. Said et al. [34] suggested the automatic teeth segmentation by using mathematical morphology approach and contrast stretching transformation.

Authorship analysis, which is the forensic analysis of online textual documents for addressing the anonymity problem, is important for cybercrime investigation on criminals to conduct malicious activities such as spamming, sending ransom e-mails, and spreading botnet malware. Iqbal et al. [35] presented a unified data mining solution to address authorship analysis problems based on the concept of frequent pattern-based writeprint. They studied on authorship identification with large training samples, authorship identification with small training samples, and authorship characterization.

Body skin marks, whose group is named as Relatively Permanent Pigmented or Vascular Skin Marks (RPPVSM) by Nurhudatiana et al. [36], have been used in forensic identification, for example, searching for a suspect in child pornography [45]. Nurhudatiana et al. developed an individuality model for independently and uniformly distributed RPPVSM patterns [36] and provided an automatic way for RPPVSM identification and implemented a fusion scheme with RPPVSM and vein patterns [37]. Also, Nurhudatiana et al. [38] provided a pilot study on the interexaminer variability of RPPVSM identification.

Although the above techniques are effective tools for criminal and victim identification, only tattoos and body skin marks are applied to deal with digital images. However, tattoos and body skin marks have their own problems. Tattoos are not universal and unique. Body skin marks is not universal and only applicable for high resolution image. Thus, new biometric traits are still needed for dealing the images from criminal cases.

2.2.2 Androgenic Hair Patterns Matching

Medical studies have indicated that androgenic hair and its follicles can be effective biometric traits. No additional follicles can be naturally formed after birth in human beings [46]. There is a cycle manifested by all androgenic hairs. Although one hair falls out, the same follicle can allow another new hair to grow [47]. It means two different hair shafts can grow in the same location. Normally, androgenic hair can have a long cycle and the leg hair cycle can have up to one year [48]. It can imply that the same hair shaft can be found within a long period of time. But androgenic hair growth [49] can have seasonal changes on different body parts. The thigh hair can grow faster in summer than in winter [50]. Nevertheless, seasonal changes can further show that the cycles of hair follicles are asynchronous, except in groups of three follicles called Demeijère trios [51-52]. Since the human hair will not fall out at the same time, some corresponding androgenic hairs can be found for matching. Although there is a permanent exit from cycling, only about 10% androgenic hair follicles can have it and the rest can cycle throughout the entire lifespan to produce hair [47]. Furthermore, the hair samples in [4] show that the androgenic hair patterns don't have much changes and some hair shafts can still find their corresponding one.

In Su et al.'s paper [4], an identification method based on androgenic hair patterns was suggested. There are three stages in the method: preprocessing, feature extraction, and matching. In the preprocessing step, Su et al. employed the semi-automatic approach which is composed of an automatic segmentation and a manual correction process. The preprocessing method is summarized below:

Step 1: To extract leg boundaries, an input color image is first converted to a grayscale image and smoothed by a two dimensional median filter. Then, the smoothed grayscale image is inputted to Sobel edge detector to obtain an edge image denoted as J . Noted that the threshold used in this edge detection is set automatically for every image.

Step 2: A skin color range is predefined to detect skin pixels in the color image. It means only the color of the pixels within the color range can be kept, the pixel value of others will be set at 0. This processed image is denoted as L .

Step 3: If a pixel in J is not an edge pixel and the corresponding pixel in L is not zero, a skin pixel is found in the corresponding pixel of the original image. If the pixel in J is an edge pixel and the corresponding neighborhood in L is not zero, a skin pixel is found in the corresponding pixel of the original image. By applying these two rules, a skin image is segmented.

Step 4: Finally, a morphological operation applies a disk-shaped structuring element to the segmented image.

Step 5: The segmented images are checked. If there is a segmentation error, manual segmentation is made.

This preprocessing step is mainly used to extract the region of interest, which is defined in [4]. After the input leg images are preprocessed, the preprocessed image is filtered by real parts of Gabor filters with different scales and orientation for computing Gabor magnitude. And then, orientation field is extracted by using Gabor magnitude and divided into small regions, which is composed of about 300 pixels. Afterwards, the local orientation histograms are formed for each small region. Finally, comparisons among these

histograms are performed in the database. The details of this feature matching method are also described in Section 2.5.1.

Su et al. further evaluated the matching method by using the database which contains the gallery set and probe set. In the gallery set, it contained 1111 images from 558 different legs of the 283 subjects. In the probe set, it contained 575 images from 283 different legs of the 283 subjects. These experimental results in [4] showed the usefulness of the Su et al.'s method because Su et al.'s method outperformed the other methods [54-55].

However, there are two problems ignored in Su et al.'s method. The first problem is the definition of the region of interest. In Su et al.'s method, the definition of the region of interest is not applicable to the legs tilted by a certain angle. The second problem is an alignment problem. Su et al.'s method is not invariant to leg pose, and they neglect the leg alignment. In Chapter 5, a solution is provided to remedy these problems and the experimental results are improved significantly.

2.2.3 Soft Biometrics and Witness Descriptions

For many criminal investigations, eyewitness descriptions are useful for searching the possible suspects. Soft biometrics, which are a new form of biometric identification that people use to describe each other, e.g. “tall”, “skinny”, and “male”, can be used in eyewitness descriptions because they have a relationship with human verbal description. Traditionally, such verbal descriptions consist of either absolute categorical labels (e.g. “tall”, “fat”) or estimates of human characteristics (e.g. 190 cm, 50 kg). However, such

descriptions are highly subjective based on eyewitness's perception. To extend the soft biometrics application to verbal identification, Reid et al. [5-7] suggested to identify human by using comparative labels from soft biometrics. Comparative labels, which are derived from comparing one person with another (e.g. "taller", "fatter"), are more objective than the absolute categorical labels.

A set of relative measurements, which is continuous values within biometric feature vector and can be inferred from comparative labels, represents the relative strength of the suspect's physical traits and is used as a biometric signature for identification. To infer the relative measurements from comparisons, Elo rating system [56], which was designed to quantify the skill of chess players, is used. In the Elo rating system, two players A and B with a rating of R_A and R_B are compared. The expected scores of A and B are

$$E_A = \frac{Q_A}{Q_A + Q_B}, \quad (2.1)$$

$$E_B = \frac{Q_B}{Q_A + Q_B}, \quad (2.2)$$

where $Q_A = 10^{R_A/U}$, $Q_B = 10^{R_B/U}$, and U is chosen to reflect how a player's current rating can affect the expected result. The ratings of A and B are updated as

$$R'_A = R_A + K(S_A - E_A), \quad (2.3)$$

$$R'_B = R_B + K(S_B - E_B), \quad (2.4)$$

where S_A and S_B are the actual scores obtained by A and B for each comparison and K is defined as the maximum rating adjustment resulting from the comparison. In practice, a set of relative measurements describing a suspect is obtained by comparing the suspect to multiple subjects, who have predefined Elo ratings. By using the Elo rating system, each

comparison is used to adjust the suspect's Elo ratings until all the comparisons have been considered.

Reid et al. [7] implemented the comparative description on facial recognition to retrieve a suspect from a 40-subject database. 27 traits, which represent as relative measurements, were used to form the biometric signatures from the original features. By selecting a suspect from the database, n randomly sampled comparisons between the suspect and other subjects were used to infer the suspect's biometric signature. Reid et al. [6] performed an experiment to retrieve a suspect from an 80 subject database with the 19 body traits, which were represented as relative measurements. The experimental results showed that the recognition accuracy increases when the number of comparisons increases.

Since this comparative representation of verbal description is useful for identifying the criminals, it can be extended to the images of other body parts, e.g. leg image, which is lack of vocabulary to describe its features. Also, when an image only contains some parts of body, the body part, like a leg, is difficult to describe and the number of soft biometric traits is very limited. In Chapter 6, a novel estimation system is developed to provide more soft biometric traits, which are deduced from the evidence image, for matching improvement of images of body parts.

2.2.4 Blood Vessel Patterns Matching

Blood vessel patterns, which are a vast network of blood vessels underneath a person's skin, were widely used in authentication and verification in the past. Blood vessel patterns recognition, which is also referred as vascular pattern recognition, is applied in different

parts of the back hand [58-59], palms [60] or fingers [61]. There are two implicit advantages for blood vessel patterns authentication and verification. Firstly, it is difficult to forge because the blood vessel patterns are inside the body parts. Secondly, it addresses the hygiene concerns because capturing blood vessel patterns is contactless.

For capturing the vein images, Wang et al. [57] showed that near-infrared imaging (NIR) is more tolerant to the external environment than far-infrared thermography (FIR). However, NIR image also keeps the defects on the skin surface, which can corrupt the structure of the vein patterns, e.g. the hair on the back of the hand.

In the past, extraction of hand vein pattern required a high-priced DSP processor for a floating-point operation and a real-time process. For doing a fixed-point operation in the preprocessing steps, Im et al. [58] introduced a shift-and-add architecture on filters. To further enhance the performance of the infrared vein patterns in the back of the hand, Wang et al. [59] suggested the minutiae features of the vein pattern can be used to perform personal verification task. They made use of 13 minutiae points on average for each vein pattern image, including 7 bifurcation and 6 ending points, which are obtained by the cross number (CN) concept [62] from the skeleton of vein patterns. Its results indicated the potential to use minutiae features for hand vein images.

In order to make the reliable and robust authentication, Lee [60] applied texture-based feature extraction techniques to palm vein authentication. Palm vein is considered as a piece of texture and its local features are extracted by 2-D Gabor filter. The features are then encoded in bit string representation as VeinCode by directional coding technique. For each pixel (x, y) , the maximum gradient image is coded to two bits, V_r and V_i . $V_r(x, y)$ is equal to 1 if $d_{Re}(x, y) \geq 0$ or 0 otherwise and $V_i(x, y)$ is equal to 1 if $d_{Im}(x, y) \geq 0$ or 0

otherwise, where $d_{\text{Re}}(x, y)$ and $d_{\text{Im}}(x, y)$ are the maximum values of real and imaginary parts on different orientations. Finally, the vein matching is performed by normalized hamming distance [62].

For the finger vein authentication, Yang et al. [61] suggested to capture finger vein image and finger-dorsa image from the same finger. By fusing a finger vein pattern and finger-dorsa texture, the authentication can be enhanced. They tested on 120 sets of samples from 60 users. The recognition rate of fused feature reached 100%, while 98.3% and 92.7% were obtained by using finger vein and finger-dorsa features separately.

Since blood vessel patterns recognition is useful for authentication and verification, some recent studies suggested that it can be used for criminal and victim identification by using vein visualization method on color images. In Chapter 7, a comparison is performed with a skin texture algorithm and state-of-the-art techniques of blood vessels.

2.2.5 Skin Texture Matching

¹Skin texture is an important topic in many research areas, including development of realistic 3D human models in computer graphic and computer-assisted diagnosis of skin disorders in dermatology. It was proposed for personal identification in a decade ago. In 2004, Cule et al. proposed a method based on bidirectional feature histograms for skin texture modeling and recognition [134]. They applied this method on a database containing

¹ Medically, many biometric traits, including fingerprint, palmprint, and face are pieces of skin. For the sake of clarity, in this chapter, skin texture refers to skin regions without obvious features, such as eyes in faces, ridges in fingerprints, and flexions in palmprints. Thus, faces, fingerprints, and palmprints are not considered as skin texture in this thesis.

skin texture images from 20 subjects collected in a clinical setting at the same day and reported accuracy of 73%. The bidirectional skin model proposed by Cule et al. drew great attention from the computer graphic community for simulating realistic 3D human models. In addition to Cule et al., in 2012, Xie et al. evaluated hand back skin texture from 160 hands collected in a cooperative and controlled environment and reported accuracy of 86.81%. Xie et al. designed a tailor-made imaging system, which uses a high resolution camera, circular light source and pegs, to capture high quality skin texture images with 450 dpi (dot per inch) for their study [135]. Cule et al.'s study, demonstrating that skin texture can be used for personal identification, was largely neglected by the biometric community. The success of other biometric technology in particular face, fingerprint and iris for governmental and commercial applications makes the development of skin texture as a biometric trait unnecessary. Furthermore, it is generally believed that skin texture is not distinctive, like most of other soft biometrics, e.g. height, BMI, and gender, because it does not have obvious features.

Though for commercial and governmental applications, various biometric systems are available in the market, in many forensic cases, skin texture may be the only option. In cases of identifying masked terrorists, gunmen, and rioters in violent crime images, and pedophiles in child sexual offense images, the traditional biometric traits, including face and tattoo, are not always available because criminals hide or cover them to avoid identification. In Chapter 7, skin texture is proposed for criminal and victim identification.

2.3 Image Alignment Methods

In this section, two different approaches are reviewed for image alignment. The first one is point registration methods, which consider two object shapes as two sets of sample points and generate the transformation for aligning these sample points. The second one is Procrustes analysis, which is used to optimize the transformation by analyzing the distribution of a set of shapes. The Procrustes analysis is applied for Chapter 6 and the point set registration is used in Chapter 3, 4, 5, and 7.

2.3.1 Point Registration

Point set registration is a key process in computer vision and medical image processing problems because most of their matching algorithms have to deal with the point sets. To check if they are identical, point set registration is here to find out how these two point sets can be matched. Its objective is to find the correspondences between two sets of points and to obtain the mapping from one point set to the other.

There are many existing point registration methods. Some well-known methods are Iterative Closest Point (ICP) [63], LM-ICP [64], TPS-RPM [65], and Coherent Point Drift (CPD) [66]. ICP is a very fast and popular method which iterates to minimize the total distance between two point sets. Nevertheless, it is also well known that ICP can generate a lot of local minima by assigning correspondences which may not be one-to-one. LM-ICP, which is a robust version of ICP, improves ICP by a general-purpose nonlinear optimization procedure. On the other hand, TPS-RPM and CPD are different from ICP and

LM-ICP because they parameterize the transformation. TPS-RPM parameterizes the transformation by Thin Plate Spline [67] and alternatively updates the transformation parameters and corresponding matrix by deterministic annealing. Nevertheless, CPD applies Gaussian mixture models (GMM) for data modeling and optimizes internal parameters by expectation-maximization (EM) method. CPD and TPS-RPM are quite similar in the non-rigid case because they both regularize the displacement field, but they use different regularization terms. However, CPD is more generalized on different cases (rigid, affine, and non-rigid) but TPS-RPM focuses on non-rigid case only. In addition, CPD is the first method to provide the complete solution for the rigid case. Furthermore, CPD can perform better than TPS-RPM in the presence of outliers [66]. Thus, CPD is chosen to be the alignment method for leg geometry in this study.

CPD aims to solve three main problems in point registration. First, the obvious problems of point set registration are noise, outliers, and missing points. Noise may appear when a point set is obtained from processes of image acquisition and feature extraction. Meanwhile, outliers can exist when some points may exist in one point set but may not have corresponding points in another point set. In addition, missing points can easily happen in the image when there is occlusion or inaccurate feature extraction in the image. Therefore, the presence of noise, outliers, and missing points can complicate the problem in the matching and may obtain the correspondences which are not one-to-one. So, when the correspondences cannot be found correctly, two point sets cannot be matched exactly.

To tackle these problems, CPD makes use of GMM models by considering one point set as the Gaussian Mixture Model (GMM) centroids and another one as the data points. The parameters of GMM models are iteratively optimized by the Expectation

Maximization (EM) method. To solve the point registration problem more precisely, CPD suggested three solutions on three cases, rigid point set registration, affine point set registration, and non-rigid point set registration. Based on the experimental results of the paper [66], the CPD had the highest tolerance of the noise and outliers when it was compared with LM-ICP and TPS-RPM. Thus, it is a very useful tool for solving the point registration problem.

2.3.2 Procrustes Analysis

Procrustes analysis [68-69] is always used to obtain a common shape template, which is also called as coordinate reference [68]. It can be divided into ordinary and generalized. The ordinary Procrustes analysis is designed to align two shapes. The generalized Procrustes analysis is used to align three or more shapes.

For ordinary Procrustes analysis, the transformation between two shapes is performed by translation, uniform scaling, and rotation. Initially, two sets of points $((x_1^{(1)}, y_1^{(1)}), (x_2^{(1)}, y_2^{(1)}), \dots, (x_n^{(1)}, y_n^{(1)}))$ and $((x_1^{(2)}, y_1^{(2)}), (x_2^{(2)}, y_2^{(2)}), \dots, (x_n^{(2)}, y_n^{(2)}))$ are extracted from two shapes. By calculating the centroid and the root mean square of each shape k ,

$$(\bar{x}^{(k)}, \bar{y}^{(k)}) = \left(\frac{1}{n} \sum_{i=1}^n x_i^{(k)}, \frac{1}{n} \sum_{i=1}^n y_i^{(k)} \right), \quad (2.5)$$

$$s^{(k)} = \sqrt{\frac{\sum_{i=1}^n (x_i^{(k)} - \bar{x}^{(k)})^2 + (y_i^{(k)} - \bar{y}^{(k)})^2}{n}}, \quad (2.6)$$

the points of each shape k are translated and scaled to $(x_i^{(k)}, y_i^{(k)}) \rightarrow ((x_i^{(k)} - \bar{x}^{(k)})/s^{(k)}, (y_i^{(k)} - \bar{y}^{(k)})/s^{(k)})$, where $i = 1, \dots, n$. After two point sets are aligned by translation and uniform scaling, the point sets of shape A and B can be stored as

$$\mathbf{A} = \begin{bmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ y_1^{(1)} & \dots & y_n^{(1)} \end{bmatrix}, \quad (2.7)$$

$$\mathbf{B} = \begin{bmatrix} x_1^{(2)} & \dots & x_n^{(2)} \\ y_1^{(2)} & \dots & y_n^{(2)} \end{bmatrix}. \quad (2.8)$$

By singular value decomposition, $\mathbf{A}^T \mathbf{B} = \mathbf{U} \mathbf{D} \mathbf{V}^T$. For the rotation, the orthogonal Procrustes problem is formed to obtain the solution of rotation $\mathbf{R}$, which is

$$\begin{aligned} \mathbf{R} &= \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{W} - \mathbf{B}\|_F^2 \quad \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \\ &= \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{W}\|_F^2 + \|\mathbf{B}\|_F^2 - 2\operatorname{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{B}) \quad \text{s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \\ &= \underset{\mathbf{W}}{\operatorname{argmax}} \operatorname{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{B}) \\ &= \underset{\mathbf{W}}{\operatorname{argmax}} \operatorname{tr}(\mathbf{W}^T \mathbf{U} \mathbf{D} \mathbf{V}^T) \\ &= \underset{\mathbf{W}}{\operatorname{argmax}} \operatorname{tr}(\mathbf{V}^T \mathbf{W}^T \mathbf{U} \mathbf{D}) \\ &= \underset{\mathbf{W}}{\operatorname{argmax}} \operatorname{tr}(\mathbf{Z} \mathbf{D}), \end{aligned} \quad (2.9)$$

where $\mathbf{Z} = \mathbf{V}^T \mathbf{W}^T \mathbf{U}$ and $\operatorname{tr}(\mathbf{Z} \mathbf{D}) = \sum_{i=1}^n Z_{i,i} D_{i,i}$. Since $\mathbf{V}^T, \mathbf{W}^T, \mathbf{U}$ are orthogonal, $\mathbf{Z}$ is also orthogonal. To maximize $\mathbf{R}$, $Z_{i,i} = 1$ which implies $\mathbf{Z} = \mathbf{I}$ and $\mathbf{W} = \mathbf{V} \mathbf{U}^T$. The optimal rotation matrix $\mathbf{R}$ is $\mathbf{V} \mathbf{U}^T$. After two shapes are aligned by translation, scaling, and optimal rotation as explained above, Procrustes distance, which is the square root of the sum of squared differences (SSD) between corresponding points, is used to measure the difference in two shapes.

Generalized Procrustes analysis (GPA) is similar to the ordinary Procrustes analysis in principle, but it is extended to compare three or more shapes to an optimally determined “mean shape”. Its iterative approach is as follows:

1. Choose a reference shape (e.g. the first one in the set or arbitrarily select one in the set).
2. Align all the remaining shapes to current reference shape by the ordinary Procrustes analysis method as explained above.
3. Re-estimate the mean from the aligned shapes.
4. If the Procrustes distance between the mean shape and the reference shape exceeds a threshold, the estimated mean is set to be a reference shape and return to step 2.

Figure 2.1 shows an example using simple mean operation and the generalized Procrustes analysis to calculate a common shape template. The template made by the generalized Procrustes analysis is obviously smoother than that formed by taking the average of all shapes and its shape is closer to the original shapes.

Figure 2.1 An example of simple mean operation and generalized Procrustes analysis for calculating a common shape template. (a) Six leg shapes in different colors are overlapped, (b) a template is obtained by taking the average of all shapes, and (c) a template is generated by the generalized Procrustes analysis.

2.4 Feature Extraction Methods

In this section, low-level feature extraction methods are reviewed because they are used by the proposed algorithms of the following chapters.

2.4.1 Shape-Based Feature Extraction

Shape feature extraction is a way to obtain shape-related features from an image. From the review in [154], there are six categories, one-dimensional function for shape representation, polygonal approximation, spatial interrelation feature, moments, scale-space methods, and shape transform domains. In one-dimensional function for shape representation, contour curvature is one way to represent the shape. Since the curvature is useful for obtaining significant points in the contour, which have high absolute curvature values, Wang et al. [155] showed its uses in shape recognition. However, the curvature is scale dependent and it has to be used with multi-scale methods or aligned images. In polygonal approximation, it is used because it can capture the whole shape by ignoring the minor variations along the edge. In spatial interrelation feature, it makes use of the relation of pixels or curves to describe the region or the contour of shape. Bounding box is one of the methods which subdivide the bounding area to fit different shapes. It is robust to the noise around the boundaries and invariant to rotation, scaling, and translation. In moments, it is computationally simple, but it is sensitive to noise in high-order moments and has information redundancy by using its non-orthogonal basis. In shape transform domains, a

conventional method is Fourier descriptor, which describes the shape by using normalized Fourier transformed coefficients. Its advantage is simple to compute, robust to noise, and compact. In scale-space methods, it can be applied with categories by changing their scales. The higher scales are used to describe global pattern information and lower scales are used to describe lower scales. Its advantage is coherence with human perception and is to obtain more information from a contour with different scales. In this thesis, different types of shape features are extracted for representing the shape more accurately.

2.4.2 Edge-Based Feature Extraction

For the earlier methods, Sobel and Prewitt detectors [156] make use of local gradient for detecting edges, but they perform poorly for the blurry and noisy edges [157]. Some well-known methods also include Canny detector [158] and Laplacian of Gaussian (LoG) edge detector [159]. Canny detector is more sensitive to weak edges, but also spurious and unstable boundaries. LoG edge detectors is to detect the edge by finding the zero crossings of the responses outputted from Gaussian filtering with the Laplacian operator. Also, it is not robust to the high density of edges and noisy environment, like hair [157]. One more advanced method is Gabor filter [90], whose frequency and orientation representations are close to human visual system. It is formed by a set of Gaussian kernel function with a sinusoidal plan wave. The advantage of Gabor filter is flexible to be tuned with different frequencies and orientations. In this thesis, Gabor filter is used to extract the edge features.

2.4.3 Texture-Based Feature Extraction

To represent different kind of texture, texton learning method [148] is one of state-of-art methods. To extract the texture features, texton dictionary is built by convolving training images with a filter bank and obtaining textons from clustering exemplar filter responses by k -means clustering. Based on texton dictionary, the texture feature is extracted from convolving an image with the filter bank and labelling each filter response with the texton which is the closest one to its filter responses. Although the filter banks can be set to be discriminative by including many orientations and scales, it is not invariant to rotation. In addition, the optimal number of filter banks is not certain of getting the best result. The shortcomings of texton learning method can be remedied by using another popular texture extraction method, local binary patterns (LBP) [53], which takes a threshold on the neighborhood of each pixel to the center value and considers the result as a binary number. LBP is controlled by number of sampling P and a circle of radius R for neighborhoods. LBP can be improved to be rotation invariant LBP ($LBP_{P,R}^{ri}$) and rotation invariance with uniform LBP ($LBP_{P,R}^{riu2}$). Uniform patterns ($LBP_{P,R}^{u2}$) mean the restriction on the number of spatial transitions (bitwise 0/1 change) based on the observed textures. Dominant local binary patterns (DLBP) [146] and completed modeling of local binary patterns (CLBP) [147] are two extensions on LBP. DLBP is designed to effectively capture the dominating patterns based on the occurrence frequencies of all rotation invariant patterns and CLBP is suggested to encode the signs and the magnitudes of the image local difference. In Chapter 7, texton learning method, DLBP, and CLBP are compared to the proposed identification algorithm with uniform LBP.

2.5 Feature Matching Methods

In this section, feature matching methods are reviewed. The first one is to make use of distance measure among features and the second one is to apply learning methods for matching.

2.5.1 Distance Based Approach

In [4], the Gabor orientation features are extracted locally to form the feature vectors. Then, the distance based or nearest-neighbor approach is applied to search the matched pair. Without considering the importance of features, two feature vectors are directly compared by using Chi-square distance.

$$X(\mathbf{H}_1, \mathbf{H}_2) = \sum_{k=1}^N \frac{(H_1(k) - H_2(k))^2}{H_1(k) + H_2(k)}, \quad (2.10)$$

where $\mathbf{H}_1$ and $\mathbf{H}_2$ are feature vectors and N is the total number of features. By calculating Chi-square distance for each pair of image, the pair with smallest distance is the matched pair. Other distance measures, e.g. Euclidean distance and Mahalanobis distance, can be considered. However, this approach assumes the most similar pair is represented by the smallest distance, but it is not always true for the nonlinear distribution of feature points. When the number of features is large, the distance becomes not useful because the feature space is very spacious for every feature points and the feature distance is not significant. Besides, this approach can be used only in cases that assume the same importance of all

features. In Chapter 5 and 7, some learning approaches are shown to outperform the distance based approach.

2.5.2 Learning Approaches

In this subsection, two learning approaches are introduced. The first one is to learn the useful information from a large feature set and the second one is to learn the ranking from a feature set. Both approaches are presented by reviewing some existing methods. The first approach is used in Chapter 5 and Chapter 7 and the second approach is applied in Chapter 6.

2.5.2.1 Partial Least Squares for Large Feature Set

Large feature set is a challenging problem for researchers. With increasing dimensionality, the complexity to design a model also increases. Since large feature set may contain a lot of irrelevant and redundant information, the learning performance can be significantly degraded without proper learning algorithms. In this section, large feature set means $p \gg n$. It is noted that a n -by- p feature set is referred to as p -dimensional. Some suggested methods using Partial Least Squares (PLS) to handle large feature set [70-73] are reviewed.

PLS [74] is a model searching for the internal relationship between a feature matrix $\mathbf{X}_{(n \times p)}$ and a response variable $\mathbf{y}_{(n \times 1)}$, where $\mathbf{X}_{(n \times p)}$ and $\mathbf{y}_{(n \times 1)}$ are mean-centered and variance-scaled, n is the number of training data and p is the dimension of the feature

vector. Latent variables are used to represent their internal relationship, which is defined as

$$\mathbf{X} = \mathbf{V}\mathbf{P}^T + \mathbf{E} , \quad (2.11)$$

$$\mathbf{y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} , \quad (2.12)$$

where $\mathbf{V}$ and $\mathbf{U}$ are the latent components, $\mathbf{P}$ and $\mathbf{Q}$ are the loading matrices, the superscript T represents a matrix transpose operator and $\mathbf{E}$ and $\mathbf{F}$ are the residue matrices.

According to [75], the derivation of PLS without latent variables is

$$\max_{\boldsymbol{\beta}} \text{cov}(\mathbf{X}\boldsymbol{\beta}, \mathbf{y}) \quad \text{s.t.} \quad \boldsymbol{\beta}^T \boldsymbol{\beta} = 1 , \quad (2.13)$$

where $\boldsymbol{\beta}$ is a $p \times 1$ regression coefficient vector of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}'$ and $\mathbf{f}'$ is a residual vector.

The original PLS, which includes latent variables and is computed by NIPALS algorithm [74], is commonly applied to extract $\boldsymbol{\beta}$ by using $\mathbf{V} = \mathbf{X}\mathbf{W}$.

To detect people in images and videos, Schwartz et al. [70] suggested to employ PLS analysis as a dimensionality reduction technique for projecting the feature set from original 170,000 to 20 dimensions. Its dimensionality reduction of a feature vector $\mathbf{Z}_{(1 \times p)}$ is performed by $\mathbf{Z}\mathbf{W}$. Based on their findings, PLS achieved better class separation than PCA in their training dataset.

For detecting vehicles with aerial images, A. Kembhavi [71] introduced a very large and rich set of feature sets (around 70,000 elements) to improve the detection performance. The feature set is reduced by using PLS and feature selection analysis. After PLS is performed on the feature set, four informative vectors, including VIP score, Regression Coefficients, Correlation, and Covariance Procedures Vector (CVP) are used to reduce the feature size instead of direct projection of the feature set by using $\mathbf{W}$. By using $\boldsymbol{\beta}$ and $\mathbf{W}$ from NIPALS algorithm, VIP score for each j^{th} regression coefficient is defined as

$$VIP_j = \sqrt{p \sum_{k=1}^d \beta_k^2 W_{jk}^2 / \sum_{k=1}^d \beta_k^2}, \quad (2.14)$$

if VIP_j score is less than 1, such coefficient is not retained. Regression Coefficients are the absolute value of regression coefficients, i.e. $|\boldsymbol{\beta}|$. Correlation is the Pearson correlation coefficients between a feature matrix $\mathbf{X}_{(n \times p)}$ and a response variable $\mathbf{y}_{(n \times 1)}$, which is defined as

$$R = \frac{\mathbf{X}^T \mathbf{Y}}{n - 1}. \quad (2.15)$$

The last one is CVP for proposed by Reinikainen and Hoskuldsson [76]. By a feature matrix $\mathbf{X}_{(n \times p)}$ and a response variable $\mathbf{y}_{(n \times 1)}$, CVP is formed as

$$CVP = \mathbf{X}^T \mathbf{Y} \mathbf{Y} \mathbf{X}. \quad (2.16)$$

By using these four informative vectors, the top K element is chosen by sorting their scores in a descending order.

Although feature selection with PLS is used in the previous two approaches, Schwartz et al. [72-73] suggested to build one-against-all PLS model directly without feature selection instead. To learn one-against-all PLS model, one subject is considered as a positive sample for each time and the remaining subjects are used as negative samples for training. They found that the imbalance of the class distributions has little influence on the performance of PLS modeling from the works of Barker and Rayens [77] and Qu et al. [78]. After the PLS model for each subject is trained to obtain different $\boldsymbol{\beta}$, the regression response of a probe sample with feature vector $\mathbf{Z}_{(1 \times p)}$ is performed by $\mathbf{Z}\boldsymbol{\beta}$ for each model. The model with the highest response is the best match with a probe subject. Since one-against-all approach provided a good result for Schwartz et al., two identification

algorithms are built by using the concept of one-against-all PLS model and an improvement of the model is also demonstrated in Chapter 5 and Chapter 7.

2.5.2.2 Learning to Rank

Learning to rank is a machine learning and information retrieval technique which is used to train a model for a ranking purpose. Normally, the training involves a number of sets, which consist of objects and label indicating their rankings. For examples, some movie lists are ordered by number of views. Each movie is an object and its label is the number of views. Then, a ranking function is trained by minimizing a certain loss function from training data. Finally, the testing stage is to find a ranked list from a new set of objects. Some ranking methods are reviewed in the following paragraphs.

Joachims [79] suggested improving retrieval function from clickthrough data, which includes user query, ranking presented to the user, and the set of links the user clicked on, by using Ranking SVM. If an item d_i is ranked higher than another item d_j for an ordering r , then $(d_i, d_j) \in r$. And then, $(d_i, d_j) \in r \Leftrightarrow \mathbf{w}\Phi(q, d_i) > \mathbf{w}\Phi(q, d_j)$, where q is a user query, $\mathbf{w}$ is a weight vector adjusted by learning, and $\Phi(q, d)$ is a mapping onto features that describes the match between a user query q and an item d . Thus, the Ranking SVM is formulated as

$$\begin{aligned} \min_{\mathbf{w}, \xi} \quad & \frac{\|\mathbf{w}\|^2}{2} + C \sum \xi_{i,j,k}, \\ \text{s.t.} \quad & \forall (d_i, d_j) \in r_1^*: \mathbf{w}\Phi(q_1, d_i) > \mathbf{w}\Phi(q_1, d_j) + 1 - \xi_{i,j,1}, \\ & \dots \end{aligned} \tag{2.17}$$

$$\begin{aligned} \forall (d_i, d_j) \in r_n^*: \mathbf{w}\Phi(q_n, d_i) &> \mathbf{w}\Phi(q_n, d_j) + 1 - \xi_{i,j,n}, \\ \forall i \forall j \forall k: \xi_{i,j,k} &\geq 0, \end{aligned}$$

where $\xi_{i,j,k}$ is a slack variable, r_k^* is the target ranking for q_k , and C is a parameter to control the margin size against training error.

Burges et al. [80] applied RankNet, which is an implementation of a pair-based cost function with neural network, on ranking toy data and data from a commercial internet search engine. For a simple neural network, its cost function can be expressed as $\sum_{i=1}^q f(\mathbf{o}_i, \mathbf{t}_i)$ where $\mathbf{o}_i$ is the outputs of the neural network and $\mathbf{t}_i$ is its targets. RankNet is accomplished by modifying the cost function as $\sum_{i=1}^{q-1} f(\mathbf{o}_{i+1} - \mathbf{o}_i, \mathbf{t}_{i+1} - \mathbf{t}_i)$ and the corresponding back-propagation equations are developed by the modified cost function.

Friedman [81] introduced to gradient boosted regression tree for regression. When the input data are $\{y_i, \mathbf{x}_i\}_{i=1}^N$, the loss function in prediction $f(x)$ for y is $L(y, f(x))$. At each iteration, the regression trees, which use only order information on the individual input variables $\mathbf{x}$, are built to best predict the current residuals g_{im} and the terminal node updates are based on the optimal solution of loss function, which is also considered as line search. The details can be found in [81].

Breiman [82] introduced the random forests for classification and regression. The main idea of random forests is to inject the right kind of randomness for accurate classifiers and regressors. In his study, bagging and random features are used for randomness. Bagging is a process to select random samples with replacement of the training set repeatedly. Then, the trees are built by fitting to these samples. Finally, predictions for unseen samples can be made by averaging or majority vote from the responses of the trees. To be more robust with respect to noise, random forests include a random of features to

split each node. This process is used to prevent very few features dominant in selection of many trees, when these features are very strong predictors for the response variable.

Freund et al. [83] suggested RankBoost algorithm for solving the problem of combining preferences, which can be combining the results of different search engines, or ranking movies for a user based on the movie rankings from other users. Freund et al. stated that “one important property of these tasks is that the most relevant information to be combined represents *relative preferences* rather than *absolute ratings*”. RankBoost makes use of the weak learner, which produces a weak ranking, to build a strong learner by updating the weight of different pairs of instances for each iteration. An important pair is supposed to get higher weight in the update. The pseudocodes of RankBoost and weak learner algorithms are shown in [83].

Metzler and Croft [84] applied coordinate ascent on parameter estimation for supervised training algorithms. They formed an evaluation function $E(\mathbf{R}_\Lambda; T)$, which produces real valued output given a set of ranked lists $\mathbf{R}_\Lambda = \{R_i(\Lambda)\}_{i=1}^N$ and the training data T . Coordinate ascent is an algorithm that iteratively optimizes a multivariate objective function by using one parameter and holding all other parameters fixed, i.e.

$$\lambda'_i = \operatorname{argmax}_{\lambda_i} E(\mathbf{R}_\Lambda; T). \quad (2.18)$$

One parameter is optimized for each time. The process continues to update the other parameters until some convergence criteria is met. If E is not partially differentiable, a line search is used to find the solution of Equation (2.18). As they mentioned, the optimization converges gradually on objective functions with long ridges.

Cao et al. [85] suggested ListNet, which is a listwise approach with neural network and gradient descent, for predicting a list of objects. In the training, a feature vector $\mathbf{x}_j^{(i)} = \Psi(q^{(i)}, d_j^{(i)})$ is formed from a pair of a query and an object $(q^{(i)}, d_j^{(i)})$, where a set of queries $\mathbf{Q} = (q^{(1)}, q^{(2)}, \dots, q^{(m)})$, a list of objects $\mathbf{d}^{(i)} = (d_1^{(i)}, d_2^{(i)}, \dots, d_{n^{(i)}}^{(i)})$, $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n^{(i)}$. Each list of features $\mathbf{x}^{(i)} = (\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{n^{(i)}}^{(i)})$ has the corresponding list of scores $\mathbf{y}^{(i)} = (y_1^{(i)}, y_2^{(i)}, \dots, y_{n^{(i)}}^{(i)})$, which can be given by human, e.g. the number of clicks on the document $d_j^{(i)}$, in the training. By using a ranking function f_ω trained by neural network model ω , a score list $\mathbf{z}^{(i)}(f_\omega) = (f_\omega(\mathbf{x}_1^{(i)}), f_\omega(\mathbf{x}_2^{(i)}), \dots, f_\omega(\mathbf{x}_{n^{(i)}}^{(i)}))$ is generated. The top k probability of corresponding documents $(d_{j_1}^{(i)}, d_{j_2}^{(i)}, \dots, d_{j_k}^{(i)})$ is obtained as

$$P_{\mathbf{z}^{(i)}(f_\omega)}(\mathbf{o}_k(j_1, j_2, \dots, j_k)) = \prod_{a=1}^k \frac{\exp(f_\omega(\mathbf{x}_{j_a}^{(i)}))}{\sum_{b=a}^{n^{(i)}} \exp(f_\omega(\mathbf{x}_{j_b}^{(i)}))}, \quad (2.19)$$

where $\mathbf{o}_k(j_1, j_2, \dots, j_k)$ is defined as all the permutations of $(j_1, j_2, \dots, j_k)$. By applying the Cross Entropy, the listwise loss function for query $q^{(i)}$ becomes

$$L(\mathbf{y}^{(i)}, \mathbf{z}^{(i)}(f_\omega)) = - \sum_{\forall g \in \mathbf{o}_k} P_{\mathbf{y}^{(i)}}(g) \log(P_{\mathbf{z}^{(i)}(f_\omega)}(g)). \quad (2.20)$$

To minimize the loss function, a learning algorithm is suggested. In the implementation, Cao et al. used a linear model as a ranking function, i.e. $f_\omega(\mathbf{x}_j^{(i)}) = \langle \omega, \mathbf{x}_j^{(i)} \rangle$.

Ranking method is also applied for age estimation. Recently, Change et al. [86] suggested CSOHR, which is a cost-sensitive ordinal hyperplanes ranking method, to improve aggregating performance of a series of binary classification results in age ranking

estimation. Let $\mathbf{x}_i$ be the feature vector extracted to represent a training image I_i and y_i be an age label of I_i . For a given age k , the dataset $\mathbf{X} = \{\mathbf{x}_i | i = 1, \dots, M\}$ is separated into two subsets, $\mathbf{X}_k^+$ and $\mathbf{X}_k^-$ into

$$\begin{aligned}\mathbf{X}_k^+ &= \{(\mathbf{x}_i, +1) | y_i \leq k\} \\ \mathbf{X}_k^- &= \{(\mathbf{x}_i, -1) | y_i > k\}.\end{aligned}\quad (2.21)$$

They defined the cost of misclassifying a data sample $\mathbf{x}_i$ in each age k as

$$cost_k(\mathbf{x}_i) = \begin{cases} C(k - y_i + 1), & \text{if } y_i \leq k \\ C(y_i - k), & \text{if } y_i > k \end{cases} \quad (2.22)$$

where y_i is an age label of $\mathbf{x}_i$ which is considered as a rank order, $C(e_i) = \mathbb{I}[e_i > L] \cdot (e_i - L)$, e_i is the error of $\mathbf{x}_i$, L is a range of tolerance, and $\mathbb{I}[\cdot]$ is 1 if the condition is true and zero otherwise. By using Equation (2.21), k th binary classification problem of SVM with cost sensitivities is defined as

$$\begin{aligned}\min_{\mathbf{w}_k, b_k, \xi} \quad & \frac{1}{2} \langle \mathbf{w}_k, \mathbf{w}_k \rangle + C \left[\sum_i cost_k(\mathbf{x}_i) \xi_i \right], \\ \text{s.t.} \quad & s_k[i] (\mathbf{w}_k^T \phi_k(\mathbf{x}_i) + b_k) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \forall i,\end{aligned}\quad (2.23)$$

where $s_k[i] = 1$ if $\mathbf{x}_i \in \mathbf{X}_k^+$, $s_k[i] = -1$ if $\mathbf{x}_i \in \mathbf{X}_k^-$, ϕ_k is an implicit mapping in the Hilbert space, and $(\mathbf{w}_k, b_k)$ are the hyperplane parameters in the implicit feature space.

The discriminating function $f_k(x)$ is then adopted as

$$f_k(x) = \langle \mathbf{w}_k, \phi_k(x) \rangle + b_k. \quad (2.24)$$

Based on $f_k(x)$, the age rank $r(x)$ is inferred by aggregating the binary decision results as

$$r(x) = 1 + \sum_{k=1}^{K-1} \mathbb{I}[f_k(x) > 0], \quad (2.25)$$

where $\mathbb{I}[\cdot]$ is 1 if the inner condition is true, and 0 otherwise. Although the CSOHR algorithm has an aggregation of binary decision, it may neglect the importance of each decision. Thus, the probability-based aggregation is used in Chapter 6.

2.6 Summary

The reviews of biometric traits, criminal and victim identification, image alignment methods, feature extraction method, and feature matching method is provided in this chapter. In the section of biometric traits, hand geometry and soft biometric traits are studied. In the section of criminal and victim identification, different kinds of existing works are given and a detailed review is given to androgenic hair patterns matching, soft biometrics and witness description, blood vessel patterns, and skin texture, which are topics closely related to the proposed algorithms in this thesis. Afterwards, image alignment methods, feature extraction methods, and feature matching methods are also provided because they are applied in the proposed algorithms. In the following chapters, the proposed algorithms are developed based on the reviews of this chapter.

Chapter 3

Proposed Feature Extraction Schemes for Soft Biometric Traits

3.1 Overview

In forensic investigation, identifying criminals and victims is always challenging but vital. Law enforcement agents have used many biometric traits, such as DNA, fingerprint, face, and palmprint. Nevertheless, they cannot be used in legal cases when only non-facial body sites of criminals or victims in evidence images are exposed for identification. Normally, these cases include riots, masked gunmen, and child pornography. In order to identify these criminals or victims, skin marks, blood vessels hidden in color images, androgenic hair patterns, and tattoos have been suggested. However, they have limitations. Tattoos are not always available. Skin marks and blood vessels are normally applied to high resolution images. Androgenic hair patterns give useful identification information even in low resolution images, but its performance is still not good enough for the real cases. Therefore, new biometric traits are still on demand, especially for low resolution evidence images. This chapter evaluates lower leg geometry as a soft biometric trait for criminal and victim identification. Since lower legs are often observable in evidence images, they are considered in this study. The evaluation algorithm first aligns two lower leg shapes from input images and extracts geometric features, including the partial sum of squared difference, the polynomial coefficients and the number of intersection points of the aligned

leg shapes. To analyze the geometric features, support vector machines, neural networks and decision trees are employed for classification. The low resolution dataset described in Section 3.2, which is 1,138 images from 283 subjects, is used for this study. The experimental results indicate that the lower leg geometry can be proposed to be an effective soft biometric trait. This study can provide a foundation for further research on body geometry of criminal and victim identification.

The aim of this study is neither developing a new segmentation method for lower leg images nor building a complete system for identification. Its aim is to evaluate the information extracted from lower leg shapes for personal identification. Some examples of the extracted lower leg shapes are shown in Figure 3.1.

3.2 Evaluation Datasets

3.2.1 Data Collection

For the data collection, 1138 images from 283 legs without tattoos were captured by two cameras – a Nikon D70s camera (the maximum resolution of 3008×2000 pixels) and a Canon EOS 500D camera (the maximum resolution of 4752×3168 pixels). The subjects of these images come from Arab, Bengali, Caucasian, Chinese, Eurasian, Indian, Indonesian, Javanese, Malay, Punjabi, and Vietnamese. The subjects were captured in two occasions with an interval of around 11 days. The distance between the cameras and the subjects was about 1.6m. On the first occasion, 575 images from 283 legs were collected and on the

second occasion, 563 images from the same 283 legs were collected. 1 to 4 images were taken for each leg. Table 3.1 shows the details.

Figure 3.1 Examples of lower leg shapes for this evaluation.

Table 3.1 The details of dataset.

Number of images captured for each subject	Number of legs (the first occasion, probe set)	Number of legs (the second occasion, gallery set)
1	5	6
2	265	274
3	12	3
4	1	0

On these two occasions, the poses of the subjects were not strictly fixed and the illumination changes were allowed, so the captured images could have different viewpoints, backgrounds, poses, illumination, and image resolutions. Some high resolution images from different viewpoints and illumination conditions are shown in Figure 3.2. The pose differences can be noted by the highlighted skin marks, scars, and other skin features. Figure 3.3 shows the images of different legs with the original scale ratios.

Figure 3.2 High resolution images in the dataset. (a)-(f) Three leg pairs. Some images are blurred because they are resized to be the same size for each pair.

Figure 3.3 18 original images of different legs in the dataset.

3.2.2 Preprocessing

Since the images in the database are not standardized, preprocessing is required to crop the interest region from the images. An automatic segmentation scheme and a manual

correction are applied for extracting the right leg from each image. Then, the semi-automatic segmentation scheme is summarized as follows:

Step 1: Each channel of the input image $(\mathbf{R}, \mathbf{G}, \mathbf{B})$ was filtered by 3×3 median filter to be

$(\mathbf{R}_M, \mathbf{G}_M, \mathbf{B}_M)$. Then, based on the suggestion in the paper [87], the ratio $I_R = \mathbf{R}_M / (\mathbf{R}_M + \mathbf{G}_M + \mathbf{B}_M)$ and $I_G = \mathbf{G}_M / (\mathbf{R}_M + \mathbf{G}_M + \mathbf{B}_M)$ were calculated to separate the skin from the background.

Step 2: Thus, the pixels were selected from the input image if $I_R > \bar{I}_R$ and $I_G < \bar{I}_G + \text{std}(I_G)$ where $\bar{x}$ and $\text{std}(x)$ are the mean and standard deviation of x .

Step 3: Afterwards, several morphological operators were applied to the image to remove the noise. Firstly, a pruning operation, which includes connecting diagonal pixel, bridging unconnected pixels, filling isolated interior pixels, and removing isolated pixels, and a dilation operation are performed on the image. Secondly, the same pruning operation and an erosion operation are applied. Thirdly, some small regions are removed by setting a threshold. Fourthly, a closing and an opening operation are applied to make the image more robust to noise.

Step 4: Then, the region of interest, which is the minimum rectangle to include these six points A, A', M, M', B, B' shown in Figure 3.4, was extracted from the image. MM' is the longest horizontal distance, then AA' and BB' are the shortest lines above and below the line MM', respectively. This definition is the same as Su et al.'s method [4] if the leg is not tilted.

Figure 3.4 shows the difference between Su et al.'s preprocessing method (1st row) and the proposed preprocessing method (2nd row). Su et al.'s method only relies on the horizontal

lines AA' and BB'. However, the proposed preprocessing scheme can handle the leg inclination and does not limit AA' and BB' to be horizontal lines. After the image is cropped, the region of interest is resized to $845 \times 1,781$ and 142×298 respectively for high resolution and low resolution study. The high resolution image size is found by taking an average of the size of images from gallery set and the low resolution image size is based on the suggestion in the Su et al.'s method [4]. The images are captured from the same group of subjects in two sessions of interval around 11 days. In these two sessions, there are different places for taking images and no strict poses are required. Some 142×298 preprocessed images from sessions 1 and 2 are shown in Figure 3.5. It shows that the images in different sessions can be very different in terms of viewpoints and illumination even though the images are preprocessed and cropped. Thus, the alignment step is unavoidable to enhance the identification results.

3.3 An Algorithm Used in the Evaluation

In this study, the geometric information in lower leg shapes is evaluated by an algorithm composed of an alignment scheme, a feature extraction scheme, and a matching scheme. The alignment scheme is involved because different poses and viewpoints of the legs are collected from the testing database with low resolution in Section 3.2. To further illustrate the alignment differences, Figures 3.6 and 3.7 show two examples. Figures 3.6(a) and (b) show two preprocessed images from the same leg and Figure 3.6(c) shows their leg shapes, which are directly overlaid. Figures 3.7(a) and (b) show two preprocessed images from different legs and Figure 3.7(c) shows their leg shapes, which are directly overlaid.

Figures 3.6(c) and 3.7(c) clearly show the unaligned shapes can be very different. Obviously, the leg shapes in Figure 3.7(c) are very different because they come from different legs, but the leg shapes in Figure 3.6(c) are very different too even though they come from the same leg. The features can be highly influenced by viewpoint and pose variations when they are extracted from the unaligned shapes. Thus, image alignment is important before the shape features can be extracted. To align the leg shapes properly, an affine CPD method, which is a point set registration method explained in Section 2.3.1, is employed. Since it is one of the state-of-the-art methods, it can be useful for the alignment of the leg shapes.

Figure 3.4 The preprocessing step: (a) the input image, (b) the region of interest defined by six edge points A , A' , M , M' , B , and B' . The horizontal distance between M and M' is the longest. AA' is the shortest distance above MM' and BB' is the shortest distance below MM' . (c) The segmented input image according to the region of interest. The first and second rows illustrate Su et al.'s [4] and our preprocessing methods.

(a)

(b)

Figure 3.5 Some examples of 142×298 preprocessed images. (a) Preprocessed images from session 1. (b) Their corresponding preprocessed images from session 2. For example, the leftmost image in the first row in (a) refers to the leftmost image in the first row in (b).

Figure 3.6 Alignment of the same legs. (a)-(b) Two images of the same leg captured at different sessions. (c) Leg shape extracted from (a) and (b) overlaid, and (d) aligned shapes.

Figure 3.7 Alignment of different legs. (a)-(b) Images of different legs, (c) leg shapes extracted from (a) and (b) overlaid, and (d) aligned shapes.

In order to present clearly, a set of notations is defined. Given leg shapes from two images, they can be represented by two sets of points in $\mathbb{R}^2$. One is denoted as a model set $\mathbf{X}$ with size N_m by 2 and the other is denoted as a data set $\mathbf{Y}$ with size of M_d by 2. An affine

transformation is represented by $\mathbf{T}_A(\cdot, \tau_A)$, which is controlled by a parametric vector τ_A . In the proposed alignment scheme, these alignment parameters τ_A can be determined by the affine CPD method. Then, the point set $\mathbf{X}$ is transformed into $\mathbf{X}_T = \mathbf{T}_A(\mathbf{X}, \tau_A)$ by using the affine transformation τ_A . In Figures 3.6 and 3.7, the alignment results from the leg shapes can be shown respectively in Figures 3.6(d) and 3.7(d). Even though the shapes from the same leg in Figure 3.6(d) are aligned perfectly, the shapes of different legs in Figure 3.7(d) still have a notable difference.

Since the number of points in $\mathbf{X}_T$ and $\mathbf{Y}$ may not be the same, sampling is taken to make them have the same size, i.e. $\mathbf{X}_T = [(x_{ms}(l), y_{ms}(l)), \dots, (x_{ms}(L), y_{ms}(L))]$ and $\mathbf{Y} = [(x_{ds}(l), y_{ds}(l)), \dots, (x_{ds}(L), y_{ds}(L))]$, where S represents a sampling function and L represents the total number of sampled points. The regular sampling method is used. Afterwards, the algorithm extracts geometric features, including the partial sum of squared difference, the polynomial coefficients and the number of the intersection points from $\mathbf{X}_T$ and $\mathbf{Y}$.

To extract the partial sum of squared difference, the point coordinates are considered. The differences between corresponding points in two aligned boundaries, which are defined as

$$A(i) = \sqrt{(x_{mr}(i) - x_{dr}(i))^2 + (y_{mr}(i) - y_{dr}(i))^2}, \quad (3.1)$$

$$B(i) = \sqrt{(x_{ml}(i) - x_{dl}(i))^2 + (y_{ml}(i) - y_{dl}(i))^2}, \quad (3.2)$$

are calculated, where $(x_{mr}(i), y_{mr}(i))$ and $(x_{ml}(i), y_{ml}(i))$ are points respectively on the left and right boundaries of $\mathbf{X}_T$ and $(x_{dr}(i), y_{dr}(i))$ and $(x_{dl}(i), y_{dl}(i))$ are points respectively on the left and right boundaries of $\mathbf{Y}$. The features of the partial sum of squared difference defined as

$$\sum_{i=1}^L A(i) , \quad (3.3)$$

$$\sum_{i=floor(KL)}^{floor((K+0.1)L)} A(i) , \quad (3.4)$$

$$\sum_{i=1}^L B(i) , \quad (3.5)$$

$$\sum_{i=floor(KL)}^{floor((K+0.1)L)} B(i) , \quad (3.6)$$

where $K \in [0.01, 0.1, \dots, 0.9]$ and $floor(x)$ is the largest integer less than or equal to x , are computed. According to the observation of Figures 3.6 and 3.7, the calculated features, which represent the global and local shape differences, are expected to have a small value for same legs and a large value for different legs. Equations (3.3)-(3.4) are the features computed from the aligned right boundaries and Equations (3.5)-(3.6) are the features computed from the aligned left boundaries.

Since the distributions of $A(i)$ and $B(i)$ are very different for the pair of same legs and the pair of different legs, polynomial fitting, which is more invariant to sampling noise, is employed for representing their distributions. Then, the features of the polynomial coefficients are derived from the distribution of $A(i)$ and $B(i)$. The polynomial coefficients can be directly computed because the leg shapes $\mathbf{X}_T$ and $\mathbf{Y}$ are aligned. $A(i)$ and $B(i)$ can be fitted on two second-order polynomials, $f_A(x) = a_2x^2 + a_1x + a_0$ and $f_B(x) = b_2x^2 + b_1x + b_0$, where a_0 , a_1 , a_2 , b_0 , b_1 , and b_2 are the polynomial coefficients. And these coefficients are normalized as $a_j/\max(a_k)$ and $b_j/\max(b_k)$, where j and $k \in \{0, 1, 2\}$.

Furthermore, when the shape differences between two aligned shapes are small, their shapes can have a large number of intersection points. Then, the number of intersection points of two aligned shapes can be used as the features. The illustration of the intersection points of two aligned shapes is shown in Figure 3.8.

In total, 30 features are obtained from aligned shapes, including 11 partial sums from aligned right boundaries (Equations (3.1)-(3.2)), 11 partial sums from aligned left boundaries (Equations (3.3)-(3.4)), 2 numbers of intersection points, and 6 polynomial coefficients from left and right boundaries. Since 6 polynomial coefficients are normalized in the extraction, the first 24 features are then normalized to zero and one in the training database and the rest are used directly. Since some features may be redundant or irrelevant, a feature selection based on the mutual information proposed by Peng et al. [88] is employed. For validating the selected features and carrying on the final evaluation, three classification methods are employed, including support vector machines (SVM), neural network (NN), and decision tree (DT).

Figure 3.8 Intersection points on the aligned leg shapes with (a) shapes from the same legs and (b) shapes from different legs.

3.4 Experimental Results

The low resolution evaluation dataset mentioned in Section 3.2 was used for evaluation. The probe set was the 575 images from the first occasion and the gallery set was the 563 images from the second occasion. In each image, leg shapes were sampled to 256 points. In order to match the lower leg shapes in the probe set with those in the gallery set, the algorithm elaborated in Section 3.3 was applied to each pair of images, one from the probe set and another from the gallery set. Then, 323,725 (575×563) feature vectors were obtained. In order to estimate the performance of a classification model more accurately, 2-fold cross validation was applied to the probe set and gallery set. In the first fold, 287 images in the probe set and 283 images in the gallery set were used for training, and then the rest were used for testing. In the second fold, the same 287 images in the probe set and the same 283 images in the gallery set were used for testing, and the rest were used for training.

Different classification methods have different parameter settings. For SVM, a radial basis function was used as a kernel function and the gamma value was set to 0.07. For NN, a feedforward network was used and the number of hidden neurons was 10. For DT, default parameters in Matlab were used, except for the parameter, minleaf, which was set to 30.

The positive (genuine) and negative (imposter) training data were found very unbalanced because mismatched pairs of images (imposter) outnumbered matched pairs of images (genuine). To solve this imbalance problem, sampling was performed. In the first (second) fold, 574 (571) genuine feature vectors from the 287 (288) images were selected

as positive training data and 574 (571) imposter feature vectors were selected as negative training data. Note that there could be more than 500 genuine feature vectors in one fold because most of the legs had 4 images in the evaluation dataset.

Due to the feature selection, the training data was further split into two parts, one was to select features and another was to validate the selected features. In this feature selection process, 10-fold cross validation was used. The selected features for different classifiers are summarized in Table 3.2. Almost all the polynomial coefficients and the numbers of intersection points were selected.

Table 3.2 A summary of the selected features.

	SVM		NN		DT	
Training data	fold 1	fold 2	fold 1	fold 2	fold 1	fold 2
Partial sum of squared difference Equation 3.3-3.6 (total 22)	20	22	16	16	15	12
Polynomial coefficients (Total 6)	6	6	6	6	6	5
Number of intersection points (Total 2)	2	2	2	2	2	2

For final evaluation, the entire training set with the selected features was used to train the classifiers. Then, the testing data was tested by the classifiers. Each classifier gave a similarity value to each testing datum and this value was used to plot cumulative match characteristic (CMC) curves shown in Figure 3.9.

Table 3.3 gives the corresponding rank-1, rank-10 and rank-20 accuracies. The rank-1 accuracies given by SVM and NN are over 40% and their rank-20 accuracies are over 80%. Even though the testing data in the two folds were formed by around 140

different legs and leg shapes of different persons can be quite similar, this performance is still surprisingly good. These experimental results further confirm the lower leg geometry can be an effective soft biometric trait.

Figure 3.9 Cumulative match characteristic curves

Table 3.3 A summary of the experimental results.

Classifier	Rank-one accuracy	Rank-10 accuracy	Rank-20 accuracy
SVM	42.96%	74.96%	81.91%
NN	43.43%	77.04%	82.78%
DT	9.91%	51.48%	72.35%

In Appendix D, Figure D.1 shows the wrongly matched images returned from NN classifier. From the wrong results, it can be observed that their shapes are very close to that of corresponding probe images even though their leg skins are very different. By using leg geometry, false positive increases because many people have similar leg geometry. In addition, some legs are partially occluded by the socks and shorts, the proposed algorithm is not robust to these cases. When the legs are occluded, their leg shapes wrongly align with other legs in the alignment steps. Since the wrong leg shapes are inputted to feature

extraction, the wrong features are obtained for classification. Then, the classifiers cannot get the correct pairs of leg. The proposed work can be further developed for dealing with occlusion on leg geometry.

3.5 Summary

Victim and criminal identification is a very challenging task when their faces are masked and the image resolution is low. To tackle this problem, tattoos, skin marks, androgenic hair patterns, and blood vessels hidden in color images have been proposed in the recent years. However, they have their own weaknesses. Skin marks and blood vessels require high resolution images. Tattoos and androgenic hair patterns are not always available. Although the forensic agents currently use a number of soft biometric traits, e.g. height, BMI, races, and genders for investigation, they ignore the body shape. Lower legs are chosen for this study because they can be clearly observed even in low resolution images. In this evaluation, the partial sum, the polynomial coefficients and the number of intersection points of aligned leg shapes are employed as a feature set. Then, 3 different classification methods (SVM, NN, and DT) are used for constructing the classifiers. Finally, the experimental results show that lower leg shapes can be a soft biometric trait. In addition, they can encourage further study on lower legs and full body geometry for forensic investigation.

Chapter 4

Proposed Alignment Algorithms on Soft Biometric Traits for Boosting Matching Performance

4.1 Using Leg Geometry to Align Androgenic Hair Patterns

4.1.1 Overview

Identifying criminals and victims in digital media can be very challenging when only their non-facial body sites are observed. These criminals and victims can be masked gunmen, paedophiles, and victims in child pornographic images. Although there are many methods suggested, like skin marks and blood vessel patterns, they have their limitations. Blood vessel patterns cannot be captured if the subjects have a high concentration of body fat or melanin. Skin marks can be exposed to the skin, but they cannot be easily observed in low resolution images. On the other hand, the paper [4], which is mentioned in Section 2.2.2, showed that androgenic hair (also called body hair) and its follicles can be used as biometric traits. Then, it further demonstrated that androgenic hair patterns can be effective for personal identification in low resolution images. However, viewpoint and pose variations, which always happen in the real cases, were ignored in that study. In order to

overcome the weaknesses of the previous method, the algorithm based on geometric information for aligning androgenic hair patterns on lower legs is proposed in this chapter. The experimental results on 1,138 low resolution images from 283 different legs show that the proposed alignment algorithm has more than 5% increase.

The schematic diagram of the whole low resolution leg identification is shown in Figure 4.1 and the dashed region shows the proposed alignment algorithm, which will be explained in the following sections.

Figure 4.1 The schematic diagram of the overall low resolution leg identification system. It includes the proposed alignment algorithm and the feature matching method in [4]. The dashed region describes the proposed alignment algorithm.

Firstly, the low resolution images of the evaluation dataset, which is 142x298 pixels, are used because the identification system targets at low resolution leg images (Figure 4.1(a)). Next, the evidence image I_x and suspect image I_y are selected from the testing database and the edge samples (the red circles in Figure 4.1(b)) are extracted from them.

These two images are aligned by an affine transformation whose parameters are estimated from the correspondence of their edge samples. Then, the feature matching method proposed in [4], which is described in Section 2.5.1, is applied to the green channel of the image pair (Figure 4.1(c)). The matching features are computed by the Gabor orientation field and divided into small grids (Figure 4.1(d)). Orientation histograms are then formed from each grid for calculating the matching distance (Figure 4.1(e)). Finally, the suspects are ranked by the matching distance in the low resolution evaluation dataset.

4.1.2 Proposed Alignment Algorithm

The proposed alignment algorithm contains edge sampling and registration. For the edge sampling, angular sampling is applied to extract edge samples. For the registration, Coherent Point Drift (CPD) is employed for image alignment. Each of them is explained as follows.

4.1.2.1 Edge Sampling

Edge is a useful feature for image alignment. Since there can be many edge points and their close neighbors may contain similar information, it is not computationally efficient to consider all edge points on leg boundaries. Instead, edge points can be sampled by some sampling methods. A simple approach is the regular sampling method, which selects every k^{th} point from a list of points sorted by increasing y-coordinates. Even though the leg shapes have different translations, rotation, and scales, the corresponding points can be found by

using this method. Regardless of its simplicity and easy implementation, it is not rotational invariant. If the leg is rotated, different sample points can be obtained. This problem is illustrated in Figure 4.2.

Figures 4.2(a) and (b) show the leg boundaries from an original image and its rotated version, which is 30° different from the original image. The sample points in Figures 4.2(a) and (b) are obtained by the regular sampling method. In Figure 4.2(c), the leg boundary in Figure 4.2(a) is overlaid with the rotated version of Figure 4.2(b), which is turned back by -30° . It is clearly shown that the sample points from Figure 4.2(a) (red crosses) and Figure 4.2(b) (blue circles) cannot match well. Since the regular sampling method is not rotational invariant, misalignment caused by this method can bring to wrong matching. To resolve this misalignment problem, another sampling scheme is proposed.

Figure 4.2 Illustration of the weakness of the regular sampling method. (a) and (b) are the boundaries of the same leg image where (b) is a rotated version of (a) by 30° . The red crosses and blue circles are the sample points given by the regular sampling method. (c) The image in (b) is turned back by -30° and overlaid with the image in (a).

The proposed scheme can effectively choose the sample points with the help of angular sampling. Thus, this scheme is necessary to find out a center point and a centerline. The center point can be estimated by the distance transform [89], which is defined as

$$D_i(p) = \min_{b \in \mathbf{B}_i} \|p - b\|_2, \quad (4.1)$$

where $i \in \{L, R\}$, $\mathbf{B}_L$ and $\mathbf{B}_R$ are the left and right boundaries of a leg, $\mathbf{D}_L$ and $\mathbf{D}_R$ are the results from the left and right boundaries, and $p = (x, y)$ is a pixel location of the leg in an input image. To make use of the information on both boundaries,

$$D_o(p) = \max(D_L(p), D_R(p)) \cdot e^{-m}, \quad (4.2)$$

where $m = \lfloor |D_L(p) - D_R(p)| \rfloor$ and $\lfloor \cdot \rfloor$ denotes a floor operator, is used. The center point is defined as $p_c = \operatorname{argmax}_p D_o(p)$. When there exist other candidate points p_k such that $D_o(p_k) = D_o(p_c)$, all candidate points (including p_c) are then sorted by their y-coordinates in ascending order and the median point is selected as the center point. If the number of candidates is even, the center point is defined as a point whose y-coordinate is the (the number of $p_k/2$)th point. Afterwards, the centerline, which passes through the center point, is determined by the points in the set $\{p_1, \dots, p_i, \dots, p_H\}$ where $p_i = \operatorname{argmax}_p D_o(p)$ and p_i is a pixel location in the i^{th} row of $\mathbf{D}_o$. Figure 4.3 illustrates the centerline and the center point by $\mathbf{D}_o$.

After the center point and centerline are found, the uniform sampling is applied to the points in terms of angular distance. Figure 4.4 illustrates it. Figure 4.4(a) is an original image, which is the same as Figure 4.2(a) and Figure 4.4(b) is its rotated version. Their sample points are taken by angular sampling. In Figure 4.4(c), the leg boundary in Figure 4.4(a) is overlaid with the rotated Figure 4.4(b), which is rotated by -30° . Clearly, the

points (cross and circle) in Figure 4.4(c) are matched better than that in Figure 4.2(c). It indicates that the proposed scheme can overcome the weaknesses of the regular sampling method. In the experiments, each image was represented by 360 sample points, which were stored as a matrix,

$$\begin{bmatrix} x_0 & x_1 & \dots & x_{358} & x_{359} \\ y_0 & y_1 & \dots & y_{358} & y_{359} \end{bmatrix}. \quad (4.3)$$

Since the leg boundaries are open, some sample points cannot be obtained in some particular angles. Then, their corresponding values in the matrix are set to zero.

Figure 4.3 Illustration of the centerline and the center point by $\mathbf{D}_o$. (a) The magnitude map of $\mathbf{D}_o$. The color is brighter if $\mathbf{D}_o$ is higher. (b) The centerline (the blue line) and the center point (the red cross) estimated using $\mathbf{D}_o$.

Figure 4.4 The proposed sampling scheme applied to the leg boundaries in Figure 4.2; (a) and (b) are the same legs from Figure 4.2(a) and Figure 4.2(b); (c) the image in (b) is turned back by -30° and overlaid with (a).

4.1.2.2 Image Registration

The coordinates of the sample points stored in the matrices, which are given in Equation (4.3), are used to align the images for matching. As explained in the previous section, some values in the matrixes are zero because the edge points cannot be found in some angles. To ensure that every sample point can have a correspondence for registration, a pruning step is applied. If any corresponding sample points in the matrixes are zero, they both are removed from the matrices. This step can still keep the property of rotational invariant because the sampling angles are based on a centerline and a center point.

For enhancing robustness to translation and scale variations, the coordinates of the sample points are normalized after pruning. Then, two sets of edge samples are aligned by an affine CPD [66], which is explained in Section 2.3.1. And then, one of the images is

mapped to another image by the transformation matrix from affine CPD and bicubic interpolation. Some examples of the aligned images are shown in Figure 4.5.

Figure 4.5 Three examples of aligned images from different pairs of legs. Images in the same row are from the same leg. The first and the second columns are preprocessed input images for alignment. Images in the third column are the alignment results.

4.1.3 Feature Matching

For the sake of the identification and evaluation, the feature matching method reported in [4] is used by the proposed alignment algorithm. The general idea of such method is mentioned in Section 2.2.2 and Section 2.5.1. Thus, this section emphasizes how such method can be applied with the alignment scheme. After two images are aligned, the green channel of each image is filtered by Gabor filters. Then, the Gabor orientation field can be calculated by

$$O(x, y) = \operatorname{argmax}_{\theta_k} |G_{\theta_k} * I|, \quad (4.4)$$

where G_{θ_k} is a Gabor filter with the orientation θ_k and I is the green channel of the input image. To divide the orientation field into blocks, a dynamic grid system which determines the block sizes based on the row width of the leg (see Figure 4.6) is employed.

Figure 4.6 Illustration of the dynamic grid system in [4]. The block size is determined by the width of each row and each block is represented by a black rectangle.

In the dynamic grid system, the number of blocks in the i^{th} row is calculated by

$$N_i = \left\lfloor \frac{E}{2} * \frac{\mu_{is}}{\mu_{ia}} + 0.5 \right\rfloor, \quad (4.5)$$

where $i = 1, \dots, E$, μ_{is} is the average number of skin pixels in the i^{th} row calculated from leg images in [4], μ_{ia} is the total number of pixels in the i^{th} row, including the background pixels, E is the number of rows to be divided and $\lfloor \cdot \rfloor$ denotes a floor operator. When the number of blocks is found, the block size can be calculated for each row easily.

In each block, a Gabor orientation histogram is created by binning the Gabor orientation field. The dissimilarity between the orientation histograms from the two inputted images is defined by the blockwise Chi-square distance, which is

$$X(\mathbf{H}_{O_i}, \mathbf{H}_{O_d}) = \sum_{t=1}^N \sum_{k=1}^K \frac{(h_{O_i B_t}(k) - h_{O_d B_t}(k))^2}{h_{O_i B_t}(k) + h_{O_d B_t}(k)}, \quad (4.6)$$

where $\mathbf{H}_{O_i} = \{h_{O_i B_1}, \dots, h_{O_i B_N}\}$ and $\mathbf{H}_{O_d} = \{h_{O_d B_1}, \dots, h_{O_d B_N}\}$ are Gabor orientation histograms from the image I_x and the aligned image I_y , respectively, K is the number of orientations and N is the total number of blocks. If the image is aligned by tilting certain degree, the first row of the aligned image may not completely appear. Thus, its blocks are not used for matching. For the sake of a fair comparison, the first row of the blocks in Su et al.'s method is discarded in the experiment.

4.1.4 Experimental Results

The low resolution evaluation dataset mentioned in Section 3.2, which contains 1,138 low resolution images from 283 different legs, was used for evaluation. To evaluate the performance of the proposed alignment algorithm for low resolution leg identification,

Cumulative Match Characteristic (CMC) curves were applied to objectively measure the effectiveness of the proposed algorithm. The CMC curves of the original Su et al.'s method [4], the feature matching method with alignment based on regular sampling, and the feature matching method with the proposed alignment algorithm are shown in Figure 4.7.

Figure 4.7 The CMC curves obtained from different alignment methods. The same matching method was used in all the methods.

Figure 4.7 shows that the proposed alignment algorithm gives a significant improvement despite the same feature matching method used by the original Su et al.'s method and the proposed algorithm. The rank-1 identification accuracy of the proposed alignment algorithm is 75.3%, in contrast to 69.6% obtained by the original Su et al.'s method in the rank-1 identification accuracy. Furthermore, the proposed alignment algorithm and the alignment method based on regular sampling are compared to demonstrate the effectiveness of the proposed sampling scheme. The rank-1 identification accuracy of the proposed alignment algorithm is 1.56% higher than that of the alignment

method based on regular sampling. These experimental results demonstrate that the proposed alignment algorithm can effectively alleviate the alignment problem, which was ignored in the paper [4].

To further illustrate the effectiveness of the proposed alignment algorithm, the alignment results of four legs given by the alignment method based on regular sampling and the proposed alignment algorithm are shown in Figure 4.8. In Figure 4.8, each row is a set of images from the same leg. The first and second columns show the probe images and gallery images. Then, the third column shows the results given by the alignment method based on regular sampling, and the last column shows the results given by the proposed alignment algorithm. To obtain the results in the third column and the last column, the images in the second column are aligned with those in the first column. Since leg images are matched without alignment by Su et al.'s method, the images in the second column are directly compared with those in the first column. Obviously, Su et al.'s method cannot match them correctly because of their pose difference. However, the results of third column also show that the alignment method based on regular sampling cannot align images in the first two columns very well. The possible reason of this inaccurate alignment is that regular sampling, which takes sample points from every k points on the boundaries, is not rotational invariant and may not find the correct correspondences as illustrated in Figure 4.2(c). Although the alignment method based on regular sampling uses the affine CPD method, which can handle rotation and variations, this error cannot be recovered and finally an inaccurate alignment is produced.

However, when the images in the first and the second columns are processed by the proposed alignment algorithm, the original images in the first column are more similar to

the shapes of the aligned images and they can be matched by the feature matching method. Since the sample points are chosen by the proposed alignment algorithm with the consideration of the rotation variations, image pairs can have the corresponding points for alignment. Therefore, the proposed alignment algorithm can perform better than the original Su et al.'s method and the alignment method based on regular sampling.

In Appendix D, Figure D.2 shows 10 cases which have wrong matches in the proposed algorithm. In these cases, there are few problems observed. Firstly, the occlusion problem exists in these wrong matches. When the upper part or lower part of gallery images are partially occluded by socks or shorts, the difference of leg shapes between probe and gallery images increases and the alignment scheme is not capable of handling this problem. Secondly, the viewpoint problem exists in the wrong matches. When the viewpoints of the probe and gallery images are different, their leg geometries become more different and their correspondences cannot be found correctly. Since the viewpoint problem is related to 3D information, the proposed alignment scheme is based on 2D affine transformation and is not capable of tackling this problem. Thirdly, feature distinctiveness is a problem in the wrong matches because wrong matches still exist in the well-aligned image pairs. After the images are aligned, Su et al.'s method, which makes use of Chi-square distance for matching and gives equal weighting for all features, are applied to obtain matching results. However, different legs have different distinctive features, which should be weighted differently. Thus, a learning approach on leg features is more appropriate to handle this issue and it is explained in Chapter 5.

Figure 4.8 The alignment results given by different alignment methods. Each row shows a set of images from the same leg. The first column shows images from the probe set. The second column shows images from the gallery set. The third column shows the results given by the alignment method based on regular sampling, and the last column shows the results given by the proposed alignment algorithm. The results in the last column are obtained by projecting the images in the second column to those in the first column. Note that the alignment results in the last column are more similar to the images in the first column than the alignment results in the third column. The red ellipses show the alignment performance by highlighting some unique features on the legs.

Although the skin color has a greater red component compared to green and blue components, the proposed algorithm makes use of the green component to extract hair patterns. To evaluate the effect of different channels on the proposed algorithm, an experiment was performed with different channels of images. The CMC curves are shown in Figure 4.9. In Figure 4.9, the algorithms with green channel and grayscale of images outperform that of red and blue channels of images.

Figure 4.9 The CMC curves obtained from the proposed alignment algorithm with Su et al.'s method [4] on different channels and grayscale.

On the other hand, it is difficult to extract the thin hair from skin in the red channel of images based on the observation in Figure 4.10. For each row, it contains an original image, its red channel, its green channel, and its blue channel. For the thin hair in the first and second rows, it is not easily observed in the red channel of images. But red and green

channels of images show clearly the thick hair in the third row. Thus, the green channel of leg images is more robust to reveal different kinds of hair.

Figure 4.10 Different channels of images. (a) Original images, (b) red channels of images, (c) green channels of images, and (d) blue channels of images.

Some may ask about the similar performance of the proposed alignment algorithm with the green channel and the grayscale of images. Since the grayscale value, which is derived from ITU-R BT.601, is calculated as $0.2989 \times \mathbf{R} + 0.5870 \times \mathbf{G} + 0.1140 \times \mathbf{B}$, where $\mathbf{R}, \mathbf{G}, \mathbf{B}$ are the RGB value of an image, the weighting of the green channel is higher than that of the red and blue channels for the calculation of the grayscale value. Therefore, they have similar performance due to the high weight of the green channel on the grayscale value.

4.1.5 Summary

Since the proposed algorithm is based on leg geometry, angular sampling, and affine transformation, it can accurately align androgenic hair patterns in low resolution leg images for criminal and victim identification. Furthermore, the experimental results on a database containing 1,138 leg images from 283 subjects demonstrate that the proposed algorithm can outperform the original Su et al.'s low resolution leg identification method. More than 5% improvement in rank-1 identification accuracy is obtained. In addition, the proposed angular sampling scheme for leg registration also outperforms the regular sampling scheme.

4.2 Using Hair Follicles with Leg Geometry to Align Androgenic Hair Patterns

4.2.1 Overview

Identifying criminals and victims, including but not limited to masked gunmen, rioters, paedophiles, and victims of sexual offenses in digital images can be a challenging forensic task because often, neither their faces nor tattoos are observable. Though skin marks and vein patterns provide new ways to identify them, new biometric traits are still demanded because skin marks on some subjects are limited and clear vein patterns are difficult to be obtained from people with high concentration of body fat or melanin. They can be covered by androgenic hair. A recent paper indicated that androgenic hair patterns extracted from lower legs can be used for personal identification and demonstrated their effectiveness. However, it neglected the importance of alignment before matching two patterns. To alleviate this alignment problem, leg geometry was proposed for low resolution androgenic hair patterns. For aligning androgenic hair patterns in high resolution images which may be obtained from child sexual abuse images and rioter images taken by reporters, in this chapter, a new feature, androgenic hair follicles, is proposed to combine with leg geometry to achieve high alignment accuracy. With this new feature, a new alignment algorithm is developed. Experiments on 1,138 high resolution images from 283 different legs show that the proposed alignment algorithm provides improvements of 5%-10% on different experimental settings.

As mentioned in Section 2.2.2, medical studies suggest that androgenic hairs and their follicles can be useful biometric traits. Using androgenic hair patterns as a biometric trait to address the identification problem mentioned previously was proposed by Su et al. [4], which is described in Section 2.2.2. However, the alignment issue was ignored. Section 4.1 describes an alignment method which makes use of leg geometry, but leg geometry is not informative enough for viewpoint change. Four examples are given in Figure 4.11. The four leg geometries are similar even though they were taken from different viewing angles. Thus, androgenic hair follicles are proposed as an additional feature with leg geometry for high resolution androgenic hair pattern alignment.

Figure 4.11 A leg is captured at different viewing angles: (a) -45° , (b) -22.5° , (c) 0° , and (d) 22.5° with respect to the subject.

4.2.2 Proposed Alignment Algorithm

Figure 4.12 shows the schematic diagram of the leg identification process using the proposed alignment algorithm. Firstly, two input color images I_x and I_y are normalized.

Then, the images are inputted to two different stages. A two-stage design is used because some legs do not have androgenic hairs, which are essential for follicle detection.

Figure 4.12 A schematic diagram of the identification process using the high resolution alignment algorithm.

Stage A extracts androgenic hair follicles and edge samples of the leg shapes denoted as blue crosses and red dots in Figure 4.12(b), while Stage B extracts edge samples of the leg shapes denoted as red dots in Figure 4.12(b). Next, the two images I_x and I_y are aligned in these two stages separately through the extracted features illustrated by Figure 4.12(c). To evaluate the effectiveness of the alignment algorithm in these stages, Su et al.'s feature matching algorithm [4], which is mentioned in Section 2.5.1, is used. Their algorithm takes two aligned images with the size of 142×298 pixels as inputs and extracts the histograms of the orientation fields as features (Figures 4.12(d) and (e)). By comparing the orientation histograms, distance describing the dissimilarity between the two input images is calculated. When two output distances $dist_1$ and $dist_2$ are compared from

Stages A and B, the minimum one is regarded as the final distance. The edge samples of the leg shapes used in the two stages are extracted by the angular sampling method presented in Section 4.1.2.1. The following sections focus on follicle extraction and registration.

4.2.2.1 Androgenic Hair Follicle Extraction

Androgenic hair follicles are small and hard to be extracted directly, but their hairs are more observable. The proposed scheme extracts androgenic hairs first and then determines their follicles based on the end points of the hairs. To extract androgenic hairs, Gabor filters [90] are used to obtain their orientation information. Since androgenic hair segments can be treated as lines, the real parts of Gabor filters [91] are applied and denoted as

$$G(x, y; \theta, \sigma, \lambda, \omega) = \frac{1}{2\pi\sigma^2\lambda} \exp\left(-\frac{\lambda^2 x'^2 + y'^2}{2\lambda^2\sigma^2}\right) \cos(2\pi\omega x'), \quad (4.7)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, θ is the orientation of the filter, σ is the standard deviation of the elliptical Gaussian along x direction, λ is the aspect ratio of the standard deviation of the elliptical Gaussian along y direction, and ω is the radial center frequency. The DC component of the filter is removed for increasing its robustness against brightness variations. In the experiments, the parameters, σ , λ , and ω were respectively set to 3, 2.5, and 0.36 for extracting hairs with a width of roughly 4 pixels and eight filters denoted as $G_{\theta_1} \dots G_{\theta_8}$ with different directions and sizes of 33×33 pixels were employed.

By convolving the image with the filters, the magnitude $M(x, y) = \min_{\theta_i} (G_{\theta_i}(x, y) * I(x, y))$ and the orientation $O(x, y) = \operatorname{argmin}_{\theta_i} (G_{\theta_i}(x, y) * I(x, y))$, where $*$ denotes as an operator of convolution, I is the green channel of the image, and θ_i is the orientation of the i^{th} filter, are obtained. Then, the pixels in the image are assigned into eight groups defined as

$$\Lambda_i = \{p_m | O(p_m) = \theta_i \text{ or } \theta_k \text{ and } M(p_m) < \bar{M}\}, \quad (4.8)$$

where $i = 1, 2, \dots, 8$, $k = i \bmod 8 + 1$, $p_m = (x_m, y_m)$ is a pixel of the i^{th} group, $m = 1, 2, \dots, |\Lambda_i|$, and $\bar{M}$ is the mean of $M(x, y)$. Note that some pixels do not belong to any group, but some pixels can appear in more than one group. Then, line segments ℓ_l formed by 4-connected pixels in each group Λ_i are extracted and their scores defined as

$$\text{score}_{\ell_l} = |\ell_l| \times \sum_{p \in \ell_l} M(p), \quad (4.9)$$

where $|\ell_l|$, the length of the line segment, is defined as the number of pixels in the skeletonized line segment, are calculated. Finally, the line segments are sorted by score_{ℓ_l} in an ascending order. The top 100 line segments are regarded as reliable androgenic hairs and their follicles are located by tracing the skeletonized hairs. Figure 4.13 illustrates the process of androgenic hair follicle extraction.

Figure 4.13 The process of androgenic hair follicle extraction: (a) magnitude $M(x, y)$ of the image, (b) orientation $O(x, y)$ of the image, (c) an ungrouped image with different orientations (represented by different colors), (d) the eight groups, (e) the selected segments, and (f) the detected androgenic hair follicles (x).

4.2.2.2 Image Registration

Since there are two stages in the identification framework, the schemes for the image registration of Stage A and Stage B are summarized below.

Stage A:

Step 1. After the androgenic hair follicles are extracted, the edge samples of the leg shapes are extracted by the angular sampling method mentioned in Section 4.1.2.1.

Step 2. Then, the androgenic hair follicles and edge samples are normalized to enhance robustness to translation and scale variations.

Step 3. And then, affine coherent point drift (CPD) [66], which builds data models by Gaussian mixture models and optimizes the result by an expectation-maximization method, is used to align two sets of androgenic hair follicles and edge samples.

Step 4. Based on the transformation matrix given by the affine CPD, one of the images is projected to the other by using bicubic interpolation. Thus, a pair of images is aligned for feature matching.

In Stage B, the above steps are repeated without using androgenic hair follicles. After the images are aligned in Stage A and Stage B, Su et al.'s feature matching method [4], which takes images with size of 142×298 pixels as inputs, is used to evaluate the alignment results. Two distances, $dist_1$ and $dist_2$ are calculated from Stage A and Stage B. The minimum of them is regarded as the final distance which describes the dissimilarity between two images. This score level fusion scheme takes advantages of the alignments from androgenic hair follicles and leg geometry to handle legs with and without androgenic hair.

4.2.3 Experimental Results

The high resolution evaluation dataset mentioned in Section 3.2, which contains 1,138 high resolution images from 283 different legs, was used for evaluation. In the experiments, the regions of the interest were extracted by the preprocessing method mentioned in Section 3.2.2. The preprocessed images were aligned and matched by the algorithms. Su et al.'s method [4] was implemented for comparison and Cumulative Match Characteristic (CMC) curve, which is used as a performance index, is shown in Figure 4.14.

Figure 4.14 shows four CMC curves: the proposed alignment algorithm, the alignment method based on edge samples only, the alignment method based on edge samples and androgenic hair follicles, and Su et al.'s method [4]. The alignment method based on edge samples and androgenic hair follicles refers to alignment scheme of Stage

A and the alignment method based on edge samples refers to alignment scheme of Stage B. The proposed alignment algorithm is the combination of the results from the alignment schemes of Stage A and B. In Figure 4.14, the matching result from the proposed alignment algorithm is the best. Although the matching result from the alignment method based on edge samples and androgenic hair follicles is worse than that from that of the alignment method based on edge samples, the proposed alignment algorithm takes the advantages of both alignment methods and achieves the best result.

Figure 4.14 CMC curves obtained from the high resolution images.

In Appendix D, Figure D.3 shows 10 failure cases of probe images. For each case, the corresponding gallery image, the aligned image based on edge samples and androgenic hair follicles, and the aligned image based on edge samples are shown accordingly. From these cases, many of them have the problem of partial occlusion at the upper part or lower part of gallery leg images. Since the alignment scheme is mainly based on the CPD method, which optimizes the solution by using the EM method, it obtains the local minimum when two leg geometries become more different in the occlusion cases. Thus, these cases are not

aligned well by using two alignment schemes. In addition, the fifth case at the right in Figure D.3 shows that the probe image doesn't have the correct match even though the gallery image is aligned well with the probe image. A possible reason is that the probe image is blurred. The blurry image makes the leg features unclear and Su et al.'s method, which extracts orientation information by using Gabor filter, is not designed for handling it. Therefore, the correct result is not obtained, although correct alignment is obtained. In order to enhance the distinctiveness of the features, the proposed algorithm in Chapter 5 is designed for learning the distinctive features from each leg image.

4.2.4 Summary

Androgenic hair patterns provide additional information for criminal and victim identification. Though Su et al. [4] showed their effectiveness for personal identification, the alignment problem was ignored. This chapter studies the feasibility of making use of androgenic hair follicles with leg geometry as features for aligning androgenic hair patterns in high resolution images. Experiments on 1,138 images from 283 different legs show encouraging results.

Chapter 5

Proposed Identification Algorithm for Enhancing the Robustness and Accuracy of Soft Biometric Traits

5.1 Overview

Soft biometric traits, which are not unique, universal, or permanent, are regularly used by law enforcement agencies all around the world for narrowing down suspect lists and combining with hard biometrics, e.g. faces and fingerprints, to achieve higher matching accuracy and speed [92]. Recently, androgenic hair also called body hair was evaluated as soft biometric traits for identifying criminals and victims in digital images [4]. These images can be collected from riots, terrorist activities, or sexual offenses. In this chapter, this method is called dynamic Gabor orientation histogram (DGOH) method, which is the same as the method mentioned in Section 2.2.2 and Section 2.5.1. Their experimental results were encouraging, but they gave an impression of low discriminative power on androgenic hair. It is suspected that this impression is due to DGOH, which is neither robust to viewpoint variation nor effective to extract discriminative features. In this chapter, an algorithm, which uses leg geometry for alignment, local orientation, color, and spatial neighborhood information from multi-directional grid systems to form a large discriminative feature vector, a class-specific model trained for each leg image to

effectively exploit the information, partial least squares regression to handle unbalanced training data, and a positive sample generation scheme to increase robustness, is proposed to examine low resolution androgenic hair patterns on lower legs. Lower legs are selected for this study because they most likely have androgenic hair, comparing with other body sites.

The preliminary work [102], which employed leg geometry as an alignment feature and Coherent Point Drift (CPD) [66] as a registration method, focused only on alignment. Its details is mentioned in Section 4.1. However, it still used the DGOH method for matching. The techniques, including multi-directional grid systems, class-specific models, PLS, and the positive sample generation scheme proposed in this chapter, were neither utilized by the preliminary work nor the DGOH method. Furthermore, the proposed algorithm is evaluated on 1,493 images, including low quality images, from 412 legs, while the preliminary work was evaluated on front pose images from 283 legs. The experimental settings in this chapter are closer to real scenarios because low quality images collected from diverse viewpoints and poses are included.

5.2 Datasets and Preprocessing

Three datasets, two front pose datasets and one low quality dataset, were collected to evaluate the proposed algorithm. They contained lower leg images from different subjects (mainly Chinese, Indian, and Vietnamese and some other Asians, Iranians, Europeans, Arabs, and South Americans) and different genders (male: 71% and female: 29%). These images were taken by a Canon EOS 500D camera (the maximum resolution of 4,752×3,168 pixels) and a Nikon D70s camera (the maximum resolution of 3,008×2,000 pixels) in two

occasions with an interval of at least one week. On each occasion, one to three front pose images were taken from each leg. The original images were collected from the left and right legs. The left leg images were flipped and considered as right leg images from new subjects to enlarge the gallery sets in the experiments.

The front pose dataset collected on the first occasion is called Image Set 1, containing 534 images from 412 legs and the front pose dataset collected on the second occasion is called Image Set 2, containing 536 images from the same 412 legs. The front pose images are to simulate images collected from ex-prisoners and other suspects and good quality images collected at crime scenes. In the image collection, the poses of the subjects and the viewpoints of the cameras were not strictly controlled. Some samples of the front pose datasets are provided in Figure 5.1. The relative scales of the images are retained to show the diversity of the images. Table 5.1 gives the distribution of the number of leg images on each occasion.

Figure 5.1 Some samples of left and right legs in the front pose datasets. The relative scale is kept for showing the sample variations.

Table 5.1 The details of the datasets.

Number of images	Number of legs (Image Set 1)	Number of legs (Image Set 2)	Number of legs (Low Quality Set)
1	290	290	139
2	122	120	142
3	0	2	0

In addition, 423 low quality images were collected from 281 legs for evaluation. This dataset is called Low Quality Set. In this dataset, the subjects posed freely when the images were taken. Thus, their images have problems of blur, out of focus, viewpoints, shadows, scaling, and illumination. These images are to simulate low quality evidence images collected at crime scenes. Also, the left and right legs of the same subject are considered as the legs of two independent subjects. Some samples in this dataset are shown in Figures 5.2(b)-(e).

Four different resolutions (25 dpi, 18.75 dpi, 12.5 dpi, and 6.25 dpi) are studied in this chapter. Their image sizes are respectively about 142×298 pixels, 106×233 pixels, 71×149 pixels, and 35×74 pixels. They simulate different distances between a subject (criminal or victim) and a camera. Their corresponding resolutions in terms of pixels per square centimeter are 97, 54, 24, and 6. Utilizing a Canon EOS 500D camera with a sensor pixel size of $2.2107 \times 10^{-5} \text{ mm}^2$ and a focal length of 55mm to capture images with these resolutions, the distances between subjects and the camera are respectively 11.9m, 15.8m, 23.8m, and 47.5m. These distances are calculated by the equation, $d = f \times (25.4 \times h_o)/h_{os}$, where f is the focal length, h_o is the object height in inch, $h_{os} = p_o \times \sqrt{\rho}$ is the height of an object on sensor in mm, p_o is the image height of the object in pixels, p_o/h_o is the resolution in dpi (dot per inch) and ρ is the sensor pixel size in mm^2 . Note that Canon EOS 500D is a six-year-old model. To demonstrate that the resolutions of the testing images can be found in real cases, Figure 5.2 lists legs collected from rioters and terrorist images on the Internet and our testing images. Their relative sizes are retained for comparison.

Figure 5.2 A comparison between the testing images and criminal images in real cases. (a) Rioter and terrorist images from the Internet and testing images with sizes of (b) 142×298 pixels, (c) 106×233 pixels, (d) 71×149 pixels, and (e) 35×74 pixels. The relative scales are retained for comparisons. The references of the Internet Images are listed in Appendix A.

Preprocessing is applied to the front pose datasets for segmenting the legs from the background. The images in the Low Quality Set collected from diverse poses and viewpoints are to simulate low quality evidence images collected at crime scenes. They are segmented manually. Manual operations are not uncommon in forensic investigations. For example, before inputting a latent print to an automatic fingerprint identification system, fingerprint experts may manually detect minutiae. The preprocessing is similar as that in Section 3.2.2, but two legs are segmented to increase the number of subjects for evaluation. After the four steps in Section 3.2.2 are finished, two legs are extracted. The segmented right legs are considered as the legs of original subjects. The segmented left legs are flipped to the right and regarded as the right legs of new subjects. Note that segmented right legs are not flipped. The region of interest is illustrated in Figure 5.3 and an illustration of these five steps is given in Figure 5.4.

Figure 5.3 The region of interest is defined by six points A, A', M, M', B, and B' (bounded by the red dotted lines).

Figure 5.4 The proposed preprocessing scheme. (a) Input images, (b) filtered images with regions of interest, (c) segmented images, and (d) the images of the right legs and the flipped left legs. The images in the first and second rows are from the same person, but different sessions.

5.3 Proposed Identification Algorithm

Figure 5.5 shows the schematic diagram of the proposed algorithm, which contains training and testing phases. The aim of the training phase is to generate a class-specific Partial Least Squares (PLS) model for each image in the training dataset. In the testing phase, input images are evaluated by the trained PLS models.

In the training phase, an image in the training dataset is first chosen to be a model image. This model image and other images from the same leg are considered as positive samples and the rest of the images in the training dataset are considered as negative samples. To increase robustness against pose variation and enhance the performance of the PLS model, a positive sample generation scheme which produces more positive samples by applying an affine transformation to the model image is proposed (Section 5.3.1). Before extracting features, all other training images, except for the images produced by the positive sample generation scheme, are aligned with the model image. In the testing phase, the input images are also aligned with the model image before matching (Section 5.3.2).

Then, local binary patterns (LBP) and Gabor features are extracted (Section 5.3.3). To enhance robustness against pose and viewpoint variations, grid systems, including the dynamic grid system reported in [4] and the proposed multi-directional grid systems are used to construct local LBP and Gabor orientation histograms (Section 5.3.4). These local histograms form large feature sets for extracting more useful information from androgenic hair. The training features are used to train a PLS model. PLS has been successfully applied to other image classification problems [71-73]. The total number of trained PLS models is equal to the number of training images because each of them is selected as a model image. The testing features are multiplied by the PLS regression coefficients β to measure their similarities to the legs of the model images (Section 5.3.5).

The following subsections explain the details of the proposed algorithm. Section 5.3.1 mentions the positive sample generation scheme. Section 5.3.2 offers the alignment method. Section 5.3.3 describes the feature extraction process. Section 5.3.4 gives the details of the grid systems and Section 5.3.5 briefly discusses the PLS regression.

Figure 5.5 A schematic diagram of the leg identification algorithm. Blue cross (+) represents the positive training samples and red minus (−) represents the negative training samples.

5.3.1 A Positive Sample Generation Scheme

The PLS model, like other classification methods, is trained by feature vectors obtained from positive and negative samples. However, the raw training set is imbalanced because the negative samples obviously outnumber the positive samples. Though PLS has been shown to be robust against imbalanced training data [78], to balance them and increase robustness against inaccurate alignment due to pose and viewpoint variations, a scheme which makes use of affine transformation to generate more positive samples is proposed. First, parameters are uniformly sampled from scale $s = \{s_{\min}, \dots, s_{\max}\}$, rotation $\alpha = \{\alpha_{\min}, \dots, \alpha_{\max}\}$, and translation $(x, y) = \{(x, y)_{\min}, \dots, (x, y)_{\max}\}$ to construct a set of affine transformation matrices, i.e.

$$A_i = \begin{bmatrix} s_i \cos \alpha_i & s_i \sin \alpha_i & x_i \\ -s_i \sin \alpha_i & s_i \cos \alpha_i & y_i \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.1)$$

Then, the affine transformation matrices and the bicubic interpolation are applied on the model image to create more images as new positive samples. And then, the LBP and Gabor features are extracted from these positive samples. Afterwards, for each positive sample, its feature vector is obtained by combining local LBP and Gabor orientation histograms outputted from grid systems. Let the number of negative samples be N_n and the number of positive samples be N_p . The scheme constructs $N_n - N_p$ affine transformation matrixes and generates $N_n - N_p$ new positive samples for each model image. It should be emphasized that the proposed algorithm aligns the negative samples with the model image, but no alignment is performed on the new positive samples. If the new positive samples are aligned with the model image, the effect of the affine transformation will be eliminated, i.e. the new positive samples and the model image will become the same. By using this

scheme, more positive samples are generated and PLS training becomes balanced and more robust against inaccurate alignment. In addition, the positive sample generation scheme is only used in training and therefore it does not increase the computational cost for matching.

5.3.2 A Low Resolution Alignment Method

To align two lower leg images, e.g. a model image with a negative sample and a model image with a testing image, a low resolution alignment method consisting of two parts, edge sampling and image registration, is proposed. This method has been described in Section 4.1 and it is briefly explained in the following paragraphs.

Edge samples of two leg boundaries are first extracted. Then, angular samples of the leg boundaries defined through the centerline and the center point illustrated in Figure 5.6 are used as registration points. In order to obtain the centerline and the center point, the distance transform [89] is defined as

$$D_i(p) = \min_{b \in B_i} \|p - b\|_2, \quad (5.2)$$

where $i \in \{L, R\}$, B_L and B_R are the left and right boundaries of a leg, D_L and D_R are the results from the left and right boundaries, and $p = (x, y)$ is a pixel location of the leg image. To combine the information from both boundaries,

$$D_o(p) = \max(D_L(p), D_R(p)) \cdot e^{-m}, \quad (5.3)$$

where $m = \lfloor |D_L(p) - D_R(p)| \rfloor$ and $\lfloor \cdot \rfloor$ denotes a floor operator, is defined. The function e^{-m} is used to ensure that the points close to the midpoints of the left and right boundaries have high responses. The center point can be found as $p_c = \operatorname{argmax} D_o(p)$. If there is more than one point p_k such that $D_o(p_k) = D_o(p_c)$ and the number of p_k (including p_c) is odd, a point whose y-coordinate is the median is chosen as the center point. If the number of p_k

is even, a point whose y-coordinate is the $(g/2)^{\text{th}}$ point is chosen as the center point, where g is the total number of p_k . The centerline, which passes through the center point, is obtained by using linear regression on the points $\{p_1, \dots, p_i, \dots, p_H\}$, where $p_i = \underset{p}{\operatorname{argmax}} D_o(p)$ and p is a pixel location in the i^{th} row of $\mathbf{D}_o$.

Figure 5.6 Illustration of the centerline and the center point defined by $\mathbf{D}_o$. (a) The magnitude map of $\mathbf{D}_o$, where brighter color represents a higher value of $\mathbf{D}_o$. (b) The centerline (the red line), the center point (the blue cross) estimated using $\mathbf{D}_o$ and the angular samples (the green dots).

After obtaining the center point and the centerline, the points on the boundaries are uniformly sampled in terms of angular distance. Afterwards, pruning and normalization processes are applied to the two point sets obtained from an image pair (e.g. a model image and a negative training sample) to ensure rotational invariant and strengthen the robustness of translation and scale variations. The affine CPD method [66] is used to align the two point sets and generates a transformation matrix. By using the transformation matrix and bicubic interpolation, one of the images is projected to the other.

For each class-specific PLS model, the positive and negative training images and the testing images are projected to the model image by using this alignment method, except for new images created by the positive sample generation scheme. The preliminary work mentioned in Section 4.1 only projects one of the color channels for feature extraction. In

this chapter, all three color channels are projected and therefore, the resultant image is still a color image.

5.3.3 Feature Extraction

The DGOH method only employs grayscale intensity for identification, but neglects information in different color channels, which is shown to be useful for other recognition problems, e.g. image segmentation and face recognition [98-100]. Thus, the Gabor and LBP features utilized in this chapter are extracted from different channels of a color space, e.g. the RGB space. First, Gabor filters are convolved with a color channel I of an image and its Gabor orientation field is calculated by

$$O(x, y) = \arg_{\theta_k} \max_{m,k} |G_{\lambda_{mk}, \theta_k, \sigma_m, \gamma} * I|, \quad (5.4)$$

where $G_{\lambda_{mk}, \theta_k, \sigma_m, \gamma}$ is the real part of a zero DC Gabor filter with the orientation $\theta_k = k\pi/8$, λ_{mk} is the wavelength of the sinusoidal component, σ_m is the standard deviation of the elliptical Gaussian window, γ is the spatial aspect ratio, $m \in \{1, \dots, M\}$ and $k \in \{1, \dots, K\}$ are the scale and orientation indices. In the implementation, 32 orientations and 4 scales are used [4].

LBP features are useful to represent image texture and invariant to monotonic gray-level changes [101]. The local binary patterns $LBP_{P,R}$ [53] are defined as

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad (5.5)$$

where P is the number of sampling points, R is the radius of the sampling circle, g_c is the center pixel value, g_p is the neighborhood pixel value and $s(x) = 1$ if $x \geq 0$, otherwise $s(x) = 0$. In the proposed algorithm, $LBP_{8,2}^{u2}$, which are the uniform LBP with at most two

bitwise transitions from 0 to 1 or vice versa when they are considered as circular patterns, are extracted from the directional grid systems and $LBP_{8,1}^{riu2}$, which are the nine rotational invariant uniform binary patterns [53], are extracted from the dynamic grid systems. Two different forms of LBP are applied because the grid size of the dynamic grid system is relatively small and the histogram of $LBP_{8,2}^{u2}$ has more bins than that of $LBP_{8,1}^{riu2}$. If $LBP_{8,2}^{u2}$ are extracted from the dynamic grid system, its grid histogram may not be stable because of the small grid size and relatively large number of bins. Thus, $LBP_{8,1}^{riu2}$ are extracted from the dynamic grid system.

5.3.4 Two Grid Systems

For effectively extracting local information, gridding the Gabor and LBP features in images is essential. The dynamic grid system and the directional grid systems are used. The dynamic grid system, which is a part of the DGOH method [4], determines the block sizes based on the row width of the leg image (see Figure 5.8(m)). The number of blocks in the i^{th} row is calculated by

$$N_i = \left\lfloor \frac{E}{2} \times \frac{\mu_{is}}{\mu_{ia}} + 0.5 \right\rfloor, \quad (5.6)$$

where $i = 1, \dots, E$, μ_{is} is the average number of skin pixels in the i^{th} row, μ_{ia} is the total number of pixels in the i^{th} row, including the background pixels, E is the number of rows to be divided and $\lfloor \cdot \rfloor$ denotes a floor operator. When the number of blocks is found, the block size can be calculated for each row easily. The corresponding histograms of Gabor feature $h_{OB_j}(k)$ and LBP feature $h_{LB_j}(k)$ respectively defined as

$$h_{OB_j}(k) = \sum_{(x,y) \in B_j} \delta(O(x,y) = k) , \quad (5.7)$$

$$h_{LB_j}(l) = \sum_{(x,y) \in B_j} \delta(LBP_{P,R}^{riu2}(x,y) = l) , \quad (5.8)$$

where B_j is a set containing coordinates of the pixels in the j^{th} block, δ is the Kronecker delta function, k is an orientation index and l is a binary pattern index, are extracted.

The dynamic grid system employed by DGOH is sensitive to viewpoint variation. When the viewpoint changes, the grids do not correspond correctly as illustrated in Figures 5.7(a) and (b). To overcome this problem, the directional grid systems are designed to increase robustness against this variation. In the directional grid systems, the blocks are rotated in different direction θ . The block height $H_{j\theta}$ and the block width $W_{j\theta}$ are respectively defined as $H_{j\theta} = \lfloor (H_\theta - 1)/N_H \rfloor$ and $W_{j\theta} = \lfloor (W_\theta - 1)/N_W \rfloor$, where H_θ is the distance between the top leg pixel and the bottom leg pixel along direction θ , W_θ is the distance between the leftmost and the rightmost leg pixels along the direction perpendicular to θ , N_W is the number of columns and N_H is the number of rows along direction θ . The corresponding histograms of Gabor feature $h_{OB_{j\theta}}(k)$ and LBP feature $h_{LB_{j\theta}}(l')$ respectively defined as

$$h_{OB_{j\theta}}(k) = \sum_{(x,y) \in B_{j\theta}} \delta(O(x,y) = k) , \quad (5.9)$$

$$h_{LB_{j\theta}}(l') = \sum_{(x,y) \in B_{j\theta}} \delta(LBP_{P,R}^{u2}(x,y) = l') , \quad (5.10)$$

where $B_{j\theta}$ is a set containing coordinates of the pixels in the j^{th} block in the direction θ , δ is the Kronecker delta function, k is an orientation index and l' is a binary pattern index, are extracted. Figures 5.8(a)-(l) illustrate the directional grid systems. In the implementation, $\theta = \{0^\circ, 30^\circ, \dots, 330^\circ\}$, $N_W = 8$, $N_H = 5$, and the first and last rows are

not used. The pairs of opposite directions (e.g. 90° and 270°) are included because the blocks are placed from one side to the other side along each direction. For example, the blocks are placed from left to right along 90° and from right to left along 270° . Due to the floor operators used in the block height and width computation, blocks placed from opposite directions are not identical. Thus, blocks in opposite directions are included to generate more features for increasing distinctiveness of the feature set.

The blocks in the directional grid systems are larger comparing with those in the dynamic grid system and therefore, they are more robust against inaccurate alignment (Figure 5.7). However, if only one grid, e.g. the one in Figure 5.8(a), is used, the feature dimension will be very low and it may not contain enough discriminative information for achieving high accuracy. With this consideration, the directional grid systems are composed of many grids in different directions. Note that the leg boundaries are also covered by the blocks, which is different from the previous work [4]. The LBP histograms and the Gabor orientation histograms in each block of the dynamic grid system and the directional grid systems are extracted and normalized. For each image m , all the corresponding histograms $h_{O_mB_j}$, $h_{L_mB_j}$, $h_{O_mB_{j\theta}}$, and $h_{L_mB_{j\theta}}$ are concatenated together to form a large feature vector with a length of 59,766, i.e.

$$X_m = \{h_{O_mB_1}, \dots, h_{O_mB_N}, h_{L_mB_1}, \dots, h_{L_mB_N}, h_{O_mB_{1,0}}, \dots, h_{O_mB_{N',330}}, h_{L_mB_{1,0}}, \dots, h_{L_mB_{N',330}}\}, \quad (5.11)$$

where N is the number of blocks from the dynamic grid system and N' is the number of blocks from the directional grid systems.

Figure 5.7 A leg with four orange dots is captured at different viewpoints. (a) is the leg images, (b) and (c) are the results given by the dynamic grid system from the DGOH method [4] and one of the directional grid systems proposed in this chapter. The first images in (a), (b), and (c) are the leg image captured from 0° with respect to the leg and the second images are the leg image captured from 15° with respect to the leg.

Figure 5.8 (a)-(l) are the directional grid systems and (m) is the dynamic grid system.

5.3.5 PLS Regression

PLS regression [74] is a model searching for the internal relationship between a feature matrix $\mathbf{X}_{(n \times p)}$ and a response variable $\mathbf{y}_{(n \times 1)}$, where $\mathbf{X}_{(n \times p)}$ and $\mathbf{y}_{(n \times 1)}$ are mean-centered and variance-scaled, n is the number of training data and p is the dimension of the feature vector. Latent variables are used to represent their internal relationship, which is defined as

$$\mathbf{X} = \mathbf{V}\mathbf{P}^T + \mathbf{E}, \quad (5.12)$$

$$\mathbf{y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \quad (5.13)$$

where $\mathbf{V}$ and $\mathbf{U}$ are the latent components, $\mathbf{P}$ and $\mathbf{Q}$ are the loading matrices, the superscript T represents a matrix transpose operator and $\mathbf{E}$ and $\mathbf{F}$ are the residue matrices. In this work, the NIPALS algorithm [74] is used and its computation details are given in Algorithm 5.1.

Algorithm 5.1. NIPALS algorithm [74]

1. for $i = 1$ to d do
2. $\mathbf{W}_i = \mathbf{X}^T \mathbf{y} / \|\mathbf{X}^T \mathbf{y}\|$
3. $\mathbf{V}_i = \mathbf{X} \mathbf{W}_i / \|\mathbf{X} \mathbf{W}_i\|$
4. $\mathbf{X} = \mathbf{X} - \mathbf{V}_i \mathbf{V}_i^T \mathbf{X}$
5. $\mathbf{y} = \mathbf{y} - \mathbf{V}_i \mathbf{V}_i^T \mathbf{y}$
6. end
7. $\boldsymbol{\beta} = \mathbf{W} \mathbf{Q}^T = \mathbf{W} (\mathbf{V}^T \mathbf{V})^{-1} \mathbf{V}^T \mathbf{y}$

where $\boldsymbol{\beta}$ is a $p \times 1$ regression coefficient vector of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}'$, $\mathbf{f}'$ is a residual vector, d is the number of latent variables, $\mathbf{W}_i$ and $\mathbf{V}_i$ are the i^{th} columns of the matrices $\mathbf{W}$ and $\mathbf{V}$ in $\mathbf{V} = \mathbf{X}\mathbf{W}$. The second and third steps are matrix projection and the fourth and fifth steps are matrix deflation. In the training phase, the target responses of all the positive samples are set to one and the target responses of all the negative samples are set to negative one. In the testing phase, an estimated regression response $\hat{y}_N$ to an input feature vector $\mathbf{x}_N$ can be calculated by $\hat{y}_N = \bar{y} + \boldsymbol{\beta}^T \mathbf{x}_N$, where $\bar{y}$ is the sample mean of the training y . The model image corresponding to the highest PLS regression response is regarded as the best match. Note that $\bar{y}$ is ignored in the response calculation because it is a constant. PLS is robust against imbalanced training data [78]. The proposed algorithm trains one class-specific PLS model for each image in gallery sets.

5.4 Experimental Results

To evaluate the performance of the proposed algorithm, experiments were conducted on the datasets, including the Image Set 1, the Image Set 2, and the Low Quality Set. Images with about 25 dpi were used as testing images for all the experiments except the last one. In the last experiment, images with about 25 dpi, 18.75 dpi, 12.5 dpi, and 6.25 dpi were used and their corresponding sizes were respectively 142×298 pixels, 106×233 pixels, 71×149 pixels, and 35×74 pixels. The Image Set 2 containing 536 images from 412 legs was regarded as a gallery set in all the experiments except the first experiment. The first experiment aimed to evaluate the performance of the proposed algorithm for the front pose leg images. The Image Set 1 and the Image Set 2 were used as a probe set and a gallery set, and vice versa. The second experiment aimed to evaluate the performance of the proposed algorithm for the low quality images. The Low Quality Set was used as a probe set and the Image Set 2 was used as a gallery set. In both of the experiments, the DGOH method [4] and the preliminary method [102], which is mentioned in Section 4.1, were used for performance comparison. The gallery sets were used for training and the probe sets were used for testing only. The Gabor and LBP features were extracted from the grayscale intensity I , the color channel C_r in the YC_bC_r color space and the color channel Q in the YIQ color space. C_r and Q were chosen based on Liu et al.'s and Choi et al.'s suggestions [98-99]. The Gabor features were obtained from the grayscale intensity I only because they are useful for locating strong edges, which are not always apparent in the C_r and Q components. In the positive sample generation scheme, the scale, rotation, and translation parameter sets were set to be $s = \{1.2, 1.1, 1.0, 0.9, 0.8\}$, $\alpha = \{-10, -6, -2, 2, 6, 10\}$, $x =$

$\{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$, and $y = \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$.

Cumulative Match Characteristic (CMC) curves are used as a performance index. CMC curves are a standard approach to evaluate biometric algorithms developed for criminal and victim identification. They have been employed by numerous researchers.

In general, probe sets and gallery sets for forensic evaluation should not be swapped, because they have different characteristics. 2-fold cross validation was adopted in the first experiment, because the Image Set 1 and the Image Set 2 have similar quality. In the first fold, the Image Set 1 was used as a probe set and the Image Set 2 was used as a gallery set. In the second fold, their roles were exchanged. Table 5.2(a) summarizes the average accuracy of these two folds. The results show clearly that the proposed algorithm performs the best and there is no significant difference between Figures 5.9(a) and (b) obtained from the two folds. Figure 5.9(c) shows the CMC curves of the three methods obtained from the Low Quality Set. In this experiment, the Image Set 2 was used as a gallery set. The performance of the DGOH method [4] is the worst in all these experiments. The preliminary work mentioned in Section 4.1 performs better than the DGOH method, but its performance is still far from the proposed algorithm. The proposed algorithm outperforms the others because it makes use of different color channels and grid systems to extract more discriminative informative. Furthermore, the large blocks relax the requirement of accurate alignment and indirectly extract leg geometry information. It also learns a class-specific PLS model for each image in the gallery sets so that the importance of different features for different legs can be taken into account. In addition, extra positive samples are generated to increase robustness against inaccurate alignment due to pose and viewpoint variations. To show clearly the difference among these three methods, Table 5.2 lists their

accuracy at different ranks. The accuracy in Table 5.2(a) is the average accuracy of the two folds. Comparing with the DGOH method, the proposed algorithm offers an improvement of over 15% in terms of rank-1 accuracy in all the cases. At the 20th rank, the proposed algorithm achieves accuracy of 100% for the Image Set 1 and the Image Set 2 and offers improvement of 24.35% for the Low Quality Set. Because the preliminary method and the proposed algorithm use the same alignment scheme, by comparing their accuracy with the accuracy of the DGOH method, improvement offered by the alignment scheme and the other computational components can be observed and they are listed in Table 5.2. Though the alignment scheme improves the performance, the other computational components play a more important role, especially for the Low Quality Set. For the rank-1 accuracy of the Low Quality Set, they offer over 14% of improvement, while the alignment scheme offers only 1.90% of improvement.

Figure 5.9 CMC curves from (a) the first fold, (b) the second fold, and (c) the probe set is the Low Quality Set and the gallery set is the Image Set 2.

Table 5.2 A summary of the experimental results. (a) The Image Set 1 with the Image Set 2 and (b) the Low Quality Set with the Image Set 2.

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	<u>94.39%</u>	<u>99.91%</u>	<u>100%</u>
The preliminary method mentioned in Section 4.1	84.49%	93.65%	96.08%
The DGOH method [4]	79.25%	91.50%	94.49%
Improvement offered by the alignment scheme	5.24%	2.15%	1.59%
Improvement offered by the other computational components	9.90%	6.26%	3.92%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.96%</u>	<u>59.10%</u>
The preliminary method mentioned in Section 4.1	10.17%	36.88%	43.97%
The DGOH method [4]	8.27%	30.26%	34.75%
Improvement offered by the alignment scheme	1.90%	6.62%	9.22%
Improvement offered by the other computational components	14.18%	16.08%	15.13%

Some may suspect that the proposed algorithm performs better than the others solely because of the large feature sets, storing more discriminative information. To study it, the proposed algorithm and the preliminary method using the large feature sets were compared on the probe sets, the Image Set 1 and the Low Quality Set. In other words, both algorithms used the same feature set and the same alignment scheme. The preliminary method still employed the Chi-square distance for matching. The corresponding CMC curves are given in Figure 5.10. Figure 5.10(a) shows that the preliminary method using the large feature sets performs even worse than the DGOH method for the Image Set 1 and Figure 5.10(b) shows that the preliminary method using the large feature sets performs slightly worse than the preliminary method for the Low Quality Set. Note that the preliminary method and the DGOH method utilize short feature vectors. These results

demonstrate clearly that without using a proper classification scheme, only large feature sets cannot increase the identification rates and even degrade the performance of the original preliminary method. The proposed algorithm can take advantage of the large feature sets because the importance of different features for different legs can be taken into account by the class-specific PLS models. The experimental results point out that without a careful design, the large feature sets, which store more information, are useless or even harmful to the performance.

This above statement can be explained by using the expected prediction error on regression [160], which can be calculated as follows, $E[(y - y_N)^2] = E(\hat{y}_N - y) + (E[\hat{y}_N^2] - E[\hat{y}_N]^2) + \tau^2$, where y represents true values, $\hat{y}_N$ represents estimated values from the learned regression function, $E(\hat{y}_N - y)$ is a bias term, $(E[\hat{y}_N^2] - E[\hat{y}_N]^2)$ is a variance term, and τ^2 is a noise term. Since the number of sample points is very few in comparison with the number of features in large feature set, a bias term can easily be minimized by a complicated function learned from a large feature set, but the complexity of this function increases the variance term in the expected prediction error, especially for learning noisy features. Thus, a careful design is necessary to minimize the expected prediction error.

Someone may suspect that PLS performs better than Chi-square distance regardless of the feature size, but it is not true for both probe sets. In Figures 5.10(a) and (b), their rank-1, rank-10 and rank-20 accuracy are 85.21%, 94.19%, 97.19% and 10.17%, 29.31%, 36.88% from the CMC curves of preliminary work using small feature sets and Chi-square distance. In Figures 5.11(a) and (b), the corresponding accuracy are 75.09%, 90.82% and 94.01% and 9.46%, 25.06%, 33.81% from the CMC curves of the Gabor features on

dynamic grid system using small feature sets and PLS. By comparing the accuracy of these CMC curves, it is shown that PLS performs worse than Chi-square distance in small feature sets. This comparison pinpoints that large feature sets are vital for the proposed algorithm.

Figure 5.10 Comparison between the proposed algorithm and the preliminary method using the large feature sets for the probe sets of (a) the Image Set 1 and (b) the Low Quality Set.

To evaluate the effectiveness of the color information, the proposed algorithms with and without color information were compared on the probe sets of the Image Set 1 and the Low Quality Set. The Image Set 2 was used as a gallery set. When color information was excluded, features were extracted from the grayscale intensity I only. When color information was included, features were extracted from the grayscale intensity I , the color channel C_r in the YC_bC_r color space and the color channel Q in the YIQ color space. Their results are summarized in Table 5.3. The proposed algorithm offers higher rank-10 and rank-20 accuracy of the Image Set 1. It also outperforms the one without color information for the Low Quality Set at the 1st, 20th, and 30th ranks. At the 30th rank, the color information provides improvement of 2.60%. These results show that color provides additional information to the proposed algorithm to achieve better performance.

Table 5.3 A summary of the experimental results of the proposed algorithm and the one without color information for the probe sets of (a) the Image Set 1 and (b) the Low Quality Set.

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	93.82%	<u>99.81%</u>	<u>100%</u>
The proposed algorithm without color information	<u>94.57%</u>	98.88%	99.25%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.25%</u>	<u>59.10%</u>
The proposed algorithm without color information	24.11%	<u>52.25%</u>	56.50%

To further understand the performance of the proposed algorithm, experiments using different algorithmic settings were conducted. Figures 5.11(a) and (b) show the CMC curves obtained from different features and different grid systems. The Image Set 1 and the Low Quality Set were respectively used as the probe sets of Figures 5.11(a) and (b). There are four different settings: the Gabor features on the dynamic grid system, the LBP features on the dynamic grid system, the Gabor features on the directional grid systems, and the LBP features on the directional grid systems. Figures 5.11(a) and (b) pinpoint clearly that the LBP features on the directional grid systems achieve the best result among the four settings and the LBP features on the dynamic grid system give the worst result. The dimension of the LBP features on the directional grid systems is 45,312. The dimension of the Gabor features on the directional grid systems is 8,192. The feature dimensions of the other two settings are only 3,232 and 3,030. The CMC curves in Figures 5.11(a) and (b) point out the importance of using large feature sets for androgenic hair pattern identification. The results in Figures 5.11(a) and (b) also agree with Su et al.'s study that the Gabor features outperform the LBP features on the dynamic grid system [4]. When they

are combined together, i.e., the proposed algorithm, it outperforms any individual setting. This result further confirms the necessity of using the features and the grid systems together in the proposed algorithm.

In order to compare the two grid systems, two experiments were performed on the Image Set 1 and the Low Quality Set. Figures 5.11(c) and (d) show their CMC curves. In these experiments, the proposed algorithm only used one type of grid systems in each of the tests, but the other computational components remained the same. Figures 5.11(c) and (d) demonstrate that the directional grid systems, which are proposed in this chapter, obviously outperform the dynamic grid system. These experimental results are expectable because the blocks in the directional grid systems are larger and more robust to inaccurate alignment and they are composed of a lot of grids to generate more informative feature vectors.

In this chapter, a positive sample generation scheme is also proposed. To evaluate its effectiveness, two experiments were performed on the Image Set 1 and the Low Quality Set. Their CMC curves are given in Figures 5.11(e) and (f). Two algorithmic settings, the proposed algorithm with and without the positive sample generation scheme, were studied in these experiments. Figure 5.11(e) shows that the performance of these two settings is close, but Figure 5.11(f) shows that the proposed algorithm with the positive sample generation scheme performs much better. The results in Figures 5.11(e) and (f) are different mainly because of alignment. Pose and viewpoint variations in the Image Set 1 (Figure 5.11(e)) are not too large and the proposed alignment scheme based on leg geometry and the directional grid systems can handle these variations well. Therefore, the positive sample generation scheme cannot offer a significant improvement. When the positive sample

generation scheme was turned off, the PLS models were trained on very imbalanced datasets, which contained only 1-3 positive samples. Figure 5.11(e) shows that the performance without the positive sample generation scheme is close to that with the positive sample generation scheme. It means that the PLS models are robust to imbalanced datasets. However, Figure 5.11(f) indicates that PLS alone is not sufficient for the Low Quality Set because pose and viewpoint variations in this dataset are very large. The PLS models trained with the samples produced by the positive sample generation scheme are more robust to inaccurate alignment because these samples are in fact distorted versions of the model images, which are generated by different scales, translation, and rotations, and are useful for training a model robust to resolution, pose, and viewpoint problems. Figure 5.11(f) demonstrates that this scheme is essential for low quality images, e.g. images collected from crime scenes.

Table 5.4 summarizes the contribution of the alignment scheme, the directional grids and the positive sample generation scheme. The directional grids make the greatest contribution for both datasets. It indicates that good feature sets are very important. The positive sample generation scheme and alignment scheme do not offer significant improvement for the Image Set 1 because the viewpoint variation of the Image Set 1 is not very large. However, they are vital for the Low Quality Set.

Figure 5.11 CMC curves were obtained from different algorithmic settings to analyze different computational components in the proposed algorithm. (a) and (b) compare the Gabor and LBP features with different grid systems. (c) and (d) compare the directional grid systems with the dynamic grid system. (e) and (f) evaluate the effectiveness of the positive sample generation scheme. The CMCs in the first column were obtained from the Image Set 1 and the CMCs in the second column were obtained from the Low Quality Set.

Table 5.4 A summary of the experimental results with different settings of the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
The proposed algorithm	93.82%	<u>99.81%</u>	<u>100%</u>
The proposed algorithm without the alignment scheme	92.51%	98.31%	99.06%
The proposed algorithm without the directional grids	86.70%	97.94%	98.50%
The proposed algorithm without the positive sample generation	<u>94.76%</u>	99.81%	100%
Improvement offered by the alignment scheme	1.31%	1.50%	0.94%
Improvement offered by the directional grids	7.13%	1.87%	0%
Improvement offered by the positive sample generation	-0.94%	0%	0%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-30 accuracy
The proposed algorithm	<u>24.35%</u>	<u>52.25%</u>	<u>59.10%</u>
The proposed algorithm without the alignment scheme	13.24%	52.01%	58.39%
The proposed algorithm without the directional grids	12.29%	33.33%	43.26%
The proposed algorithm without the positive sample generation	17.97%	38.06%	45.15%
Improvement offered by the alignment scheme	11.11%	0.24%	0.71%
Improvement offered by the directional grids	12.06%	18.92%	15.84%
Improvement offered by the positive sample generation	6.38%	14.19%	13.95%

The proposed algorithm makes use of PLS instead of linear SVM, which is commonly employed by other researchers, for training the models. An experiment was conducted to compare PLS and linear SVM on the proposed algorithm. In this experiment, all features and schemes in the proposed algorithm were the same, except that one algorithm used PLS and another applied linear SVM. Note that positive sample generation scheme was also used with PLS and linear SVM. The linear SVM is implemented by using L2-regularized L2-loss support vector regression (primal) with the default parameters from

LIBLINEAR [103]. The results are given in Table 5.5. In the first ten ranks, the proposed algorithm using PLS performs better than that using linear SVM for both probe sets. The greatest difference is about 1.31% for the Image Set 1 and 3.55% for the Low Quality Set. It shows that PLS is a reasonable choice of the proposed algorithm.

Table 5.5 A comparison between PLS and linear SVM on the proposed algorithm of (a) the Image Set 1 and (b) the Low Quality Set.

(a)

	Rank-1 accuracy	Rank-5 accuracy	Rank-10 accuracy
The proposed algorithm with PLS	<u>93.82%</u>	<u>99.81%</u>	<u>99.81%</u>
The proposed algorithm with linear SVM	92.51%	98.88%	<u>99.81%</u>

(b)

	Rank-1 accuracy	Rank-5 accuracy	Rank-10 accuracy
The proposed algorithm with PLS	<u>24.35%</u>	<u>38.77%</u>	<u>43.74%</u>
The proposed algorithm with linear SVM	22.22%	35.22%	43.03%

The last experiment was to test the performance of the proposed algorithm on very low resolution images. To keep the parameters in the proposed algorithm unchanged, images in the probe sets and gallery set were first resized to 142×298 pixels (25 dpi), 106×233 pixels (18.75 dpi), 71×149 pixels (12.5 dpi), and 35×74 pixels (6.25 dpi), and were then scaled back to 142×298 pixels. The Image Set 2 was used to generate the new gallery sets and the other two datasets were used to generate the new probe sets. Figures. 5.2(b)-(e) show images in these four resolutions. The proposed algorithm was applied to these new probe and gallery sets. The experimental results are given in Figure 5.12. Figure 5.12(a) shows that in most of the ranks, when the resolution decreases, the performance drops. However, the results in Figure 5.12(b) are different. The CMC curves obtained from

the images with 18.75 dpi and 12.5 dpi are better than those obtained from the images with 25 dpi and 6.25 dpi. The algorithm is more sensitive to inaccurate alignment for higher resolution images. Due to the low image quality of this dataset, accurate alignment is more difficult to be achieved. Thus, the CMC curve obtained from the images with 25 dpi is not better than those from the images with 18.75 dpi and 12.5 dpi. In the lowest resolution images (6.25 dpi), even though the influence from inaccurate alignment diminishes, the CMC curve still drops because of the reduction of discriminative information in the images. For these very low resolution images, which can be found in real cases (Figure 5.2), the proposed algorithm gives the rank-15 accuracy of 49.65% and the rank-30 accuracy of 57.45%. It demonstrates the effectiveness of the proposed algorithm for low resolution images. Nevertheless, the proposed algorithm performs similarly on images with different resolutions.

Figure 5.12 CMC curves for different image resolutions of (a) the Image Set 1 and (b) the Low Quality Set.

5.5 Discussion

The previous study exposed the potential of androgenic hair patterns in low resolution images as a soft biometric trait for forensic applications [4]. However, the previous results gave an impression that androgenic hair is not a discriminative biometric trait, like other soft biometric traits such as gender, age, and race. To further study androgenic hair as a soft biometric trait, in this chapter, a new algorithm, which is composed of an alignment scheme based on leg geometry, multi-directional grid systems, class-specific PLS models, and a positive sample generation scheme, is designed. The proposed algorithm achieves accuracy of over 99% at the 10th rank for the frontal leg image database and over 59% at the 30th rank for the low quality image database. These results demonstrate that the impression is due to the previous method, DGOH, but not androgenic hair.

In forensic applications, no single biometric trait is suitable for all cases. For example, face is not applicable for identifying masked rioters. Even though hard biometric traits are available, the quality of signals may not be enough for accurate matching. Traditional soft biometrics have been used to combine hard biometrics and to narrow down suspect lists. Comparing with other soft biometric traits such as gender, race and age, androgenic hair even in low resolution images has valuable discriminative information and is worth for further study.

The Gabor and LBP features, which are effective for representing orientation, local density, and textural information [53,101], are applied to androgenic hair patterns. The medical studies [46-47] implied that androgenic hair density is relatively stable because androgenic hair follicles are very stable. The directions of their follicles partially determine

androgenic hair orientations. By considering androgenic hairs as line segments with different orientations and scales, a set of Gabor filters, which offer spatial locality, spatial frequency, and orientation selectivity, is used to extract the orientation and density information. Although androgenic hairs are movable locally, their overall characteristics do not change naturally. For example, curly hairs would not turn to straight hairs and dense hair regions would not turn to sparse hair regions without manual operations like shaving. By considering androgenic hair as a pattern on the skin, i.e., skin texture, local LBP features, which are well-known texture descriptors and invariant to monotonic gray-level changes [101], are employed to extract androgenic hair patterns.

The testing database in this study contains 1,493 images from 412 legs, which is the largest one comparing with all the previous studies. Its size is not large compared with face recognition databases, which have been studied nearly for a half-century and by numerous researchers. However, comparing with other public biometric databases such as UBIRIS.v1 iris database collected from 241 irises [104], CMU keystroke dynamic database collected from 51 subjects [105] and CASIS multi-spectral palmprint database collected from 100 subjects [106], the size of our database is comparable. In some forensic cases, the gallery set is not necessary very large because other information can significantly narrow down suspect lists. John Jihadi's case is an example. His British accent and terrorist activities gave important hints to law enforcement agencies by limiting the suspects of U.K. terrorists in Middle East. In some extreme cases, it is in fact a verification problem, one-to-one matching. For example, gait recognition was applied to a legal case for verifying a suspect [92].

To achieve higher accuracy, a large feature set is used in the proposed algorithm. For example, three color channels are used in this chapter instead of one channel in the previous work [102], which is mentioned in Section 4.1. It would increase the computation time. The leg models can be trained in offline and forensic investigation does generally not demand real-time identification. In addition, the development of GPU for scientific computation and other parallel computation resources, e.g. cloud, can support using large feature sets for offering more accurate identification. Since an incremental PLS algorithm has been developed, the scalability of the proposed algorithm is not a problem [107]. Note that we do not claim that PLS is an optimal classifier, but our experimental results show that it is a reasonable choice. It should be emphasized that the aim of this chapter is to demonstrate that androgenic hair contains rich information for forensic applications.

The rank-1 accuracy of the proposed algorithm for the Low Quality Set is not very high. However, its rank-20 and rank-30 accuracy is respectively more than 50% and 59%, which are 20% more than the previous work. Some may comment that this accuracy is still low. It should be noted that accuracy of forensic identification is generally low because of quality of signals. Paulina et al. reported that VerFinger, a commercial fingerprint matcher, achieves accuracy of 30.6% for NIST SD27 latent fingerprints with ugly quality [108]. Comparing with other soft biometric traits such as gender and age, the accuracy reported in this chapter has demonstrated the distinctive power of androgenic hair. It should be re-emphasized that images captured by surveillance cameras are not within the scope of this study.

To balance the number of positive and negative, PLS can effectively handle it based on [78]. However, the positive sample generation scheme not only increases the number of

positive samples, but also it can be robust against inaccurate alignment because of pose and viewpoint variations. Due to these two reasons, the positive sample generation scheme is included. Fig. 5.11(e) shows that the PLS is effective to handle imbalance problem, but Fig. 5.11 (f) demonstrates that the robustness of imbalance data is not sufficient to handle the problems with pose and viewpoint variations. By including the positive sample generation scheme, the identification rate can be improved. Even though the results in Figure 5.11(f) may have many factors behind, the proposed algorithm with the positive sample generation scheme can be useful to handle them. In addition, the positive sample generation scheme is used to generate positive samples by using affine transformation, which includes translation, rotation, and scale. Thus, the improvement is related to pose and view variations.

In Appendix D, Figure D.4 shows 10 failure cases for Image Set 1 at rank-1. Six-female sets are shown in the first three rows and four-male sets are shown in the last two rows. In Figure D.4, it shows that the wrongly matched female legs also look similar to each other. Since many females have a habit to shave their leg hair, their legs may have no hair or only few hair and their legs tend to be smoother. It is relatively difficult to obtain the discriminative features on their legs. Besides, some statistics are collected for the failure cases in the first experiment. By using the data in two folds, 8.97% female legs and 4.22% male legs failed to get the correct matches at rank-1. Although the percentage of failure cases on female legs is higher than that of male legs, the difference of the number of wrong cases is not very significant. The further study is still required on the relationship between hair density and identification performance.

Figure D.5 shows 10 failure cases for Low Quality Set at rank-1. In these failure cases, the illumination is very different from probe images and gallery images. When the probe images are quite dark, the hair and skin texture on the probe images are very different from the gallery images. Therefore, the probe images and gallery images failed to match in these cases. To tackle these issues, illumination normalization can be used for image preprocessing. However, it is not capable of solving partial illumination problem, like the first case at the left in Figure D.5. These cases reveal that the illumination problem is also a factor to affect the performance of proposed algorithm and a further study is required for this problem.

5.6 Summary

Criminal and victim identification is always an important forensic task. The previous study done by Su et al. [4] exposed the potential of using androgenic hair patterns in low resolution images for criminal and victim identification. However, they neglected the importance of alignment and their Gabor features were not distinctive enough. To overcome these weaknesses and achieve more accurate androgenic hair pattern identification, the new algorithm, which is composed of the alignment method based on leg geometry, the Gabor and LBP features, the positive sample generation scheme, the dynamic and directional grid systems, and the PLS models, is proposed. The algorithm addresses pose and viewpoint variations by using the alignment method, the directional grid systems, and the positive sample generation scheme. To enhance the discriminative power, it utilizes large feature sets with color, orientation, and texture information. In order

to handle these large feature sets, the PLS regression is employed for final identification. The experimental results show that for the both front pose and low quality datasets, the proposed algorithm outperforms the DGOH method proposed by Su et al. [4] with a significant margin of more than 15% of identification accuracy. Besides, the experimental results also demonstrate that the proposed algorithm is effective for very low resolution images, e.g., 6.25 dpi. The proposed algorithm and the experimental results expose the discriminative power of androgenic hair patterns in low resolution images. According to my best knowledge, it is the first study of androgenic hair patterns in low resolution images with large viewpoint and pose variations.

Chapter 6

Proposed Estimation Algorithm for Deducing Soft Biometric Traits

6.1 Overview

The correlation between soft biometric traits and leg has been studied by physiologists, anthropometrists, forensic anthropologists, and medicolegal investigators for diverse proposes, e.g. forensic investigation, aging, obesity, and sport analysis. Ahmed et al. [109] found that leg dimensions, including tibia length, are sexually dimorphic in a Sudanese population and Chow et al. [110] found that the difference of the gastrocnemius and soleus muscles between males and females is statistically significant in terms of fiber bundle length, pennation angle and muscle thickness. Satorio et al. [111] showed that lower limb maximal power is negatively influenced by age, but positively correlated with BMI for obese people. Jones et al. [112] deduced that a large portion of the strength loss associated with aging may be accounted for a loss of muscle cross sectional area. Gupta et al. [113] showed a positive correlation between the tibia length and the estimated height in living north Indian males. Duyar et al. [114] further suggested a group-specific regression model based on tibia length for height estimation. Some of these research results are applicable to forensic investigation in cases of wars, mass disasters and criminal assaults, where the physical body of victims and criminals are available for direct measurement. Leg appearance in images is related to muscle and size of legs as well as length of tibia. For

example, an obese person should have fat legs. These research results imply that it is possible to estimate soft biometric traits from leg appearance in images.

In order to identify/verify criminals and victims, law enforcement agencies make use of DNA, fingerprint, palmprint, tattoo, iris, face, ear, skin mark, and gait [93, 115]. However, these traits are not always observable in evidence images from cases of sexual offenses, child sexual abuse, and voyeur. Investigation of child sexual abuse cases becomes more complex and difficult because of the recent rise of online grooming [117]. Over the past four years, there was a 60% increase in child sexual abuse reported to the police in England and Wales, but the number of people arrested dropped by 9% and the number of prosecutions and convictions for child abuse offences fell [117]. According to the U.S. Immigration and Customs Enforcement (ICE), some paedophiles distribute images and videos recorded during their sexual offenses in the Web [118-120]. Figure 6.1 shows a simulation of ¹an evidence photo from a sexual offense case handled by U.S ICE. The subject in the original evidence photo was a ²minor victim, whose identity and soft biometrics, including height and age, were not known. Furthermore, voyeurism increases with the advance of technology [121-122]. Some voyeurism photos are given in Figure 6.2. In some cases, suspects are caught, but victims cannot be found because evidence photos only show non-facial body parts. In 2012, Santa Barbara police arrested a voyeurism suspect and needed to search the victims for further investigations [123]. Although neither faces nor tattoos of criminals and victims are always observable, parts of their bodies are exposed in these cases.

¹ U.S. ICE showed the original evidence photo to the last author in their meeting.

² The age of the victim was not known. Based on the last author estimation, she should be a minor.

Figure 6.1 A simulation of an evidence photo obtained by enforcement agency for child sexual abuse case.

Figure 6.2 Examples of voyeur images collected randomly from the Internet. The orange triangles are used to prevent showing the underwears of the females.

To extract comparative features and to estimate soft biometrics from leg images, an approach, including image alignment, raw feature extraction, comparative feature derivation, and rank estimation, is proposed. Training images are first aligned with a leg

model and raw features are extracted from the aligned images. The raw features are grouped and assigned to different ranks (e.g. the 5th rank indicating the interval, [170 cm, 175cm)) according to their value metrics. The training features in the same rank are further clustered by the k -means algorithm and the means are used as references. Given a testing image, it is first aligned with the model and its raw feature is extracted and compared with the references to derive comparative features. The comparative features are inputted to the proposed OpRank algorithm, which is a novel probability model to optimize the correlation between the ranks of soft biometrics and the comparative features with the regularization of the ranking information. In this study, the proposed algorithm is used to estimate the soft biometric traits, height, BMI, and age from leg images. It is found that the proposed OpRank algorithm is more effective than other related algorithms and the estimated soft biometric traits can be combined with other biometric traits to enhance matching performance.

The rest of this chapter is organized as follows. Section 6.2 presents an overview of the proposed approach. Section 6.2.1 and Section 6.2.2 describe the alignment and feature extraction schemes. Section 6.2.3 and Section 6.2.4 give the computational details of comparative feature derivation and the proposed OpRank algorithm. Section 6.3 reports the experimental results. Section 6.4 gives conclusive remarks.

6.2 Proposed Approach

Figure 6.3 illustrates the proposed approach for estimating soft biometric feature based on comparative features. Initially, leg images are segmented by using the scheme reported in Section 3.2.2. Then, legs are aligned to a template and resized (Section 6.2.1). Shape,

Gabor, and texture features are extracted as raw features (Section 6.2.2). The raw feature is converted to a set of comparative features. By inputting the set of comparative features and the rank patterns which are expected outputs, OpRank algorithm is used to generate a model optimizing the correlation between the ranks of soft biometrics and comparative features. The definition of rank patterns is given in Section 6.2.3. Based on OpRank model, a set of comparative features from a testing image is used to infer its rank, which is considered as a soft biometric feature for the corresponding soft biometric trait (Section 6.2.4). This approach is applied to estimate three soft biometric traits, including height, BMI, and age. Only shape feature is used for height and BMI estimation, but all features are included for age estimation.

Figure 6.3 A schematic diagram of the proposed approach.

6.2.1 Template Generation and Image Alignment Scheme

All the training and testing images are aligned to a leg template, which is generated from the training set. To generate the leg template, regular sampling is applied to leg shapes.

Note that leg shape of each leg refers to two leg boundaries. Then, generalized Procrustes analysis (GPA), which is mentioned in Section 2.3.2, is applied to generate the leg template. If the mean of the leg shapes is used as a template, the template will be distorted by pose and viewpoint variations associated with the leg images. An example displayed in Figure 2.18 shows the difference between a simple mean template and GPA template. The GPA template is smoother than the mean template. GPA initially aligns all the leg shapes to a leg shape, which is chosen arbitrarily, to reduce their initial geometric distortion. In this alignment, non-reflective similarity transformation is used. Since the differences between all the leg shapes are reduced, a new leg template is obtained by averaging all the transformed leg shapes. If the differences between all the leg shapes and the new leg template are still large, the leg shapes are aligned to the new leg template. The previous averaging and alignment steps repeat until the difference is less than a threshold. By using the obtained leg template, all the legs in the training and testing sets are aligned with non-reflective similarity transformation T , which can be obtained by solving the following equation

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = T \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (6.1)$$

where (u, v) and (x, y) are the normalized coordinates of the sample points on leg template and legs in the training and testing set.

6.2.2 Feature Extraction

In this section, shape, Gabor and texture features are extracted from leg images. They are used to form comparative features for estimating the soft biometric traits. The shape

features are used for all estimations. The Gabor and texture features are applied on age estimation only because aging would affect skin quality and its texture.

To normalize the images, the height of all the leg images are resized to 298 pixels. Seven different shape features, which contain global and local features, are extracted. In the global features, the first one is the ratio of height and width. Although the heights of the resized images are the same, their widths are not. If a subject is taller or slimmer, this ratio is expected to be higher. The second one is the correlation between two leg boundaries, which indicates their similarity. The third one is the ratio of the top width and the bottom width of the leg. If a subject is fat, this ratio is expected to be larger. The fourth one is Fourier descriptors of two leg boundaries, which are used to describe the geometrical information of the leg shape. The fifth one is the coefficients of the cubic polynomial fitting on each leg boundary, which is extracted to obtain the spatial information of the leg shape.

In the local features, the first one is the regional differences of the leg. The leg image is divided into 10 horizontal regions. Then, the row widths in the b^{th} region are summed up to be W_b . The regional differences of the leg are defined as $W_b - W_{10}$, where $b = 1, 2, \dots, 9$. If the height is not divisible by 10, some regions can be overlapped to make sure that every region has the same width. In the implementation, height was set at 298 pixels. To make use of all the rows of an image, region size was set to be 30 in the implementation. Then, 11 evenly spaced points in the interval $[1, 298]$ were generated and rounded and the first 10 points were used to be the starting row of the regions, i.e. some regions were overlapped by one row. The second one is the local curvatures, which can reflect local leg shape. To get the local curvatures, the leg image is cut up into 2 and 4 horizontal regions. For each leg boundary of each region, a circle equation $z^2 =$

$(x_i - a)^2 + (y_i - b)^2$ is formed to approximate one leg boundary, where x_i and y_i are the x -coordinate and y -coordinate of the i^{th} sample on each leg boundary. The least square solution of radius z is

$$z = \sqrt{\frac{\sum_{i=1}^N ((x_i - a)^2 + (y_i - b)^2)}{N}}, \quad (6.2)$$

where N is the number of samples on each leg boundary. By substituting Equation (6.2) into the circle equation, a system of linear equations can be solved to obtain a and b . When they are found, the radius z is determined by Equation (6.2) and the local curvature $1/z$ is obtained for each leg boundary of each region.

To obtain Gabor and texture features, all the leg images are resized to 142×298 pixels, same as [4] mentioned in Section 2.2.2. The Gabor features are extracted by two operations in two layers suggested by Guo et al. [127]. According to Guo et al. [127], these operations simulate human visual processing in cortex. The parameters used in the two operations, which are chosen empirically, are given in Table 6.1. The rest of the parameters (wavelength, spatial aspect ratio) in the Gabor filters are the same as those in [4] except for scale. Note that the number of bands is 8 and the number of orientation is 4.

In the filtering layer, two Gabor filters, which are defined in [4], are applied with specific Gabor scales for each band and each orientation. In the merging layer, the two Gabor response maps with the same orientation, but different scales are merged into one by using the “MAX” operation defined as

$$F_i = \max(x_i^j, x_i^{j+1}), \quad (6.3)$$

where x_i^j and x_i^{j+1} are the Gabor responses at the position i from the Gabor filters with scales j and $j+1$, respectively. Then, the merged map is divided into blocks, whose size for

each band is defined in Table 6.1, and standard deviation of each block is calculated. These standard derivations represent the Gabor features for leg images. Figure 6.4 illustrates the Gabor feature extraction process.

Table 6.1 Gabor parameters used in the two layers.

Merging layer			Filtering layer
Scale band	Number of blocks in column	Number of blocks in row	Gabor scale
Band 1	49	23	0.2
			0.5
Band 2	37	17	0.7
			0.9
Band 3	29	14	1.1
			1.3
Band 4	24	11	1.5
			1.7
Band 5	21	10	1.9
			2.1
Band 6	18	8	2.3
			2.5
Band 7	16	7	2.7
			2.9
Band 8	14	7	3.1
			3.3

Figure 6.4 Illustration of Gabor feature extraction process. Only one band is shown in this illustration. (a) An input image, (b) response maps from Gabor filters with the same orientation but different scales, (c) the maps merged by “MAX” operation, (d) the blocks produced from (c) and (e) the Gabor feature vector constituted by the standard derivations from the blocks .

Finally, the texture feature is obtained by using local binary patterns (LBP) with directional grid systems. The local binary patterns $LBP_{P,R}$ [53] are defined as

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad (6.4)$$

where P is the number of sampling points, R is the radius of the sampling circle, g_c is the center pixel value, g_p is the neighborhood pixel value and $s(x) = 1$ if $x \geq 0$, otherwise $s(x) = 0$. The uniform LBP $LBP_{8,2}^{u2}$, which are with at most two bitwise transitions from 0 to 1 or vice versa when they are considered as circular patterns, are extracted from the directional grid systems, which is proposed in Chapter 5. Figure 6.5 illustrates the directional grid systems. Their block height $H_{j\theta}$ and width $W_{j\theta}$ are respectively defined as $H_{j\theta} = \lfloor (H_\theta - 1)/N_H \rfloor$ and $W_{j\theta} = \lfloor (W_\theta - 1)/N_W \rfloor$, where H_θ is the distance between the top leg pixel and the bottom leg pixel along direction $\theta = \{0^\circ, 30^\circ, \dots, 330^\circ\}$, W_θ is the distance between the leftmost and the rightmost leg pixels along the direction perpendicular to θ , $N_W = 8$ is the number of columns, $N_H = 5$ is the number of rows along direction θ and the first and last rows are not used. The pairs of opposite directions (e.g. 90° and 270°) are included because of the floor operators used in the block height and width computation. Thus, the features from these opposite directions are not identical. For example, the blocks are placed from left to right for 90° and from right to left for 270° .

Figure 6.5 (a)-(l) are the directional grid systems.

6.2.3 Comparative Feature and Rank Pattern Generation Scheme

The features extracted in the previous section are the raw feature f_i for an image I_i with the rank r_i from a soft biometric trait (height, BMI, and age). The rank r_i represents that the value metric (e.g. 172cm) of the soft biometric trait is in a particular interval (e.g. [170cm 175cm)). For height and BMI estimation, only shape feature is used as a raw feature. For age estimation, 100 features from Gabor and texture features are chosen as raw features through the minimum redundancy maximal relevance (mRMR) feature selection [88].

To obtain the comparative features, the raw features extracted from images with the same rank are grouped first. After grouping the data, k -means clustering is applied to get three clusters from each group to reduce the bias of the groups with many data. The centroids of the group from the n^{th} rank are used as a reference ($\mathbf{m}_n^1, \mathbf{m}_n^2, \mathbf{m}_n^3$) for deriving

comparative features. If the number of images with the same rank is less than three, k -means clustering is set to generate one cluster ($k = 1$) and its centroid is duplicated three times to obtain the references. Let the raw feature and the element of the reference be $\mathbf{f}_i \in \mathbb{R}^{N_f \times 1}$ and $\mathbf{m}_n^k \in \mathbb{R}^{N_f \times 1}$ for $k = 1, 2, 3$. The comparative features of the image I_i are computed as

$$\mathbf{c}_i = (\boldsymbol{\phi}_1^i, \boldsymbol{\phi}_2^i, \dots, \boldsymbol{\phi}_O^i)^T \in \mathbb{R}^{O \times 3N_f}, \quad (6.5)$$

where $\boldsymbol{\phi}_n^i = ((\mathbf{f}_i - \mathbf{m}_n^1)^T (\mathbf{f}_i - \mathbf{m}_n^2)^T (\mathbf{f}_i - \mathbf{m}_n^3)^T)^T \in \mathbb{R}^{3N_f \times 1}$, $n = 1, 2, \dots, O$, O is the number of ranks, and N_f is the dimensionality of the raw feature vector.

Since a set of comparative features is formed by the comparisons with different ranks, these comparisons may provide extra information for learning. In order to use the comparisons with different ranks effectively, the rank pattern of the n^{th} rank is defined as a set of labels $\mathbf{b}_n = (b_1, b_2, \dots, b_O)^T$, which represents the expected results of the comparisons of the n^{th} rank with different ranks $1, 2, \dots, O$, i.e., $b_a = 1$ if $n \leq a$ and -1 otherwise, where T is a transpose operator, O is the number of ranks, and $n = 1, 2, \dots, O$.

For each leg image I_i , it has its own rank pattern. If the leg image I_i belongs to the n^{th} rank, ($r_i = n$), its rank pattern is $\mathbf{b}_{r_i} = \mathbf{b}_n$. For example, the rank pattern of $r_i = 3$ is an ordered sequence $\{-1, -1, 1, 1, 1\}^T$ provided that there are five ranks ($O = 5$) in total. Figure 6.6 illustrates the concept of the rank pattern.

Figure 6.6 The concept of the rank pattern.

6.2.4 OpRank Algorithm

An algorithm OpRank, which takes comparative features $(\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N)$ and rank patterns $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_O)$ as inputs, is used to obtain the models for rank $n = 1, 2, \dots, O$. The set of comparative features $\mathbf{c}_i \in \mathbb{R}^{O \times 3N_f}$ is reorganized as a column vector $\mathbf{c}_i' \in \mathbb{R}^{3N_f O \times 1}$ and the comparative features of N training data form $\mathbf{X} = (\mathbf{c}_1', \mathbf{c}_2', \dots, \mathbf{c}_N')^T \in \mathbb{R}^{N \times 3N_f O}$. The label y_i is set to 1 if the image I_i belongs to the n^{th} rank ($r_i = n$); otherwise it is equal to -1. Since PLS is effective for handling imbalanced data [78], it can be used as a one-against-all approach for training a model to classify data from a specific rank against data from all other ranks. A derivation of PLS without latent variables, which is suggested by Bennett et al. [75], is defined as

$$\max_{\mathbf{w}} \text{cov}(\mathbf{X}\mathbf{w}, \mathbf{y}) \text{ s.t. } \mathbf{w}^T \mathbf{w} = 1, \quad (6.6)$$

where $\mathbf{X}$ is a zero-mean matrix, $\mathbf{w}$ is the weight, $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ is the data labels and T is a transpose operator. By using Equation (6.6), the optimal weight $\mathbf{w}_n^*$ is obtained for each rank n . Note that one-against-all approach is used in this study and therefore, for different ranks, the data labels are different. For a new image I' with a set of comparative

features $\mathbf{c}'$, its estimated rank is $\underset{j}{\operatorname{argmax}} \mathbf{c}' \mathbf{w}_j^*$. However, the ranking information in this

formulation is not fully utilized because both y_k and y_l can be set to -1 even if their ranks are different ($r_k \neq r_l \neq n$).

To make use of such information effectively, the comparative features of N training data and the weight are reorganized as $\mathbf{X} \in \mathbb{R}^{O \times 3N_f \times N}$ and $\mathbf{w} = (\mathbf{w}_1^T, \dots, \mathbf{w}_j^T, \dots, \mathbf{w}_O^T)^T \in \mathbb{R}^{O \times 3N_f}$, where a rank pattern index $j = 1, \dots, O$, $\mathbf{w}_j = (w_1 \dots w_k \dots w_{3N_f})$, $\mathbf{X}_j = (\boldsymbol{\phi}_j^1, \dots, \boldsymbol{\phi}_j^i, \dots, \boldsymbol{\phi}_j^N)$, $\boldsymbol{\phi}_j^i \in \mathbb{R}^{3N_f \times 1}$ is a set of comparative features of the image I_i defined in Section 6.2.3, a feature index $k = 1, \dots, 3N_f$, and an image index $i = 1, \dots, N$. By introducing a regularization term, which contains the rank patterns and the sets of comparative features, the optimization for each rank n from Equation (6.6) is modified as follows:

$$\max_{\mathbf{w}, \boldsymbol{\alpha}} \sum_j (\mathbf{w}_j \mathbf{X}_j \mathbf{y} + \alpha_j \mathbf{w}_j \mathbf{X}_j \hat{\mathbf{y}}_j^n) \quad \text{s.t. } \operatorname{tr}(\mathbf{w}^T \mathbf{w}) = 1, \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1, \text{ and } \alpha_j \geq 0, \quad (6.7)$$

where $y_i = 1$ if the image I_i belongs to rank n and -1 otherwise, $\hat{\mathbf{y}}^n = (\hat{\mathbf{y}}_1^n, \dots, \hat{\mathbf{y}}_j^n, \dots, \hat{\mathbf{y}}_O^n) = (\mathbf{b}_{r_1} \circ \mathbf{b}_n, \dots, \mathbf{b}_{r_i} \circ \mathbf{b}_n, \dots, \mathbf{b}_{r_N} \circ \mathbf{b}_n)^T$, $\hat{\mathbf{y}}_j^n = (\mathbf{b}_{r_1}(j) \mathbf{b}_n(j), \dots, \mathbf{b}_{r_i}(j) \mathbf{b}_n(j), \dots, \mathbf{b}_{r_N}(j) \mathbf{b}_n(j))^T$ is the difference between the rank pattern $\mathbf{b}_n$ of the rank n and rank patterns of all images for a specific j , and $\circ$ is Hadamard product. In Equation (6.7), the first term is the same as the term in the Equation (6.6) and the second term is used as a regularization term to give a penalty based on the difference between the training rank n and the rank of image I_i , i.e. the difference between $\mathbf{b}_{r_i}$ and $\mathbf{b}_n$. Note that the rank pattern in the Equation (6.7) is relative to the training rank, i.e. the rank pattern of $r_i = 3$ with an ordered sequence $\mathbf{b}_{r_i=3} = \{-1, -1, 1, 1, 1\}^T$ becomes $\mathbf{b}_{r_i=3} \circ \mathbf{b}_{n=3} = \{1, 1, 1, 1, 1\}^T$ provided that there are five ranks ($O = 5$) and $n = 3$ in total. If the pattern of the training

data does not follow the rank pattern relative to the training rank, it will be penalized. For example, a rank pattern relative to the training rank is $\mathbf{b}_{r_i=2} \circ \mathbf{b}_{n=3} = \{1, -1, 1, 1, 1\}^T$, whose second one is negative. If the corresponding element in $\mathbf{w}_{j=2}\mathbf{X}_{j=2}$ is not negative, it is considered as a violation because negative value is expected. The violation can lead to a penalty by reducing its corresponding weight in $\boldsymbol{\alpha}$, which is used to combine the components of different rank index j by different weights. Besides, the Equation (6.7) can be rewritten as

$$\max_{\mathbf{w}, \boldsymbol{\alpha}} \sum_j \mathbf{w}_j \mathbf{X}_j (\mathbf{y} + \alpha_j \hat{\mathbf{y}}_j^n) \quad \text{s.t. } \text{tr}(\mathbf{w}^T \mathbf{w}) = 1, \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1, \text{ and } \alpha_j \geq 0. \quad (6.8)$$

By imposing the conditions on the Equation (6.8), it becomes constrained optimization

$$\min_{\mathbf{w}, \boldsymbol{\alpha}} (-R_n) = \min_{\mathbf{w}, \boldsymbol{\alpha}} \left(- \sum_j \frac{\mathbf{w}_j}{\sqrt{\text{tr}(\mathbf{w}^T \mathbf{w})}} \mathbf{X}_j (\mathbf{y} + \frac{\alpha_j}{\sqrt{\boldsymbol{\alpha}^T \boldsymbol{\alpha}}} \hat{\mathbf{y}}_j^n) \right) \quad \text{s.t. } \alpha_j \geq 0, \quad (6.9)$$

which can be solved by Sequential Quadratic Programming (SQP) algorithm [132]. The gradients of Equation (6.9) are

$$\frac{\partial(-R_n)}{\partial w_{jk}} = - \left(\frac{\text{tr}(\mathbf{w}^T \mathbf{w}) - w_{jk} w_{jk}}{\text{tr}(\mathbf{w}^T \mathbf{w})^{3/2}} \mathbf{X}_{jk} \left(\mathbf{y} + \frac{\alpha_j}{\sqrt{\boldsymbol{\alpha}^T \boldsymbol{\alpha}}} \hat{\mathbf{y}}_j^n \right) \right), \quad (6.10)$$

$$\frac{\partial(-R_n)}{\partial \alpha_j} = - \left(\frac{\mathbf{w}_j}{\sqrt{\text{tr}(\mathbf{w}^T \mathbf{w})}} \mathbf{X}_j \left(\frac{\boldsymbol{\alpha}^T \boldsymbol{\alpha} - \alpha_j \alpha_j}{(\boldsymbol{\alpha}^T \boldsymbol{\alpha})^{3/2}} \hat{\mathbf{y}}_j^n \right) \right). \quad (6.11)$$

After obtaining $\mathbf{w}$ and $\boldsymbol{\alpha}$ for each rank, the response R_n' of each testing data $\mathbf{X}' = [\mathbf{X}'_1, \dots, \mathbf{X}'_j, \dots, \mathbf{X}'_O]^T \in \mathbb{R}^{O \times 3N_f}$ with the rank n is

$$R_n' = \sum_j \frac{\mathbf{w}_j}{\sqrt{\text{tr}(\mathbf{w}^T \mathbf{w})}} \mathbf{X}'_j \left(1 + \frac{\alpha_j}{\sqrt{\boldsymbol{\alpha}^T \boldsymbol{\alpha}}} \right), \quad (6.12)$$

where $\mathbf{X}'_j = \left[(\mathbf{f}' - \mathbf{m}_j^1)^T (\mathbf{f}' - \mathbf{m}_j^2)^T (\mathbf{f}' - \mathbf{m}_j^3)^T \right]^T \in \mathbb{R}^{3N_f \times 1}$, $\mathbf{f}'$ is the raw feature of testing image I' , and a rank index $j = 1, \dots, O$. Normally, the rank of testing data $\mathbf{X}'$ is the

one with maximum response from all responses $\mathbf{R}' = (R'_1, R'_2, \dots, R'_O)$, which are obtained by using Equation (6.12). But Chang et al. [86] showed that the age rank can be inferred by aggregating a series of binary classification results. In this chapter, an aggregation of probabilities inferred from different responses of the ranks based on each image is considered instead. The probability $p(n|I')$, of rank n based on I' is calculated as

$$p(n|I') = \frac{NR'_n}{\sum NR'_n}, \quad (6.13)$$

where $NR'_n = (R'_n - \min(\mathbf{R}'))/(\max(\mathbf{R}') - \min(\mathbf{R}'))$ is a normalized response. In order to remove the noise, the probabilities of all ranks based on I' are sorted in a descending order and its cumulative distribution is calculated. By selecting the ranks, whose cumulative probabilities are within 95%, the rank of each leg image n' is

$$n' = \text{ROUND} \left(\sum_{s \in S} n_s \cdot p(n_s|I') \right), \quad (6.14)$$

where $\text{ROUND}(\cdot)$ is an operator to round the value to the nearest integer, the indices of selected ranks are $S \subset \{1, \dots, O\}$, O is the total number of rank, and $p(n_s|I')$ is the posterior probability of a selected rank n_s based on a leg image I' . The estimated rank of each leg image is used to be the soft biometric feature of each leg image and this estimation is applied to height, BMI, and age correspondingly.

6.3 Experimental Results

Two databases, containing front leg images and side leg images, were employed for testing. Each database was divided into training and testing sets, each of which has its own probe set and gallery set. The images in these databases were collected in two sessions with an

interval of at least one week. The subjects' poses and cameras' viewpoints were not strictly controlled. One to three images were taken for each pose in each session and the left and right legs of the subjects were captured. To increase the size of gallery sets, the left leg images were flipped and considered as right leg images from new subjects. Some examples are displayed in Figure 6.7. The gender, height, body mass index (BMI), and age distributions are reported in Table 6.2 and Figure 6.8.

Figure 6.7 Some examples in the (a) front leg and (b) side leg database. The relative scales of the images are retained to show the diversity of the images

Table 6.2 Gender of two databases.

Gender	Front leg database		Side leg database	
	No. of images	Percentage	No. of images	Percentage
Male	758	70.84%	670	71.58%
Female	312	29.16%	264	28.27%

Figure 6.8 The first column and the second column are height, BMI, and age distributions of the front and side leg databases. (a-b) Height, (c-d) BMI, and (e-f) age.

Based on the region of the interest defined in Section 3.2.2 and the color ratios suggested by Crowley et al. [87], the backgrounds of leg images were removed by using the preprocessing method in Section 3.2.2. Then, left leg images were flipped to be right

and all the images were resized to form two sets with different scales. The images in the first set had different widths but a fixed height of 298 pixels and the images in the second set had fixed size of 142×298 pixels, same setting as Su et al. [4]. If a Canon EOS 500D camera with a sensor pixel size of $2.2107 \times 10^{-5} \text{ mm}^2$ and a focal length of 55mm was used to capture images in this resolution, the estimated distance between subjects and camera is about 11.9m [4]. Images with different widths were used to extract shape feature and images with fixed size were used to extract Gabor and texture features.

Before the experiments were performed, the ranks were assigned to different soft biometric traits (height, BMI, age). Each of them has its own value metrics and ranks. Their ranks are obtained by sorting the value metrics ascendingly and grouping them by a specific interval $[v_n, v_n + u)$ for each rank n and $u > 0$. For example, when $u = 2$ and the value metrics of subjects' heights are (159, 160, 161, 163, 164, 166), their ranks are (1, 1, 2, 3, 3, 4). For BMI, the subjects' BMIs are rounded to the nearest integer before their ranks are calculated. In this implementation, $u = 2$ was applied for height, BMI, and age estimation on leg images.

To compare the performance of the proposed OpRank algorithm with different methods, mean absolute error (MAE) and cumulative score (CS), which are two widely adopted performance measures [86, 128], are used. They are defined as follows:

$$\text{MAE} = \sum_{i=1}^M e_i / M, \quad (6.15)$$

$$\text{CS}(L) = 100\% \times \sum_{i=1}^M h(e_i, L) / M, \quad (6.16)$$

where $e_i = |\hat{l}_i - l_i|$, $\hat{l}_i$ and l_i are the predicted and original ranks of the i^{th} testing sample, M is the number of testing data and $h(e_i, L) = 1$ if $e_i \leq L$ and 0 otherwise. Since the value

metrics of height, BMI, and age from the leg images in the databases are not consecutive, the intervals of consecutive ranks, each of which is estimated by taking the difference between average value metrics of two consecutive ranks, are not uniformly distributed. To increase the readability of the experimental results, the median of intervals between two ranks is displayed as a reference. For example, when the value metrics of subjects' heights (in cm) are (159, 160, 161, 163, 164, 166) and their ranks are (1, 1, 2, 3, 3, 4), the average heights for four ranks are (159.5, 161, 163.5, 165). Thus, the intervals between two consecutive ranks are (1.5, 2.5, 1.5) and the median interval is 1.5cm. On the other hand, all the following experimental results were performed by two-fold validation and presented by taking the average of two-fold results in the following experimental results.

The first experiment was to evaluate the effectiveness of OpRank by comparing with nine existing methods on height, BMI, and age estimation. According to my best knowledge, it is the first study of soft biometric estimation on leg images. So, the comparison methods were chosen from related areas of soft biometric estimation. They are Elo ratings system [5-7], Ranking SVM [79], RankNet [80], Regression Tree [81], Random Forest [82], RankBoost [83], Coordinate Ascent [84], ListNet [85], and CSOHR [86]. The Elo ratings system is a method applied to verbal comparative descriptions of soft biometrics [5-7]. CSOHR is one of the latest works from facial age estimation [86]. The rest of them are popular methods of learning to rank [133]. In addition, the implementation of Ranking SVM was from Joachims [129] and the implementations of RankNet, Regression Tree, Random Forest, RankBoost, Coordinate Ascent, and ListNet were from RankLib [133]. The Elo ratings system and CSOHR were also implemented. The Elo ratings system requires a binary classifier to perform comparisons and kernel SVM, whose number of

features and cost were chosen by using two-fold cross validation, were employed. Figures 6.9 and 6.10 show the CS curves and MAEs from the experimental results. A higher score percent means a better CS result, but a lower MAE means a better MAE result. The CS graphs also show that the proposed OpRank algorithm performs well. It reaches more than 80% at the rank-2 of BMI and age of the front and side leg databases. Although it reaches about 80% at the rank-4 of the height of the front leg database, it is still better than other methods. For the height of the side leg database, CSOHR performs better than the proposed algorithm only in the first few ranks, and the MAE difference between the proposed algorithm (3.45cm/unit) and CSOHR (3.42 cm/unit) is very small. For the other bar charts of MAE, the proposed OpRank algorithm performs better than the others because it achieves the lowest mean square error. In summary, the proposed algorithm generally outperforms all other methods in terms of MAE and CS curves, except for one test.

Figure 6.9 CS curves and MAEs of the proposed OpRank and different methods from the front leg database. (a-b), (c-d) and (e-f) are results of height, BMI, and age estimation. One unit on the x-axes of the first column and y-axes of the second column represents 2cm for height, 2kg/m² for BMI, and 2 years for age. The error level of CS curves means tolerance of error. If error level is 2 in CS curves, the real error tolerance is ± 4 cm, ± 4 kg/m² and ± 4 years for height, BMI, and age.

Figure 6.10 CS curves and MAEs of the proposed OpRank and different methods from the side leg database. (a-b), (c-d) and (e-f) are results of height, BMI, and age estimation. One unit on the x-axes of the first column and y-axes of the second column represents 2cm for height, 2kg/m² for BMI and 2 years for age. The error level of CS curves means tolerance of error. If error level is 2 in CS curves, the real error tolerance is ± 4 cm, ± 4 kg/m² and ± 4 years for height, BMI, and age.

The previous results were done by single leg, but two legs of the same person may be captured in one image also. An additional experiment was done to evaluate the performance between one leg and both legs by using MAE. By removing the rounding operation in the Equation (6.14), the rank for each leg was obtained. Then, the new rank of both legs was calculated by taking the average of the ranks of two legs and rounding the average to the nearest integer. To show the improvement more clearly, the MAEs of height, BMI and age are given in cm, kg/m^2 and years by multiplying the measured units with the medians of height, BMI and age per unit (i.e. 2). The corresponding results of 240 left legs, 240 right legs, and 240 both legs are reported in Table 6.3. When both legs of the same person are used, the height and age of the front leg database can be improved by at least 0.05cm, 0.03kg/m^2 , and 0.01years. It shows that two legs provide extra information for improvement of the estimation of height, BMI, and age.

Table 6.3 The MAEs between one leg and two legs of front leg database.

	Left leg	Right leg	Two legs
Height	6.53cm	6.57cm	6.48cm
BMI	2.62kg/m^2	2.60kg/m^2	2.57kg/m^2
Age	3.21years	3.15years	3.14years

The third experiment was to demonstrate the usefulness of the estimated features (height, BMI, and age) for improving two existing works [131] and [4]. The details of Islam et al.'s method [131] are explained in Chapter 3. To use the estimated soft biometric features to improve such work, a scheme, learning the feature weight $\mathbf{w}$, was used to combine the estimated features with the matching scores given by the Islam et al.'s method.

The result of Islam et al.'s method ($s_{i,j}$) was formed as a feature vector with three estimated traits (height, BMI and age) of a probe image h_i^k and ground truth of three traits of gallery image g_j^k , where $k = 1, 2, 3$, i and j are indices of probe set and gallery set. For each probe image i and gallery image j , its feature vector was formed as $\mathbf{D}_\varepsilon = (s_{i,j} \ d_{i,j}^1 \ d_{i,j}^2 \ d_{i,j}^3)^T$, where $\varepsilon = 1, \dots, E$ is an index for each pair (i,j) , $\mathbf{D} \in \mathbb{R}^{E \times 4}$, E is the number of pairs from probe and gallery set, and $d_{i,j}^k = (h_i^k - g_j^k)^2 / (h_i^k + g_j^k)$. The feature weights $\mathbf{w} \in \mathbb{R}^{1 \times 4}$ were learned from

$$\min_{\mathbf{w}} \left(\frac{\mathbf{X}_0 \mathbf{w}^T}{P} - \frac{\mathbf{X}_1 \mathbf{w}^T}{Q} \right) \quad s.t. \quad \mathbf{w}^T \mathbf{w} = 1, \mathbf{w}_{k'} \geq 0, \quad (6.17)$$

where $\mathbf{X}_0 \in \mathbb{R}^{P \times 4}$ and $\mathbf{X}_1 \in \mathbb{R}^{Q \times 4}$ are disjoint subsets of $\mathbf{D}$ and $P + Q = E$, $\mathbf{X}_0$ contains features from the pairs whose i and j belong to the same object and P is the number of these pairs, $\mathbf{X}_1$ contains features from the pairs whose i and j belong to a different object and Q is the number of these pairs, and $k' = 1, 2, 3, 4$ is a weight index. Equation (6.17) was solved by Sequential Quadratic Programming (SQP) algorithm. After the feature weight $\mathbf{w}$ was obtained, the response is obtained by $\mathbf{X}' \mathbf{w}^T$, where $\mathbf{X}'$ is the feature vector extracted from a gallery image and a testing image I' . The smallest response was chosen to be a matched pair of a gallery image and the testing image I' . Cumulative Match Characteristic (CMC) curves, which is a standard approach to evaluate biometric algorithms for criminal and victim identification [25,28,30,32,92], is used to report the performance of the merged features. In Figure 6.11(a) and Figure 6.12(a), Islam et al.'s method was implemented by using SVM ($s_{i,j} = \exp(-s_{SVM})$), neural network ($s_{i,j} = \exp(-s_{NN})$), and their combination $s_{i,j} = \exp(-s_{SVM}) + \exp(-s_{NN})$. Their combination

(SVM and neural network) achieves the best from these three results and it is further improved by combining with estimated features, i.e. 24.9% at rank 20 is improved for the front leg database and 7.16% at rank 20 is improved for the side leg database. Besides, its combination with estimated features was used as extra features for the work of Su et al. [4] in the next experiment.

Su et al. suggested using androgenic hair patterns on legs to identify criminals and victims [4], which is explained in Section 2.2.2. Gabor features with Chi-square distance χ^2 were used. Its distance matrix is calculated as $\chi^2(i, j) = \sum_{k=1}^K \left(h_i(k) - h_j(k) \right)^2 / (h_i(k) + h_j(k))$, where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the histograms of the Gabor features from a probe image i and a gallery image j , respectively and K is the number of bins in the histograms. By using the weights $\mathbf{w}$ obtained from the Equation (6.17), the final distance of a testing image I' with different gallery data was found by $\mathbf{X}'\mathbf{w}_s^T$, where $\mathbf{w}_s = (1 \ \mathbf{w}^T)$ and $\mathbf{X}' = (\chi^2(i, j) \ s_{i,j} \ d_{i,j}^1 \ d_{i,j}^2 \ d_{i,j}^3)^T \in \mathbb{R}^{E \times 5}$ is a feature vector obtained by a gallery image and the testing image I' . Thus, the gallery image having minimum distance with the testing image I' was a matched one. In Figure 6.11(b) and 6.12(b), Su et al.'s method improves 4.4% at rank 1 for front leg database and 0.77% for side leg database. Although the estimated features in Figure 6.12(b) cannot contribute significant improvement at Rank 1, there is an improvement for the first 10 ranks.

Figure 6.11 CMC curves of the estimated soft biometric features with (a) Islam et al.'s method [131] and (b) Su et al.'s method [4] on the front leg database.

Figure 6.12 CMC curves of the estimated soft biometric features with (a) Islam et al.'s method [131] and (b) Su et al.'s method [4] on the side leg database.

In this chapter, rank patterns are proposed for ranking optimization. An experiment was done on the front leg and side leg databases to evaluate its usefulness. To show the effectiveness of the optimization with rank patterns (OpRank), optimization without rank patterns (Op), PLS regression with NIPALS method (PLS) and SVR with RBF kernel

(SVR) were implemented for comparison. All the settings were the same except for the optimization scheme in the proposed algorithm. The Op method was optimized by Quasi-Newton method with BFGS approximation because it can be considered as unconstrained optimization by removing rank patterns in Equation (6.9). In the implementation, 25 latent variables were used for PLS and LIBSVM [130] was employed for SVR with default parameters. For reducing the influence of other factors, ranking was directly determined by the responses of Equation (6.12) instead of using aggregation with probabilities in this experiment. The CS and MAE results are shown in Figure 6.13 and Figure 6.14. Since OpRank has the highest CS percentages and lowest MAE values for all CS and MAE results, it outperforms all the methods. OpRank can be better than Op because of the rank patterns. The rank patterns, which make use of the comparative concepts in [5-7], are applied to model the comparison pattern with different ranks and regularize the ranking optimization. If the ranking optimization is used without the rank pattern, the internal relationship among different ranks is not fully utilized. On the other hand, the performance of Op is close to PLS in several cases because Op is the derivation of PLS without using latent variable.

Figure 6.13 CS curves and MAEs of the proposed OpRank with rank patterns and other methods for (a-b) height, (c-d) BMI, and (e-f) age estimations from the front leg database. One unit on the x-axes of the first column and y-axes of the second column represents 2cm for height, 2kg/m² for BMI and 2 years for age. The error level of CS curves means tolerance of error. If error level is 2 in CS curves, the real error tolerance is ± 4 cm, ± 4 kg/m² and ± 4 years for height, BMI and age.

Figure 6.14 CS curves and MAEs of proposed OpRank with rank patterns and other methods for (a-b) height, (c-d) BMI, and (e-f) age estimations from side leg database. One unit on the x-axes of the first column and y-axes of the second column represents 2cm for height, 2kg/m² for BMI and 2 years for age. The error level of CS curves means tolerance of error. If error level is 2 in CS curves, the real error tolerance is ± 4 cm, ± 4 kg/m² and ± 4 years for height, BMI and age.

In Appendix D, Figures D.6 and D.7 show the front and side leg images whose CS error levels are more than 2 in height, BMI, or age estimation. In height and BMI estimation, Figures D.6 (a-b) show that the viewpoint is very different from different front leg images and Figures D.7 (a-b) show that the occlusion problem of upper and lower leg parts is serious from different side leg images. Since the height and BMI estimation rely on the leg geometry information, the proposed algorithm is not invariant to the viewpoint and occlusion problems. In order to tackle this problem, training images can include the leg images with different viewpoints and occlusions, but the data collection is required for obtaining more images.

In the age estimation, Figures D.6(c) and D.7(c) show that the leg images with large errors have a lot of hair. Since the proposed algorithm includes the texture of leg skin for age estimation, the occlusion of leg skin can affect the performance of the proposed algorithm. Therefore, hairy leg is a problem for age estimation because the leg skin information is covered by hair. To handle this problem, one solution is to make weighting for different features. When the hair density is high for some legs, the weighting of texture feature is reduced. However, the hair density classifier has been developed first and it is left for further development.

6.4 Summary

Nowadays, child sexual abuse cases and voyeurism become more frequent and serious in the world. The criminal and victim identification is necessary to deal with these cases effectively. Soft biometrics, which can be close to human descriptions [124-126], provide useful clues for enforcement agents, especially when the biometric identifiers of the

criminal or the victim have not been recorded. Its importance cannot be ignored for criminal and victim identification.

In this chapter, a novel approach is proposed to estimate soft biometric features (height, BMI, and age) by using OpRank algorithm, which is a probabilistic model to optimize the correlation between the ranks of soft biometrics and the comparative features with the regularization of the ranking information, and it is combined with the existing works to improve the performance of soft biometrics and identification on leg images. The proposed approach was tested on front leg and side leg database, which contain more than 2000 leg images in total. From the experimental results, it can be seen that the estimated features are useful for enhancing the soft biometrics and identification on leg images. Furthermore, the proposed OpRank algorithm is shown to outperform the evaluated ranking methods. These experimental results indicate that the estimation of soft biometrics in leg images by comparative features has potential to use on criminal and victim identification.

Chapter 7

Proposed Identification Algorithm for Comparison with Soft and Hard Biometric Traits

7.1 Overview

Skin texture is an important topic in many research areas, including development of realistic 3D human models in computer graphic and computer-assisted diagnosis of skin disorders in dermatology. Skin texture is suitable for low resolution images and is usually captured in evidence images, such as riot, terrorist activities, and child sexual abuse images. Figure 7.1 gives examples of these images. Note that images obtained from surveillance cameras are not within the scope of this chapter.

Recently, some research results demonstrate that blood vessels hidden in color images can be visualized [136-138]. Watanabe et al. [137], Song et al. [138], and Tang et al. [136] target respectively on commercial biometric, medical, and forensic applications. Figure 7.2 shows some ¹blood vessels visualized from images collected from the Internet [136]. Furthermore, not all images taken by consumer cameras have vascular information for visualization. The quality of the visualized blood vessels highly depends on illuminant,

¹ In this chapter, the term blood vessel, instead of vein, is used because some arteries, such as ulnar artery are close to skin surface and they may be visualized by the methods [136-138].

concentration of fat and melanosomes of the subjects, resolution, poses, and viewpoints. All these factors are not controllable in forensic environments. Figure 7.3 shows images collected from the Internet without clear blood vessels. Thus, development of skin texture is essential. Figure 7.4 shows skin texture from different forearms and thighs.

Figure 7.1 Examples of terrorist and rioter images. (References can be found in Appendix B)

Figure 7.2 Blood vessels visualized from images collected from the Internet [136].

Figure 7.3 Images without clear blood vessels by using the method [136].

The aim of this chapter is to evaluate the distinctiveness of body skin texture in low resolution images for forensic investigation. To study its distinctiveness in low resolution

images, an algorithm, which is composed of a positive sample generation scheme, dynamic and directional grids, a large feature set generation scheme, and partial least square regression, is proposed. Over 6000 images collected from 869 inner forearms and 753 inner thighs are employed in this evaluation. In this study, blood vessels, which is considered as a hard biometric trait, from the same body sites are used to compare skin texture.

Before presenting the technical details, the difference between this study and the previous studies are highlighted [134-135]. Though Cule et al. and Xie et al. studied skin texture for personal identification before, none of them investigated low resolution skin texture for forensic applications. According to our best knowledge, no similar study has been performed. Table 7.1 summarizes the difference between Cule et al.'s, Xie et al.'s and this study. Figure 7.5 compares sample images used in these studies.

Figure 7.4 Images skin texture from different forearms and thighs. (a) Forearms and (b) thighs.

The rest of the chapter is organized as follows. Section 7.2 presents the testing databases. Section 7.3 describes the algorithm designed for this study. Section 7.4 reports the experimental results, including a comparison between the proposed algorithm and the state-of-the-art texture recognition methods, a comparison between skin texture and blood vessels from color and near infrared images, and a comparison between skin texture of low and very low resolution images. Section 7.5 provides a discussion of the findings and their impacts.

Table 7.1 Summary of the difference between the previous studies and this study.

	Cule et al.'s study [134]	Xie et al.'s study [135]	This study
Target applications	Mainly for skin modelling, but the proposed model can be applied to personal identification.	Commercial and governmental applications	Forensic applications
Body sites	Forehead, cheek, chin and nose	hand back	Inner forearm and thigh
Number of subjects	20	80	474 (forearm) and 400 (thigh)
Imaging environments	A clinical setting with control on light directions and camera directions.	A tailor-made system with pegs for positioning and circular LED for capturing high quality images. The images were taken in a semi-closed environment.	DSRL cameras were used to collect a part of the testing images. The others were collected from the Internet, no pose, viewpoint and illumination control.
Resolution	Very high resolution, but the resolution in dpi was not given.	450 dpi	6.63 ~ 30.57 dpi

Figure 7.5 Skin texture sample images from different studies. (a) Cula et al.'s study [142] (b) Xie et al.'s study (shared by the first author in the paper [135]), and (c) this study.

7.2 Testing database

Six image datasets, NTU color inner forearm datasets, NTU near infrared (NIR) inner forearm dataset, NTU color inner thigh dataset, NTU NIR inner thigh dataset, an Internet inner forearm dataset, and an Internet inner thigh dataset are used to perform this study. Inner forearms and inner thighs are selected for this study because in general, they do not have a lot of hair; inner forearms can be captured in terrorist and riot images and inner thighs always appear in sexual offense images. Except for the images in the two NIR datasets, all the other images were collected by consumer cameras. For the sake of presentation convenience, the term inner in the dataset names will not be used in the following sections as long as the message is clear. The NIR images were collected by a near infrared camera. Images in the NTU color forearm (thigh) datasets have corresponding images in the NTU NIR forearm (thigh) dataset. These corresponding images were collected almost at the same time from the same forearms (thighs). Each NTU dataset was collected in two sessions, whose interval is 14.53 days on average. The NIR datasets are

used to compare skin texture and blood vessels. Figure 7.6 shows sample images in the four NTU datasets. The NTU color forearm and thigh datasets were respectively collected from 344 and 272 subjects, who are mainly Chinese, Malay, Indian, Vietnamese, and Caucasian in Singapore. 298 and 234 subjects offered forearm and thigh images are males. Totally, 656 forearms and 524 thighs were photographed. Each forearm (thigh) has two images in NTU color forearm (thigh) dataset and two images in NTU NIR forearm (thigh) dataset. One was collected in the first session and the other was collected in the second session. These four datasets contain totally 4720 images. No strict restriction on pose was imposed on the imaging process. To increase the sizes of the gallery sets in the experiments, the left forearm and thigh images were flipped and considered as right forearm and thigh images from different subjects. In the experiments, all the segmented forearm images were resized to 142×298 pixels and 35×74 pixels and all the segmented thigh images were resized to 142×298 pixels and 35×74 pixels to evaluate skin texture in low resolution images. Their resolutions in average correspond to 30.57, 7.51, 26.81, and 6.63 dpi. The corresponding distances between subjects and the camera are 9.72m, 39.56m, 11.08m, and 44.80m by using a Canon EOS 500D camera with a focal length of 55mm and a sensor pixel size of $2.2107 \times 10^{-5} \text{ mm}^2$. The distances are obtained by the equation, $d = f \times (25.4 \times h_0)/h_{0s}$, where f is the focal length, h_0 is the object height in inch, $h_{0s} = p_0 \times \sqrt{\rho}$ is the height of an object on sensor in mm, p_0 is the image height of the object in pixels, p_0/h_0 is the resolution in dpi (dot per inch) and ρ is the sensor pixel size in mm^2 . Note that one inch is equal to 25.4 mm.

The Internet datasets containing images collected from the Internet are much more challenging because they have problems of blur, shadows, out of focus, and compression,

and also viewpoint, pose, illumination, and scaling variations. Figures 7.7 and 7.8 show sample images. The left forearm and thigh images were also flipped and considered as right forearm and thigh images from different subjects. In the experiments, 213 and 640 forearm images collected from the Internet were put into the probe and gallery sets, respectively and 229 and 504 thigh images collected from the Internet were put into the probe and gallery sets, respectively. The original sizes of these images range from 35×50 pixels to 903×1150 pixels, but their original resolution in dpi is not known. To increase the sizes of the gallery sets, the gallery images of the NTU color forearm and thigh datasets were also put into the gallery sets. Thus, the probe and gallery sets for testing the Internet forearm images contain 213 and 1296 images respectively and the probe and gallery sets for testing the Internet thigh images contain 229 and 1028 images. In total, the forearm and thigh datasets contain respectively 1509 images and 1257 images, which were taken from 869 forearms and 753 thighs. These datasets are called Internet inner forearm dataset and Internet inner thigh dataset. All the images were resized to 142×298 and 35×74 pixels to evaluate skin texture in low resolution images. Table 7.2 summarizes the Internet datasets.

Figure 7.6 Sample images in the four NTU datasets. (a) NTU color forearm dataset, (b) NTU color thigh dataset, (c) NTU NIR forearm dataset, and (d) NTU NIR thigh dataset. The relative scales are retained for comparison within the same subfigure.

Table 7.2 The details of the Internet datasets.

Number of images	Number of forearms in the probe set	Number of forearms in the gallery set	Number of thighs in the probe set	Number of thighs in the gallery set
1	213	658	229	529
2	0	76	0	176
3	0	64	0	45
4	0	61	0	3
5	0	10	0	0

Figure 7.7 Examples from the Internet forearm dataset. The images in the probe set and some of their corresponding images in the gallery set are shown in (a) and (b). The images from the same forearm are marked with the same number in (a) and (b). Relative scales are retained for comparison.

Figure 7.8 Examples from the Internet thigh dataset. The images in the probe set and some of their corresponding images in the gallery set are shown in (a) and (b). The images from the same thigh are marked with the same number in (a) and (b). Relative scales are retained for comparison.

7.3 A Skin Texture Matching Algorithm

Before presenting the proposed skin texture matching algorithm for this study, two properties related to skin texture and forensic application should be discussed first. Skin texture on different locations of the same body site is not the same, illustrated in Figure 7.9. In forensic application, image acquisition is uncontrollable. The previous skin texture studies [134-135] relied on an accurate image acquisition process to collect high quality

images and neglected the local difference within skin texture. When designing the skin texture matching algorithm, these two properties have been taken into account. Figure 7.10 gives a schematic diagram to illustrate the proposed algorithm, which uses body boundaries for alignment, a positive sample generation scheme for robustness enhancement, multi-grid systems for high dimensional feature vector extraction, and a linear classifier for final classification. An image in the training set is first selected as a model image. Except for the model image and the images from the same body site, the rest training images are regarded as negative samples. Since the negative samples outnumber the positive samples, extra positive samples are generated by the positive sample generation scheme. All the samples, except for the samples generated by the scheme, are aligned to the model image. Then, high dimensional feature vectors are computed from all the samples and used to train a classifier. In the testing phase, testing images are first aligned with the model image. Then, features are extracted from the aligned images and submitted to the trained classifier.

Figure 7.9 Skin texture from different locations of the same (a) forearms and (b) thighs. The skin textures in the same row are from the same person, but skin textures in different rows are from different persons.

Figure 7.10 A schematic diagram of the proposed skin texture matching algorithm.

7.3.1 Alignment and Positive Sample Generation

This study assumes that body sites in images, e.g. inner forearms and inner thighs, have been segmented. In forensic applications, gallery sets can be formed by ex-prisoners, who have higher chances committing another offense after several years of their release [143] and other suspects arrested by police before. These gallery images are collected in a controlled environment under supervision of law enforcement officers. Thus, they can be easily segmented [144]. For probe images collected from crime scenes, they are generally processed semi-automatically or manually even in the current fingerprint identification process.

This alignment and positive sample generation scheme are the same as Section 5.3.1 and 5.3.2. In the alignment, the boundaries of body sites in the forearm and thigh images are sampled. In the positive sample generation scheme, images from the same body site of the model image are considered as positive samples and the rest of the training images are considered negative samples. In the experiments, 1080, 1080, 2880, and 2880 images were generated respectively for each model image for NTU color forearm, NTU color thigh, Internet forearm, and Internet thigh datasets. Since the body boundaries from

the Internet images are not reliable, the image alignment is not performed, but more images are generated for the Internet datasets to handle the change of viewpoint.

7.3.2 Feature Extraction Based on Two Gridding Systems and Classification

After alignment and positive sample generation, features are extracted and classified. By following Section 5.3.3, the Gabor and LBP features are extracted from dynamic grid system, which makes use of blocks, whose sizes are based on the number of horizontal pixels of thigh (vertical pixels of forearm) (see Figure 7.10(b)), and directional grid system, which applies larger blocks in different directions (see Figures. 7.10(c)-(n)). The extracted Gabor and LBP features form a skin texture feature vector in a high dimensional space. Note that these features are extracted from the R, G, and B channels because red, green, and blue light have different penetration capability and capture different information from skin.

Figure 7.12 shows different information in the R, G, and B channels. The dimension of the feature vector is 57118. Because it is already in a high dimensional space, kernel methods, which project feature vectors to a high dimensional space for classification, are not necessary and linear classifiers are suitable. Partial least squares (PLS) regression is used in the current implementation. One PLS model is trained for one model image. The PLS model providing the highest response is regarded as a match.

Figure 7.11 Illustration of the gridding systems. (a) An original image, (b) the dynamic grid system, and (c)-(n) the directional grid system with $\theta = \{0^\circ, 30^\circ, \dots, 330^\circ\}$, $N_H = 5$, and $N_W = 8$.

Figure 7.12 Skin texture samples: (a) original images, (b) red channels, (c) green channels, and (d) blue channels

7.4 Experimental Results

To evaluate skin texture in low resolution images, blood vessels hidden in color images visualized by Watanabe et al.'s method [137], Song et al.'s method [138], and Tang et al.'s method [136] are employed for comparison. Blood vessels are selected for comparison because according to our best knowledge, no one studied low resolution body skin texture before and the previous results show that blood vessels can be visualized from high resolution color images for personal identification [136-138]. It should be emphasized that blood vessels are generally considered as a hard biometric trait, but skin texture is generally considered as a soft biometric trait. The image sizes in evaluating skin texture are 142×298 pixels for forearm and thigh. For Watanabe et al.'s method, Song et al.'s method, and Tang et al.'s method, the size of the testing images is 284×596 pixels because blood vessels can only be visualized from high resolution images. Furthermore, skin texture in RGB images and blood vessels in NIR images are compared to evaluate their distinctiveness and the proposed algorithm is compared with three state-of-the-art texture recognition methods, dominant LBP (DLBP) [146], completed LBP (CLBP) [147], and the texton learning method using the k-means clustering (TL_KM) [148], which were used as benchmarking methods in Xie et al.'s study [135].

Some parameters have to set for the blood vessel visualization methods. Watanabe et al. use

$$b_o = b_i^\alpha, \quad (7.1)$$

where b_o , b_i , and α are respectively an input of the saturation component, an output of the saturation component and an exponent parameter, to visualize blood vessels. α is set to 0.3,

which is estimated from Figure 3 in Watanabe et al.'s paper [137]. For Song et al.'s method, the corresponding vector of the wavelength 650 nm mentioned in Table II of their paper [138] is used because the corresponding vector of the wavelength 620 nm is not provided and the closest wavelength given in the table is 650 nm. Song et al.'s method was proposed for medical applications and therefore, it does not include a matching scheme. The blood vessel matching scheme proposed by Tang et al. [136] is applied to match their blood vessel images. For Tang et al.'s method, the parameter setting is the same as that of the original paper [136]. For the proposed skin texture matching algorithm, the pixel coordinates of image are normalized to $[-1 \ 1]$ when the extra positive samples are generated by using affine transformation with parameters, $s \in \{1.2, 1.1, 1.0, 0.9, 0.8\}$, $\alpha \in \{-10^\circ, -6^\circ, -2^\circ, 2^\circ, 6^\circ, 10^\circ\}$, $t_x \in \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$, $t_y \in \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$. The parameters are sampled regularly to ensure that the number of positive samples, including the original positive samples, the generated samples, and negative samples are balanced. The Gabor features are extracted using 8 orientations, 4 scales (0.2 0.5 0.7 0.9), spatial aspect ratio $s = 2$, bandwidth $\phi = 1.5$ in octaves, and radial frequency $\omega_0 = 1$ in radians per unit length. In dynamic grid system, $E = 16$ and 108 blocks in total are applied. The directional grid system with $\theta = \{0^\circ, 30^\circ \dots, 330^\circ\}$, $N_H = 5$, and $N_W = 8$, is applied. For classification, partial least squares (PLS) regression with 25 latent variables is used. Cumulative Match Characteristic (CMC) curves and matching accuracy in different ranks, which are common approaches to evaluate biometric methods for forensic applications, are employed as performance indexes.

The first experiment was designed to compare skin texture and blood vessels hidden in color images. The proposed algorithm was used for matching skin texture in the red,

green, and blue channels and Watanabe et al.'s, Song et al.'s, and Tang et al.'s methods [136-138] were used for visualizing and matching blood vessels. NTU color forearm and thigh datasets were used as the testing datasets in this experiment. Both forearm probe and gallery sets contained 656 images and both thigh probe and gallery sets contained 524 images. The images in the probe and gallery sets were collected in different sessions. Two-fold cross validation was applied to this experiment because the probe sets and the gallery sets have similar quality. The probe sets were used as testing sets and the gallery sets were used as training sets in the first fold and their roles were exchanged in the second fold. The CMC curves are given in Figure 7.13, where the identification rate in different ranks is the average accuracy of the two folds. Figure 7.13 indicates clearly that skin texture processed by the proposed algorithm outperforms inner forearm and inner thigh blood vessels visualized by the three methods with very great margins. To show the performance difference more clearly, identification rates of rank-1, rank-10, and rank-20 are recorded in Table 7.3. In terms of rank-1 accuracy, the proposed skin texture matching algorithm outperforms Tang et al.'s method, which is the second best method in this comparison, by almost 40% for inner forearms and more than 50% for inner thighs. These results clearly show that skin texture is effective for identification if a suitable matching algorithm is used. Watanabe et al.'s method is the worst possibly because its blood vessel visualization process is mainly based on Equation (7.1). The difference between Tang et al.'s method and Song et al.'s method is the visualization schemes because they employed the same matching method in the experiment. Tang et al.'s method outperforms Song et al.'s method because Tang et al. used multiple models to extract information from images with different properties and optimization schemes to adjust parameters for each image. The thigh images

are more challenging and the performance difference between skin texture and blood vessels increases (Figure 7.13(b)). The proposed skin texture recognition method outperforms all the blood vessel visualization methods significantly.

Figure 7.13 The CMC curves from (a) NTU color forearm dataset and (b) NTU color thigh dataset.

Table 7.3 Experimental results on (a) NTU color forearm dataset and (b) NTU color thigh dataset.

(a)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
Tang et al.'s method	53.81%	70.27%	74.54%
Song et al.'s method	38.95%	58.84%	66.01%
Watanabe et al.'s method	8.61%	22.79%	30.49%
Skin texture algorithm	92.15%	97.18%	98.09%

(b)

	Rank-1 accuracy	Rank-10 accuracy	Rank-20 accuracy
Tang et al.'s method	15.84%	32.92%	40.46%
Song et al.'s method	5.06%	17.56%	24.05%
Watanabe et al.' method	1.91%	7.73%	13.26%
Skin texture algorithm	68.89%	87.02%	89.98%

The second experiment was to test the performance of the four methods on the Internet datasets, whose images have viewpoint, scaling, illumination, pose, blurry, out of focus, and shadow problems. These images also have occlusion problem and therefore, the body boundaries are not a stable feature for alignment. The alignment step based on body boundaries in the skin texture matching algorithm was skipped. To compensate it, the positive sample generation scheme used a large parameter set, $\alpha \in \{-30^\circ, -26^\circ, -22^\circ, -18^\circ, -14^\circ, -10^\circ, -6^\circ, -2^\circ, 2^\circ, 6^\circ, 10^\circ, 14^\circ, 18^\circ, 22^\circ, 26^\circ, 30^\circ\}$, $s \in \{1.2, 1.1, 1.0, 0.9, 0.8\}$, $t_x \in \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$, $t_y \in \{-0.1, -0.06, -0.02, 0.02, 0.06, 0.1\}$, to generate more distorted samples and increase robustness. The sizes of the testing images in this experiment were the same as the previous one. Figure 7.14 shows the resultant CMC curves and Table 7.4 lists the corresponding rank-1, rank-20, and rank-50 accuracy. As with the previous experiment, the skin texture matching algorithm outperforms all the blood vessel visualization methods. In terms of rank-1 accuracy, the skin texture matching algorithm outperforms the blood vessel visualization methods by 40% for Internet forearm dataset and by 44% for Internet thigh dataset. At the rank-50, their performance difference becomes 38% for Internet forearm dataset and 70% for Internet thigh dataset. These results show clearly that skin texture is much more effective for images taken in uncontrolled environments, comparing with blood vessels extracted from color images. In uncontrolled environments, it is hard to guarantee that images contain vascular information because this hidden information is hard to be captured in weak lighting environments, skin with a high concentration of fat or melanin and non-ideal viewpoints. Inner forearm skin is thin and therefore, Tang et al.'s and Song et al.'s methods still provide some useful matching results at the rank-50. For inner thighs, where

skin is thick, the three blood vessel visualization methods almost do not provide any useful matching results. Figure 7.14 and Table 7.4 once again show that skin texture processed by the proposed algorithm performs significantly better than blood vessels extracted by the visualization methods. Figure 7.15 provides some retrieval results given by the proposed skin texture matching algorithm.

Figure 7.14 The CMC curves from (a) Internet forearm dataset and (b) Internet thigh dataset.

Table 7.4 A summary of the experimental results. (a) Internet forearm dataset and (b) Internet thigh dataset.

(a)

	Rank-1 accuracy	Rank-20 accuracy	Rank-50 accuracy
Tang et al.'s method	5.63%	23.94%	31.46%
Song et al.'s method	7.98%	30.52%	43.66%
Watanabe et al.'s method	3.29%	8.45%	24.41%
Skin texture algorithm	48.36%	77.46%	81.69%

(b)

	Rank-1 accuracy	Rank-20 accuracy	Rank-50 accuracy
Tang et al.'s method	0.87%	4.80%	10.48%
Song et al.'s method	0.44%	5.24%	6.99%
Watanabe et al.'s method	0.00%	3.06%	5.24%
Skin texture algorithm	45.85%	73.80%	81.22%

Figure 7.15 Top ten retrievals in a search of (a) inner forearms and (b) inner thighs given by the skin texture matching algorithm. Their relative scales are retained for comparison. The query images and the matched ones are highlighted in white and red, respectively. For example, the matched one of the first probe image is the image at the rank 1. More examples are shown in Appendix C.

The previous experimental results indicate that skin texture processed by the proposed algorithm is more distinctive than blood vessels in color images processed by the visualization methods. However, we cannot conclude that skin texture is more distinctive than blood vessels completely because visible light is not an ideal spectrum to capture blood vessels and the matching algorithms are different. In this experiment, the proposed algorithm was applied to skin texture in color images and blood vessels in NIR images. Since the same algorithm was applied and the color and NIR images were taken from the same body sites, the performance of these two biometric traits can be directly compared. Note that near infrared is a standard spectrum to take blood vessel images for commercial biometric and medical applications. The NTU color forearm and thigh datasets were used to examine skin texture and the NTU NIR forearm and thigh datasets were used to examine blood vessels. The Internet datasets were not used in this experiment because the corresponding NIR images are not available. The sizes of the probe sets and the gallery sets were the same as those in the first experiment. The color and NIR images were captured by a digital camera and a NIR camera at almost the same time from the same body site. Thus, the numbers of images in the color and NIR datasets were the same. Skin texture in the color images and blood vessels in the NIR images are respectively extracted from the RGB and grayscale channels. The resultant CMC curves given in Figure 7.16 demonstrate that skin texture from color images outperforms blood vessels from NIR images significantly. In terms of rank-1 accuracy, skin texture outperforms blood vessels by 6.86% for the inner forearm datasets and by 18.71% for the inner thigh datasets. These experimental results match the previous experimental results. However, they may be out of some people's expectation because blood vessels are generally considered as a hard

biometric trait and skin texture is considered as a soft biometric trait. These experimental results indicate that the distinctiveness of body skin texture was underestimated.

Figure 7.16 Comparison of skin texture from color images and blood vessels from NIR images. (a) NTU color and NIR forearm datasets and (b) NTU color and NIR thigh datasets.

The fourth experiment was to test the robustness of the skin texture matching algorithm for low resolution images. The images tested in the first experiment were resized to 35×74 pixels, which is about a quarter of the original size (142×298 pixels). The low resolution images were up-sized to the original size. Thus, all the algorithmic parameters were remained the same. Figures 7.17(a)-(d) give the CMC curves from NTU color forearm, NTU color thigh, Internet forearm, and Internet thigh datasets. Even though the image size was significantly reduced, the performance did not drop dramatically. These results show clearly that skin texture is effective even for lower resolution images. It should be emphasized that blood vessels cannot be extracted from images in this resolution.

Figure 7.17 The comparisons between two image resolutions (142×298 pixels and 35×74 pixels) from (a) NTU color forearm dataset, (b) NTU color thigh dataset, (c) Internet forearm dataset, and (d) Internet thigh dataset.

The last experiment was to compare the proposed algorithm with the state-of-the-art texture recognition methods, DLBP [146], CLBP [147], and TL_KM [148]. These methods were used as benchmarking methods in Xie et al.'s study [135]. The feature vectors of those texture recognition methods were formed by considering the whole body site image as a texture region, as with Xie et al.'s implantation. Using the DLBP implementation in [149], different features, including $LBP_{8,1}^{ri}$, $LBP_{8,1}^{riu2}$, $LBP_{16,2}^{ri}$, $LBP_{16,2}^{riu2}$, $LTP_{8,1}^{ri}$, $LTP_{8,1}^{riu2}$, $LTP_{16,2}^{ri}$, and $LTP_{16,2}^{riu2}$ and 1-NN classification method were compared. For CLBP, the implementation given in [150] was used. It utilized CLBP_S, CLBP_M, CLBP_C, CLBP_M/C, CLBP_S_M/C, CLBP_S/M, and

CLBP_S/M/C as features and Chi-square distance as a matching function. For TL_KM, features, including LM, MR8, and S and Chi-square distance were applied [151]. For each feature used by TL_KM, a dictionary was formed by k -means clustering with the parameters used in [151]. The details of these features and implementation can be found in [149-151]. Figures 7.18(a)-(b) show CMC curves obtained from the proposed algorithm and the state-of-the-art texture methods. For NTU color forearm and thigh datasets, the result of CLBP_M is reported because it is the best feature for CLBP. $LBP_{8,1}^{ri}$ and $LTP_{8,1}^{ri}$ are reported separately because they are the best DLBP features for NTU color forearm and thigh datasets, respectively. For TL_KM, all features are reported. Figure 7.18 shows clearly that the proposed algorithm outperforms the texture recognition methods on both NTU color forearm and thigh datasets with significant margins. From these experimental results, it not only shows that the proposed algorithm significantly outperforms the existing works mentioned in this chapter, but it also demonstrates that the proposed algorithm is more robust than these methods to the images with different resolutions, viewpoints, poses, illuminations, and scaling factors. These results encourage further study on skin texture for forensic investigation.

Figure 7.18 Comparison of the proposed algorithm and the state-of-the-art texture recognition methods. (a) NTU color forearm dataset and (b) NTU color thigh dataset, (c) Internet forearm dataset, and (d) Internet thigh dataset.

7.5 Discussion

In this study, the skin texture matching algorithm is proposed to extract skin texture information and skin texture and blood vessels are empirically compared for studying the distinctiveness of skin texture. Some may expect that uniqueness models should be developed for this study [36,152-153]. In general, uniqueness models are based on a predefined feature representation, e.g. minutiae for fingerprints, bits for irises and points

for skin marks. Body skin texture in low resolution images is a new biometric trait and not a widely used representation and therefore, uniqueness models are not considered.

Comparing with the previous skin texture studies [134-135], the testing datasets in this study containing 6306 images from 1622 different inner forearms and inner thighs are larger than Cula et al.'s database with 2560 images from 20 faces and Xie et al.'s database with 1920 images from 160 hands. The database size in this chapter is also comparable with other public biometric databases, such as NIST tattoo detection database with 2349 images [145], UBIRIS.v1 iris database collected from 241 irises [104], CMU keystroke dynamic database collected from 51 subjects [105] and CASIS multi-spectral palmprint database collected from 100 subjects [106]. To further study skin texture for forensic investigation, more images will be collected for algorithm development and evaluation.

In forensic cases, suspect databases are not always very large because other information can be used to reduce their sizes. Taking terrorist identification as an example, many countries have a list of terrorist suspects, who are friends or relatives of confirmed terrorists or posed terror messages online. Thus, a large number of other suspects can be ruled out. In some cases, it is in fact one-to-one verification [116].

Though skin texture processed by the proposed algorithm outperforms blood vessels from color and NIR images and the proposed algorithm performs better than the state-of-the-art texture recognition methods [146-148], some may still think that the reported accuracy, in particular that from the Internet image databases, is low. It should be emphasized that in forensic applications, accuracy is usually low even for mature biometric traits, such as fingerprint because of low quality signals collected in uncontrolled and uncooperative environments. Paulina et al. reported that VerFinger, a commercial

fingerprint matcher, achieves accuracy of 30.6% for NIST SD27 latent fingerprints with ugly quality [108]. In general, forensic investigation uses all available biometric traits and law enforcement officers would not solely rely on the top 1 suspect returned by biometric systems. Hard and soft biometric traits are always used together and suspects at the top of the list, e.g. top 30 suspects, will be further investigated. The proposed skin texture matching algorithm achieves rank-30 accuracy of 79.34% for the Internet forearm dataset and rank-30 accuracy of 77.73% for the Internet thigh dataset. This accuracy demonstrates the potential use of skin texture for forensic investigation.

The proposed algorithm is designed based on several principles and the properties of low resolution body skin texture. Skin texture on different locations even on the same body site of the same person is different. RGB channels have different information because red, green, and blue light have different penetration capability. Thus, local features are extracted from the three channels, which are different from the previous studies and body boundaries are used when they are reliable. To perform classification effectively, the algorithm uses the dynamic and directional grid systems to obtain feature vectors in a high dimensional space. Because the feature vectors are already in a high dimensional space, it is not necessary to use kernel methods to project them to another higher dimensional space and linear classifier is a suitable choice. Though PLS is selected in the current implementation, it does not mean that it is the best classifier. However, the experimental results demonstrate that it is a suitable choice for this study. With the recent development of incremental partial least squares method [107], the proposed algorithm is scalable.

The contributions of different components in the proposed algorithm to the matching performance are not reported in this chapter because the aim of this chapter is to

study distinctiveness of body skin texture. The extra information about contributions of different components would distract the focus of this chapter.

Based on the promising results in this chapter, some may think that skin texture is capable of taking over the existing biometric traits, like patterns, blood vessel pattern, and pigmented and vascular skin marks. Even though the experiments show the promising results in skin texture, androgenic hair patterns, blood vessel pattern, and pigmented and vascular skin marks are still needed for further exploration because of the limitations of skin texture. For androgenic hair patterns, it is useful to identify people when they have hairy body parts, whose skin texture is covered by a lot of hair. Since skin texture is not observed, it is preferred to make use of androgenic hair patterns. Also, the dataset can contain hairy and non-hairy body parts in the real application, so androgenic hair patterns can be applied with skin texture for enhancing the matching result.

For blood vessel pattern, it is still be used for medical and forensic applications. For medical applications, visualizing blood vessels from smart phones is a cost-effective approach comparing with tailor-made medical devices based on NIR. For forensic applications, once complete blood vessel patterns are visualized, they can be presented in courts as evidence against or for suspects because comparing clear blood vessel patterns is easier than comparing skin texture for human vision. In these two applications, blood vessel pattern can still be useful.

For pigmented and vascular skin marks, it is used for identification in the high resolution images. Although it is not universal and not robust to hairy body parts and resolution, Nurhudatiana et al. [37] suggested the skin marks fused with vein patterns to

improve vein identification. Thus, they can be combined with skin texture for further improvement on the high resolution images.

In Appendix D, Figures D.8 and D.9 show 20 failure cases on experiments of NTU color forearm dataset and NTU color thigh dataset. In Figure D.8, the illumination and viewpoint problem exist in the wrong cases. For the illumination problem, uneven illumination causes some parts of skin region to be dark and some to be bright. Since the proposed algorithm makes use of skin texture for training, the local change of skin color can affect the distinctiveness of its texture feature, especially on uneven illumination. For the viewpoint problem, some probe and gallery images cannot be matched when they are captured from two viewpoints. Since the proposed algorithm makes use of affine transformation to handle the change of viewpoints, it is not sufficient to deal with the prospective transformation from different viewpoints. The general solution is to use 3D model for generating different viewpoints, but it requires data collection for further development. In Figure D.9, the occlusion problem is a main issue for matching thigh images. Since the upper part of thigh can be partially occluded by shorts, the thigh images cannot be aligned correctly using thigh boundaries. In addition, its boundaries can be changed due to different poses. Thus, the occlusion problem causes incomplete alignment of thigh images and the probe and gallery images cannot be matched well.

In addition, some may suspect that the female skin is easier to have wrong matching result than male skin. In the experiment of NTU color forearm dataset, wrong matched results come from 6.22% males and 17.74% females. In experiments of NTU color thigh dataset, wrong matched results comes from 31.87% males and 27.47% females. Since female skin is expected to be smoother, its skin feature may not be discriminative.

However, it is surprising that there are more failure cases on thigh images of males. Since thigh images have occlusion problem from shorts and females tend to wear shorter pant than males, wrong matched results of thigh images can not only be caused by gender, but it can also be caused by occlusion problem.

7.6 Summary

Traditionally, skin texture is regarded as a soft biometric trait and not considered as a distinctive biometric trait, especially in low resolution images. To study its distinctiveness, a skin texture matching algorithm, which includes an alignment scheme based on body boundaries, a positive sample generation scheme, a feature extraction scheme based on Gabor orientation histograms, LBP, dynamic grid system and directional grid system, and one-against-all PLS models, is proposed. Skin texture is compared with blood vessels from color and NIR images, which are generally regarded as a hard biometric trait. Skin texture in low resolution images outperforms blood vessels in all the experiments. These results challenge the traditional belief that hard biometric traits should be more distinctive than soft biometric traits and expose the potential of skin texture for forensic investigation. The proposed algorithm is also compared with the state-of-the-art texture recognition methods and outperforms them with very significant margins. These experimental results demonstrate the effectiveness of the proposed skin texture matching algorithm.

Chapter 8

Conclusion

This thesis proposes novel techniques for forensic investigations. Their applications include but not limited to child sexual abuse, voyeurism, terrorists, and riots. In addition, their performance is evaluated through different datasets, including Internet images and the experimental results show the feasibility of these novel techniques. In this chapter, the main contributions of this thesis are concluded and directions for future work are suggested.

8.1 Summary of Main Contributions

A summary of the main contributions of this thesis is as follows:

In Chapter 3, new soft biometric traits are proposed.

- The dataset, which contain 1,138 leg images with 283 subjects, was collected and preprocessed as low resolution and high resolution images for evaluation. Since no strict instructions were given and illumination changes were allowed in the data collection, the images have illustration and pose distortions. 575 images from the 283 subjects collected from the first occasion form a probe set and 563 images from the same 283 subjects collected from the second occasion form a gallery set.
- The evaluation on lower leg geometry as a soft biometric trait is performed for criminal and victim identification. A feature set comprised of the partial sum, the polynomial coefficients, and the number of intersection points of aligned leg shapes and three

classification methods are used in this evaluation. The experimental results indicate that lower leg shapes contain quality information as a soft biometric trait and encourage further study on lower legs and full body geometry for forensic investigation.

In Chapter 4, use of soft biometric traits on image alignment is proposed.

- To accurately align androgenic hair patterns in low resolution leg images for criminal and victim identification, an algorithm based on leg geometry, angular sampling, and affine transformation is proposed in that chapter. The experimental results demonstrate that the proposed algorithm outperforms Su et al.'s low resolution leg identification method [4].
- Androgenic hair patterns provide additional information for criminal and victim identification. The feasibility of making use of androgenic hair follicles with leg geometry as features for aligning androgenic hair patterns is evaluated in high resolution images. The experiments show encouraging results.

In Chapter 5, a new algorithm is proposed to increase robustness against viewpoint and pose variations and achieve more accurate identification of androgenic hair patterns.

- A new algorithm, which makes use of leg geometry to align lower leg images, large feature sets (about 60,000 features) extracted through multi-directional gridding systems to increase the discriminative power and robustness and the partial least squares (PLS) method to handle imbalanced training data and perform the multi-grid feature fusion, is proposed.
- To further increase robustness against viewpoint and pose variations, a scheme is designed to generate more positive samples from a single image.

- The experimental results show that the proposed multi-grid feature fusion algorithm outperforms the method suggested by Su et al. [4] with a significant margin of more than 15% of identification accuracy for the both front pose and low quality datasets, which contain 1,493 low resolution leg images with large viewpoint and pose variations from 412 legs. The experimental results also demonstrate that the proposed algorithm is effective for very low resolution images, e.g. 6.25 dpi, which can be easily found in real cases.
- According to my best knowledge, it is the first study of androgenic hair patterns in low resolution images with large viewpoint and pose variations.

In Chapter 6, some soft biometrics, which are not visible from leg images, are estimated by using comparative features.

- The usage of comparative descriptions is extended to non-verbal comparative features extracted from legs for soft biometrics estimation.
- A novel approach is proposed to estimate soft biometric features (e.g. height, BMI, and age) by using the proposed algorithm, which is a probabilistic model to optimize the correlation between the ranks of soft biometrics and the comparative features with the regularization of the ranking information.
- The estimated soft biometrics can be applied to the existing works [4,131] to improve the performance of soft biometrics and identification on leg images.
- The proposed approach was evaluated on front leg and side leg databases, which contain more than 2000 leg images in total. The proposed algorithm outperforms the ranking methods which are involved in the evaluation. The experimental results

indicate that the estimation of soft biometrics in leg images by comparative features are useful in criminal and victim identification.

In Chapter 7, the performance of the existing vein visualization methods and the proposed skin texture recognition algorithm is evaluated on color and NIR images for forensic analysis.

- The existing vein visualization methods and a skin texture recognition algorithm are evaluated on a skin dataset, which contains 1312 inner arm images and 1048 inner thigh image, and an Internet dataset, which contains 1509 inner arm images and 1257 inner thigh images.
- Since the skin texture recognition algorithm can significantly outperform the state-of-the-art vein visualization methods, it indicates that the potential use of skin texture for criminal and victim identification instead of the vein visualization.

8.2 Future Work

As the research area is new, various technical and operational aspects are still not covered. Some topics for future research include tackling the occlusion, viewpoint, and illumination problems on soft biometric estimation and person identification, using the features extracted from leg geometry and androgenic hair follicles for identification, and extending soft biometric estimation to different body parts.

Low quality problems, which are due to low resolution, partial occlusion, change of viewpoint and illumination, are common issues in computer vision. In this study, low

resolution issue exists because a subject is normally captured at a certain distance and low resolution image can save memory space. Partial occlusion happens when some body parts are covered by the clothes. For example, part of the leg is covered by the socks or shorts. Viewpoint issue appears because a subject is not fixed on a certain pose for taking the pictures in the real case. Illumination issue exists because the subject's images can be taken from outdoor or indoor with multiple lighting sources. In Chapter 3, 4, and 6, low quality problems are not specifically addressed. In Chapter 5 and 7, even though the proposed algorithms perform well on low resolution images and are robust to some changes of viewpoint and illumination, further improvement is still necessary when the low quality problems become serious.

To deal with these problems, the individual studies of illumination, partial occlusion, and viewpoint are necessary for androgenic hair patterns, soft biometric traits estimation, and skin texture. In order to achieve these goals, different kinds of dataset for illumination, partial occlusion, and change of viewpoint and illumination have been collected. However, the number of samples is limited in data collection. It is possible to make use of CAD models for generating more samples with different illuminations, partial occlusions, and viewpoints. Based on the new datasets, the algorithm on androgenic hair patterns, soft biometric traits estimation, and skin texture can be enhanced to be robust with different situations.

The features extracted from leg geometry and androgenic hair follicles are possible to be used for identification in Chapter 4. Two soft biometric traits, which are leg geometry and androgenic hair follicles, are introduced for image alignment. In image alignment, leg geometry is extracted as sample points of leg boundary and androgenic hair follicles are

extracted as sample points at the top of the hairs. When they are used for image alignment, their patterns of sample points are effective to align the images. Since they are useful for image alignment, they can be also extracted as features, like curve fitting parameters of leg geometry and density of androgenic hair follicles, for identification. To use them for identification, feature fusion and noise robust feature extraction methods are possible ways for further investigation.

In Chapter 6, the soft biometric estimation can be extended to involve more body parts. Currently, only leg image is used to estimate the soft biometric traits, which cannot be directly observed. Other body parts, like thigh and lower arm, can be included for estimation. There are two possible ways to conduct the study. The first way is to estimate soft biometric traits by individual body part and the second way is to estimate them by combining different body parts.

In Chapter 7, skin texture algorithm is proposed. Even though the corresponding experiments give the promising results, the performance may be further improved by combining with existing biometric traits, including androgenic hair patterns, blood vessel pattern, and pigmented and vascular skin marks. The fusion with them is possible for further study. Also, skin aging is one of the factors to downgrade the performance of the skin texture algorithm. Although skin texture is a soft biometric trait which is lack of permanence, the study of skin texture on aging can provide additional information for improvement of the identification rate on the proposed algorithm.

References

- [1] “Child abuse images – statistics,” *Beyond Borders Au-delà des frontières ECPAT CANADA*, June 2012 [Online].
Available: <http://www.beyondborders.org/wp/wp-content/uploads/Child-Abuse-Images-Statistics.pdf>
- [2] A. Nurhudatiana, A.W.K. Kong, K. Matinpour, S.Y. Cho, and N. Craft, “Fundamental statistics of relatively permanent pigmented or vascular skin marks for criminal and victim identification,” in *Proc. IJCB*, Washington, DC, USA, 2011, pp. 1-6.
- [3] J.E. Lee, A.K. Jain, and R. Jin, “Scars, marks and tattoos (SMT): soft biometric for suspect and victim identification,” in *Proc. Biometrics Symposium*, Tampa, FL, USA, 2008, pp. 1-8.
- [4] H. Su and A.W.K. Kong, “An evaluation on low resolution androgenic hair patterns for criminal and victim identification,” *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 666-680, 2014.
- [5] D.A. Reid and M.S. Nixon, “Using comparative human descriptions for soft biometrics,” in *Proc. IJCB*, Washington, DC, USA, 2011, pp. 1-6.
- [6] D.A. Reid and M.S. Nixon, “Soft biometrics; Human identification using comparative descriptions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1216-1228, 2014.
- [7] D.A. Reid and M.S. Nixon, “Human identification using facial comparative descriptions,” in *Int. Conf. on Biometrics*, Madrid, Spain, 2013, pp. 1-7.
- [8] C. Tang, A.W.K. Kong, and N. Craft, “Uncovering vein patterns from color skin images for forensic analysis,” in *Proc. IEEE CVPR*, Colorado Springs, CO, USA, 2011, pp. 665-672.
- [9] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, “Biometric identification through hand geometry measurements”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no.10, pp. 1168-1171, 2000.
- [10] C. Oden, A. Ercil, and B. Buke, “Combining implicit polynomials and geometric features for hand recognition,” *Pattern Recognition Letters*, vol. 25, no. 13, pp. 2145-2152, 2003.
- [11] “London riots spread as police lose control,” *The Age* Aug. 9, 2011 [Online].

- Available: <http://www.theage.com.au/world/london-riotsspread-as-police-lose-control-20110809-1ijmm.html>
- [12] “Bahrain clears 21 medics of protest charges,” *Yahoo News* Mar. 28, 2013 [Online]. Available: <http://news.yahoo.com/bahrain-clears-21-medics-protest-charges-104533403.html>
 - [13] “Bahrain Air grounded by mounting debts,” *Yahoo News* Feb. 13, 2013 [Online]. Available: <http://news.yahoo.com/bahrain-air-grounded-mounting-debts-073854094--finance.html>
 - [14] “The £25bn bombshell: rioters stone MP after Greece pushes through austerity package,” *Daily Mail* June 30, 2011 [Online]. Available: <http://www.dailymail.co.uk/news/article-2009437/Greek-vote-Rioters-stone-MP-Greece-pushes-austerity-package.html>
 - [15] “Huge protests continue in Brazil — small bands of radicals break into stores and loot,” *Business Insider* Jun. 19, 2013 [Online]. Available: <http://www.businessinsider.com/huge-protests-continue-in-brazil--small-bands-of-radicals-break-into-stores-and-loot-2013-6>
 - [16] “Hand geometry,” *National Science and Technology Council (US)* Aug 7, 2006 [Online]. Available: <http://www.biometrics.gov/documents/handgeometry.pdf>
 - [17] Y. Bulatov, S. Jambawalikar, P. Kumar, and S. Sethia, “Hand recognition system using geometric classifiers,” *DIMACS Workshop on Computational Geometry*, Piscataway, NJ, 2002, pp. 14-15.
 - [18] M. Bača, P. Grd, and T. Fotak, “Basic principles and trends in hand geometry and hand shape biometrics,” in *New Trends and Developments in Biometric*, InTech, 2013, pp. 77-99.
 - [19] A. de-Santos-Sierra, C. Sánchez-Ávila, G. Bailador-del-Pozo, and J. Guerra-Casanova, “Unconstrained and contactless hand geometry biometrics,” *Sensors*, vol. 11, no. 11, pp. 10143-10164, 2011.
 - [20] A. Dantcheva, C. Velardo, and A. D'Angelo, “Bag of soft biometrics for person identification,” *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739-777, 2011.
 - [21] A.K. Jain, S.C. Dass, and K. Nandakumar, “Soft biometric traits for personal recognition systems,” in *Proc. Int. Conf. Biometric Authentication*, Hong Kong, 2004, pp. 731-738.
 - [22] “‘Soft’ biometrics is the new way to monitor people,” *NewScientist* Nov. 28, 2013 [Online].

Available: <http://www.newscientist.com/article/mg22029455.200-soft-biometrics-is-the-new-way-to-monitor-people.html#.UxR8zvmSxK4>

- [23] F. Chen, J. Feng, A.K. Jain, and J. Zhou, "Separating overlapped fingerprints," *TIFS*, vol. 6, no. 2, pp. 346-359, 2011.
- [24] S. Yoon, J. Feng, and A.K. Jain, "Latent fingerprint enhancement via robust orientation field estimation," in *IJCB*, 2011, pp. 1-8.
- [25] D. Manger, "Large-scale tattoo image retrieval," in *9th Conf. on Computer and Robot Vision*, 2012, pp. 454-459.
- [26] B. Heflin, W.J. Scheirer, and T.E. Boulton, "Detecting and classifying scars, marks and tattoos found in wild," in *BTAS*, 2012, pp. 31-38.
- [27] H. Han and A.K. Jain, "Tattoo based identification: sketch to image matching," in *ICB*, 2013, pp. 1-8.
- [28] P.C. Yuen and C.H. Man, "Human face image searching system using sketches," *IEEE Trans. SMC, part A*, vol. 37, no. 4, pp. 495-504, 2007.
- [29] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955-1967, 2009.
- [30] H. Han, B. Klare, K. Bonnen, and A.K. Jain, "Matching composite to face photos: A component based approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 191-204, 2012.
- [31] H. Nejati, L. Zhang, and T. Sim, "Eyewitness face sketch recognition based on two-step bias modeling," in *Computer Analysis of Images and Patterns, part 2*, 2013, pp. 26-33.
- [32] Y. Tang, S.N. Shrihari, H. Kasiviswanathan, and J.J. Corso, "Footwear print retrieval system for real crime scene marks," in *4th Int. Workshop on Computation Forensics*, 2011, pp. 88-100.
- [33] D.E. Nassar, A. Abaza, X. Li and H. Ammar, "Automatic construction of dental charts for postmortem identification," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 2, pp. 234-246, 2008.
- [34] E.H. Said, D.E.M. Nassar, G. Fahmy, and H. Ammar, "Teeth segmentation in digital dental X-ray films using mathematical morphology," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 178-189, 2006.
- [35] F. Iqbal, H. Binsalleeh, B.C.M. Fung, and M. Debbabi, "A unified data mining solution for authorship analysis in anonymous textual communications,"

Information Sciences (INS): Special Issue on Data Mining for Information Security, vol. 231, no. 1, pp. 98-112, 2013.

- [36] A. Nurhudatiana, A.W.K. Kong, K. Matinpour, D. Chon, L. Altieri, S.Y. Cho and N. Craft, "The individuality of relatively permanent pigmented or vascular skin marks (RPPVSM) in independently and uniformly distributed patterns," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 998-1012, 2013.
- [37] A. Nurhudatiana and A.W.K. Kong, "On Criminal Identification in Color Skin Images Using Skin Marks (RPPVSM) and Fusion with Vein Patterns," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 916-931, 2015.
- [38] A. Nurhudatiana, A.W.K. Kong, N. Craft, and H.L. Tey, "Relatively permanent pigmented or vascular skin marks (RPPVSM) for Identification: A Pilot Reliability Study," *Journal of Forensic Sciences*, vol. 61, no. 1, pp. 52-58.
- [39] J. Matas and O. Chum, "Randomized RANSAC with $t_{d,d}$ test," *Image and Vision Computing*, vol. 22, no. 10, pp. 837-842, 2004.
- [40] A. Sapkota, B. Parks, W.J. Scheirer, and T.E. Boult, "FACE-GRAB: Face recognition with general region assigned to binary operator," in *Computer Vision and Pattern Recognition Workshops*, 2010, pp. 82-89.
- [41] S. Liao, A.K. Jain, and S.Z. Li., "Partial face recognition: Alignment free approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1193-1205, 2012.
- [42] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955-1967, 2009.
- [43] A. Sanfeliu and K.S. Fu, "A distance measure between attributed relational graphs for pattern recognition," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 13, no. 3, pp. 353-362, 1983.
- [44] Y. Rubner, C. Tomasi, and L.J. Guibas, "The earth movers distance as a metric for image retrieval," *Int. J. of Comput. Vision*, vol. 40, no. 2, pp. 99-121, 2000.
- [45] "HSI seeks public's help identifying 'Jane Doe' producer of child pornography, rescuing young child from ongoing sexual abuse," *ICE News Releases* Dec 19, 2012 [Online]. Available: <http://www.ice.gov/news/releases/hsi-seeks-publicshelp-identifying-jane-doe-producer-child-pornography-rescuing-young>
- [46] R.E. Billingham, "A reconsideration of phenomenon of hair neogenesis with particular reference to the healing of cutaneous wounds in adult mammals." In: W. Montagna and RA Ellis (eds) *The biology of hair growth*. Academic Press, New York, 1958.

- [47] K.S. Stenn and R. Paus, "Controls of hair follicle cycling," *Physiological Reviews*, vol. 81, pp. 449-491, 2001.
- [48] A. Vogt, K.J. McElwee, and U. Blume-Peytavi, *Biology of the hair follicle*. in U. Blume-Peytavi, A. Tosti, D.A. Whiting and R.M. Trüeb (Eds) Hair Growth and Disorders, Springer, 2008.
- [49] V.A. Randall and N.V. Botchkareva, *The biology of hair growth*. in Cosmetic Applications of Laser and Light-based system, G.S. Ahluwalia (ed.), William Andrew Inc. 2009.
- [50] V.A. Randall and E.J.G. Ebling, "Seasonal changes in human hair growth," *Br. J. Dermatol.*, vol. 124, pp. 146-151, 1991.
- [51] H.B. Chase, "Growth of hair," *Physical Review*, vol. 34, pp. 113-126, 1954.
- [52] N.V. Botchkareva, M. Khlgatian, B.J. Longley, V.A. Botchkarev and B.A. Gilchrest, "SCF/c-kit signaling is required for cyclic regeneration of the hair pigmentation unit," *The Journal of the Federation of American Societies for Experimental Biology*, vol. 15, no. 3, pp. 645-659, 2001.
- [53] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in *ECCV*, 2000, pp. 404-420.
- [54] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *ICCV*, 2005, pp. 786 - 791.
- [55] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005, pp. 886 - 893.
- [56] A.E. Elo, *The rating of chessplayers, past and present*. Bastsford, 1978.
- [57] L. Wang and G. Leedham, "Near- and far- infrared imaging for vein pattern biometrics," in *IEEE Int. Conf. on Video and Signal Based Surveillance*, 2006, pp. 52.
- [58] S.K. Im, H.M. Park, Y.W. Kim, S.C. Han, S.W. Kim, and C.H. Kang, "An biometric identification system by extracting hand vein patterns," *J. of the Korean Physical Society*, vol. 38, no. 3, pp. 268-272, 2001.
- [59] L. Wang, G. Leedham, and D.S.Y. Cho, "Minutiae feature analysis for infrared hand vein pattern biometrics," *Pattern Recognition*, vol. 41, no. 3, pp. 920-929, 2008.

- [60] J.C. Lee, "A novel biometric system based on palm vein image," *Pattern Recognition Lett.*, vol. 33, no. 12, pp. 1520-1528, 2012.
- [61] W. Yang, X. Yu, and Q. Liao, "Personal authentication using finger vein pattern and finger-dorsa texture fusion," in *17th ACM Int. Conf. on Multimedia*, 2009, pp. 905-908.
- [62] J.G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148-1161, 1993.
- [63] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239-256, 1992.
- [64] A. Fitzgibbon, "Robust registration of 2D and 3D point sets," *Image and Vision Computing*, vol. 21, no. 13-14, pp. 1145-1153, 2003.
- [65] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, nos. 2-3, pp. 114-141, 2003.
- [66] A. Myronenko and X. Song, "Point set registration: coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262-2275, 2010.
- [67] F.L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567-585, 1989.
- [68] M.B. Stegmann and D.D. Gomez, *A Brief Introduction to Statistical Shape Analysis*. Technical University of Denmark, Lyngby, 2002.
- [69] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*. Wiley, Chichester, 1998.
- [70] W.R. Schwartz, A. Kembhavi, D. Harwood, and L.S. Davis, "Human detection using partial least squares analysis," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 24-31.
- [71] A. Kembhavi, D. Harwood, and L.S. Davis, "Vehicle detection using partial least squares," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1250-1265, 2011.
- [72] W.R. Schwartz, H. Guo, and L.S. Davis, "A robust and scalable approach to face identification," in *ECCV*, 2010, pp. 476-489.

- [73] W.R. Schwartz, H. Guo, J. Choi, and L.S. Davis, "Face identification using large feature sets," *IEEE Trans. on Image Process.*, vol. 21, no. 4, pp. 2245-2255, 2012.
- [74] R. Rosipal and N. Kramer, "Overview and recent advances in partial least squares," in *Proc. Int. Conf. Subspace, Latent Structure and Feature Selection*, 2005, pp. 34-51.
- [75] K.P. Bennett and M.J. Embrechts, "An optimization perspective on kernel partial least squares regression," in *Advances in Learning Theory: Methods, Models and Applications*, 2003, pp. 227-250.
- [76] S. Reinikainen and A. Hoskuldsson, "Covproc method: Strategy in modeling dynamic systems," *J. Chemometrics*, vol. 17, pp. 130-139, 2003.
- [77] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemometrics*, vol. 17, no. 3, pp. 166-173, 2003.
- [78] H.N. Qu, G.Z. Li, and W.S. Xu, "An asymmetric classifier based on partial least squares," *Pattern Recognition*, vol. 43, no. 10, pp. 3448-3457, 2010.
- [79] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. of the ACM Conf. on Knowledge Discovery and Data Mining*, New York, NY, 2002, pp. 133-142.
- [80] C.J.C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton and G. Hullender, "Learning to rank using gradient descent," in *Proc. of ICML*, Germany, 2005, pp. 89-96.
- [81] J.H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. of Stat.*, vol. 29, no. 5, pp. 1189-1232, 2001.
- [82] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [83] Y. Freund, R. Iyer, R. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *The J. of Mach. Learning Research*, vol. 4, pp. 933-969, 2003.
- [84] D. Metzler and W.B. Croft, "Linear feature-based models for information retrieval," *J. Inform. Retrieval*, vol. 10, no. 3, pp. 257-274, 2007.

- [85] Z. Cao, T. Qin, T.Y. Liu, M.F. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise Approach," in *Proc. of ICML*, Corvallis, OR, 2007, pp. 129-136.
- [86] K.Y. Chang and C.S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. on Image Process.*, vol. 24, no. 3, pp. 785-798, 2015.
- [87] J.L. Crowley and F. Berard, "Multi-modal tracking of faces for video communications," in *Proc. IEEE CVPR*, Puerto Rico, 1997, pp. 640-645.
- [88] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226-1238, 2005.
- [89] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 265-270, 2003.
- [90] T.S. Lee. "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959-971, 1996.
- [91] J. Chen, Y. Sato, and S. Tamura, "Orientation space filtering for multiple orientation line segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 5, pp. 417-429, 2000.
- [92] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 441-476, 2016.
- [93] I. Bouchrika, M. Goffredo, J. Carter, and M. Nixon, "On using gait in forensic biometrics," *J. of Forensic Science*, vol. 56, no. 4, pp. 882-889, 2011.
- [94] "A new kind of riot? From Brixton 1981 to Tottenham 2011," *The Guardian* Dec. 9, 2011 [Online]. Available: <http://www.guardian.co.uk/uk/2011/dec/09/riots-1981-2011-differences>
- [95] "A tale of two riots," *CBC News* Jun. 16, 2011 [Online]. Available: <http://www.cbc.ca/news/canada/story/2011/06/16/f-vancouver-riot-1994-2011.html>

- [96] "Scenes of destruction: 80 arrests, businesses torched, looted in Ferguson," *Fox News*, Nov. 25, 2014 [Online]. Available: <http://insider.foxnews.com/2014/11/25/scenes-destruction-80-arrests-businesses-torched-looted-ferguson>
- [97] "2 billion consumers worldwide to get smart (phones) by 2016," *eMarketer* Dec. 11, 2014 [Online]. Available: <http://www.emarketer.com/Article/2-Billion-Consumers-Worldwide-Smartphones-by-2016/1011694>
- [98] Z. Liu and C. Liu, "Fusion of color, local spatial and global frequency information and face recognition," *Pattern Recognition*, vol. 43, no. 12, pp. 2882-2890, 2010.
- [99] J.Y. Choi, Y.M. Ro, and K.N. Plataniotis, "Color local texture features for color face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1366-1380, 2012.
- [100] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 780-787, 2008.
- [101] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *Proc. European Conf. Comput. Vis.*, 2004, pp. 469-481.
- [102] F.K.S. Chan and A.W.K. Kong, "Using leg geometry to align androgenic hair patterns in low resolution images for criminal and victim identification," in *Int. Conf. on Pattern Recognition*, 2014, pp. 495-500.
- [103] LIBLINEAR – A library for large linear classification. [Online] Available: <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
- [104] H. Proenca and L.A. Alexandre, "UBIRIS: a noisy iris image database," in *Int. Conf. on Image Anal. and Process.*, 1996, pp. 790-777.
- [105] K.S. Killourhy and R.A. Maxion, "Comparing anomaly detectors for keystroke dynamics," in *Annu. Int. Conf. on Dependable Syst. and Networks*, 2009, pp. 125-134.
- [106] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multi-spectral palm image fusion for accurate contact-free palmprint recognition," in *ICIP*, 2008, pp. 281-284.
- [107] X.Q. Zeng and G.Z. Li, "Incremental partial least squares analysis of big streaming data," *Pattern Recognition*, vol. 47, no. 11, pp. 3726-3735, 2014.

- [108] A.A. Paulino, J. Feng, and A.K. Jain, "Latent fingerprint matching using descriptor-based Hough transform," in *Int. Joint Conf. on Biometrics*, 2011, pp. 1-7.
- [109] A.A. Ahmed, "A study of correlations within the dimensions of lower limb parts for personal identification in a Sudanese population," *The Scientific World Journal*, vol. 2014, 2014.
- [110] R.S. Chow, M.K. Medri, D.C. Martin, R.N. Leekam, A.M. Agur, and N.H. McKee, "Sonographic studies of human soleus and gastrocnemius muscle architecture: gender variability," *European J. of Applied Physiology*, vol. 82, no. 3, pp. 236-244, 2000.
- [111] A. Sartorio, M. Proietti, P.G. Marinone, F. Agosti, F. Adorni, and C.L. Lafortuna, "Influence of gender, age and BMI on lower limb muscular power output in a large population of obese men and women," *Int. J. of Obesity*, vol. 28, no. 1, pp. 91-98, 2004.
- [112] E.J. Jones, P.A. Bishop, A.K. Woods, and J.M. Green, "Cross-sectional area and muscular strength: A brief review," *Sports Medicine*, vol. 38, no. 12, pp. 987-994, 2008.
- [113] P. Gupta, P. Kumar, A. Gaharwar, H. Ansari and M. Hussein, "Correlation of percutaneous length of tibia with body height and estimation of stature in living north Indian males," *Scholars J. of Appl. Medical Sci.*, vol. 2, no. 2D, pp. 848-852, 2014.
- [114] I. Duyar and C. Pelin, "Body height estimation based on tibia length in different stature groups," *Amer. J. of Physical Anthropology*, vol. 122, no. 1, pp. 23-27, 2003.
- [115] M.S. Nixon, I. Bouchrika, B. Arbab-Zavar, and J.N. Carter, "On use of biometrics in forensics: gait and ear," in *18th European Signal Process. Conf.*, Aalborg, Denmark, 2010, pp. 1655-1659.
- [116] I. Bouchrika, M. Goffredo, J. Carter, and M.S. Nixon, "On using gait in forensic biometrics," *J. Forensic Sci.*, vol. 56, no. 4, pp. 882-889, 2011.
- [117] "Reported child sexual abuse has risen 60% in last four years, figures show," *The Guardian* Apr. 9, 2015 [Online]. Available: <http://www.theguardian.com/society/2015/apr/09/reported-child-sexual-abuse-has-risen-60-in-last-four-years-figures-show>

- [118] “Washington state man sentenced to 15 years for producing child pornography,” *U.S. Immigration and Customs Enforcement News Release* Aug. 15, 2014 [Online]. Available: <https://www.ice.gov/news/releases/washington-state-man-sentenced-15-years-producing-child-pornography>
- [119] “Arkansas man sentenced for producing child pornography of 8-year-old,” *U.S. Immigration and Customs Enforcement News Release* Aug. 15, 2014 [Online]. Available: <http://www.ice.gov/news/releases/1408/140815fayetteville.htm>
- [120] “School disk jockey sentenced to 40 years for sexually exploiting minors,” *U.S. Immigration and Customs Enforcement News Release* Sep. 3, 2014 [Online]. Available: <http://www.ice.gov/news/releases/1409/140903sanjuan.htm>
- [121] “Eyes behind the mirror: The rising menace of voyeurism,” *NewsGram* Jul. 15, 2015 [Online]. Available: <http://www.newsgram.com/eyes-behind-the-mirror-the-rising-menace-of-voyeurism/>
- [122] “Tech-savvy peeping Toms,” *TheStarTV.com* Dec. 27, 2015 [Online]. Available: <http://www.thestar.com.my/news/nation/2015/12/27/techsavvy-peeping-toms-smart-phones-spy-cameras-in-all-shapes-and-sizes-and-silent-mode-options-are/>
- [123] “Santa Barbara Police Arrest Video Voyeurism Suspect,” *Noozhawk* Dec. 14, 2012 [Online]. Available: http://www.noozhawk.com/article/121412_santa_barbara_police_arrest_video_voyeurism_suspect
- [124] S. Samangooei, B. Guo, and M.S. Nixon, “The use of semantic human description as a soft biometric,” in *2nd IEEE Int. Conf. on Biometrics: Theory Applications and Systems*, Arlington, VA, 2008, pp. 1-7.
- [125] S. Samangooei, M.S. Nixon, “On semantic soft-biometric labels,” in *Proc. of the BIOMET 2014*, Sofia, Bulgaria, 2014, pp. 3-15.
- [126] D.A. Reid and M.S. Nixon, “Imputing human descriptions in semantic biometrics,” in *Proc. of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, NY, USA, 2010, pp. 25-30.
- [127] G. Guo, G. Mu, Y. Fu, and T.S. Huang, “Human age estimation using bio-inspired features,” in *Proc. IEEE CVPR*, Miami, FL, 2008, pp. 112-119.

- [128] J. Lu, V.E. Liong, and J. Zhou, "Cost-sensitive local binary feature learning for facial age estimation," *IEEE Trans. on Image Process.*, vol. 24, no. 12, pp. 5356-5368, 2015.
- [129] Support Vector Machine for Ranking. [Online]. Available: https://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html
- [130] LIBSVM [Online]. Available: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [131] M.R. Islam, F.K.S. Chan, and A.W.K. Kong, "A preliminary study of lower leg geometry as a soft biometric trait for forensic investigation," in *Int. Conf. on Pattern Recognition*, Stockholm, Sweden, 2014, pp. 427-431.
- [132] J. Nocedal and S. J. Wright. *Numerical Optimization, Second Edition*. Springer Series in Operations Research, Springer Verlag, 2006.
- [133] RankLib [Online]. Available: <https://people.cs.umass.edu/~vdang/ranklib.html>
- [134] O.G. Cula, K.J. Dana, F.P. Murphy, and B.K. Rao "Skin texture modeling," *Int. J. of Comput. Vis.*, vol. 62, no. 1, pp. 97-119, 2003.
- [135] J. Xie, L. Zhang, J. You, D. Zhang, and X. Qu, "A study of hand back skin texture patterns for personal identification and gender classification," *Sensors*, vol. 12, pp. 8691-8709, 2012.
- [136] C. Tang, H. Zhang, and A.W.K Kong, "Using multiple models to uncover blood vessel patterns in color images for forensic analysis," *Inform. Fusion*, vol. 32, part B, pp. 26-39, 2016.
- [137] T. Watanabe and T. Tanaka, "Vein authentication using color information and image matching with high performance on natural light," in *ICCAS-SICE*, 2009, pp. 3625-3629.
- [138] J.H. Song, C. Kim, and Y. Yoo, "Vein visualization using a smart phone with multispectral Wiener estimation for point-of-care applications," *IEEE J. of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 773-778, 2015.
- [139] "England riots lead to 1,500 convictions so far," BBC, Feb. 23, 2012 [Online]. Available: <http://www.bbc.com/news/uk-17141883>

- [140] “Occupy protest in Rome hijacked by rioters who cause damage put at Euro2m,” *The Guardian* Oct. 16, 2011 [Online]. Available: <http://www.guardian.co.uk/world/2011/oct/16/rome-riot-damage-ringleaders-police>
- [141] “More protests in Brazil,” *3 News* Jun. 21, 2013 [Online]. Available: <http://www.3news.co.nz/More-protests-in-Brazil/tabid/417/articleID/302213/Default.aspx>
- [142] Computational Skin Texture [Online]. Available: <http://www.ece.rutgers.edu/~kdana/Research2014.html>
- [143] A. Blumstein and K. Nakamura, “Redemption in the presence of widespread criminal background checks,” *Criminology*, vol. 47, no. 2, pp. 372-359, 2009.
- [144] N.Q. Huynh, X. Xu, A.W.K. Kong, and S. Subbiah, “A preliminary report on a full-body imaging system for effectively collecting and processing biometric traits of prisoners,” *IEEE Symp. Series on Computational Intell.*, 2014, pp. 167 – 174.
- [145] M. Ngan, G.W. Quinn, and P. Grother, “Tattoo recognition technology — Challenge (Tatt-C) Outcomes and Recommendations,” *National Institute of Standards and Technology, U.S. Department of Commerce*, Sept 2015 [Online]. Available: <http://dx.doi.org/10.6028/NIST.IR.8078>
- [146] S. Liao, M.W.K. Law, and A.C.S. Chung, “Dominant local binary patterns for texture classification,” *IEEE Trans. Image Process.* vol. 18, no. 5, pp. 1107–1118, 2009.
- [147] Z. Guo, L. Zhang and D. Zhang. “A completed modeling of local binary pattern operator for texture classification,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [148] M. Varma and A. Zisserman, “Classifying images of materials: achieving viewpoint and illumination independence,” in *Proc. of European Conf. on Comput. Vision*, Copenhagen, 2002, pp. 255–271.
- [149] Source code of DLBP [Online]. Available: <http://dismac.dii.unipg.it/ldlbp/code.html>.
- [150] Source code of CLBP [Online]. Available: <http://www.comp.polyu.edu.hk/~cslzhang/code/CLBP.rar>.

- [151] Source code of Textons [Online]. Available: [http://www.robots.ox.ac.uk/~vgg/publications/2002/ Varma02/](http://www.robots.ox.ac.uk/~vgg/publications/2002/Varma02/)
- [152] S. Pankanti, S. Prabhakar and A.K. Jain, "On the individuality of fingerprints," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1010-1025, 2002.
- [153] R.M. Bolle, S. Pankanti, J.H. Connell and N.K. Ratha, "Iris individuality: a partial iris model," in *Proc. of the 17 the Int. Conf. on Pattern Recognition*, vol. 2, 2004, pp. 927-930.
- [154] M. Yang, K. Kpalma, and J. Ronin, "A survey of shape feature extraction techniques," *Pattern Recognition*, IN-TECH, pp. 43-90, 2008.
- [155] Y.P. Wang, K.T. Lee, and K. Toraichi, "Multiscale curvature-based shape representation using B-spline wavelets," *IEEE Trans. Image Process*, vol. 8, no. 10, pp. 1586-1592, 1999.
- [156] W.K. Pratt. *Digital Image Processing*. New York: Wiley-Interscience, 1978.
- [157] M. Basu, "Gaussian-based edge-detection methods-a survey," *IEEE Trans. Man Cybern. C, Appl. Rev.*, vol. 32, no. 3, pp. 252-260, 2002.
- [158] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679-698, 1986.
- [159] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London Series B Biological Sciences*, vol. 27, no. 1167, pp. 187-207, 1980.
- [160] T. Hastie, R. Tishirani, and J. Friedman, *An Introduction to Statistical learning*, Springer, 2009.

Appendices

Appendix A: References of the Internet Images in Figure 5.2

- “The £25bn bombshell: rioters stone MP after Greece pushes through austerity package,” *Daily Mail* Jun. 30, 2011 [Online]. Available: <http://www.dailymail.co.uk/news/article-2009437/Greek-vote-Riotersstone-MP-Greece-pushes-austerity-package.html>
- “Protesting skateboarder shreds for riot police in Brazil,” *NBC News* Mar. 28, 2014 [Online]. Available: <http://www.nbcnews.com/news/photo/protesting-skateboarder-shreds-riot-police-brazil-n66121>
- “Greek austerity protests turn violent,” *NBC News* Jun. 28, 2011 [Online]. Available: <http://www.nbcnews.com/slideshow/news/greek-austerity-protests-turn-violent-43575292>
- “Greece: general strike 29.06.2011 – photos,” *IMC UK* Jun. 29, 2011 [Online]. Available: <http://www.indymedia.org.uk/en/2011/06/481438.html>
- “Photos: riots, fire, destruction after Vancouver’s loss,” *National Post* Jun. 16, 2011 [Online]. Available: <http://news.nationalpost.com/news/canada/photos-riots-fire-destruction-after-vancouvers-loss>
- “25 killed, 30 wounded as Islamic terrorist attacks increase in Niger,” *Examiner.com* May 24, 2013 [Online]. Available: <http://www.examiner.com/article/25-killed-30-wounded-as-islamic-terrorist-attacks-increase-niger>
- “Venezuela's Street Protests Are More Terrifying Than Ever,” *Business Insider* Apr. 8, 2014 [Online]. Available: <http://www.businessinsider.sg/venezuelan-protests-are-getting-intense-2014-4/#.VSIUNPnLcU0>
- “Activists defend Maidan tent city,” *A Red Rover* Aug. 7. 2014 [Online]. Available: <http://aredrover.com/europe/ukraine/ukraine-activists-defend-maidan-squares-tent-city/>
- “Riot Police Fighting,” *imgkid.com* [Online]. Available: <http://imgkid.com/riot-police-fighting.shtml>
- “Filosofando en la barricada,” *flickr.com* Nov. 25, 2010 [Online]. Available: <https://www.flickr.com/photos/rafaelbernal77/5207283493/in/photostream/>
- “‘Paro’ nacional convocado por la CUT,” *flickr.com* Apr. 15, 2009 [Online]. Available: <https://www.flickr.com/photos/cosmopolita1/3448685157/in/photostream/>

- “Voy a dar mi opinion,” *flickr.com* May 19, 2009 [Online]. Available: <https://www.flickr.com/photos/cosmopolita1/4477647383/in/photostream/>
- “Cat scratch fever,” *flickr.com* Apr. 15, 2009 [Online]. Available: <https://www.flickr.com/photos/cosmopolita1/5353037825/in/photostream/>
- “Athens Greece Syntagma Square 28 29 June 2011_03,” *flickr.com* Jun. 28, 2011 [Online]. Available: <https://www.flickr.com/photos/efthymios-gourgouris/5889357310/in/photostream/>
- “Syntagma Square 15 June 2011DSC_3794_resize,” *flickr.com* Jun. 15, 2011 [Online]. Available: <https://www.flickr.com/photos/efthymios-gourgouris/5842176771/in/photostream/>
- “Syntagma Square 15 June 2011DSC_3742_resize,” *flickr.com* Jun. 15, 2011 [Online]. Available: <https://www.flickr.com/photos/efthymios-gourgouris/5842169333/in/photostream/>
- “Greek protesters clash with police in Syntagma Square, Athens following austerity vote,” *flickr.com* Jun. 29, 2011 [Online]. Available: <https://www.flickr.com/photos/teacherdudebbq2/5888126403/in/photostream/>
- “Miedo,” *flickr.com* May 1, 2012 [Online]. Available: <https://www.flickr.com/photos/unaqueotraalucinacion/7133339465/in/pool-riots/>
- “Con todo,” *flickr.com* Aug. 22, 2012 [Online]. Available: <https://www.flickr.com/photos/titovega89/7841674958/in/pool-riots/>
- “Warzone,” *flickr.com* Oct. 25, 2012 [Online]. Available: <https://www.flickr.com/photos/82147437@N08/8121545056/in/pool-riots/>
- “Anarchist riot,” *imgkid.com* [Online]. Available: <http://imgkid.com/anarchist-riot.shtml>
- “Blueprint for peace: what Ferguson can learn from Cincinnati,” *NEC News* Aug. 31, 2014 [Online]. Available: http://media1.s-nbcnews.com/i/newscms/2014_35/641931/140829-cincinnati-riot-main_3c02cc0874697f28aeeeca41e25fe2fb.jpg

Appendix B: References of the Internet Images

in Figure 7.1

- “Protests spread across Brazil over cost of hosting the World Cup,” *The Times* d18 Jun. 2013 [Online]. Available: http://www.thetimes.co.uk/tto/multimedia/archive/00423/130553351__423426b.jpg
- “In Brazil, unrest is simmering again, over bus fares, Olympic golf and Nazi cops,” *The Washington Post* Jan 12, 2015 [Online]. Available: https://www.washingtonpost.com/world/the_americas/in-brazil-unrest-is-simmering-again-over-bus-fares-olympic-golf-and-nazi-cops/2015/01/11/2c7518a0-968d-11e4-8385-866293322c2f_story.html
- “What the hell is happening in Brazil?” *The Daily Dish* Jun. 18, 2013 [Online]. Available: <http://dish.andrewsullivan.com/2013/06/18/what-the-hell-is-happening-in-brazil/>
- “demonstrator, Protester, Protests, marches, Brazil,” *est100* Jun. 19, 2013 [Online]. Available: <http://okok1111111111.blogspot.hk/2013/06/demonstrator-protester-protests-marches.html>
- “Greek financial crisis tied to country’s rising rates of HIV, depression, infant Deaths,” *The Huffington Post* Feb 20, 2014 [Online]. Available: http://www.huffingtonpost.ca/2014/02/20/greece-financial-crisis-impact_n_4827063.html
- “Greece protest photos 3,” *CRYTOME* Jun. 28, 2011 [Online]. Available: <http://cryptome.org/info/greece-protest3/greece-protest3.htm>
- “Female Le Pen to stand trial for insulting Nazism,” *Breaking News Blast* Sep. 23, 2015 [Online]. Available: <http://www.breakingnewsblast.com/2015/09/22/>
- “The invisible face of Hams,” *Israel Defense Forces* Aug. 7, 2014 [Online]. Available: <https://www.idfblog.com/blog/2014/08/07/invisible-face-hamas/>
- “Palestinians in Gaza said to test long-range rocket,” *The Times of Israel* Dec. 31, 2013 [Online]. Available: <http://www.timesofisrael.com/palestinians-in-gaza-test-improved-long-range-rocket/>
- “10 tips from Hamas to ISIS: a terrorist’s love letter,” *Inquisitr.com* Aug. 19, 2014 [Online]. Available: <http://www.inquisitr.com/1421749/10-tips-from-hamas-to-isis-a-terrorists-love-letter/>
- “Human appeal and Hamas: the evidence,” *The Telegraph* Oct. 19, 2015 [Online]. Available: <http://www.telegraph.co.uk/news/uknews/terrorism-in-the-uk/11940489/Human-Appeal-and-Hamas-the-evidence.html>

- “Pics of training camp Jabhat al Nusra front, Syria. Photo’s were recently posted on AlQaeda related website.” *haralddoornbos* Dec 11, 2012 [Online]. Available: <https://haralddoornbos.wordpress.com/2012/12/11/pics-of-activities-at-training-camp-jabhat-al-nusra-front-syria/>

Appendix C: Some Results of the Skin Texture Algorithm

(a)

(b)

Figure C.1 Ten retrieval examples in a search of (a) inner arm and (b) inner thigh by the skin texture algorithm. Their relative scales are kept for comparison. The query image and the matched one are highlighted in white and red. For example, the matched one of the first probe is the image at the rank 1.

Appendix D: Some Failure Cases

Figure D.1 10 failure cases of 142×248 images at rank-1 in Chapter 3. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: expected match, right: wrong result)

Figure D.2 10 failure cases of low resolution images at rank-1 in Section 4.1. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: gallery image, right: aligned gallery image) Note that the green channel of images are shown.

Figure D.3 10 failure cases of high resolution images at rank-1 in Section 4.2. 5 sets are on the left and 5 sets are on the right. (For each set, first from the left: probe image, second from the left: gallery image, second from the right: aligned gallery image based on edge samples and androgenic hair follicles, first from the right: aligned gallery image based on edge samples) Note that the green channel of images are shown and the aligned images are resized for matching.

Figure D.4 10 failure cases of Image Set 1 at rank-1 in Chapter 5. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: gallery image, right: wrong result)

Figure D.5 10 failure cases of Low Quality Set at rank-1 in Chapter 5. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: gallery image, right: wrong result)

(a)

(b)

(c)

Figure D.6 Some front legs whose error levels of (a) height, (b) BMI, or (c) age estimation are more than 2 in CS curves of Chapter 6 are shown. Note that the ratio of height and width are kept for the images.

(a)

(b)

(c)

Figure D.7 Some side legs whose error levels of (a) height, (b) BMI, or (c) age estimation are more than 2 in CS curves of Chapter 6 are shown. Note that the ratio of height and width are kept for the images.

Figure D.8 10 failure cases from the experiment on NTU color forearm dataset at rank-1 in Chapter 7. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: gallery image, right: wrong result)

Figure D.9 10 failure cases from the experiment on NTU color thigh dataset at rank-1 in Chapter 7. 5 sets are on the left and 5 sets are on the right. (For each set, left: probe image, middle: gallery image, right: wrong result)