

Hand vein-based biometric authentication with limited training samples

by

Emile Beukes



*Thesis presented in partial fulfilment of the requirements for
the degree of Master of Science (Applied Mathematics) in the
Faculty of Science at Stellenbosch University*



Supervisor: Dr. J. Coetzer

Co-supervisor: Dr. J. Swanepoel

March 2018

The financial assistance of the Ball Family towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the Ball Family.

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 2018/03/01

Copyright © 2018 Stellenbosch University
All rights reserved.

Abstract

Hand vein-based biometric authentication with limited training samples

E. Beukes

*Department of Applied Mathematics,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.*

Thesis: MSc (Applied Mathematics)

March 2018

A number of novel hand vein-based biometric authentication systems are proposed. Said systems are non-intrusive and may for example assist with user authentication at automated teller machines. An infrared image of either the dorsal or ventral surface of an individual's hand is acquired through specialised equipment, after which the geometrical properties of the hand are used to extract a suitable region of interest (ROI). A novel protocol, which is based on morphological reconstruction, is employed for the purpose of isolating the veins within the ROI. Feature vectors are extracted from the isolated veins through the calculation of the discrete Radon transform. The feature vectors are appropriately normalised in order to ensure rotational, translational and scale invariance. The dissimilarity between the corresponding feature vectors extracted from a questioned image and a *reference* image belonging to the claimed client are represented by an average Euclidean or dynamic time warping-based distance. A score-based or rank-based classifier is subsequently employed for authentication purposes. It is demonstrated that, when *only one* training sample (of *arbitrary* quality) is available per client, and the client is granted six opportunities for authentication, an average error rate (AER) of 2.85% is achievable for a data set that contains *dorsal* hand vein patterns from 100 individuals. In a scenario where the *single* training sample is *guaranteed* to be of very high quality and the client is granted *only three* opportunities for authentication, the AER may be reduced to 0.77%.

Uittreksel

Handbloedvatgebaseerde biometriese verifikasie met beperkte afrigvoorbeelde

E. Beukes

*Departement Toegepaste Wiskunde,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.*

Tesis: MSc (Toegepaste Wiskunde)

Maart 2018

'n Aantal nuwe handbloedvatgebaseerde biometriese verifikasiesistels word voorgestel. Hierdie sistels is nie-indringend en kan byvoorbeeld met gebruikersverifikasie by outomatiese tellermasjiene help. 'n Infrarooi-beeld van die dorsale of ventrale oppervlak van 'n individu se hand word met behulp van gespesialiseerde toerusting verkry, waarna die geometriese eienskappe van die hand gebruik word om 'n geskikte fokusgebied (FG) te onttrek. 'n Nuwe protokol, wat op morfologiese rekonstruksie gebaseer is, word gebruik om die bloedvate in die FG te isoleer. Kenmerkvektore word vanuit die geïsoleerde bloedvate met behulp van die berekening van die diskrete Radon-transform onttrek. Die kenmerkvektore word vervolgens sinvol genormaliseer ten einde rotasie-, translasie- en skaal-invariansie te verseker. Die verskil tussen die ooreenstemmende kenmerkvektore wat vanuit 'n bevrage tekende beeld en 'n *verwysings*beeld wat aan die beweerde kliënt behoort, onttrek is, word deur 'n gemiddelde Euklidiese of dinamiese tydsverbuigingsafstand voorgestel. 'n Telling- of ranggebaseerde klassifiseerder word vervolgens vir verifikasiedoeleinde aangewend. Daar word aangetoon dat, wanneer *slegs een* afrigvoorbeeld (van *lukrake* kwaliteit) per kliënt beskikbaar is, en die kliënt ses geleenthede vir verifikasie gebied word, 'n gemiddelde foutkoers (GFK) van 2.85% haalbaar is vir 'n datastel wat *dorsale* handbloedvatpattrone van 100 individue bevat. In 'n scenario waar die *enkele* afrigvoorbeeld *gewaarborg* is om van 'n *baie hoë kwaliteit* te wees en die kliënt *slegs drie* geleenthede vir verifikasie gebied word, kan die GFK na 0.77% verminder word.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations:

1. Dr. Hanno Coetzer, for his unwavering support and supervision regarding the research presented in this thesis, as well as for the general guidance he provided that greatly contributed to my development as a programmer, researcher, applied mathematician and as a human being.
2. Dr Jacques Swanepoel, for his co-supervision and fundamental insights that contributed to the quality of the research presented in this thesis.
3. The Bogazici University, for allowing the use of their hand vein data set in this study.
4. The Institute of Control and Information Engineering, also for allowing the use of their hand vein data set.
5. The Postgraduate Funding department of Stellenbosch University, for their patience and effort regarding the financial assistance of this study.
6. The Ball family, for their financial support that made the completion of this study possible.
7. My family and close friends, for their love and support, especially in hard times.

Contents

Declaration	i
Abstract	ii
Uittreksel	iii
Acknowledgements	iv
Contents	v
List of Figures	vii
List of Tables	xii
List of Abbreviations	xv
Nomenclature	xvi
1 Introduction	1
1.1 Motivation	1
1.2 Background	2
1.3 Scope and objectives	5
1.4 Thesis overview	6
1.5 Contributions	10
2 Literature Study	11
2.1 Introduction	11
2.2 Data acquisition	11
2.3 Preprocessing and hand vein segmentation	12
2.4 Feature extraction	13
2.5 Feature matching and verification	14
2.6 Concluding remarks	14
3 Image segmentation	15
3.1 Introduction	15
3.2 Extracting the ROI	18

<i>CONTENTS</i>	vi
3.3 Extracting the veins	32
3.4 Concluding remarks	41
4 Feature extraction and matching	42
4.1 Introduction	42
4.2 Feature extraction	45
4.3 Feature matching	51
4.4 Concluding remarks	57
5 Experiments	58
5.1 Introduction	58
5.2 The Bosphorus data set	60
5.3 The CIE data set	61
5.4 Statistical performance measures	62
5.5 Experimental protocol: the <i>k-one</i> scenario	63
5.6 Experimental protocol: the <i>k-six</i> scenario	64
5.7 Results	66
5.8 Concluding remarks	79
6 Conclusion and future work	81
6.1 Conclusion	81
6.2 Future work	82
Appendices	83
A Experimental conceptualisations	84
B Authentication examples	89
List of References	91

List of Figures

1.1	Categorisation of authentication factors. The authentication factor employed in this study is depicted in boldface.	2
1.2	A graphic depicting the stages of an automatic identity verification system.	4
1.3	Different matching strategies. The matching strategy employed in this study is depicted in boldface.	5
1.4	The proposed hand vein segmentation protocol.	7
3.1	(a) A typical NIR image from the Bosphorus data set of size 300×240 pixels. (b) A typical NIR image from the CIE data set of size 1024×768 pixels.	15
3.2	(a) The image in Figure 3.1 (a). (b) The targeted veins in (a) manually annotated and superimposed in yellow onto (a). (c) The image in Figure 3.1 (b) after being converted into grey-scale format and resized to 320×240 pixels. (d) The targeted veins in (c) manually annotated and superimposed in yellow onto (c).	16
3.3	Conceptualisation of the proposed hand-vein extraction protocol.	17
3.4	Manually annotated boundaries (in yellow) of the targeted ROIs for the images depicted in Figures 3.2 (a) and (c).	18
3.5	(a) A grey-scale image corrupted by Gaussian noise. (b) The histogram of the image in (a), where the location of the appropriate threshold value as determined by Otsu's algorithm is denoted by the red vertical line.	19
3.6	(a) The image in Figure 3.1 (a). (b) The binarised version of (a) after the application of Otsu's threshold.	20
3.7	(a) A convex object. (b) A non-convex or <i>concave</i> object.	21
3.8	(a) The image in Figure 3.6 (b). (b) The convex hull of (a).	21
3.9	Figure 3.8 (a) after the pixels that belong to the outer black background regions were negated. Note that the boundary of the image is not explicitly shown.	22
3.10	The detected hand.	22

3.11	(a) The boundary of an object. (b) Conceptualisation of the morphological opening of the object in (a) with the blue disc-shaped SE. (c) The result of opening (a) with the disc-shaped SE depicted in (b), where it is clear that the brown regions in (b) are removed.	23
3.12	(a) The image in Figure 3.10. (b) The result of morphologically opening the hand in (a) with a disc-shaped SE.	24
3.13	(a) A 300×200 image containing an annular object. (b) The morphological erosion of (a) by the 16×16 square-shaped SE shown in white. (c) The morphological erosion of (a) by the 16×16 disc-shaped SE shown in white.	24
3.14	(a) The object in Figure 3.12 (b). (b) The current ROI, where dark shadows are clearly visible in close proximity to the hand boundary. (c) The morphological erosion of (a) by a 16×16 square shaped SE. (d) The new ROI. (e) The eroded part of (a). (f) The dark boundary region removed from (b).	25
3.15	(a) The image depicted in Figure 3.12 (b), where the line that is used to approximate the orientation of the body of the hand is manually annotated in red. (b) The vertically oriented version of (a).	26
3.16	(a) The boundary of the detected hand. (b) The opened hand. (c) The union of (a) and (b). (d) The intersection of (a) and (b). (e) The webbing between the four fingers. (f) The targeted reference points: the centre of mass of (b) and the webbing between the middle and ring fingers.	27
3.17	(a) The vertically oriented version of the object depicted in Figure 3.14 (c). (b) The result of applying morphological opening to (a) by a vertical, one-dimensional SE.	28
3.18	Conceptualisation of the proposed procedure for estimating the width of the object depicted in Figure 3.17 (b).	28
3.19	(a) The morphological opening of the object in Figure 3.17 (b) by a tailor-made disc-shaped SE. (b) The final ROI for the vertically oriented version of the image depicted in Figure 3.2 (a). The non-ROI pixels are darkened for effect.	29
3.20	(a) A sample from the Bosphorus data set that contains jewellery in the vicinity of the wrist. (b) The final ROI for the vertically oriented version of the image shown in (a). The non-ROI pixels are darkened for effect.	29
3.21	(Left) A random selection of samples from the Bosphorus data set. (Right) The final ROIs for the vertically oriented versions of the corresponding images on the left.	30
3.22	(Left) A random selection of NIR images from the CIE data set after being converted into grey-scale format. (Right) The final ROI for the corresponding samples on the left.	31

3.23	The ROIs for selected images in Figures 3.21 and 3.22. These ROIs are presented again in order to highlight key differences between vein and non-vein regions.	32
3.24	(a) Figure 3.21 (d) that depicts a ROI that contains a local shadow. (b) Figure 3.21 (f) that depicts a ROI that contains hair. (c) Figure 3.22 (d) that depicts a ROI that contains palm lines and a local shadow. (d) Figure 3.22 (h) that depicts a ROI that contains palm lines and a local shadow.	33
3.25	The black top-hat transforms of the corresponding images in Figure 3.23, after binarisation through Otsu's method.	34
3.26	The LoG transformation of the corresponding images in Figure 3.23, after binarisation through Otsu's method.	35
3.27	The intersection of the corresponding images depicted in Figures 3.25 and 3.26. The detected pixels serve as seed pixels for morphological reconstruction by dilation.	36
3.28	(a) The image in Figure 3.13 (a). (b) The pixels that are appended to (a), when dilated by the white 16×16 square-shaped SE, are shown in yellow. (c) The object in (a) after being dilated by said SE.	37
3.29	The mask images for the purpose of morphological reconstruction by dilation. These images correspond to the respective images depicted in Figure 3.23.	38
3.30	The morphologically reconstructed versions of the corresponding seed images shown in Figure 3.27. The relevant mask images are shown in Figure 3.29.	38
3.31	The final segmentation results for selected NIR images from the Bosphorus data set. (Left) The ROIs depicted in Figure 3.21. (Right) The extracted hand veins.	39
3.32	The final segmentation results for selected NIR images from the CIE data set. (Left) The ROIs depicted in Figure 3.22. (Right) The extracted hand veins.	40
4.1	(a) An image containing a binarised hand vein pattern. (b) An image containing a binarised Oriental handwritten signature.	42
4.2	Conceptualisation of the proposed feature extraction protocol.	44
4.3	(a) A binarised hand vein image associated with a specific individual from the Bosphorus data set. (b) A binarised hand vein image from the same individual as the one referred to in (a), but captured on a different occasion. (c) A binarised hand vein image associated with a specific individual from the CIE data set. (d) A binarised hand vein image from the same individual as the one referred to in (c), but captured on a different occasion.	45
4.4	The results after retaining only the cropped sub-images within the bounding boxes for the images shown in Figure 4.3.	46

4.5	Conceptualisation of the discrete Radon transform being applied to a typical binarised hand vein image.	47
4.6	The result of applying the discrete Radon transform to the corresponding images in Figure 4.4, using 64 equally distributed angles within the interval $\theta \in [0^\circ, 180^\circ)$	48
4.7	Conceptualisation of the decimation of the leading and trailing zeros for a selected unnormalised feature vector from the image depicted in Figure 4.6 (a). The leading zeros appear before the first red vertical line, while the trailing zeros appear after the second red vertical line.	49
4.8	The feature sets for the corresponding images depicted in Figure 4.4.	50
4.9	Conceptualisation of the difference between two features within a pair of corresponding feature vectors, as quantified by the length of the red vertical line.	51
4.10	Conceptualisation of the proposed feature matching protocol.	52
4.11	Conceptualisation of the squared Euclidean distance between two corresponding feature vectors.	53
4.12	(a) A questioned feature set. (b) The feature set in (a) shifted towards the right by exactly half of its number of columns, followed by a wrap-around procedure and the subsequent flipping of the wrapped columns in an up-down fashion. It is clear that said columns constitute the last half of the columns of the feature set depicted in (a).	54
4.13	An illustration of the possibility that two prominent features may be associated with <i>different</i> components of corresponding feature vectors.	55
4.14	A conceptualisation of the restricting bandwidth associated with the calculation of a DTW-based distance between two corresponding feature vectors.	56
5.1	A typical NIR image from the Bosphorus data set.	60
5.2	A typical palmar NIR image from the CIE data set.	61
A.1	Conceptualisation of the experimental protocol for the <i>k-one</i> scenario within the context of a score-based classifier.	85
A.2	Conceptualisation of the experimental protocol for the <i>k-one</i> scenario within the context of a rank-based classifier.	86
A.3	Conceptualisation of the experimental protocol for the <i>k-six</i> scenario (PES2) within the context of a score-based classifier.	87
A.4	Conceptualisation of the experimental protocol for the <i>k-six</i> scenario (PES2) within the context of a rank-based classifier.	88

- B.1 (a) A relatively high quality template (reference sample) for a specific individual. (b) A *positive* evaluation sample that is *accepted* due to its similarity to the template in (a). (c) The image depicted in (a). (d) A *positive* evaluation sample that is *accepted* despite small differences in hand vein structure. (e) The image depicted in (a). (f) A *positive* evaluation sample *rejected* due to the significant structural differences of the hand veins. (g) A relatively low quality reference sample for the *same* individual as the one associated with the image depicted in (a). (h) The *positive* sample depicted in (b) is now *rejected* due to the low quality of the template depicted in (g). 90

List of Tables

1.1	Experimental results for the benchmark systems. The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.	8
1.2	The ideal benchmark results (within the context of the <i>k-one</i> scenario). The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.	9
1.3	Experimental results for the most proficient systems when implemented on the Bosphorus data set, as well as the corresponding results when the same systems are implemented on the CIE data set. The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.	9
5.1	Bosphorus data set: EER. Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	68
5.2	Bosphorus data set: EER. Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>three</i> opportunities for authentication. Note that the benchmark results are indicated in boldface and that the results for the <i>ideal</i> (benchmark) experiments are indicated in italics.	68
5.3	Bosphorus data set: EER. Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>six</i> opportunities for authentication.	68
5.4	Bosphorus data set: FAR_{zero}. Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	70
5.5	Bosphorus data set: FAR_{zero}. Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>three</i> opportunities for authentication.	70

5.6	Bosphorus data set: FAR_{zero} . Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>six</i> opportunities for authentication. The best results within the context of the <i>k-one</i> scenario (<i>excluding</i> those for the <i>ideal</i> scenario) are boxed.	70
5.7	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	72
5.8	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>three</i> opportunities for authentication.	72
5.9	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-one</i> scenario in which an individual is granted <i>six</i> opportunities for authentication.	72
5.10	Bosphorus data set: EER . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	73
5.11	Bosphorus data set: EER . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>three</i> opportunities for authentication. Note that the benchmark results are indicated in boldface. The results for PES0 and PES1 within the context of a score-based classifier are also included.	73
5.12	Bosphorus data set: EER . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>six</i> opportunities for authentication.	74
5.13	Bosphorus data set: FAR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	75
5.14	Bosphorus data set: FAR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>three</i> opportunities for authentication.	75
5.15	Bosphorus data set: FAR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>six</i> opportunities for authentication. The best results within the context of the <i>k-six</i> scenario are boxed.	75
5.16	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>only one</i> opportunity for authentication.	76
5.17	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>three</i> opportunities for authentication.	76
5.18	Bosphorus data set: FRR_{zero} . Experimental results for the <i>k-six</i> scenario in which an individual is granted <i>six</i> opportunities for authentication.	77

5.19	CIE data set: <i>k-one</i> scenario. The results for the benchmark experiments, the ideal (benchmark) experiments, as well as for the system for which the best results are achieved when the Bosphorus data set is considered.	77
5.20	CIE data set: <i>k-six</i> scenario. The results for the benchmark system and for the system for which the best results are achieved when the Bosphorus data set is employed.	78

List of Abbreviations

AER	Average error rate
ANN	Artificial neural network
ATM	Automated teller machine
CLAHE	Contrast limited adaptive histogram equalisation
DTW	Dynamic time warping
DRT	Discrete radon transform
EER	Equal error ate
FAR	False acceptance rate
FRR	False rejection rate
FAR _{zero}	Zero false acceptance rate
FRR _{zero}	Zero false rejection rate
FIR	Far infra-red
GMF	Gaussian matched filter
IGMF	Improved Gaussian matched filter
ICA	Independent component analysis
LEM	Line-edge mapping
LoG	Laplacian of Gaussian
NIR	Near infra-red
NMF	Non-negative matrix factorisation
PIN	Personal identification number
POSM	Point of sale machine
PES	Proficiency enhancement strategy
PRES	Precision
REC	Recall
ROI	Region of interest
SE	Structuring element

Nomenclature

N_T	Number of independent <i>trials</i> within an experiment
N_O	Number of individuals used for optimisation purposes
N_E	Number of individuals used for evaluation purposes
N_P	Number of permutations of individuals in data set
N_F	Number of folds over which data partitioning is cross-validated
N_R	Number of cross-validations of the selected reference sample
n^+	Number of positive samples used in a specific trial
n^-	Number of negative samples used in a specific trial
t^+	An instance of a true positive
t^-	An instance of a true negative
f^+	An instance of a false positive
f^-	An instance of a false negative
T^+	Total number of true positive instances in a specific trial
T^-	Total number of true negative instances in a specific trial
F^+	Total number of false positive instances in a specific trial
F^-	Total number of false negative instances in a specific trial

Chapter 1

Introduction

1.1 Motivation

In the modern economic framework, reliable identity verification systems are becoming increasingly relevant in preventing identity fraud. According to a recent survey by the American Banking Association (2015), fraudulent transactions involving debit accounts caused an estimated \$1.910 billion in losses to the industry in 2014 alone. It was furthermore determined that 66% of these losses may be attributed to debit *card* fraud, where either a handwritten signature or a personal identification number (PIN code) is employed as an *authentication factor*. The aforementioned survey also mentions that additional fraudulent transactions amounting to \$11 billion were prevented by existing fraud prevention measures imposed by commercial banks, which indicates that the aforementioned authentication factors are prone to falsification.

The utilisation of the aforementioned authentication factors within the context of debit card transactions at automated teller machines (ATMs) or point of sale machines (POSMS) is fast gaining popularity, since said factors are both non-intrusive *and* efficient. Both of these factors are however fragile and prone to falsification (as reflected by the statistics of the aforementioned survey), since a PIN may be memorised, while a handwritten signature may be imitated with relative ease.

A number of *novel* identity verification systems, that are both non-intrusive and efficient, *and* significantly *reduce* the opportunity for fraudulent transactions, are proposed in this thesis. Said systems acquire an image of an individual's prominent hand veins (in real time) for the purpose of verifying a claim of identity when a bank card is presented at an ATM or POSM, or when access to a facility is requested.

In the remainder of this chapter a concise background to the broad field of identity verification, as well as the objectives and contributions of this study, are presented.

1.2 Background

We now introduce some key concepts within the field of identity verification, which forms part of the broader field of *pattern recognition*. The research conducted in this thesis is also placed into perspective.

1.2.1 Pattern recognition: a preliminary discussion

We first emphasise the subtle but important distinction between the concepts of *pattern recognition* and *pattern verification*. A *pattern recognition* system is prompted to recognise (identify) the pattern class to which an unlabelled (questioned) pattern belongs to. A *pattern verification* system, on the other hand, receives a *claim* pertaining to the class a questioned sample belongs to, and is therefore expected to simply verify whether said claim is valid or not.

The *recognition* of the pattern class that a questioned pattern belongs to involves the iterative matching of the features associated with said pattern with models that represent the candidate (enlisted) pattern classes. The *verification* of a questioned pattern simply constitutes the matching of the features associated with the questioned sample with a model of the *claimed* pattern class. A pattern verification system may therefore be viewed as a *bi-class* pattern recognition system, since the questioned pattern is either *accepted* or *rejected*.

We propose a number of hand vein-based identity *verification* systems, which constitute bi-class pattern recognition systems.

1.2.2 Identity verification: authentication factors

Within the context of identity verification, the pattern classes correspond to the identities of the individuals enrolled into the system. The patterns involved are associated with a so-called *authentication factor*, of which a categorisation is provided in Figure 1.1.

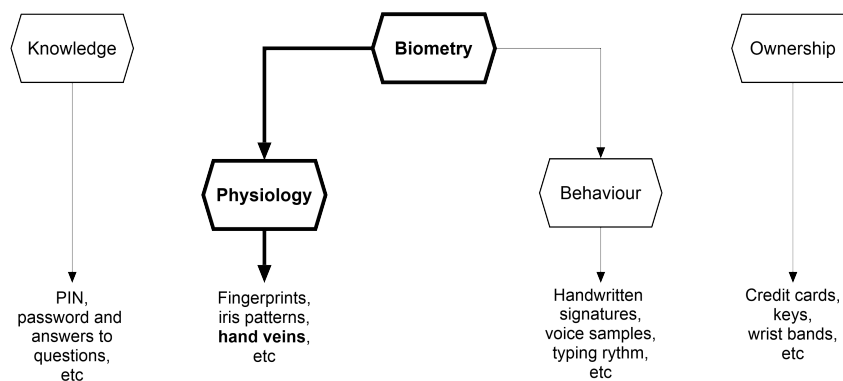


Figure 1.1: Categorisation of authentication factors. The authentication factor employed in this study is depicted in boldface.

The main objective of an identity verification system is to prevent identity theft. The selected authentication factor(s) must therefore be (1) difficult to forge (imitate) and (2) difficult to obtain illegally. A well-designed identity verification system should also be (1) time-efficient and (2) non-intrusive.

A knowledge-based factor constitutes an object of memory such as a PIN, password, answers to questions etc. (see Figure 1.1). Possession or ownership-based factors include, for example, tokens, physical keys, wrist bands, bank cards etc. Although knowledge-based and possession-based factors are popular in a fast paced modern economy, mainly due to their non-intrusiveness and time-efficiency, most of these factors are also relatively easy for criminals to obtain.

Biometric authentication factors are inherent to the behaviour or physiology of a certain individual and include fingerprints and hand veins. In order to acquire the majority of biometric factors, the individual in question is required to be physically present. This renders biometric factors superior to knowledge-based and possession-based factors from a security point of view. In rare cases however, handwritten signatures and fingerprints may be acquired and falsified on one site and then illegally presented somewhere else. The main disadvantages associated with the implementation of the majority of biometric factors involve the relatively intrusive and time-consuming nature of the acquisition process. This leads to the fact that factors such as iris and voice patterns are almost exclusively used in high-security scenarios, but *not* in public.

A trade-off therefore exists between the advantages associated with knowledge-based and possession-based factors, and those associated with most biometric factors. A factor in this context may either be (1) fast, non-intrusive *and* relatively easy for criminals to acquire, or (2) intrusive and time-consuming, but *also* difficult to obtain (or imitate) illegally. We therefore opt to employ an image of an individual's hand veins as an authentication factor, since it (1) may be acquired in a relatively non-intrusive and time-efficient manner (when compared to iris patterns for example) and (2) is very difficult to obtain illegally.

An engineer named Joseph Rice presented the first biometric identification system in Rice (1987) that specifically employs hand vein patterns as a personal authentication factor. The use of hand veins as an authentication factor has grown in popularity ever since. The main reason for this trend is the fact that the successful falsification of an individual's hand vein pattern is yet to be accomplished. Interesting scenarios in which hand veins may be used as an authentication factor include authentication through devices such as cell phones, laptops or vehicles. The main challenges associated with the utilisation of hand veins as an authentication factor include (1) the effect of skin diseases on the quality of the acquired image, and the presence of (2) hair, jewellery and other artefacts.

In the next section we discuss the typical architecture of an automated identity verification system.

1.2.3 Identity verification: typical system architecture

Figure 1.2 depicts the main stages of an automatic identity verification system, which is followed by a brief discussion of each stage.

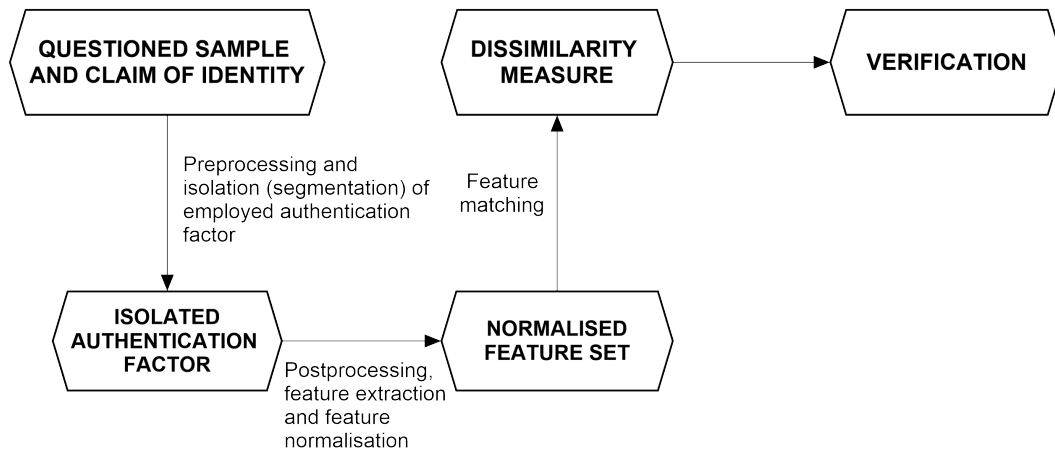


Figure 1.2: A graphic depicting the stages of an automatic identity verification system.

1.2.3.1 Data acquisition

The format in which the data is acquired depends largely on the choice of authentication factor. A well-designed acquisition procedure results in minimal redundant information (to the extent possible).

1.2.3.2 Preprocessing and segmentation

A number of *preprocessing* techniques are usually required in order to isolate the pertinent information associated with the targeted object from the background, and is referred to as image segmentation. The complexity of the preprocessing phase depends on the type of redundant information that has to be removed, as well as the choice of authentication factor.

1.2.3.3 Feature extraction

A well-constructed feature set constitutes robust measurements of the targeted object that should be as *invariant* as possible with respect to noise, as well as changes in scale, orientation and position of said object across several samples of the same authentication factor. The aforementioned measurements are therefore typically preceded *and* followed by appropriate normalisation procedures. A well-constructed feature extraction protocol should also be geared

towards the (1) minimisation of *intra*-class variation, as well as the (2) maximisation of *inter*-class variation.

1.2.3.4 Feature matching and classification

The purpose of an identity verification system is to either accept or reject a questioned sample that is claimed to belong to a certain individual (class). In order to achieve this, the system has to compare (match) the features associated with the questioned sample to a model of the claimed identity. There are two main matching strategies that a system may employ in order to classify a questioned pattern, namely (1) decision-theoretic matching or (2) structural matching. Said matching strategies are depicted in Figure 1.3. The decision-theoretic-based matching strategy employed in this study involves template matching through dynamic time warping (DTW), and is denoted in boldface.

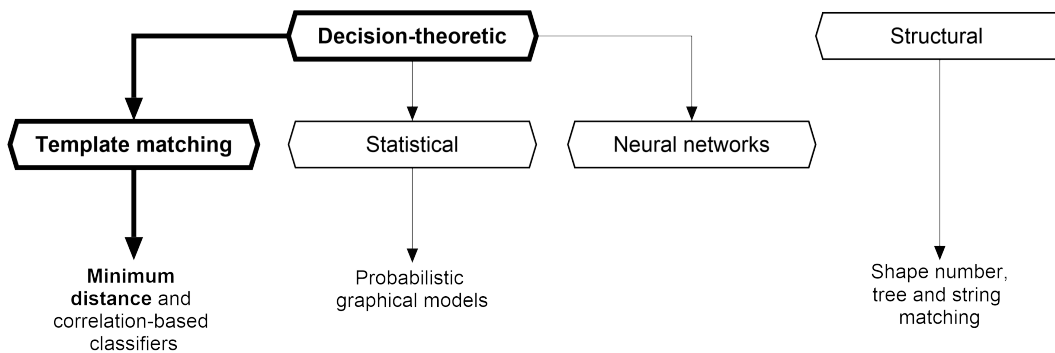


Figure 1.3: Different matching strategies. The matching strategy employed in this study is depicted in boldface.

1.3 Scope and objectives

The utilisation of the discrete Radon transform (DRT) for feature extraction, in conjunction with DTW for feature matching, proved to be successful within the context of identity verification based on *behavioural* biometrics, particularly static handwritten signatures (Coetzer (2005)). However, to the best of our knowledge, the above-mentioned combination of techniques has not been investigated within the context of identity verification based on *physiological* biometrics.

This study investigates the feasibility of utilising the above-mentioned techniques for the purpose of identity verification based on infrared images of hand veins on the dorsal and ventral surfaces of human hands. The successful development of such a system may prove beneficial within the context of automated access control to buildings and the validation of debit cards at automated teller machines (ATMs).

This study therefore aims to:

1. Develop an efficient and proficient *stand-alone* system for the purpose of authenticating an individual based on an infrared image of the hand veins on the dorsal or ventral surface of his/her hand. If authentication is unsuccessful after the first attempt, the individual may be allowed to present his/her hand again by, for example, re-orientating it.
2. Develop a system that performs adequately despite a lack of training data. In extreme scenarios, only *one* training sample may be available.
3. Develop a system that is robust in the sense that very little re-calibration of the system parameters is required when a different database is considered (provided that a high quality reference sample (template) is available for every enrolled individual).
4. Develop a system that performs well despite the presence of prominent grooves on the palmar surface, non-uniform illumination (shadows) and other artefacts like hair, wrist watches, arm bands, etc.
5. Appropriately conduct feature extraction and feature matching so that scale, rotation, and translation invariance are ensured. Intra-class variation is therefore minimised.
6. Employ suitable rank-based and score-based classifiers for the purpose of identity verification.

1.4 Thesis overview

In this section the reader is provided with a brief overview of (1) the proposed identity verification systems, (2) the experiments conducted in order to evaluate the performance of said systems within specific real-world scenarios, and (3) the experimental results.

1.4.1 System design

The designs of the proposed systems are based on the objectives listed in Section 1.3.

1.4.1.1 Data acquisition

Two existing data sets are considered for this research, namely (1) the Bosphorus data set (see Bogazici University (2010)) and (2) the CIE data set (see The Institute Of Control And Information Engineering (2013)). Samples in the Bosphorus data set constitute *near infra-red (NIR)* grey-scale images of hand

veins associated with a number of individuals captured from a *dorsal* (back-hand) point of view, while samples in the CIE data set constitute NIR images of the hand veins on the *palmar surface* of hands associated with a different set of individuals. Note that the dorsal surface of a hand is opposite to the ventral (palmar) surface.

1.4.1.2 Preprocessing and segmentation

The detection of the pixels associated with (prominent) hand veins is achieved in two stages, that is (1) the extraction of the *region-of-interest* (ROI) from the acquired image, followed by (2) the segmentation of the targeted hand vein pixels from said ROI. The protocol for extracting the ROI differs substantially for the two data sets employed (see Section 3.2). The novel hand vein segmentation protocol proposed in this thesis (within the context of *both* data sets considered) is conceptualised in Figure 1.4.

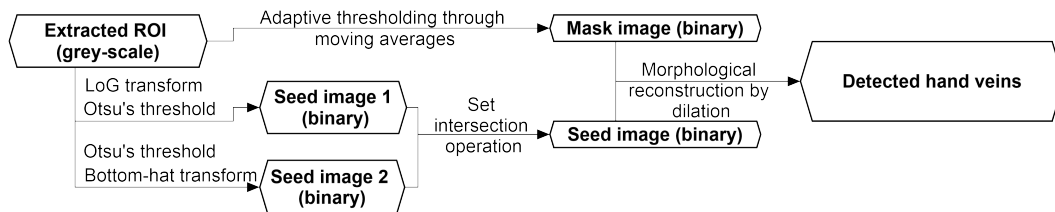


Figure 1.4: The proposed hand vein segmentation protocol.

1.4.1.3 Feature extraction

We propose to employ the discrete Radon transform (DRT) for the purpose of extracting features from a binary image of the ROI in which one-valued pixels are associated with hand veins, and zero-valued pixels with the background. Said transform involves the calculation of projection profiles from different projection angles (see Figure 4.5). A number of steps are taken towards achieving rotational, translational and scale invariance.

1.4.1.4 Feature matching and classification

We opt to evaluate the proficiency of the proposed systems when (1) a rank-based classifier and (2) a score-based classifier is employed for the purpose of feature matching and classification. The aforementioned classifiers may either employ (1) an Euclidean-based distance measure or (2) a DTW-based distance measure for the purpose of obtaining the dissimilarity between the feature sets associated with a questioned and reference sample.

1.4.2 Abbreviated results

In this section we provide a concise overview of the most pertinent results. A detailed discussion of the experimental protocols, as well as a detailed analysis of the experimental results, is reserved for Chapter 5.

The proficiency of the proposed identity verification systems is estimated by considering two different hand vein data sets. The so-called *Bosphorus* data set (see Section 5.2) contains 12 *dorsal* hand vein samples from 100 different individuals, while the so-called *CIE* data set (see Section 5.3) contains 12 *palmar* (ventral) hand vein samples from 50 different individuals.

The experiments conducted in this study simulate two main scenarios, which we refer to as the *k-one* and *k-six* scenario respectively. The *k-one* scenario constitutes the specific real-world situation in which a company, that aims to employ one of the proposed systems, acquires *only one* positive sample per individual during enrollment (see Section 5.5). The *k-six* scenario on the other hand constitutes the specific real-world situation in which said company acquires a total of *six* images per enrolled individual (see Section 5.6). Within the context of the *k-six* scenario it should therefore (in theory) be possible to enhance the proficiency of the employed system by extracting (and utilising) information from all six available images.

1.4.2.1 The benchmark experiments

In this section we present the results for the benchmark experiments that are conducted on the Bosphorus and CIE data sets. The benchmark experiments (1) employ the EER criterion for selecting an optimal decision threshold, (2) employ the Euclidean distance for obtaining a dissimilarity value between two feature sets and (3) grant an individual 3 opportunities for authentication.

Table 1.1: Experimental results for the benchmark systems. The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.

(%)	BOSPHORUS				CIE			
	<i>K-ONE</i>		<i>K-SIX</i>		<i>K-ONE</i>		<i>K-SIX</i>	
	Rank	Score	Rank	Score	Rank	Score	Rank	Score
FAR	24.05	26.35	16.02	16.14	43.94	46.7	38.77	38.65
FRR	5.43	8.36	1.08	1.86	9.62	18.51	5.03	5.49
AER	14.74	17.35	8.55	9.00	26.78	32.60	21.9	22.07
PRES	79.73	77.67	86.06	85.88	67.29	63.57	71.01	70.98
REC	94.57	91.64	98.92	98.14	90.38	81.49	94.96	94.51

1.4.2.2 The *ideal* benchmark experiments

In this section we present the results for the *ideal* benchmark experiments *only* conducted within the context of the *k-one* scenario. Said experiments are based on the *assumption* that a company that decides to employ a template-based authentication system is able to *ensure* that a *single* excellent quality template (reference sample) is acquired for *each* enrolled individual.

Table 1.2: The ideal benchmark results (within the context of the *k-one* scenario). The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.

(%)	BOSPHORUS		CIE	
	Rank	Score	Rank	Score
FAR	1.42	1.62	3.10	3.28
FRR	0.11	0.17	0.16	0.52
AER	0.77	0.89	1.63	1.90
PRES	98.58	98.4	80.21	96.81
REC	99.89	99.83	97.51	99.48

1.4.2.3 The most proficient systems

In this section we present the results for the most proficient systems within the context of the Bosphorus data set. Said systems specifically utilise (1) the Euclidean distance for dissimilarity measurement, (2) a rank-based classifier for verification purposes and (3) the FAR_{zero} criterion for threshold selection. An individual is also granted six opportunities for verification. The results when said systems are implemented on the CIE data set are also presented.

Table 1.3: Experimental results for the most proficient systems when implemented on the Bosphorus data set, as well as the corresponding results when the same systems are implemented on the CIE data set. The relevant statistical performance measures (FAR, FRR, AER, PRES, and REC) are discussed in detail in Section 5.4. Also see the list of abbreviations on page xiv.

(%)	BOSPHORUS		CIE	
	<i>K-ONE</i>	<i>K-SIX</i>	<i>K-ONE</i>	<i>K-SIX</i>
FAR	0.76	2.43	0.56	0.77
FRR	4.95	3.03	42.01	28.54
AER	2.85	2.73	21.29	14.66
PRES	99.21	97.55	99.04	98.93
REC	95.05	96.97	57.99	71.46

1.5 Contributions

The use of hand vein patterns as an authentication factor has recently become increasingly popular as newer and more compact imaging devices are being developed, since it is virtually resistant to fraud. If the intrusive and time-consuming nature of large imaging devices is eliminated, hand vein technology may replace many of the standing authentication factors such as bank cards at ATMs and POSMs. The main contributions of this research are listed below.

1.5.1 A number of novel, stand-alone hand vein-based identity verification systems

Although several hand vein-based identity verification systems have been successfully developed in the past, this study contributes towards the standing literature in the sense that a number of novel *stand-alone* hand vein-based identity verification systems are proposed. The proposed systems should therefore be complementary to any existing system when appropriately combined.

In scenarios where a company is able to ensure that a *high* quality template (reference sample) is acquired for all clients, the employed system parameters require very little re-calibration across different hand vein databases. The robustness and proficiency of the proposed systems are demonstrated in Chapter 5 by employing a number of popular statistical performance measures. The proficiency of the proposed system compares favourably with that of existing systems.

1.5.2 A novel hand vein segmentation protocol

A novel segmentation protocol for detecting hand vein patterns in NIR images is proposed, by defining a vein-region as a *dark* and *narrow* structure that is located within the vicinity of *well-defined* edges. It is demonstrated in Chapter 3 that the proposed hand vein detection (segmentation) protocol is accurate despite the presence of prominent grooves on the palmar surface, non-uniform illumination (shadows) and other artefacts like hair, wrist watches, arm bands, etc.

1.5.3 A novel feature extraction and matching protocol for hand vein-based identity verification

To the best of our knowledge, the utilisation of the DRT and DTW (in combination) for the specific purpose of extracting and comparing hand vein patterns is novel and constitutes one of the major contributions of this study.

Chapter 2

Literature Study

2.1 Introduction

In this chapter we present some of the leading hand vein-based authentication systems proposed in recent years. The pertinent components of the above-mentioned systems may be categorised into (1) the mathematical techniques proposed for the purpose of isolating (segmenting) the hand veins within an acquired image (see Section 2.3), (2) the algorithms proposed for extracting features from the isolated hand veins (see Section 2.4) and (3) the feature matching and classification protocols proposed for authentication purposes (see Section 2.5). We report on the proficiency of each of the aforementioned hand vein-based identity verification systems in Section 5.7.9.

The countries that are currently leading the research on employing hand vein patterns as an authentication factor include Turkey, China, Greece, Poland, India, Tunisia and Mauritius.

2.2 Data acquisition

An *infra-red* (IR) imaging device is typically required for the purpose of obtaining an image of an individual's hand veins, since hand veins cannot be penetrated by natural (visible) light. Two types of IR imaging devices are mainly employed for the aforementioned purpose, that is (1) *near* infra-red (NIR) and (2) *far* infra-red (FIR) devices. According to a study by Wang *et al.* (2007), FIR imaging devices operate by detecting electromagnetic radiation emitted from a human hand, and may therefore exploit the fact that blood vessels in the hand are generally warmer than the surrounding tissue. NIR imaging devices on the other hand exploit the fact that the *hemoglobin* in blood vessels absorbs incident (near) infra-red radiation, and therefore appear darker in a NIR image than the surrounding tissue (which reflects almost all NIR radiation).

The above-mentioned study investigates the quality of hand vein images when obtained with NIR and FIR imaging devices under specified conditions. It was firstly determined that a FIR imaging device may successfully capture prominent hand veins from a dorsal (back-hand) point of view, since the temperature of prominent (large) hand veins is significantly different from surrounding tissue. However, since FIR imaging devices are sensitive to changes in temperature (by design), they may not be able to acquire a good quality hand vein image in excessively hot and humid environments. For the same reason, said devices may also fail to discriminate between the surrounding tissue and hand veins that are located close to the surface of the hand.

Wang *et al.* (2007) also reported that NIR imaging devices are proficient in capturing veins from a dorsal *and* palmar point of view due to their discriminative nature, and the fact that NIR imaging devices are (by design) robust with respect to variations in the environmental and body temperature. However, the downside of employing NIR imaging devices lies in the fact that prominent local shadows, skin lesions, hair and palm lines are often also acquired, since said artefacts may typically absorb a similar amount of NIR light as the hemoglobin in the blood vessels within the hand.

Ai *et al.* (2016) proposed a novel (proclaimed) state-of-the-art hand vein imaging system. Said system employs two industrial cameras that are equipped with two (extra reflective) NIR light sources. The input from the aforementioned cameras are combined with that of a number of cameras that operate in the visible spectrum of light through a process referred to as camera *multiplexing*. This significantly improves the discriminative ability of the proposed imaging system. The proposed system also captures an individual's hand veins from a number of different angles, and subsequently creates a 3D reconstruction of the veins. In doing so, many unwanted artefacts remain undetected, while the structure of the reconstructed hand veins is robust across many samples of a specific individual's veins.

2.3 Preprocessing and hand vein segmentation

A number of *preprocessing* techniques are typically required for the purpose of reducing the effect of noise and enhancing the targeted hand veins. Trabelsi *et al.* (2013) propose that a median filter and a *Wiener* filter are employed for noise reduction purposes. Said study furthermore suggests that the *contrast-limited-adaptive-histogram-equalisation* (CLAHE) algorithm be employed for the purpose of enhancing the targeted hand vein regions within the filtered image.

Prasanna *et al.* (2012) showed that the application of *highboost* filtering followed by histogram equalisation constitutes a superior image enhancement protocol (within the context of hand veins) when compared to several other preprocessing techniques that include unsharp masking and grey-level slicing.

Note that the scope of the aforementioned study is limited to the preprocessing of hand veins.

Rossan *et al.* (2014) tested the proficiency of combining the CLAHE algorithm with a number of adaptive thresholding techniques for segmentation purposes, which include *mean-C* adaptive thresholding, *median-C* adaptive thresholding and *Wellner's* adaptive thresholding method. All of the aforementioned methods have been shown to produce trustworthy results.

Yuksel *et al.* (2011) employed *Yasuda's* binarisation method (see Sezgin and Sankur (2004)). This entails (1) a normalisation process, followed by (2) a non-linear smoothing operation, (3) an adaptive thresholding protocol and finally (4) the morphological *skeletonisation* of the thresholded image. They compared the proficiency of *Yasuda's* method to that of *Bernsen's* thresholding method, as well as *Niblack's* thresholding method (see Sezgin and Sankur (2004)). Yuksel *et al.* (2011) subsequently showed that *Yasuda's* method is superior to that of *Bernsen* and *Niblack*, since they claim that the non-linear smoothing operation preserves edges while also eliminating noise.

Ai et al. (2016) proposed a local adaptive vein segmentation method that is based on multiple feature clustering. Said method reportedly produces a more accurate segmentation of peripheral veins, bifurcation structures, and vein edges as compared to a Gaussian matched filter (GMF).

2.4 Feature extraction

Trabelsi et al. (2013) used a *Gaussian matched filter* (GMF) for the purpose of extracting features associated with hand veins. The aim was to develop a so-called *improved GMF* (IGMF) that constitutes a fused version of the standard GMF and its first derivative. Two unique GMF responses are therefore obtained; one associated with the original GMF and one associated with its first derivative. Said responses are subsequently fused in order to eliminate false positive results associated with either of the two responses.

Yuksel *et al.* (2011) employed (1) independent component analysis (ICA), (2) non-negative matrix factorisation (NMF) and (3) line-edge mapping (LEM) as feature extraction techniques within the context of dorsal hand veins. The use of ICA is motivated by the fact that its goal is to extract statistically independent variables from a number of (binarised hand vein) samples, and subsequently outputs a number of so-called *separation matrices*. Said matrices are therefore, by design, maximally separated, which is the mark of a well-designed feature extraction algorithm. When NMF is applied, the (binary hand vein) image matrix is factorised into basis vectors and *feature vectors*. LEM constitutes a technique that is used for the purpose of extracting prominent lines in a (binary hand vein) image as features. The author concluded that ICA is superior to NMF and LEM.

The features proposed by Rossan *et al.* (2014) constitute the coordinates of the white (1-valued) pixels within a binarised hand vein image associated with a specific individual. Said coordinates are stored in a 2D matrix that is used for correlation purposes during feature matching. These coordinate matrices are normalised across all of the coordinate matrices associated with the individuals enrolled into the system so as to ensure proper alignment and consistency.

2.5 Feature matching and verification

Trabelsi *et al.* (2013) employed an *artificial neural network* (ANN) for feature matching and classification purposes. Said ANN is constructed from features extracted via the IGMF from all images in a training set. During training, an optimal weight is assigned to each feature. Several different activation functions are considered. The classification errors associated with the respective activation functions are subsequently estimated, after which the activation function associated with the minimum average error rate is selected for testing. The author concludes that only 5 neurons are required in order to obtain acceptable results. The use of an ANN for the purpose of hand vein-based authentication is motivated (1) by the fact that the design of an ANN is fundamentally based on the biological nervous system, and (2) by the fact that an ANN ensures a robust transition between coding and decision space.

A number of minimum distance-based decision-theoretic classification strategies are proposed by Yuksel *et al.* (2011). The dissimilarity (distance) between a template and reference sample is obtained by employing three distance functions, that is (1) the L_1 -norm, (2) the Euclidean norm and (3) a cosine-based dissimilarity measure. Said study further proposes that three score-level and two decision-level fusion strategies are employed in an attempt to improve the performance of the individual classifiers. The aforementioned score-level fusion strategies include the (1) *product*, (2) *sum* and (3) *weighted sum* of the individual normalised scores. Said scores are normalised by employing the well known z-score. The decision-level fusion strategies include the so-called *Borda-count* (which constitutes a rank-based classifier) and *majority voting*.

Rossan *et al.* (2014) opted to employ the Euclidean distance for the purpose of obtaining the dissimilarity value between a reference and questioned sample. Said dissimilarity is simply compared to a threshold value in order to authenticate a questioned sample.

2.6 Concluding remarks

In the next chapter we discuss the novel hand vein segmentation protocol developed during the course of this study.

Chapter 3

Image segmentation

3.1 Introduction

In this chapter we propose a novel strategy for the extraction of prominent veins in *near infra-red* (NIR) images of human hands. The proposed strategy is robust and can be applied to NIR images that contain prominent veins on either the dorsal *or* palmar surface of human hands.

We demonstrate the feasibility of the proposed hand-vein segmentation protocol by considering two different data sets that are discussed in detail in Sections 5.2 and 5.3. Samples from the so-called Bosphorus data set constitute NIR images of the dorsal surface of human hands (see Figure 3.1 (a)), while samples from the so-called CIE data set constitute NIR images of the palmar surface of human hands (see Figure 3.1 (b)).

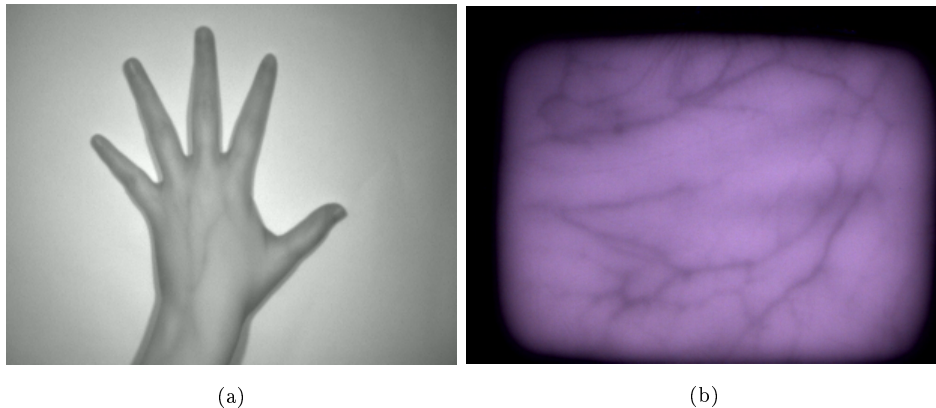


Figure 3.1: (a) A typical NIR image from the Bosphorus data set of size 300×240 pixels. (b) A typical NIR image from the CIE data set of size 1024×768 pixels.

The NIR images from the CIE data set are acquired in RGB format, but are subsequently converted into grey-scale format (see Figure 3.2 (c)), since hue-related information is insignificant within the context of the proposed protocol.

The main objective of said protocol is therefore to extract the veins indicated in yellow in Figures 3.2 (b) and (d). Said veins were manually annotated and may therefore serve as a *ground truth*.

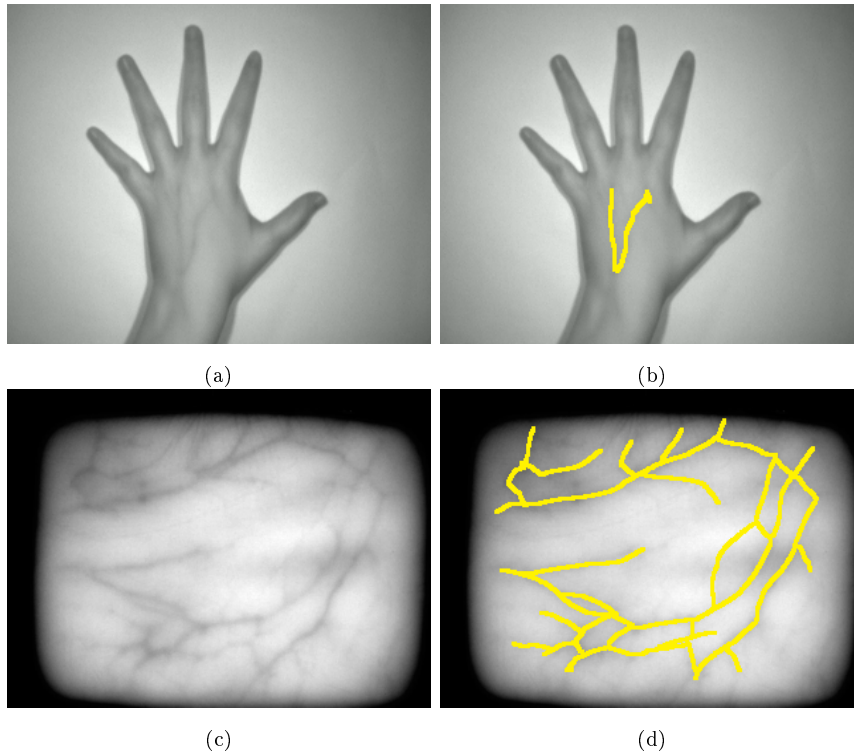


Figure 3.2: (a) The image in Figure 3.1 (a). (b) The targeted veins in (a) manually annotated and superimposed in yellow onto (a). (c) The image in Figure 3.1 (b) after being converted into grey-scale format and resized to 320×240 pixels. (d) The targeted veins in (c) manually annotated and superimposed in yellow onto (c).

The main steps involved in achieving the above-mentioned objective can be summarised as follows:

1. Extract the *region of interest* (ROI), which is the first major relevant sub-image within a questioned sample. This step is detailed in Section 3.2, and is mainly applicable to samples in the Bosphorus data set, since the targeted ROI for both data sets is located on the body of the hand and not the fingers or wrist.
2. Separate the pixels within the extracted ROI into *object* pixels and *background* pixels, where the targeted object pixels are denoted in yellow in Figures 3.2 (b) and (d). This step is equally important within the context of both data sets, and is detailed in Section 3.3.

The proposed hand-vein extraction protocol is graphically conceptualised in Figure 3.3 and discussed in detailed in the remainder of this chapter.

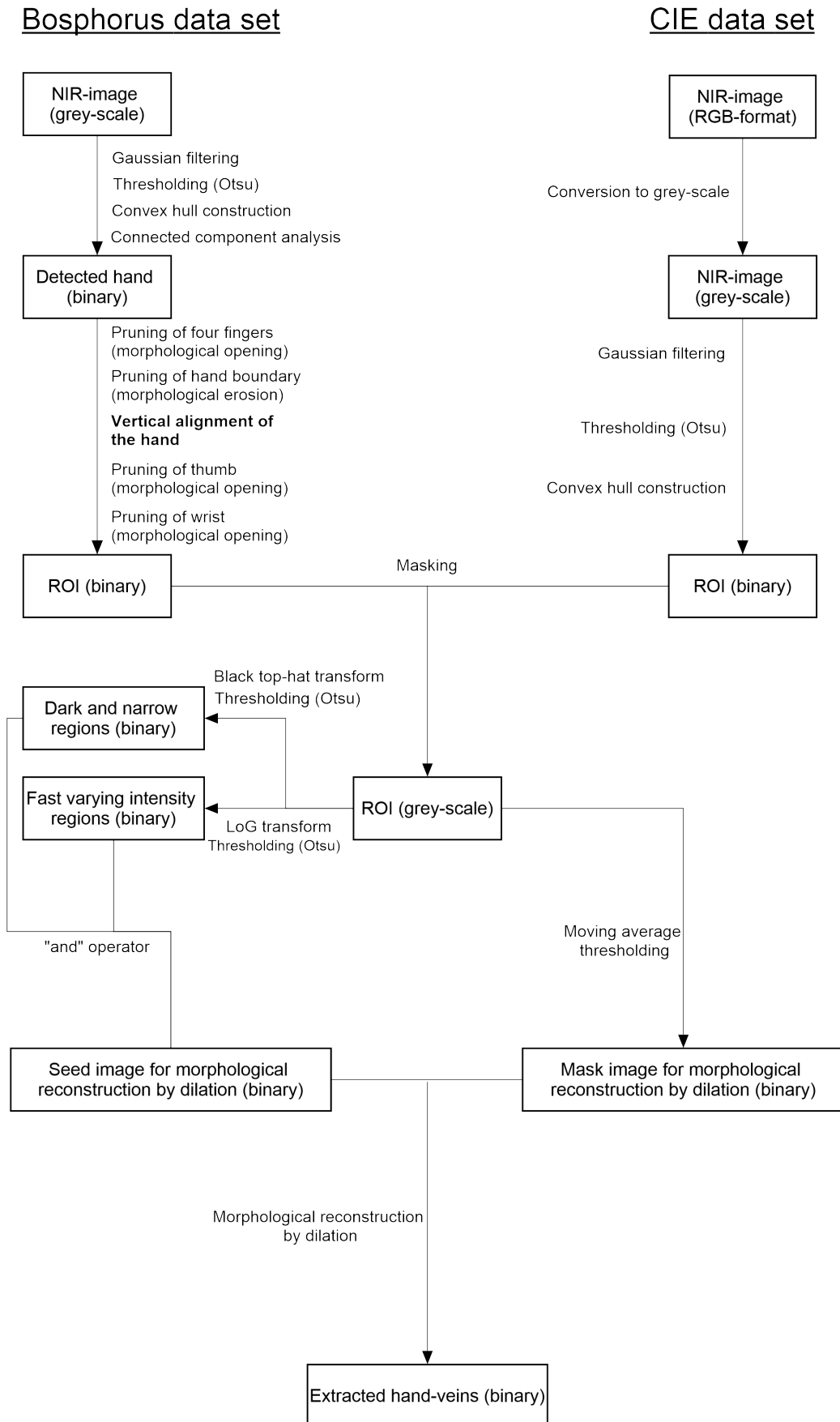


Figure 3.3: Conceptualisation of the proposed hand-vein extraction protocol.

3.2 Extracting the ROI

The prominent veins in the context of the Bosphorus data set are located on the body of the *dorsal* surface of the hand (see Figure 3.2 (b)). This surface, without the fingers and wrist, therefore represents a suitable ROI for samples in said data set. A suitable ROI for samples in the CIE data set is much simpler to obtain (see Section 3.2.6), since only the palmar surface of the body of the hand is originally captured (see Figure 3.2 (c)). Figure 3.4 shows manually annotated boundaries of the targeted ROIs for typical samples in the aforementioned data sets.

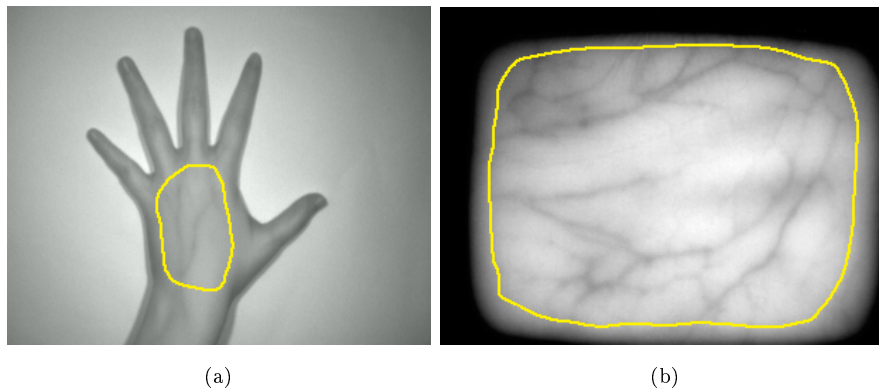


Figure 3.4: Manually annotated boundaries (in yellow) of the targeted ROIs for the images depicted in Figures 3.2 (a) and (c).

Note that the boundaries of the targeted ROIs are chosen to be marginally inside the actual boundary of the dorsal or palmar surface of the body of the hand. This strategy is followed to (1) ensure that shadows in close proximity to the actual boundary of the hand are not misinterpreted as prominent veins, and (2) to implement prior knowledge that the targeted prominent veins are not located in the immediate vicinity of the actual boundary of the hand. The task of defining the ROI within the context of the Bosphorus data set may be broken down into the following simpler tasks:

1. The detection of the hand (see Section 3.2.1).
2. The pruning of the four fingers (see Section 3.2.2).
3. The pruning of the boundary of the hand (see Section 3.2.3).
4. The vertical alignment of the hand (see Section 3.2.4).
5. The pruning of the thumb and wrist (see Section 3.2.5).

3.2.1 Detecting the hand

When the images in Figures 3.2 (a) and (c) are compared, it is clear that this stage is only applicable to samples from the Bosphorus data set, for which the following observations will aid in the detection of the hand:

1. The majority of pixels that form part of the hand are much darker than the pixels in the immediate vicinity of the hand. This is a result of the NIR lighting equipment employed during image acquisition.
2. The hand is represented by the majority of the darker pixels in the image, while the background is represented by the majority of the lighter pixels.

These two observations guide us towards adopting a *thresholding* approach for detecting the hand. Thresholding involves the application of a specific threshold value to a grey-scale image $f(x, y)$, where $f \in [0, 255]$, thereby converting it into a *binary* image. In the remainder of this thesis we shall refer to a set of white (1-valued) 8-connected pixels within a binary image as an *object*. If a certain pixel within $f(x, y)$ therefore has a value greater than the specified threshold value, it is assigned a value of 1. The remaining pixels are assigned a value of 0 (rendered black) and form part of the *background*.

We opt to employ a well-known thresholding method referred to as *Otsu's* method (Otsu (1979)). Said method considers each possible threshold value $t \in [0, 255]$ and selects the one that maximizes the *inter-class variance*. In doing so, said algorithm ensures the best possible separation between the actual object and its background. The principles involved in Otsu's method are conceptualised in Figure 3.5, where the optimal threshold is located roughly midway between the two peaks of a bimodal histogram.

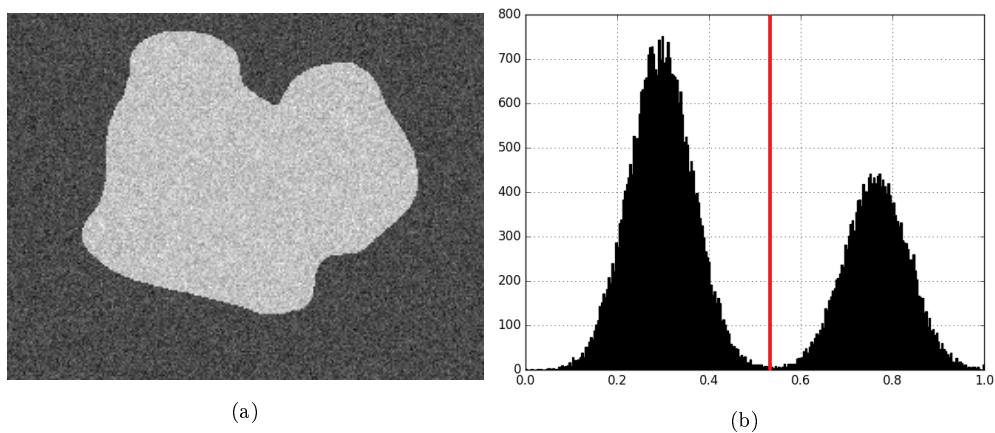


Figure 3.5: (a) A grey-scale image corrupted by Gaussian noise. (b) The histogram of the image in (a), where the location of the appropriate threshold value as determined by Otsu's algorithm is denoted by the red vertical line.

Otsu's algorithm proves useful in our case since the hand and its immediate background are well contrasted (see Figure 3.6 (a)), which leads us to expect at least a well-defined hand boundary.

The result of applying Otsu's method to the image depicted in Figure 3.6 (a) is shown in Figure 3.6 (b). This constitutes the first step towards the detection of the hand. Note that within the context of the system proposed in this thesis, a Gaussian smoothing filter is applied to all grey-scale images prior to the application of a thresholding operation.

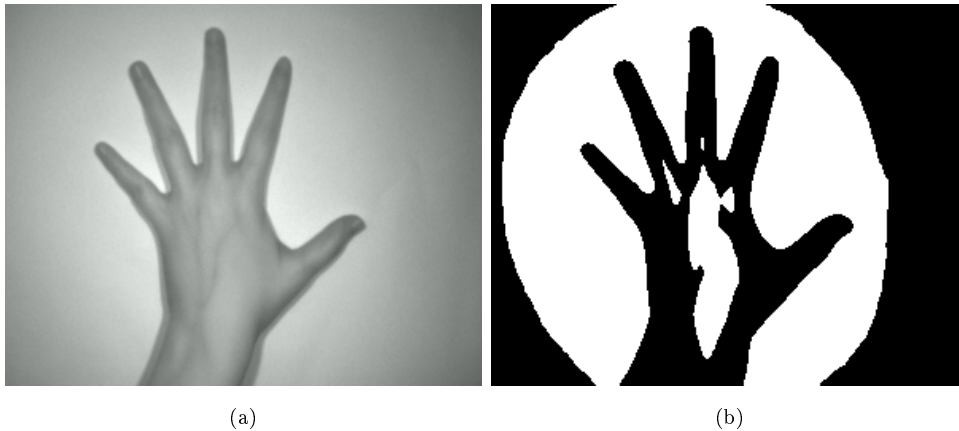


Figure 3.6: (a) The image in Figure 3.1 (a). (b) The binarised version of (a) after the application of Otsu's threshold.

At first glance, the application of Otsu's threshold may appear to have caused certain complications, since arbitrary objects are now present near the center of a black, hand-shaped region, while large black regions are also present in all four corners (see Figure 3.6 (b)). Both of these complications are however the result of the lighting source used during image acquisition. This brings us to the next useful observation that enables us to unify the large black corner regions with the actual immediate background of the hand (The white regions within the black, hand-shaped region will be removed afterwards.) Note how the application of Otsu's threshold invariably results in a solid white object that surrounds a black, hand-shaped region (see Figure 3.6 (b)). This is due to the nature of the light source used for NIR imaging. The hand allows some penetration of the NIR light energy, but the surface upon which the hand rests is solid and reflects almost all NIR light energy (see Figure 3.6 (a)). The pixels surrounding the actual hand are therefore much brighter than the hand and the black corner regions which are located furthest from the light source.

In order to isolate all the pixels that belong to the large black corner regions, we exploit the fact that said white object will invariably surround a black, hand-shaped region after the application of Otsu's threshold within the context of the Bosphorus data set. Said background regions are therefore detected by using the morphological concept of a *convex hull*.

Per definition, a convex object is an object for which a line between any two pixels that belong to the object also belongs to said object (see Figure 3.7 (a)). A non-convex object is depicted in Figure 3.7 (b) where it is clear that a segment of line L_1 between points P_1 and P_2 lies outside said object.

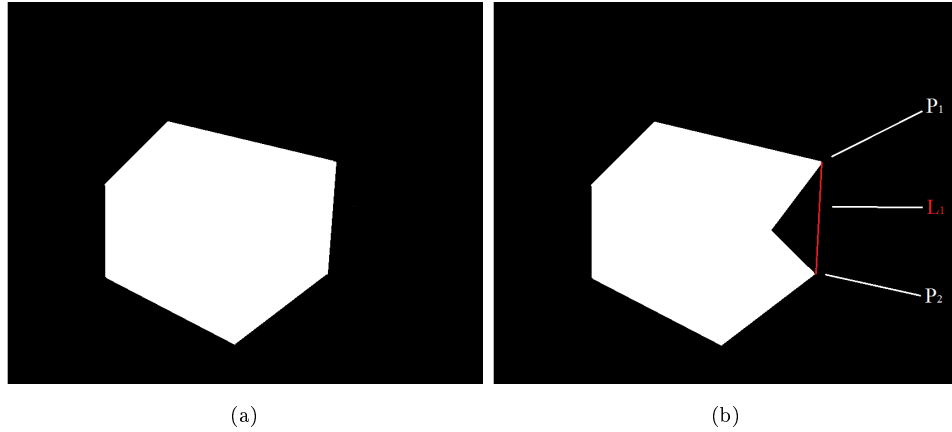


Figure 3.7: (a) A convex object. (b) A non-convex or *concave* object.

The convex *hull* of an object is the smallest convex object that contains said object. When line L_1 is added to the boundary of the object in Figure 3.7 (b), the boundary of the convex hull of said object is obtained.

The convex hull of the largest white object in Figure 3.6 (b) therefore constitutes both the large white object and the pixels associated with the black, hand-shaped region (see Figure 3.8 (b)).



Figure 3.8: (a) The image in Figure 3.6 (b). (b) The convex hull of (a).

The location of all the pixels that belong to the outer black background regions is now known. These pixel values can therefore be negated and consequently merged with the large white object that surrounds the black, hand-shaped region in Figure 3.8 (a). The result is shown in Figure 3.9.



Figure 3.9: Figure 3.8 (a) after the pixels that belong to the outer black background regions were negated. Note that the boundary of the image is not explicitly shown.

All that remains is to remove the white artefacts within the remaining black, hand-shaped region. This is accomplished by retaining only the two largest connected components within the image depicted in Figure 3.9, namely the surrounding white object and the black, hand-shaped region. The small white objects are therefore negated and consequently merged with their immediate surroundings. The resulting image is finally negated so that the only remaining object corresponds to the detected hand (see Figure 3.10).



Figure 3.10: The detected hand.

3.2.2 Pruning the four fingers

We accomplished the task of detecting the hand by exploiting the fact that the hand is *dark* when compared to its immediate surroundings. The removal of the fingers must be approached differently, since they have similar grey-scale intensities when compared to the rest of the hand.

This motivates us to consider *morphological* image processing techniques, which exploit prior knowledge about the shape and size (morphology) of the desired region. Within said context a clear distinction can therefore be made between the dorsal surface of the body of the hand and the fingers.

The language of image morphology is based on set theory. Most morphological operations employ an appropriately shaped and sized sub-image, called a *structuring element* (SE), in conjunction with set operations in order to extract certain regions within an image. The nature of the SE therefore determines the regions that are retained or removed. Note that we mainly employ morphological operations that are suited for binary images.

In order to distinguish the fingers from the body of the hand, we exploit the fact that the fingers are relatively *narrow* when compared to the main dorsal surface of the hand. The operation of choice for removing narrow parts of objects within binary images is morphological *opening*. In particular, morphological opening removes the parts of objects that are smaller than the SE.

Said operation can be demonstrated by moving the SE in question along the inside of the boundary of the object in question and retaining only those parts of the boundary for which the SE fits entirely. In other words, if the SE cannot be moved alongside certain parts of the inner boundary of an object, said parts are removed through negation.

In Figure 3.11 a simple example of morphological opening is presented, in which case the relevant SE is disc-shaped and depicted in blue. The brown regions in Figure 3.11 (b) are therefore removed (see Figure 3.11 (c)).

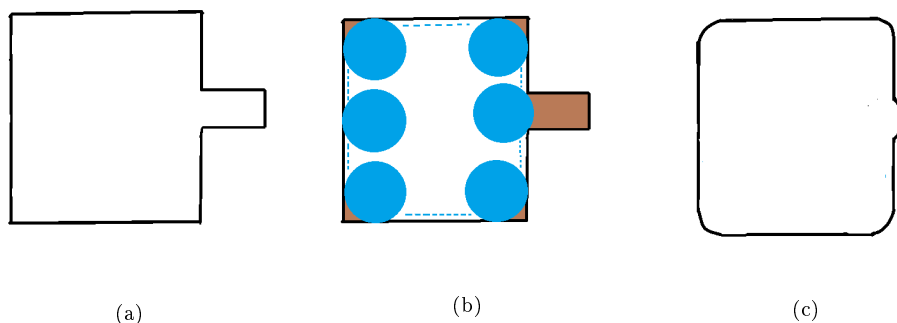


Figure 3.11: (a) The boundary of an object. (b) Conceptualisation of the morphological opening of the object in (a) with the blue disc-shaped SE. (c) The result of opening (a) with the disc-shaped SE depicted in (b), where it is clear that the brown regions in (b) are removed.

The result of applying morphological opening to the image depicted in Figure 3.10, by a disc-shaped SE of which the diameter slightly exceeds the estimated width of a finger, is shown in Figure 3.12 (b). It is clear from Figure 3.12 (b) that the four fingers are successfully pruned.



Figure 3.12: (a) The image in Figure 3.10. (b) The result of morphologically opening the hand in (a) with a disc-shaped SE.

3.2.3 Pruning the hand boundary

The boundary region of the remaining part of the body of the hand invariably contains dark shadows that may be mistaken for hand-veins (see Figure 3.14 (b)). It is therefore advisable that said region is also removed.

This task is accomplished through morphological *erosion*, which effectively shrinks objects within a binary image. The morphological erosion of an object by a certain SE is achieved by specifying that the center of the employed SE forms part of the eroded object if and only if the entire SE fits into the object. Figures 3.13 (b) and (c) illustrate the effect of morphologically eroding an annular object (depicted in Figure 3.13 (a)) by a square and disc-shaped SE respectively, where the eroded regions are depicted in yellow.

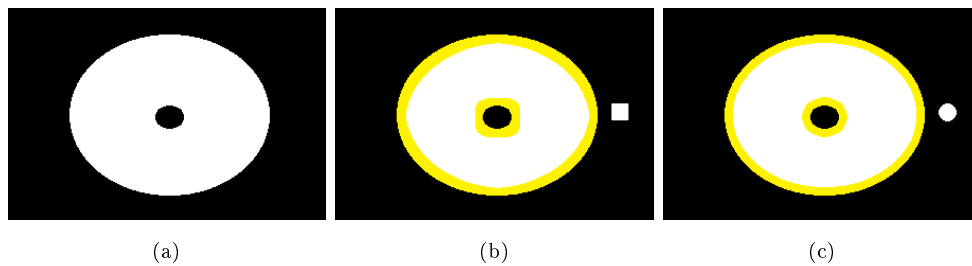


Figure 3.13: (a) A 300×200 image containing an annular object. (b) The morphological erosion of (a) by the 16×16 square-shaped SE shown in white. (c) The morphological erosion of (a) by the 16×16 disc-shaped SE shown in white.

From Figure 3.13 (c) it is clear that the object is shrunk by roughly half of the width of the square-shaped SE. Said SE is therefore well-suited for the removal of the dark boundary of more or less uniform thickness that is clearly visible in Figure 3.14 (b). The estimated width of said boundary region is 8 pixels, and therefore morphological erosion by a 16×16 square-shaped SE successfully removes the relevant region, as is clearly the case in Figures 3.14 (c) and (d).

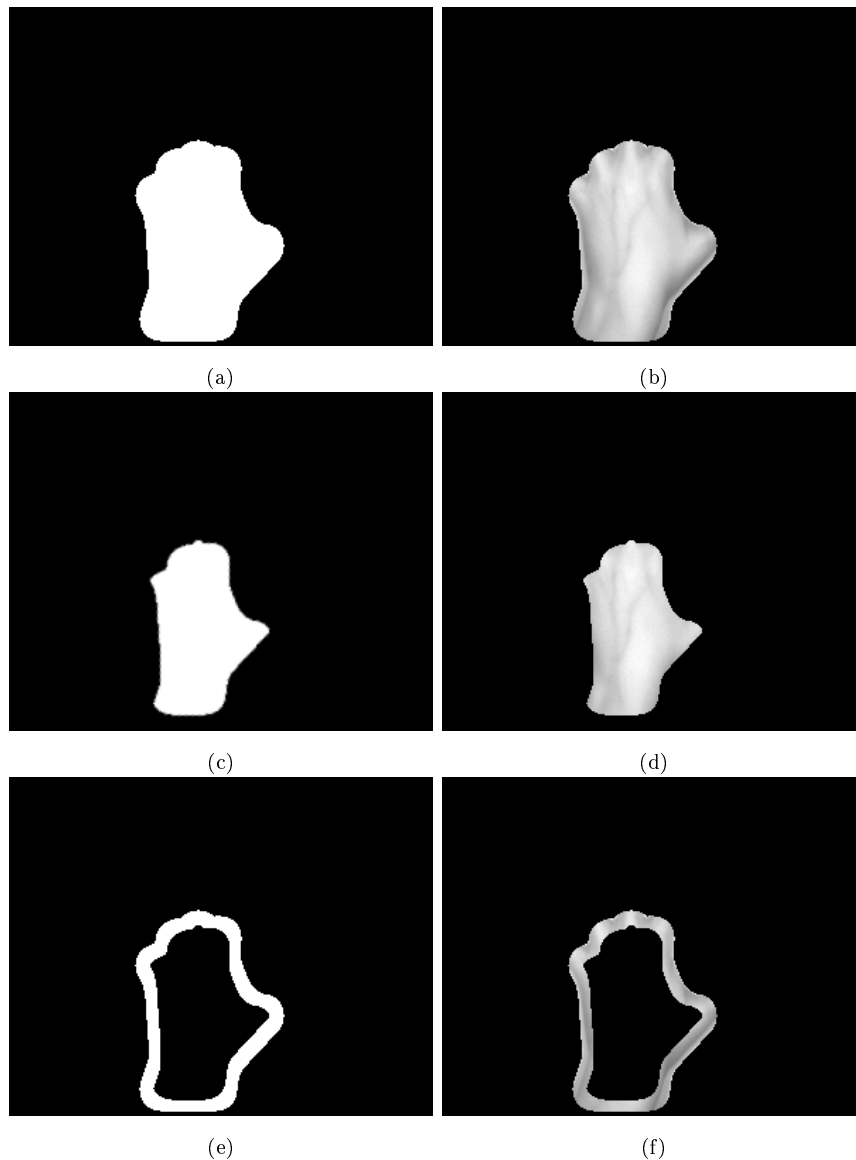


Figure 3.14: (a) The object in Figure 3.12 (b). (b) The current ROI, where dark shadows are clearly visible in close proximity to the hand boundary. (c) The morphological erosion of (a) by a 16×16 square shaped SE. (d) The new ROI. (e) The eroded part of (a). (f) The dark boundary region removed from (b).

3.2.4 The vertical alignment of the hand

It is important to note that all of the SEs employed up to this stage are *isotropic* in the sense that their effect is independent of the orientation of the input image. Also note that the orientation of the hands within the context of the Bosphorus data set may vary substantially from sample to sample.

At this stage of system design, that is after the removal of the hand boundary and before the removal of the thumb, all images within the Bosphorus data set are rotated (re-orientated) in such a way that they are approximately aligned in the vertical (upright) direction. The orientation of the body of a hand is specified to be approximately perpendicular to the axis along the length of the thumb, and approximately parallel to the line connecting the *center of mass* of the object depicted in Figure 3.12 (b), and the webbing between the pruned middle and ring finger of the same object. This line is indicated in red in Figure 3.15 (a), from which it is clear that the body of the hand is not perfectly vertical, although the hand may appear to be vertical at first glance. This robust method is used to vertically align all the hands in the Bosphorus data set.

The aforementioned re-orientation procedure forms part of the normalisation techniques discussed in Chapter 4, but is however applied at this stage of system design for the purpose of convenience. Note that in the remainder of this chapter it is therefore assumed that all the hands in the Bosphorus data set are vertically aligned (see Figure 3.15 (b)).



Figure 3.15: (a) The image depicted in Figure 3.12 (b), where the line that is used to approximate the orientation of the body of the hand is manually annotated in red. (b) The vertically oriented version of (a).

The purpose of the above-mentioned re-orientation procedure within the context of ROI extraction is to facilitate the pruning of the thumb and wrist, which is detailed in Section 3.2.5. The robust method for finding the red line depicted in Figure 3.15 (a) is discussed in the following section.

3.2.4.1 Estimating the orientation of the hand

The boundary of the detected hand is first obtained through morphological boundary extraction (see Figure 3.16 (a)). The morphological boundary of an object is defined as the difference between said object and its erosion by a 3×3 square-shaped SE.

The intersection of the boundary of the hand and the opened hand (as depicted in Figure 3.16 (d)) is subsequently calculated. Information pertaining to the location and morphology of the objects contained within said intersection is then considered so as to retain only the object associated with the webbing between the middle and ring fingers (see Figure 3.16 (f)).

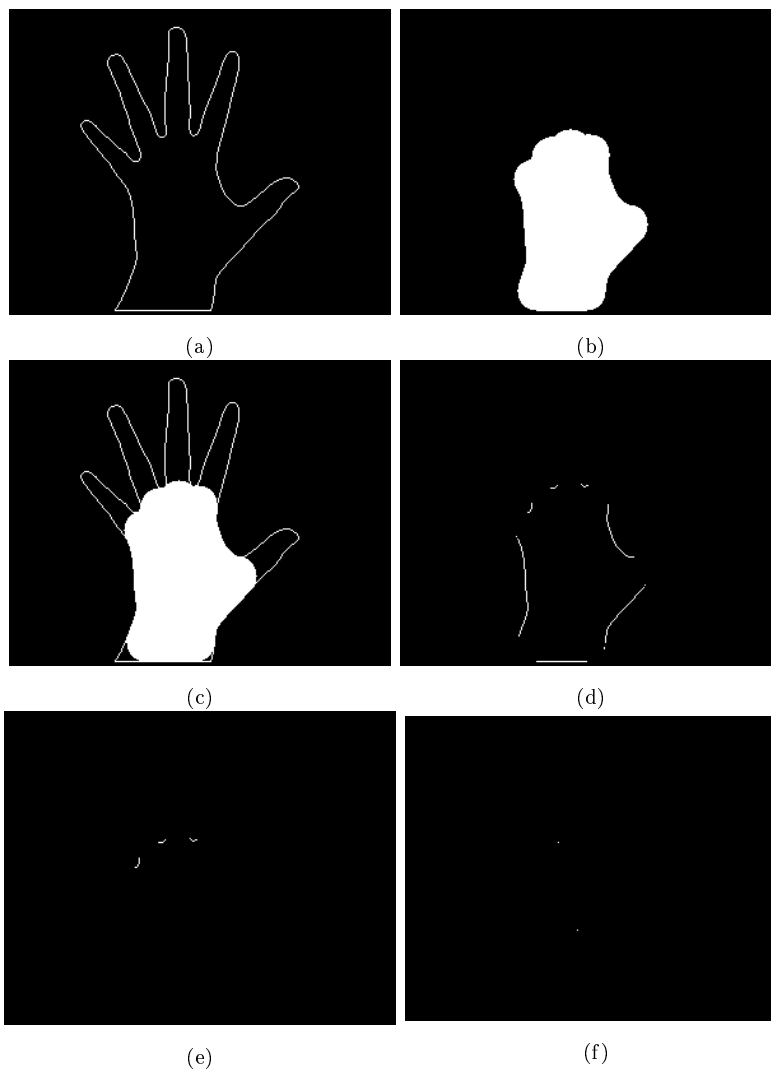


Figure 3.16: (a) The boundary of the detected hand. (b) The opened hand. (c) The union of (a) and (b). (d) The intersection of (a) and (b). (e) The webbing between the four fingers. (f) The targeted reference points: the centre of mass of (b) and the webbing between the middle and ring fingers.

3.2.5 Pruning the thumb and wrist

In order to remove the thumb, the vertically oriented version of the object depicted in Figure 3.14 (c) is morphologically opened by a vertical, one-dimensional SE (see Figure 3.17 (b)) of which the length is 80% of the estimated height of said object. This height is estimated to be twice the length of the red line depicted in Figure 3.15 (b).



Figure 3.17: (a) The vertically oriented version of the object depicted in Figure 3.14 (c). (b) The result of applying morphological opening to (a) by a vertical, one-dimensional SE.

The wrist and the remainder of the four fingers that form part of the object depicted in Figure 3.17 (b) are subsequently removed through the morphological opening by a tailor-made disc-shaped SE, of which the diameter is approximately equal to 80% of the width of said object. This width is estimated to be twice the length of the red line connecting the center of mass of said object and the pixel on the boundary of the object towards its right (see Figure 3.18).

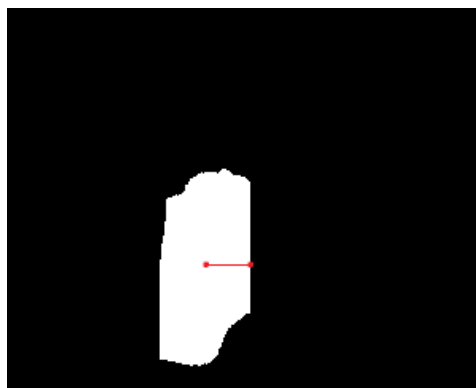


Figure 3.18: Conceptualisation of the proposed procedure for estimating the width of the object depicted in Figure 3.17 (b).

Figure 3.19 (a) shows the result of morphologically opening the object in Figure 3.17 (b) by the aforementioned disc-shaped SE. Figure 3.19 (b) presents the final ROI associated with the vertically oriented version of the image shown in Figure 3.2 (a). The non-ROI pixels are darkened for effect.

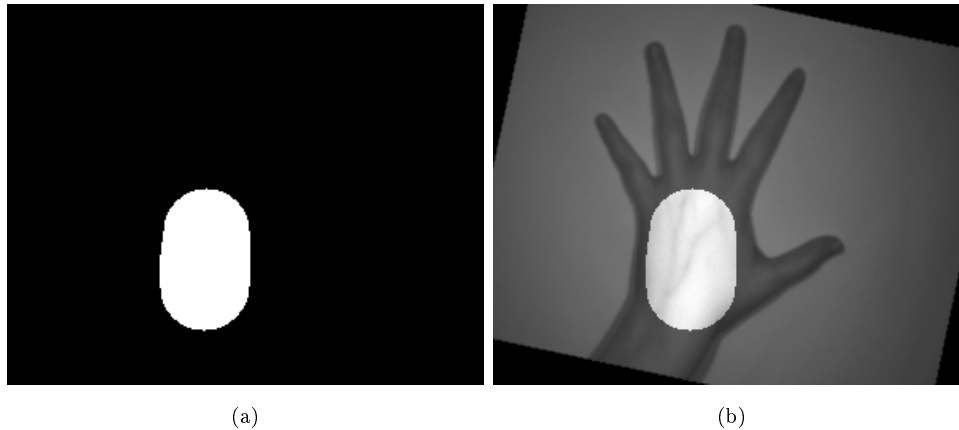


Figure 3.19: (a) The morphological opening of the object in Figure 3.17 (b) by a tailor-made disc-shaped SE. (b) The final ROI for the vertically oriented version of the image depicted in Figure 3.2 (a). The non-ROI pixels are darkened for effect.

The proposed protocol for extracting suitable ROIs from images in the Bosphorus data set is robust with respect to variations in orientation, position and scale. It is also resilient to noise and artefacts such as jewellery in the vicinity of the wrist (see Figure 3.20). This concludes the detailed description of the proposed protocol for extracting the ROI within the context of the Bosphorus data set.

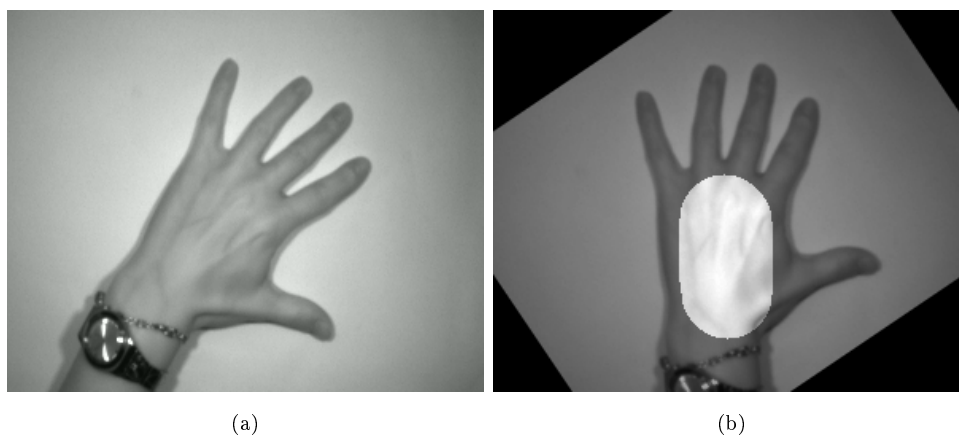


Figure 3.20: (a) A sample from the Bosphorus data set that contains jewellery in the vicinity of the wrist. (b) The final ROI for the vertically oriented version of the image shown in (a). The non-ROI pixels are darkened for effect.

The final ROIs for the vertically oriented versions of a random selection of NIR images from the Bosphorus data set are presented in Figure 3.21.

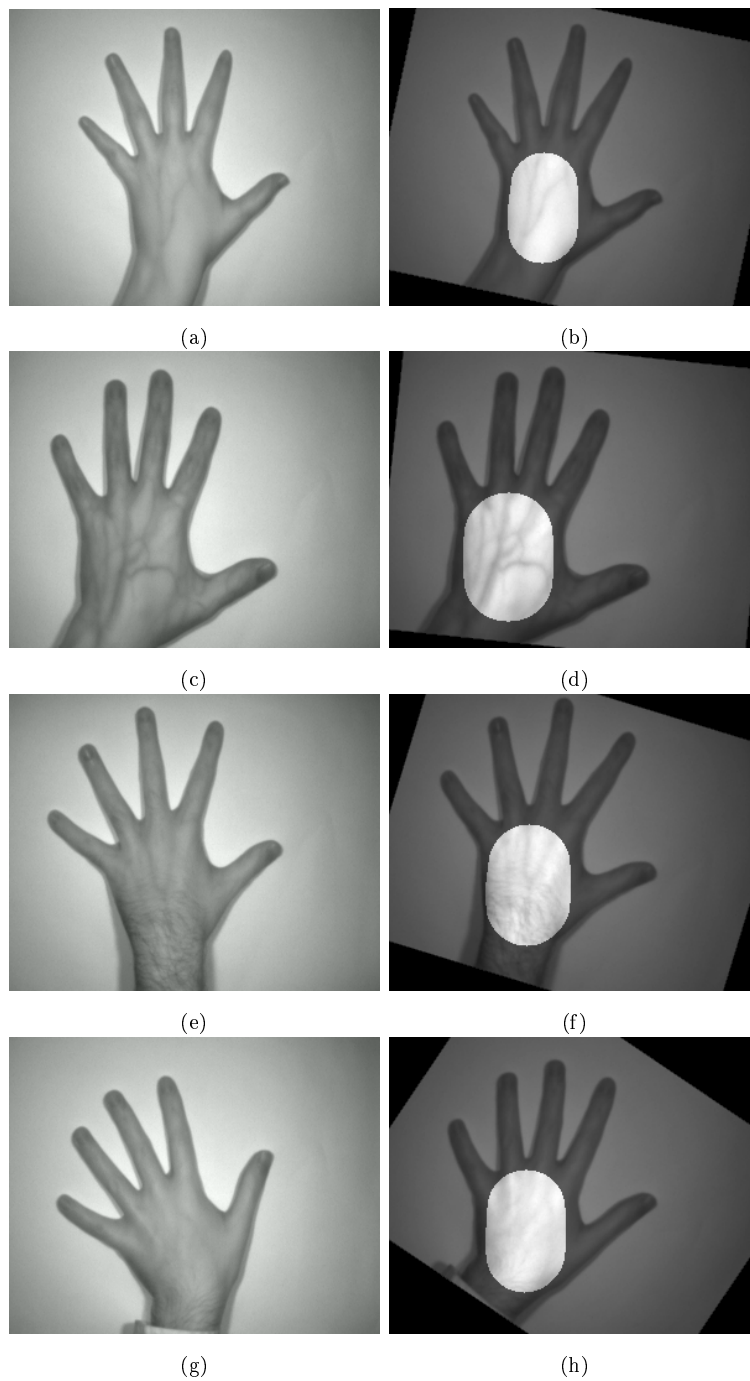


Figure 3.21: (Left) A random selection of samples from the Bosphorus data set. (Right) The final ROIs for the vertically oriented versions of the corresponding images on the left.

3.2.6 Extracting the ROI: The CIE data set

The ROI for the samples in the CIE data set is extracted in two steps: (1) The application of Otsu's algorithm to the original sample in grey-scale format results in a large white object surrounded by a thick black frame. (2) The convex hull of said object constitutes the final ROI. We present the final ROI for a random selection of samples in the CIE data set in Figure 3.22.

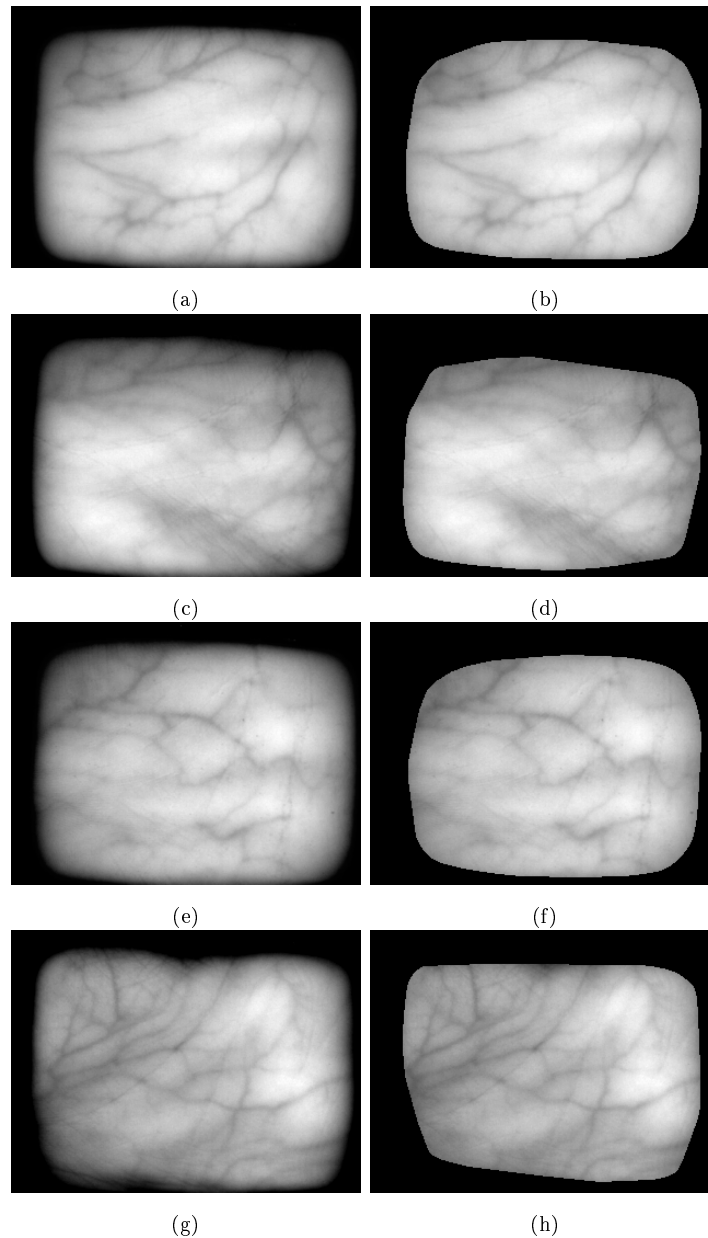


Figure 3.22: (Left) A random selection of NIR images from the CIE data set after being converted into grey-scale format. (Right) The final ROI for the corresponding samples on the left.

3.3 Extracting the veins

The key to the successful extraction of prominent veins from NIR images as proposed in this chapter lies in the simplicity of the individual steps.

In Figure 3.23 the final ROIs depicted in Figures 3.21 (d) and (f), as well as those depicted in Figures 3.22 (d) and (h) are presented again in order to draw attention to certain robust and reliable attributes typically associated with prominent veins.

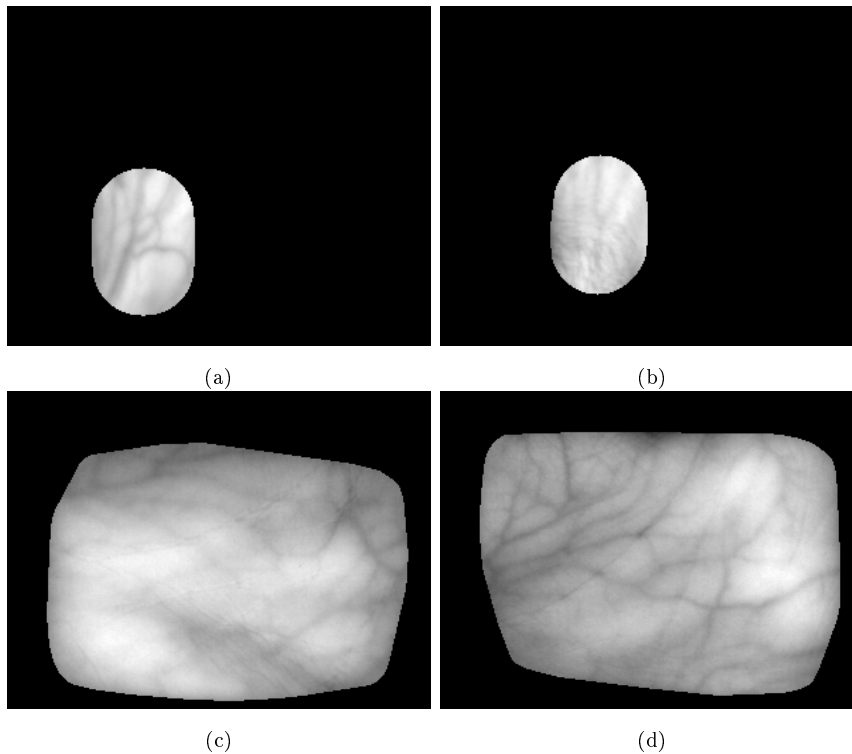


Figure 3.23: The ROIs for selected images in Figures 3.21 and 3.22. These ROIs are presented again in order to highlight key differences between vein and non-vein regions.

The first noticeable attribute is that veins are relatively *dark* when compared to their immediate vicinity. Furthermore, in a morphological sense, veins typically constitute *narrow* structures which are located in the immediate proximity of *fast varying spatial intensities*. Said attributes enable us to stipulate the following three key criteria that a region in a NIR image has to satisfy in order to be a good hand-vein candidate:

1. Relatively dark pixel values.
2. Narrow morphology.
3. In the immediate vicinity of fast varying spatial intensities.

Strategies for exploiting the first two criteria are detailed in Section 3.3.1, while a strategy for exploiting the third criterion is discussed in Section 3.3.2.

Exceptions do however occur where non-vein regions may satisfy the criteria outlined above, as is often the case with hair, dark localised shadows and palm lines, where a palm line constitutes a groove in the skin of the palmar surface. Examples of hair and palm lines are manually annotated in yellow in Figures 3.24 (b), (c) and (d) and local shadows are encircled in green in Figures 3.24 (a), (c) and (d).

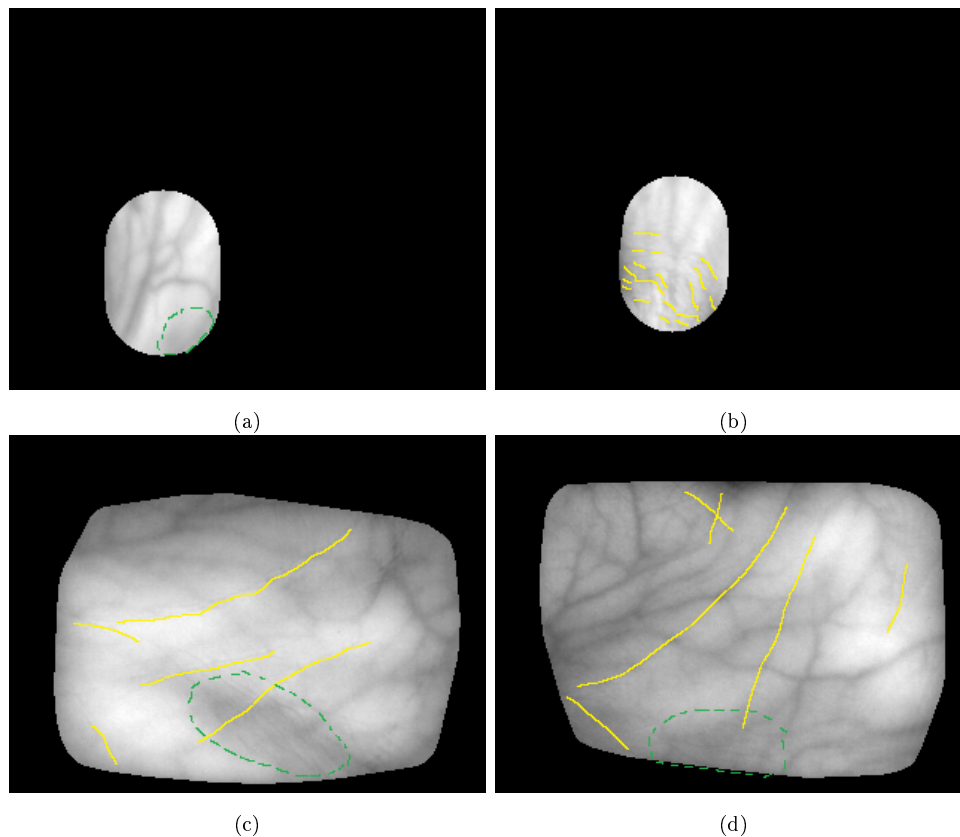


Figure 3.24: (a) Figure 3.21 (d) that depicts a ROI that contains a local shadow. (b) Figure 3.21 (f) that depicts a ROI that contains hair. (c) Figure 3.22 (d) that depicts a ROI that contains palm lines and a local shadow. (d) Figure 3.22 (h) that depicts a ROI that contains palm lines and a local shadow.

The removal of said non-vein regions are beyond the scope of this research. We rather opt to design our hand-vein segmentation protocol in such a way that it is robust with respect to noise, which may to a certain extent account for some of the above-mentioned artefacts. The inclusion of palm lines may in fact *add* to the discriminative potential of said protocol, since palm lines are also unique to a specific individual.

3.3.1 Extracting dark and narrow regions

In order to identify dark and narrow regions within a grey-scale image, we opt to employ the *black top-hat* transform (Gonzales (2010)). Said transform constitutes a morphological operator that extracts dark regions within a grey-scale image that are smaller than the SE being employed. This transform is therefore well-suited for detecting dark and narrow regions.

However, since this operator is applied to a grey-scale image, the pixel values in the resultant image constitute weighted responses of how well each pixel fits the targeted attributes. The result is therefore binarised through Otsu's algorithm, so that only the strongest responses are retained.

The black top-hat transform of the images depicted in Figure 3.23 are binarised through Otsu's method and presented in Figure 3.25. The SE in question is disc-shaped with a diameter that marginally exceeds the estimated thickness of a vein.

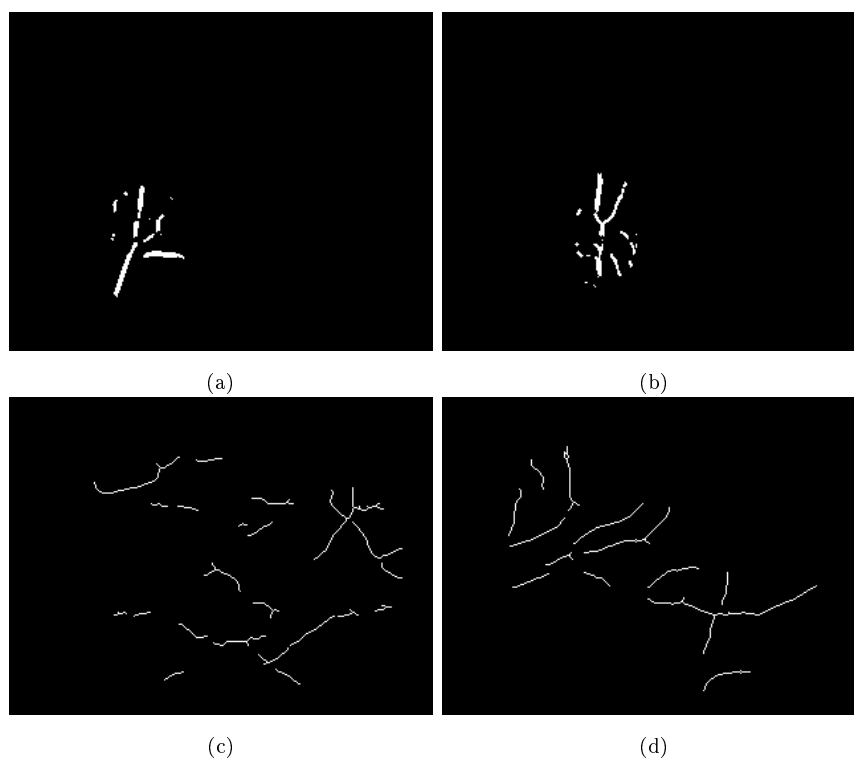


Figure 3.25: The black top-hat transforms of the corresponding images in Figure 3.23, after binarisation through Otsu's method.

The results presented in Figure 3.25 appear to be accurate even for those cases where local shadows are present (see Figures 3.24 (a), (c) and (d)), since local shadows are typically associated with weaker responses to the black top-hat transform than is the case for hand veins. However, some of the prominent hairs indicated in Figure 3.24 (b) are unavoidably detected.

3.3.2 Extracting regions that contain fast-varying intensities

The *Laplacian-of-Gaussian* (LoG) transform is well suited for detecting pixels that satisfy the third criterion outlined at the start of this section. The LoG approximates the second derivative of an image at a given location so that higher responses are typically associated with pixels in regions that are characterised by large abrupt changes in grey-scale intensity.

Similar to the approach followed in the previous section, we again select the strongest responses to the LoG transform through the application of Otsu's algorithm. The images in Figure 3.26 shows the binarised versions of the LoG transformations of the corresponding images in Figure 3.23.

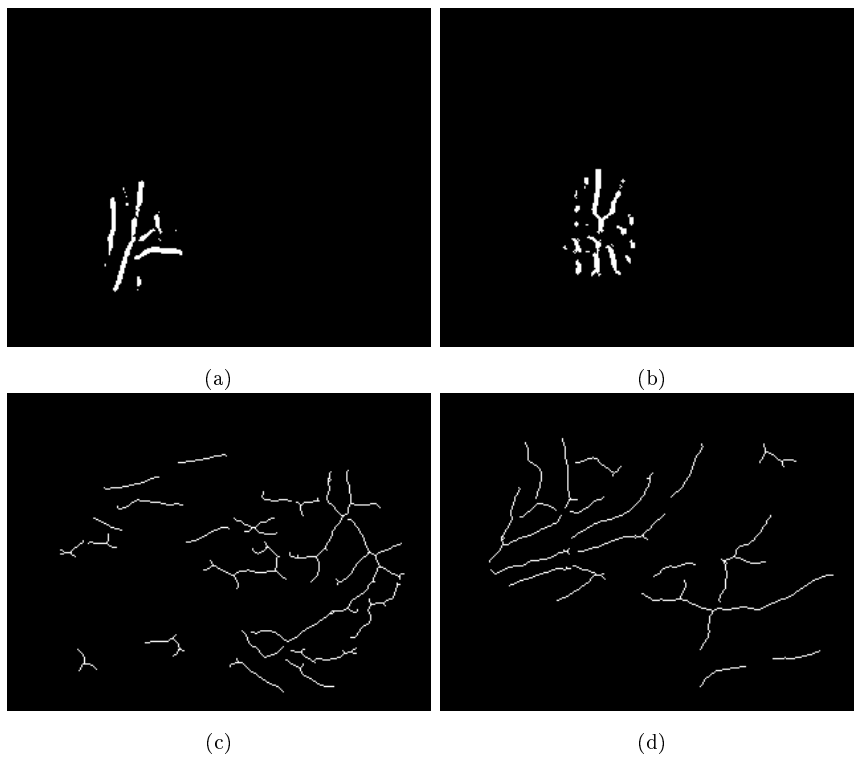


Figure 3.26: The LoG transformation of the corresponding images in Figure 3.23, after binarisation through Otsu's method.

It is clear from Figure 3.26 (a) that the local shadow encircled in green in Figure 3.24 (a) has a sufficiently strong response to the LoG transform, as is also the case with hair (see Figure 3.26 (b)). The results presented in Figure 3.26 are however merged with the results presented in Figure 3.25 in order to facilitate the combination of all three key criteria stipulated at the start of this section.

3.3.3 Combining all three criteria

The three key criteria stipulated at the start of this section are combined by taking the intersection of the corresponding binary images depicted in Figures 3.25 and 3.26 through the *and* operator. The results are presented in Figure 3.27. It is clear from Figure 3.27 (a) that even though the local shadow encircled in green in Figure 3.24 (a) is detected by the LoG transform, it does not satisfy the dark and narrow criterium (see Figure 3.25 (a)), and therefore does not appear in the corresponding combined image (see Figure 3.27 (a)).

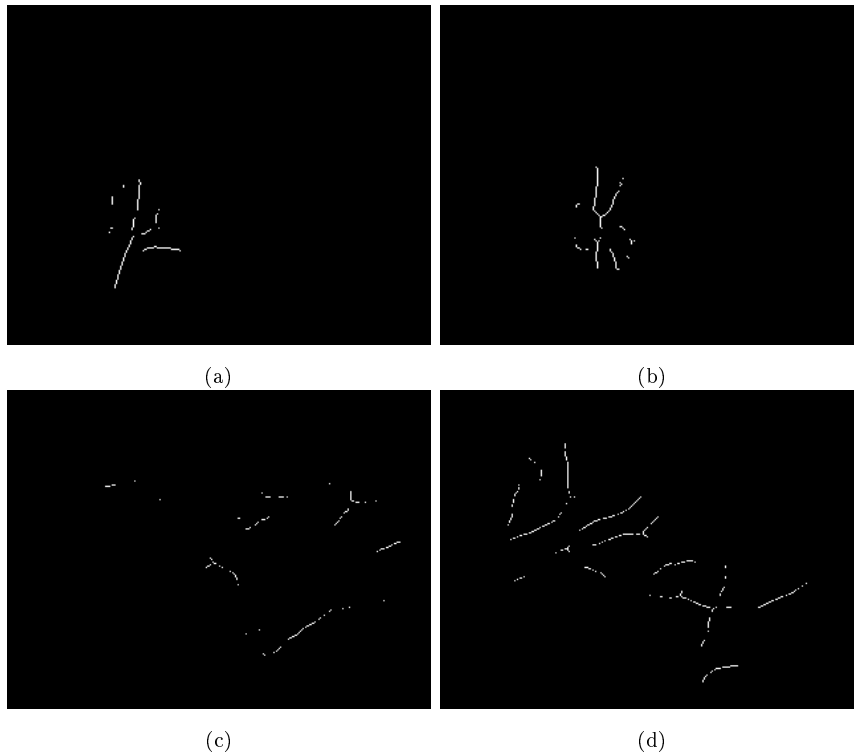


Figure 3.27: The intersection of the corresponding images depicted in Figures 3.25 and 3.26. The detected pixels serve as seed pixels for morphological reconstruction by dilation.

The detected objects depicted in Figure 3.27 are therefore excellent *candidates* for hand veins since they satisfy all three key criteria stipulated at the start of this section, but it is clear that said pixels do not represent *all* of the targeted hand veins. The detected pixels in Figure 3.27 however constitute ideal *seed* pixels for finding the *other* hand vein pixels through morphological reconstruction by dilation, as explained in Section 3.3.5. We first introduce the concept of morphological dilation in the following section.

3.3.4 Morphological dilation

Morphological dilation is conceptually explained in Figure 3.28 by employing the same annular object depicted in Figure 3.13 (a). The morphological dilation of an object is characterised by the enlargement of said object within a binary image. The morphological dilation of an object by a certain SE is achieved by specifying that whenever the center of the SE forms part of the original object, all of the pixels in the transformed image that overlap with the SE form part of the dilated object.

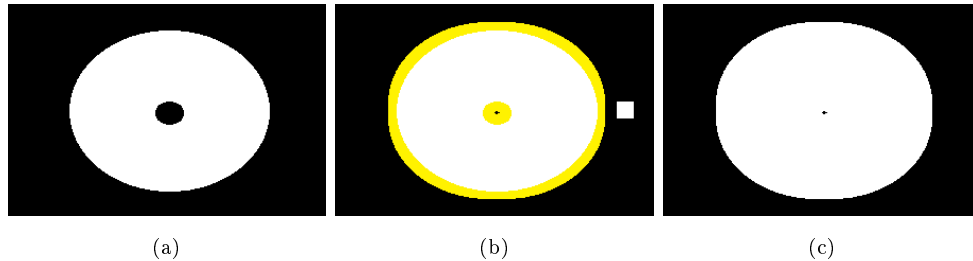


Figure 3.28: (a) The image in Figure 3.13 (a). (b) The pixels that are appended to (a), when dilated by the white 16×16 square-shaped SE, are shown in yellow. (c) The object in (a) after being dilated by said SE.

Morphological dilation is therefore well-suited for the purpose of finding other hand vein pixels that are connected to one or more of the seed pixels in the images depicted in Figure 3.27.

3.3.5 Morphological reconstruction by dilation

Morphological reconstruction by dilation involves the *iterative* dilation of the objects in a so-called *seed* image until the boundaries of the objects that constitute the so-called *mask* image are reached. The seed image contains carefully selected pixels that very probably form part of the regions that are to be reconstructed, while the *mask* image (1) restricts the so-called growth of the seed pixels, and (2) determines the final morphology of the detected regions.

An appropriate mask image should contain as many targeted vein pixels as possible and have well-defined object boundaries. In order to construct such a mask image we opt to employ a textural segmentation method called thresholding by *moving averages* (Gonzales (2010)). This method is well-suited for extracting an optimal amount of textural information from a given image. The method of moving averages determines a local threshold value for a pixel by considering the average intensity within a certain local *neighbourhood* of said pixel. Figure 3.29 shows the result of applying moving average thresholding to the images depicted in Figure 3.23 by considering a 10×10 pixel neighbourhood. The images shown in Figure 3.29 subsequently serve as suitable mask images for morphological reconstruction by dilation.

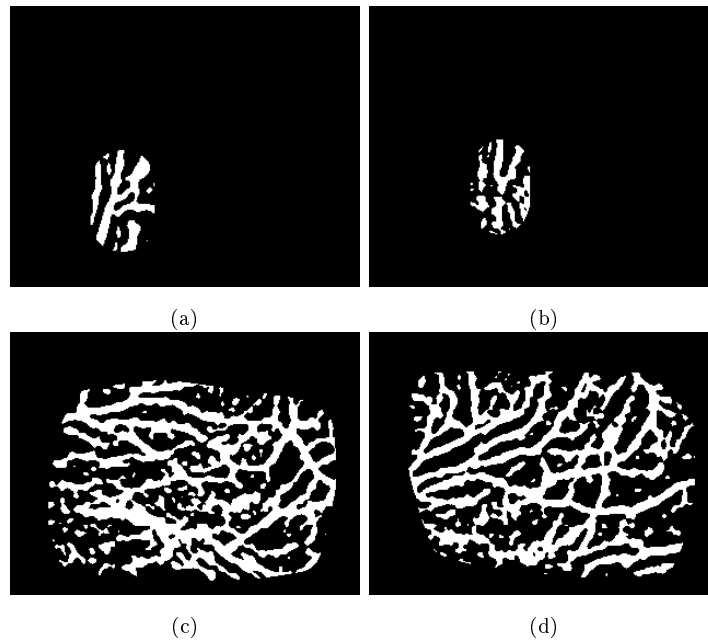


Figure 3.29: The mask images for the purpose of morphological reconstruction by dilation. These images correspond to the respective images depicted in Figure 3.23.

The results presented in Figure 3.30 are obtained by applying morphological reconstruction by dilation to the seed images shown in Figure 3.27. The reconstruction process is restricted by the mask images shown in Figure 3.29.

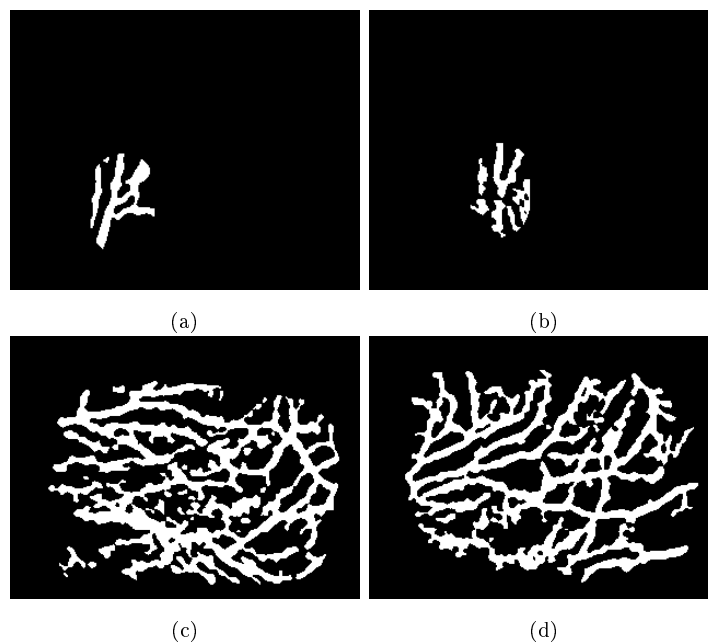


Figure 3.30: The morphologically reconstructed versions of the corresponding seed images shown in Figure 3.27. The relevant mask images are shown in Figure 3.29.

3.3.6 Final segmentation results: Bosphorus data set

Figure 3.31 depicts the final segmentation results for the NIR images depicted in Figure 3.21. The proposed hand-vein segmentation protocol is therefore sufficiently accurate and robust within the context of the Bosphorus data set.

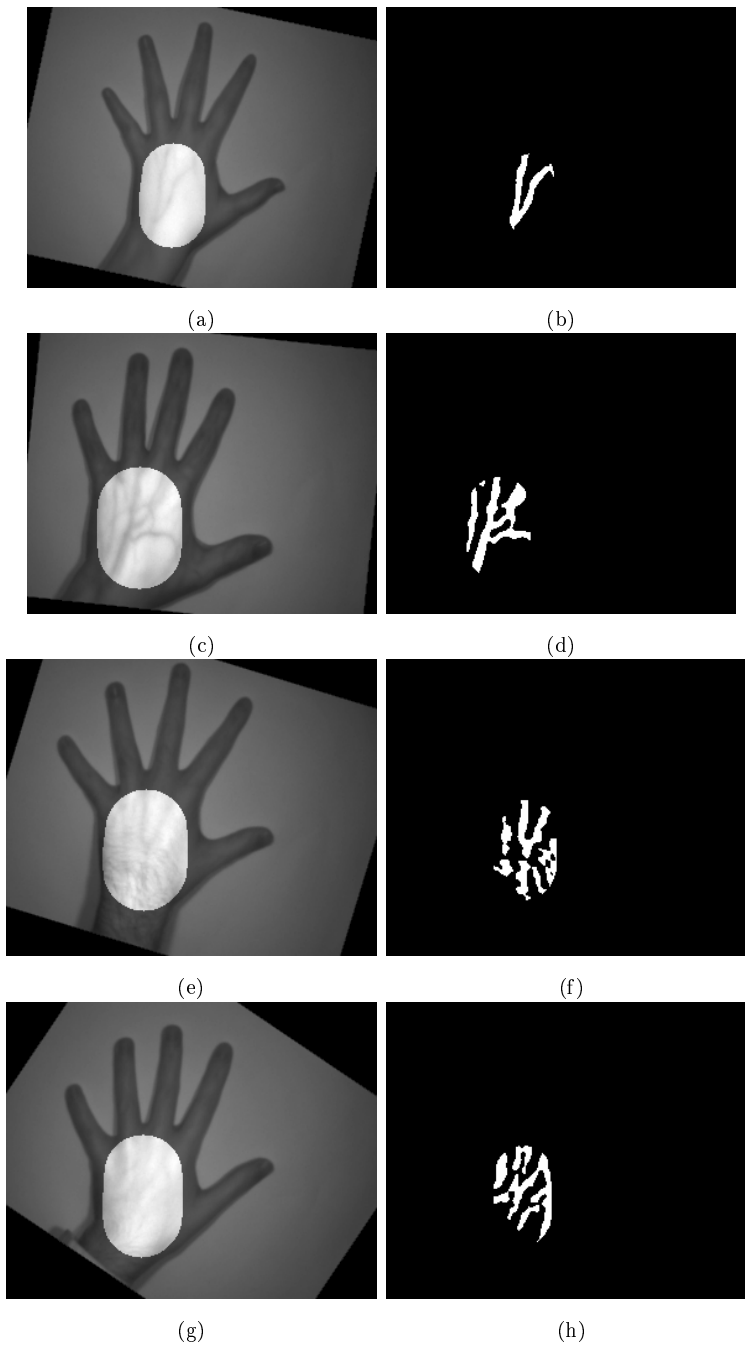


Figure 3.31: The final segmentation results for selected NIR images from the Bosphorus data set. **(Left)** The ROIs depicted in Figure 3.21. **(Right)** The extracted hand veins.

3.3.7 Final segmentation results: CIE data set

Figure 3.32 depicts the final segmentation results for the NIR images depicted in Figure 3.22. The proposed hand-vein segmentation protocol is therefore sufficiently accurate and robust within the context of the CIE data set.

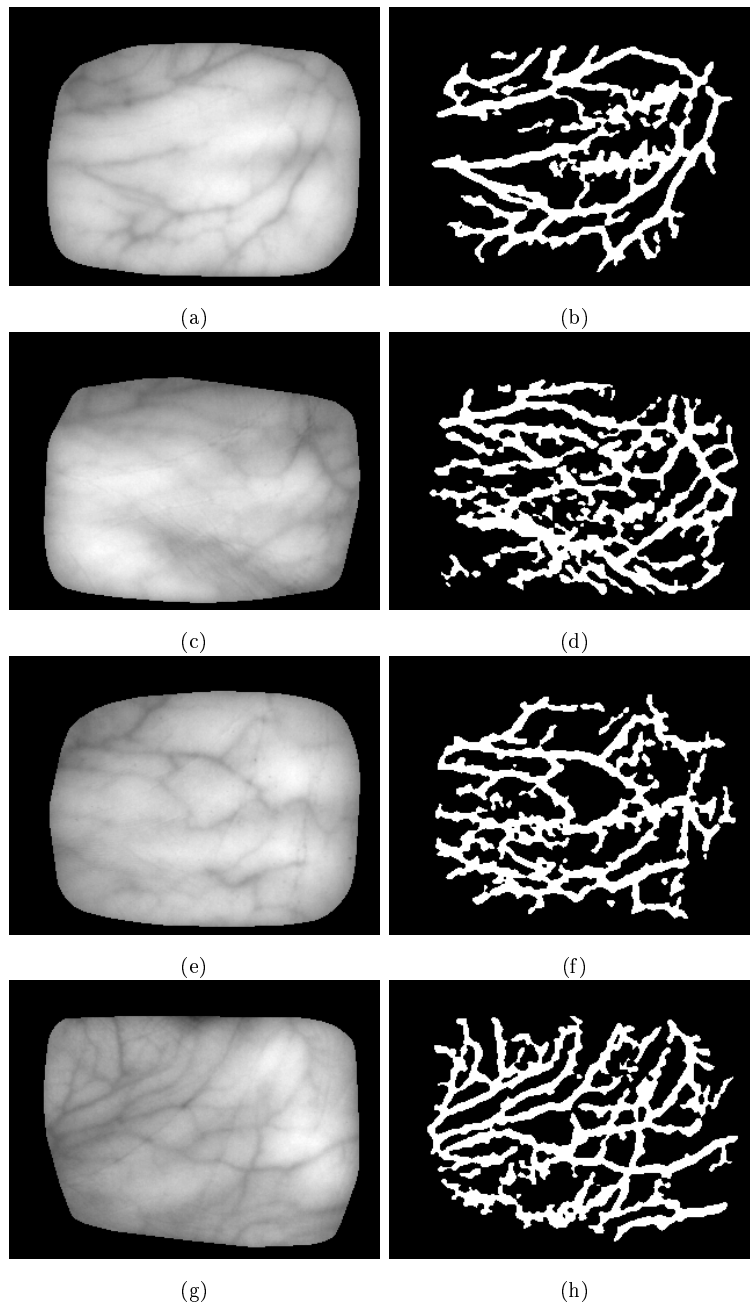


Figure 3.32: The final segmentation results for selected NIR images from the CIE data set. (Left) The ROIs depicted in Figure 3.22. (Right) The extracted hand veins.

3.4 Concluding remarks

In this chapter we proposed a novel hand-vein segmentation protocol that extracts prominent veins from either the dorsal or palmar surface of the body of a human hand. We demonstrated that this protocol is robust with respect to noise and artefacts like hair and localised shadows. We also demonstrated that said protocol is proficient in both detecting targeted hand veins and identifying a sufficient number of targeted hand vein pixels.

In the next chapter we propose a robust strategy for extracting features from a binary image in which the objects represent detected hand veins. We also explain how this may be achieved in a rotational, translational and scale invariant way.

Chapter 4

Feature extraction and matching

4.1 Introduction

In this chapter we present a robust feature extraction protocol that is able to accurately portray prominent characteristics of binarised hand veins like those depicted in Figure 4.1 (a). We also present a feature matching technique that is able to quantify the difference between features extracted from a questioned hand vein pattern and those extracted from a reference pattern. Coetzer (2005) concluded that the combination of the aforementioned feature extraction and feature matching techniques are proficient in detecting skilled forgeries within the context of off-line handwritten signature verification. An investigation into the feasibility of the above-mentioned combined protocol for the purpose of hand vein-based biometric verification is therefore inspired by the fact that binarised hand vein patterns and binarised handwritten signatures (especially Oriental signatures) have a remarkably similar morphology (see Figure 4.1).



Figure 4.1: (a) An image containing a binarised hand vein pattern. (b) An image containing a binarised Oriental handwritten signature.

In this thesis we opt to represent a hand vein pattern by a sequence of un-quantised feature vectors. Said feature vectors are obtained by applying the discrete *Radon* transform to a binarised hand vein image like the ones depicted in Figure 4.1 (a). The above-mentioned feature extraction protocol is detailed in Section 4.2. The difference between a feature vector extracted from a questioned sample and the corresponding feature vector extracted from a reference sample that is known to belong to the claimed identity, is obtained by calculating either (1) the squared Euclidean distance or (2) a *dynamic time warping-based (DTW-based)* distance between the feature vectors in question. The aforementioned matching strategies are detailed in Section 4.3.

It is important to recall that a handwritten signature constitutes a *behavioural* biometric that is typically associated with relatively large intra-class variations. This is unavoidable due to the very nature of handwritten signatures. Intra-class variations may also differ substantially from one individual (signer) to the next.

A hand vein pattern on the other hand constitutes a *physiological* biometric that is typically associated with relatively small intra-class variations. The intra-class variations associated with hand vein patterns are however mainly due to variations in scale (resulting from variations in the distance between the hand and the camera), location and orientation of the hand, as well as variations in lighting which, in turn, may result in variations in the location of shadows. Said types of variation are due to external factors and are not intrinsic to the physiology of the hand vein pattern. It is therefore not unreasonable to expect that hand vein samples that belong to the same individual, but were captured at different occasions, should be very similar. The feature extraction and matching protocol proposed in this chapter is based on said philosophy.

The influence of external factors on the scale, orientation and location of a hand vein pattern can be avoided by, for example, controlling the image acquisition process. The subject may be instructed to place his/her hand under or above a camera in such a way that the hand is (1) appropriately orientated and (2) located at the correct distance from the camera. The environmental temperature, as well as the lighting conditions may also be controlled. Even when the image acquisition process is not strictly regulated, it should be possible to remove almost all intra-class variations by, for example, rotating all questioned samples in such a way that they are all similarly orientated. The goal of the proposed feature extraction and matching protocol is therefore to *minimise* intra-class variation. Said objective is achieved through data *normalisation*, which reduces the effect of external influences on variations in scale, location and orientation of the morphology of the binarised hand veins. Data normalisation in this context involves the application of several different operations to the intermediate results associated with the proposed feature extraction and matching protocol. Said protocol is conceptualised in Figures 4.2 and 4.10 and detailed in Sections 4.2 and 4.3 respectively.

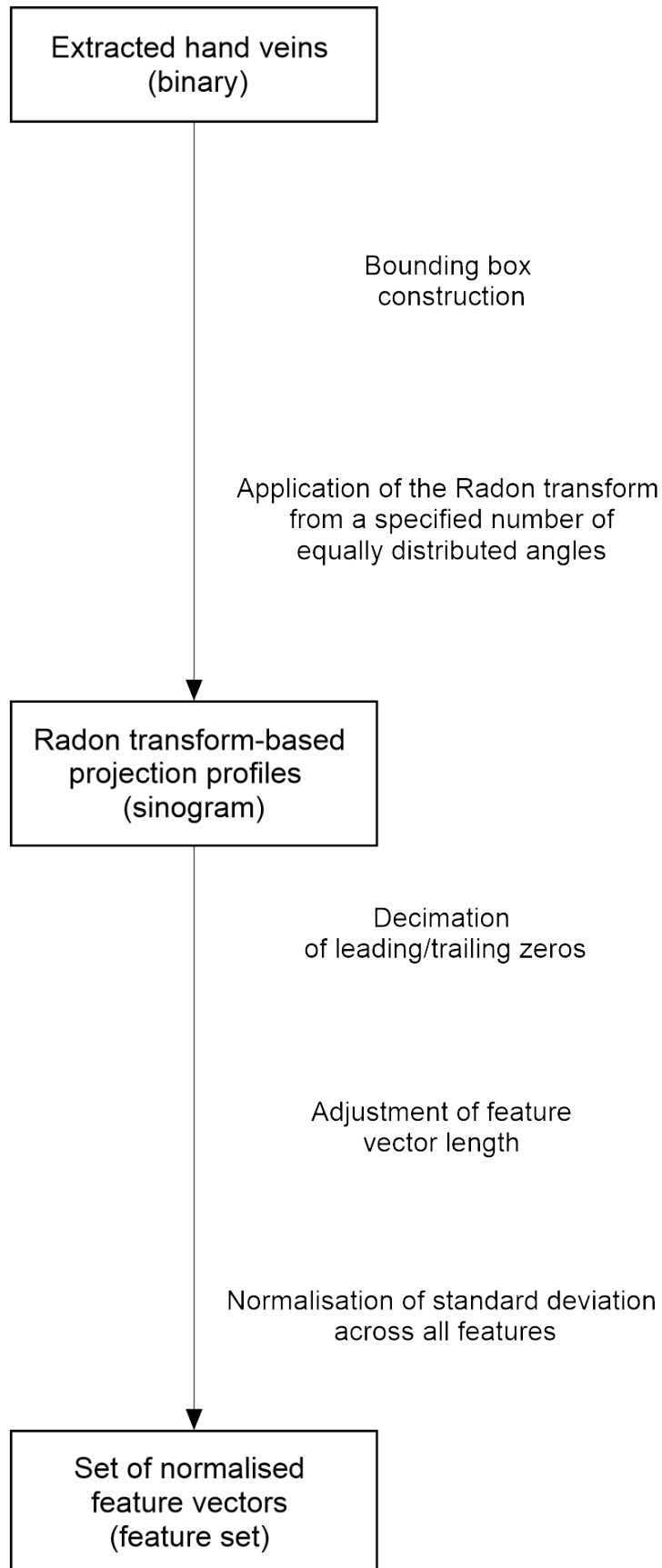


Figure 4.2: Conceptualisation of the proposed feature extraction protocol.

4.2 Feature extraction

In Figures 4.3 (a) and (b), two binarised hand vein images that belong to a single individual from the Bosphorus data set are shown. Similarly, Figures 4.3 (c) and (d) depict two binarised hand vein images that belong to a single individual from the CIE data set. We present these images in order to emphasise the effect of variations in scale, location and orientation on the morphology of a hand vein pattern within the context of a specific individual. For example, when Figures 4.3 (a) and (b) are compared, it is clear that the two hand vein patterns are located at slightly different positions. Said spatial variation is also visible when Figures 4.3 (c) and (d) are compared.

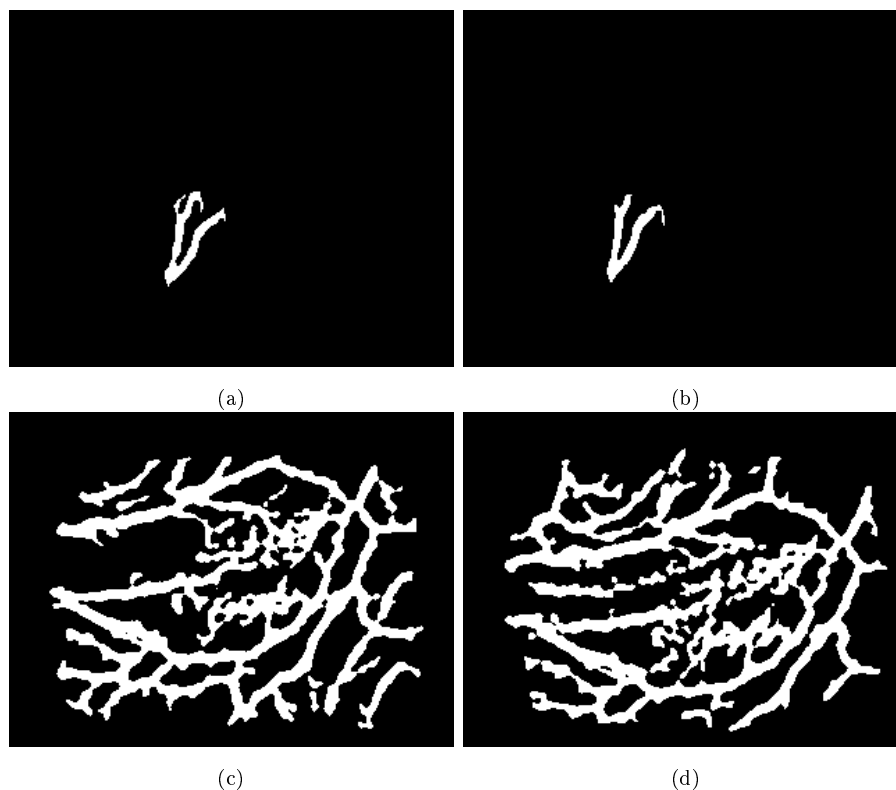


Figure 4.3: (a) A binarised hand vein image associated with a specific individual from the Bosphorus data set. (b) A binarised hand vein image from the same individual as the one referred to in (a), but captured on a different occasion. (c) A binarised hand vein image associated with a specific individual from the CIE data set. (d) A binarised hand vein image from the same individual as the one referred to in (c), but captured on a different occasion.

Recall that the first round of reduction in orientational variation within the context of the Bosphorus data set already occurred during the image segmentation stage (see Section 3.2.4). The difference in orientation between the images depicted in Figures 4.3 (a) and (b) is therefore not significant.

4.2.1 Bounding box construction

The first round of spatial variation normalisation constitutes the elimination of the surrounding black background in the images depicted in Figure 4.3. This is accomplished by constructing a *bounding box* around each binarised hand vein pattern.

The concept of a bounding box is similar to that of a convex hull (see Section 3.2.1), except for the fact that the shape of a bounding box is limited to a square or rectangle. The results from constructing bounding boxes for the images depicted in Figure 4.3 are shown in Figure 4.4. It is important to note that the dimensions of the bounding boxes are dependent on the morphology of the hand vein pattern in question.

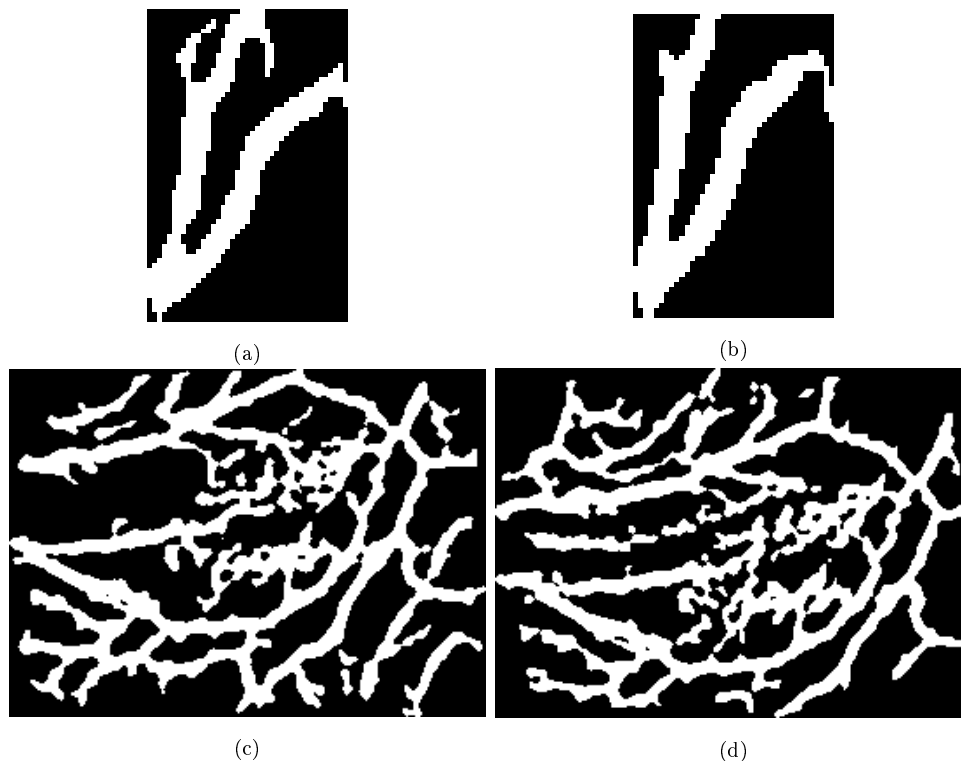


Figure 4.4: The results after retaining only the cropped sub-images within the bounding boxes for the images shown in Figure 4.3.

The application of a bounding box eliminates most of the spatial variation. Some spatial variation among the hand vein pixels within a specific bounding box may however still be present due to slight movements during image acquisition, or due to a variation in the position or angle of the lighting source. The remaining spatial variation is reduced in two additional steps that are detailed in Sections 4.2.3 and 4.3.2. At this stage of the feature extraction protocol we proceed to extract the feature vectors by applying the Radon transform to the images depicted in Figure 4.4.

4.2.2 The Radon transform

The discrete Radon transform is obtained by calculating a set of one-dimensional projection profiles of a two-dimensional image (matrix) from a number of equally distributed angles within the interval $\theta \in [0^\circ, 180^\circ)$. Each projection constitutes an unnormalised feature vector (see Figure 4.5), wherein each component (feature) represents a *beam-sum*. The number of angles, as well as the number of beams utilised to calculate a projection profile from any of these angles, are predetermined. Within the context of the system proposed in this thesis, each unnormalised feature approximates the number of hand-vein pixels within the relevant beam. The calculation of a projection profile (unnormalised feature vector) of a typical hand vein image from a specific angle θ is conceptualised in Figure 4.5.

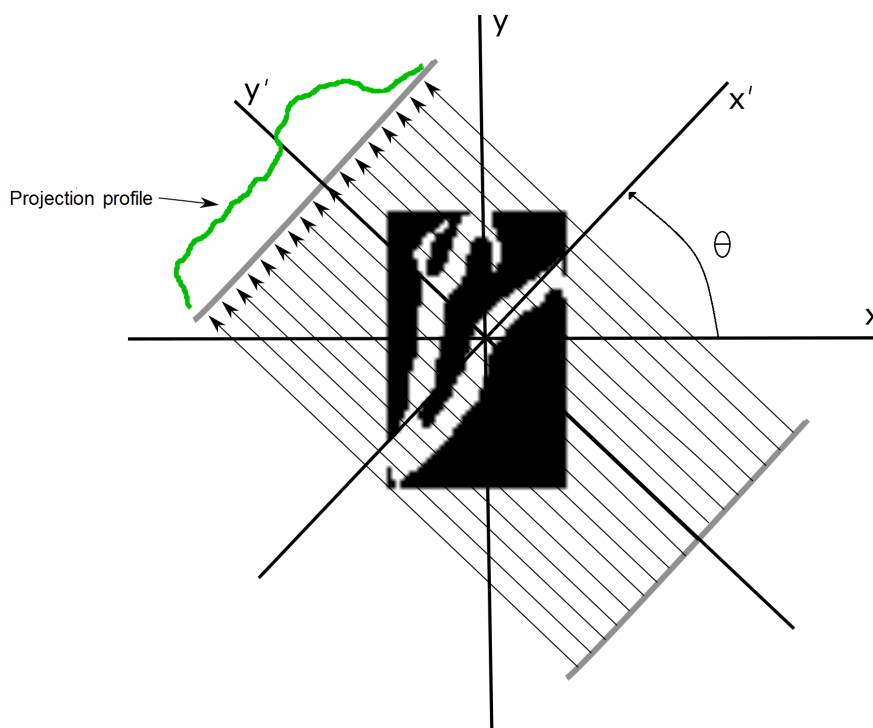


Figure 4.5: Conceptualisation of the discrete Radon transform being applied to a typical binarised hand vein image.

When all the projection profiles obtained from equally distributed angles within the interval $\theta \in [0^\circ, 180^\circ)$ are packed into the columns of a matrix, the result is often displayed as a grey-scale image, which is referred to as a *sinogram*. The respective sinograms for the images depicted in Figure 4.4 are shown in Figure 4.6. These projections were obtained from 64 equally distributed angles within the interval $\theta \in [0^\circ, 180^\circ)$.

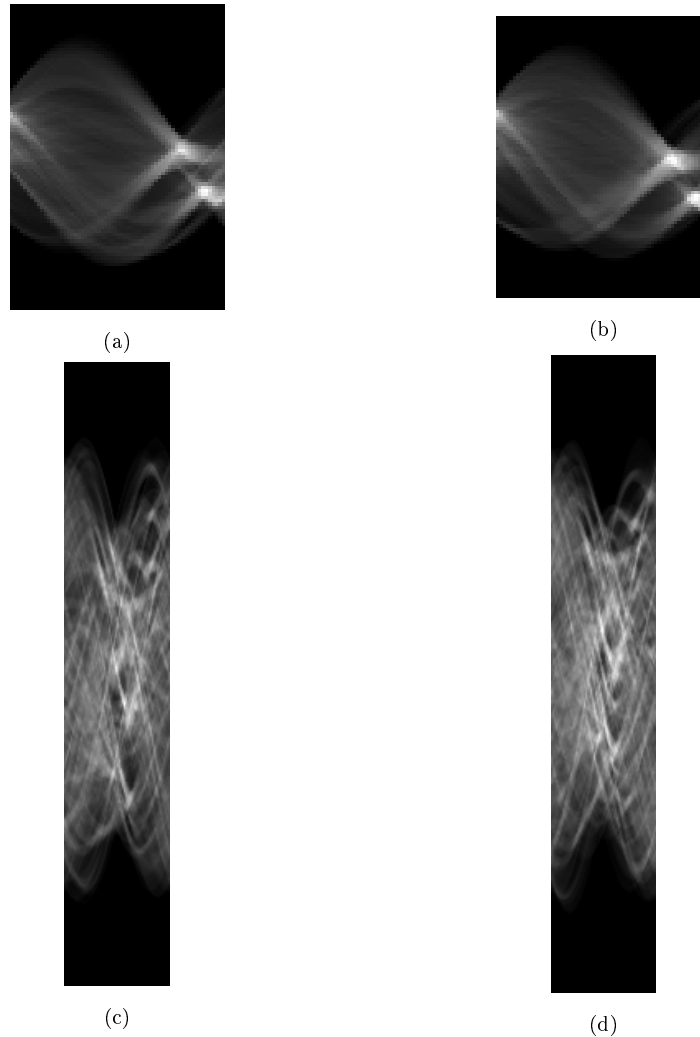


Figure 4.6: The result of applying the discrete Radon transform to the corresponding images in Figure 4.4, using 64 equally distributed angles within the interval $\theta \in [0^\circ, 180^\circ)$.

Note that the width of all of the sinograms depicted in Figure 4.6 constitutes 64 pixels, which corresponds to 64 equally distributed projection angles within the interval $\theta \in [0^\circ, 180^\circ)$. The number of beams per angle is dependent on the largest dimension of the input image. The height of the sinograms for the samples from the CIE data set (Figures 4.6 (c) and (d)) is therefore significantly larger than that of the sinograms for samples from the Bosphorus data set (Figures 4.6 (a) and (b)). We therefore slightly re-scaled the images depicted in Figures 4.6 (c) and (d) for visualisation purposes. We now proceed to explain the second round of spatial variation normalisation, which involves the removal of the leading and trailing zero-valued entries from all of the projection profiles like those depicted in Figure 4.6.

4.2.3 Removal of leading and trailing zeros

From Figure 4.6 it is clear that the non-zero entries towards the middle of each sinogram are surrounded by zero-valued entries towards the top and bottom. Said zero-valued entries are referred to as leading and trailing zeros. They are to be removed (decimated) in order to further reduce external influences on the spatial variation of the extracted feature vectors (as explained in Section 4.2.1). This constitutes the second round of spatial variation normalisation. The "trimmed" feature vectors differ in length and can therefore not constitute the columns of a matrix. The removal of the leading and trailing zeros for a specific projection profile is conceptualised in Figure 4.7.

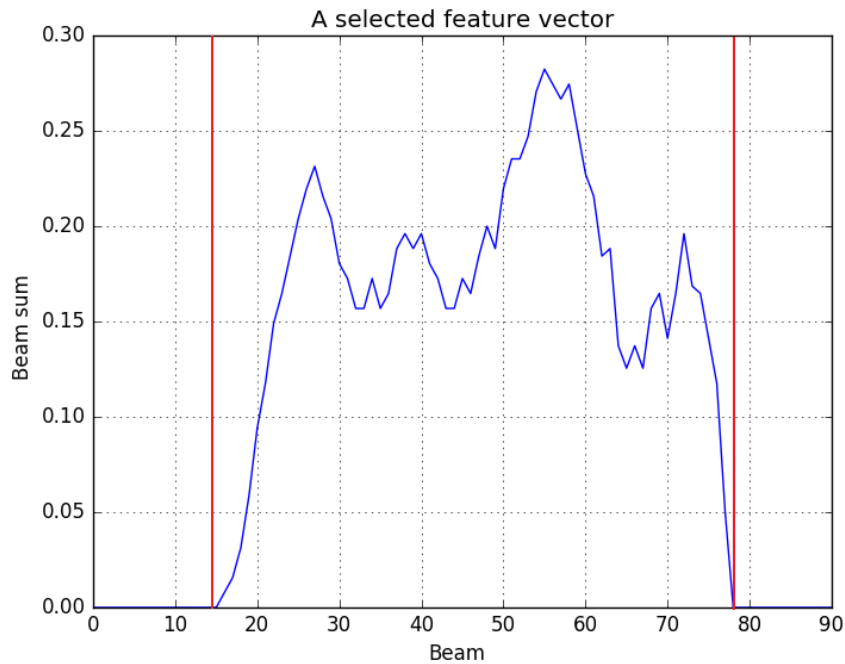


Figure 4.7: Conceptualisation of the decimation of the leading and trailing zeros for a selected unnormalised feature vector from the image depicted in Figure 4.6 (a). The leading zeros appear before the first red vertical line, while the trailing zeros appear after the second red vertical line.

In Figure 4.7 it is demonstrated how the zero-valued entries that lie outside the vertical red lines are removed from an unnormalised feature vector. Only the non-zero entries are retained. The dimension of each "trimmed" feature vector is subsequently adjusted (through linear interpolation) to a predefined global target dimension, as is explained in the next section.

4.2.4 Scale normalisation of features

When Figures 4.6 (a) and (b) are compared, it is clear that the height (the number of beams per projection angle) of the respective sinograms differ slightly. Recall that the number of beams per projection angle is dependent on the dimensions of the ROI. It is therefore necessary to adjust the length (dimension) of each of the "trimmed" feature vectors to a predefined target dimension (through linear interpolation), in order to complete the first round of scale normalisation. Each of these adjusted feature vectors are subsequently packed into the columns of a new matrix. The second round of scale normalisation is accomplished by dividing said matrix by the *standard deviation* across all of its entries. The resulting matrix is henceforth referred to as a *feature set*, of which the columns constitute normalised feature vectors.

The results of (1) removing the leading and trailing zeros from the images depicted in Figure 4.6, (2) adjusting the length of the resulting trimmed feature vectors to a predefined target dimension through linear interpolation, (3) packing the adjusted feature vectors into the columns of a new matrix and (4) dividing said matrix by the standard deviation across all of its entries are shown in Figure 4.8.

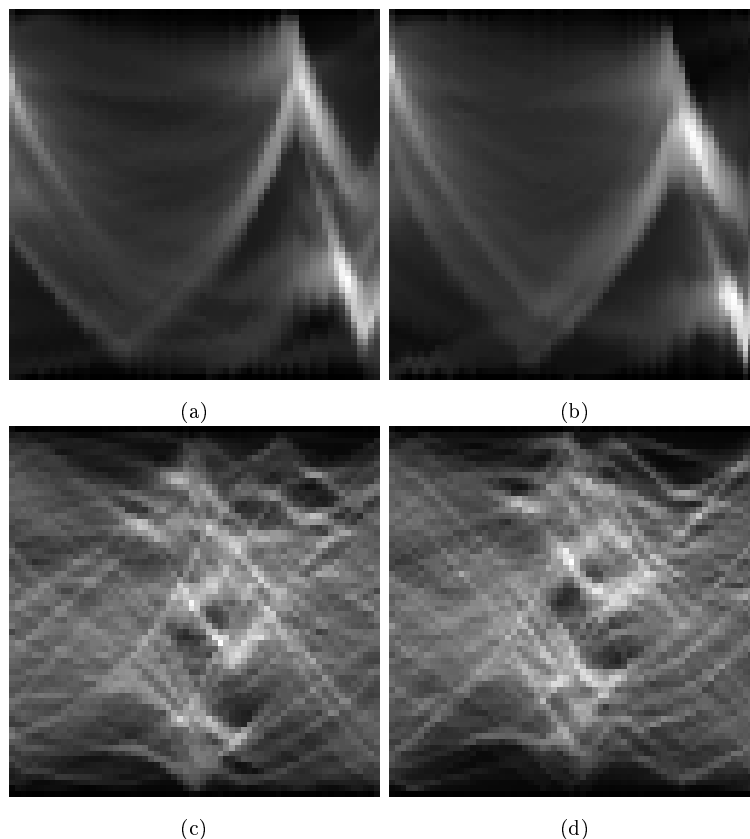


Figure 4.8: The feature sets for the corresponding images depicted in Figure 4.4.

4.3 Feature matching

In this section we present robust techniques for quantifying the difference between two feature sets for the purpose of deciding whether or not the matched feature sets belong to the same individual. Recall that *corresponding* feature vectors are obtained by calculating projection profiles from the same angle, followed by appropriate normalisation. We therefore propose that the difference between the two feature sets in question is quantified by the *average* difference between the corresponding feature vectors. The difference between two individual features (entries) within two corresponding feature vectors may be quantified by the length of the red vertical line in Figure 4.9.

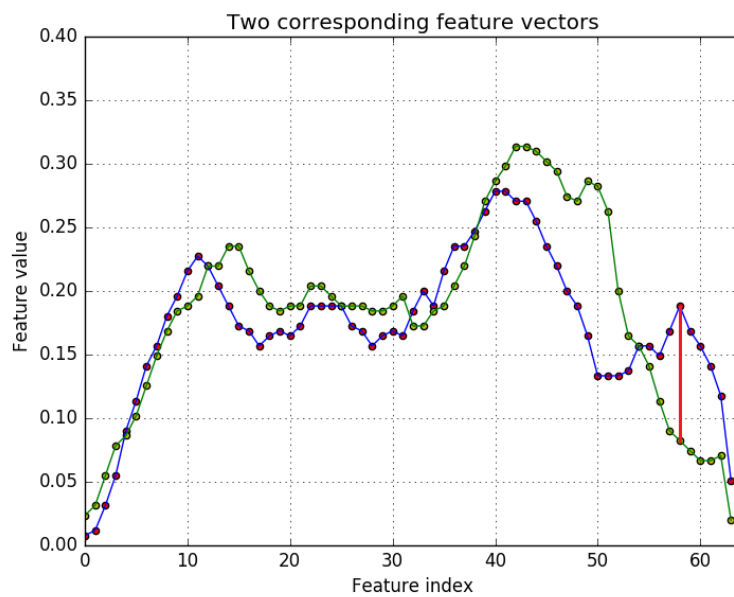


Figure 4.9: Conceptualisation of the difference between two features within a pair of corresponding feature vectors, as quantified by the length of the red vertical line.

Since each feature (yellow or red dot) within a feature vector, like those depicted in Figure 4.9, is represented by a scalar value, we opt to employ either (1) the squared Euclidean distance (the L_2 -norm) or (2) a DTW-based distance in order to quantify the difference between any two feature vectors. As mentioned earlier, the difference between two feature sets is quantified by the *average* of the aforementioned distances across all projection angles. The application of the aforementioned distance functions within the context of this thesis is conceptualised in Figure 4.10 and detailed in Sections 4.3.1 and 4.3.2 respectively.

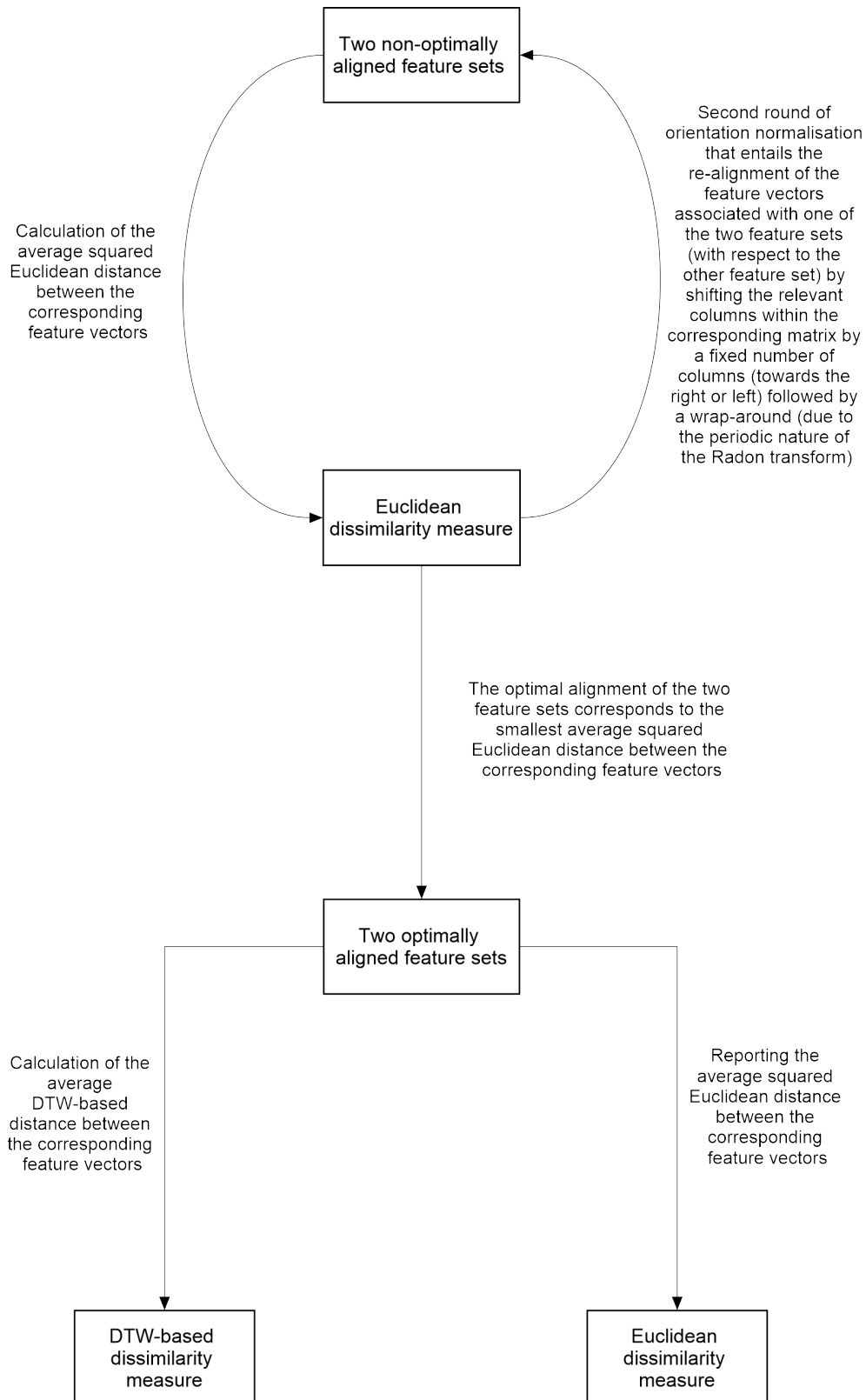


Figure 4.10: Conceptualisation of the proposed feature matching protocol.

4.3.1 The squared Euclidean distance

In this section we explain how the squared Euclidean distance is employed to quantify the difference between two feature sets. The squared Euclidean distance between two corresponding feature vectors is equal to the sum of the squared distances between all of the corresponding features, that is the sum of the squared lengths of the red vertical lines depicted in Figure 4.11.

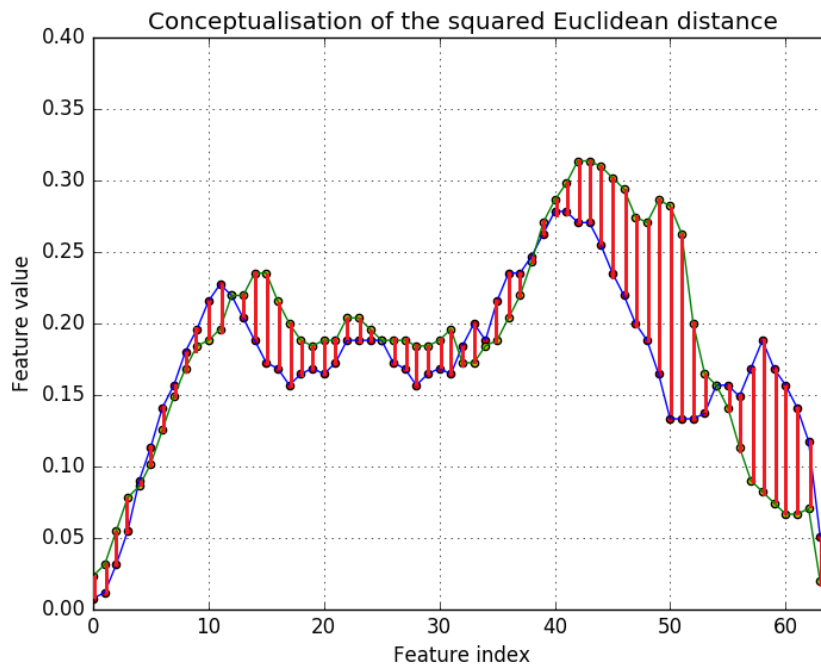


Figure 4.11: Conceptualisation of the squared Euclidean distance between two corresponding feature vectors.

As indicated in Figure 4.10, the second round of orientation normalisation (orientation adjustment) is achieved by matching two feature sets by employing the squared Euclidean distance. We subsequently explain how said adjustment is achieved. Recall that the columns of a feature set (when represented as a matrix) represent the feature vectors associated with unique projection angles in the interval $\theta \in [0^\circ, 180^\circ)$. The re-alignment procedure referred to in Figure 4.10 involves the re-assignment of the angles associated with different feature vectors within a questioned feature set.

By shifting the columns of the questioned matrix by a certain number of columns towards the right or left, a *rotated* version of the questioned feature set is obtained. Recall that the discrete Radon transform is periodic over the interval $\theta \in [0^\circ, 360^\circ)$, and that the projection profiles obtained from angles in the interval $\theta \in [180^\circ, 360^\circ)$ constitute reflected versions of those projection

profiles obtained from angles in the interval $\theta \in [0^\circ, 180^\circ)$. This enables one to apply a subsequent wrap-around procedure to the shifted matrix, without having to obtain *new* information. For example, when a questioned matrix (feature set) is shifted by 2 columns towards the right, the feature vector within the last column of the original matrix is first flipped in an up-down fashion (due to symmetry) and then packed into the second column of the re-aligned matrix and so forth. This allows for orientation adjustment without having to obtain new information.

In Figure 4.12 we demonstrate said adjustment procedure by shifting a questioned feature set (depicted by the image in Figure 4.12 (a)) towards the right by exactly half of its number of columns. The re-aligned (adjusted) feature set is depicted by the image in Figure 4.12 (b). Said adjustment is equivalent to rotating the original questioned image through an angle of 90° in a clockwise direction.

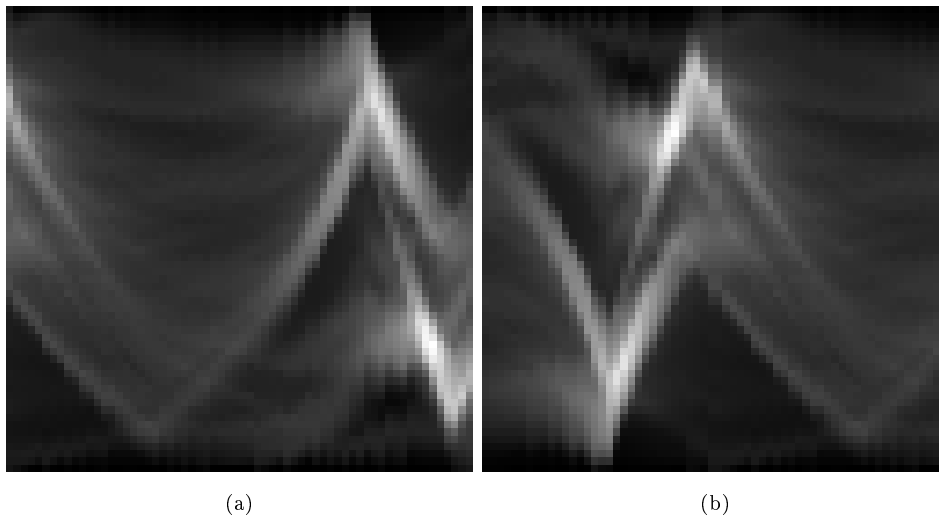


Figure 4.12: (a) A questioned feature set. (b) The feature set in (a) shifted towards the right by exactly half of its number of columns, followed by a wrap-around procedure and the subsequent flipping of the wrapped columns in an up-down fashion. It is clear that said columns constitute the last half of the columns of the feature set depicted in (a).

The purpose of the aforementioned adjustment procedure is to iteratively search for the best possible alignment between the two feature sets being matched. For each iteration of the adjustment procedure, the average squared Euclidean distance between the re-aligned feature sets is obtained. The smallest of said average distances therefore corresponds to the best possible alignment between the feature sets in question. Note that said alignment is applied to an entire feature set, and *not* to an individual feature vector like those depicted in Figure 4.11.

4.3.2 A DTW-based distance

In this section we explain how a better alignment of the features within two corresponding feature vectors may be obtained. Recall that the protocol developed for achieving the second round of orientation normalisation was explained in the previous section, from which the best alignment (orientation-wise) between two feature sets is obtained. The protocol explained in this section is therefore only applied to each corresponding pair of feature vectors within the optimally-aligned feature sets in question.

Note from Figure 4.11 that all the red lines connecting corresponding features are vertical. This is facilitated by the fact that the dimension of every feature vector had been adjusted (through linear interpolation) to a fixed global dimension, causing an appropriate alignment of the prominent features. The squared Euclidean distance therefore constitutes the sum of the squared distances between corresponding components (features).

A *prominent* feature may constitute a prominent peak or valley (that coincide with the same region within the corresponding binarised hand vein images like those depicted in Figure 4.4) within a certain feature vector. In Figure 4.13 we illustrate the possibility that prominent features may be associated with *different* components of two corresponding feature vectors by manually connecting prominent peaks and valleys with red lines. It is clear from Figure 4.13 that many of these red lines are *not* vertical as conceptualised in Figure 4.11.

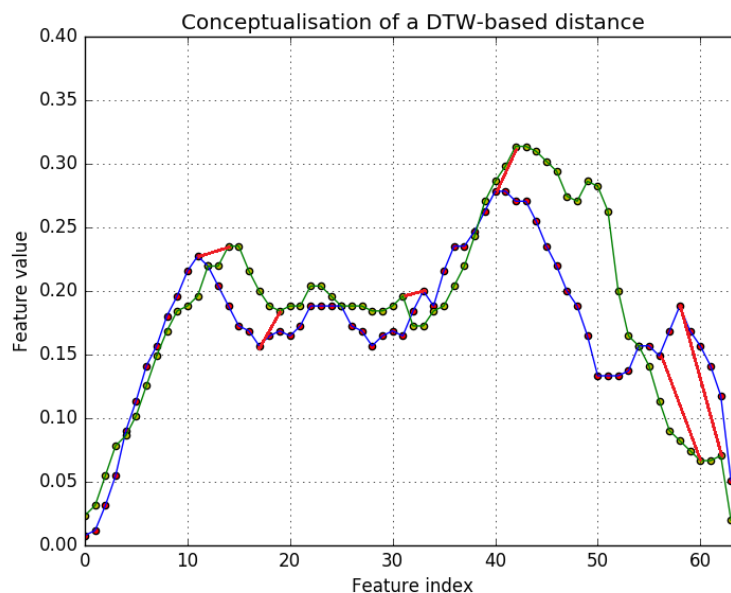


Figure 4.13: An illustration of the possibility that two prominent features may be associated with *different* components of corresponding feature vectors.

Recall how spatial variation of the intrinsic structure of binarised hand veins like those depicted in Figure 4.4 may still be possible. It is therefore necessary to employ a DTW-based distance for the purpose of possibly aligning actual *prominent* features. This constitutes the second round of spatial variation reduction, and its main goal is to reduce intra-class variation without reducing inter-class variation to the same extent.

The philosophy behind employing a DTW-based distance for feature matching is that such a distance is expected to align actual prominent features, based on the fact that (1) the aforementioned (remaining) spatial variation is expected to be relatively small (less than the specified restricting bandwidth), and (2) feature sets representing hand-vein patterns from different individuals are expected to differ substantially (more than the specified restricting bandwidth).

Note however that said philosophy only holds (within the context of hand veins) when the *search space* associated with the DTW-based distance is restricted to a certain bandwidth. Said bandwidth constitutes the total number of features in the neighbourhood of the current feature that are allowed to be paired up with said feature. A conceptualisation of a DTW-based distance with a restricting bandwidth of 20 components (features) is shown in Figure 4.14.

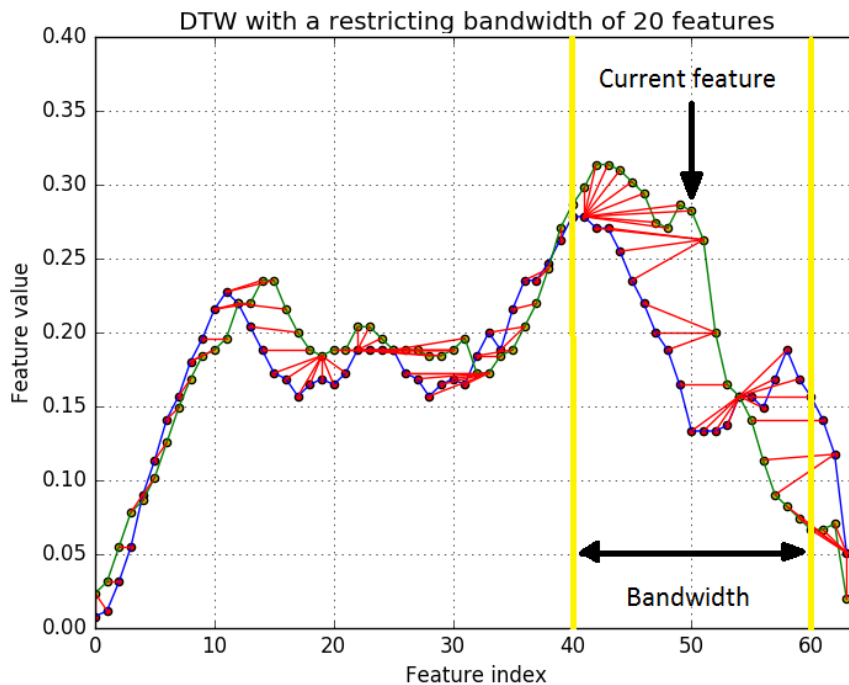


Figure 4.14: A conceptualisation of the restricting bandwidth associated with the calculation of a DTW-based distance between two corresponding feature vectors.

The *average* DTW-based distance may therefore be calculated between the optimally aligned feature vectors within any two feature sets being matched.

Alternatively, the average squared Euclidean distance between the optimally aligned feature vectors may be considered. The aforementioned protocol is equivalent to employing a DTW-based distance measure with a restricting bandwidth of zero.

The performance of classifiers that employ a DTW-based dissimilarity measure within the context of a non-zero bandwidth may subsequently be compared to that of classifiers that employ an Euclidean distance-based dissimilarity measure. Such a comparison is conducted and reported on in the next chapter.

4.4 Concluding remarks

In the next chapter we conduct experiments that simulate a number of real-world scenarios. Said experiments are designed for the purpose of estimating the proficiency of the hand vein-based identity verification systems proposed in this thesis.

Chapter 5

Experiments

5.1 Introduction

This chapter describes the experiments that are conducted in this study. The purpose of an experiment is to obtain a set of five statistical performance measures (see Section 5.4) that constitute estimations of the proficiency of a *specific* system within the context of a *specific* real-world scenario.

Number of available positive samples. The above-mentioned experiments simulate two real-world scenarios, namely (1) a scenario in which only *one* positive sample is available for each enrolled individual, that is the so-called *k-one* scenario as discussed in Section 5.5 and (2) a scenario in which *six* positive samples are available for each enrolled individual, that is the so-called *k-six* scenario as discussed in Section 5.6.

Data sets. Two unique hand-vein data sets are employed in the aforementioned experiments, namely (1) the *Bosphorus* data set (as outlined in Section 5.2) and (2) the *CIE* data set (as outlined in Section 5.3). When employed, a data set is partitioned according to (1) the specific parameters of the employed system and (2) the type of scenario that is simulated.

Score-based and rank-based classifiers. The specific parameters for each experiment are determined by a number of factors. For example, a particular system may either employ (1) a *score*-based classifier or (2) a *rank*-based classifier for the purpose of verifying the authenticity of a questioned hand vein pattern. The score-based classifier either accepts or rejects a questioned sample by considering either (1) the average Euclidean distance or (2) the average DTW-based distance between the feature vectors associated with the questioned sample and the corresponding feature vectors associated with a selected *reference* sample positively linked to the claimed individual ¹. The

¹Any raw distance may be converted into a score by simply negating it. In other words, the larger the distance, the smaller the score.

rank-based classifier employs a *holdout* set, which constitutes a set of samples that are known (by the system) *not* to belong to the claimed individual. For each individual, a ranked *list* of distances (dissimilarities) between the feature set associated with the selected positive *reference* sample and those associated with the samples in the holdout set is first constructed. The distance between the feature set associated with the selected *positive* reference sample and the one associated with the *questioned* sample is subsequently *inserted* into the aforementioned ranked list, after which its rank is reported. The rank-based classifier then either accepts or rejects the questioned sample based on its *rank*.

Trials. Each experiment is performed by conducting N_T independent *trials*. During each trial, hand vein patterns from N_O individuals are used for performance *optimisation* purposes, while hand vein patterns from N_E *other* individuals are used for performance *evaluation* purposes. During each experiment, N_P *permutations* (unique, random orderings) of the individuals in the relevant data set are considered. For each permutation, cross-validation is conducted over N_F *folds*, resulting in $N_T = N_P N_F$ trials.

A trial within the context of this thesis therefore comprises of two main stages, namely (1) an *optimisation stage* and (2) an *evaluation stage*. During the optimisation stage the system is iteratively prompted to verify whether a number of *labelled* samples belong to certain individuals. Recall that within the context of this thesis said verification is either score-based or rank-based. It is therefore reasonable to presume that, within a suitable interval, a specific score or rank exists for which the system performs *optimally* (based on a specific statistical performance measure). An optimal decision threshold τ may subsequently be *selected* based on one of the following (discrete) statistical performance measures (see Section 5.4):

1. The *equal error rate* (EER) (where the false *acceptance* rate (FAR) equals the false *rejection* rate (FRR)).
2. The *zero false rejection rate* (FRR_{zero}) (where the FRR = 0).
3. The *zero false acceptance rate* (FAR_{zero}) (where the FAR = 0).

The evaluation stage outputs a set of performance measures when *only* the selected optimal score-based or rank-based decision threshold is employed for the purpose of authenticating a number of *unlabelled* (questioned) samples. We investigate scenarios in which an individual is granted 1, 3 and 6 opportunities respectively to present his/her hand to the system for the purpose of authentication. The proficiency of the employed system is consequently quantified by a set of *average* performance measures across *all* conducted trials.

5.2 The Bosphorus data set

The Bosphorus hand-vein data set (acquired by Bogazici University (2010)) contains a total of 1575 data samples from 100 different individuals. Each data sample constitutes a *near infra-red (NIR)*, grey-scale image of the *dorsal* surface of the left (or right) hand of a specific individual (see Figure 5.1).



Figure 5.1: A typical NIR image from the Bosphorus data set.

It is clear from Figure 5.1 that the prominent (well-defined) veins on the dorsal surface of the hand have been successfully acquired. The acquisition of said veins is made possible by employing a suitable NIR camera, as well as two infra-red light sources that irradiate the back of the hand.

The following categorisation of the 12 *left-hand* images associated with a particular individual from the Bosphorus data set is applicable within the context of this thesis:

- Three images are acquired under normal conditions (room temperature, atmospheric pressure etc.),
- three images are acquired after a 3-kg bag has been carried for a period of 1 minute,
- three images are acquired after an elastic ball has been repetitively squeezed for a period of 1 minute and
- three images are acquired after an ice pack has been held to the back of the hand for a period of 1 minute.

A grand total of 1200 left-hand images belonging to 100 individuals that form part of the Bosphorus data set are therefore considered. Note that right-hand and time-lapsed samples are not considered.

5.3 The CIE data set

The CIE hand-vein data set (acquired by The Institute Of Control And Information Engineering (2013)) comprises of 2400 data samples acquired from the left and right hands of 50 different individuals. One half of the data samples constitutes *near infra-red (NIR)* grey-scale images of the *palmar* surface of the left and right hands of the aforementioned individuals, while the remaining half constitutes NIR grey-scale images of the left and right *wrists* of said individuals. Figure 5.2 shows a typical *palmar* NIR grey-scale image. Prominent veins are highlighted by employing a suitable NIR-based camera.

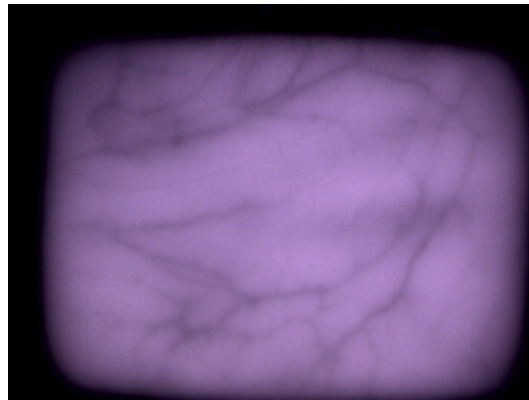


Figure 5.2: A typical palmar NIR image from the CIE data set.

The following categorisation of the 48 images associated with a particular individual from the CIE data set is applicable within the context of this thesis:

- 12 images of the palmar surface of the left hand,
- 12 images of the palmar surface of the right hand,
- 12 images of the palmar surface of the left wrist, and
- 12 images of the palmar surface of the right wrist.

We opt to only consider the images in the first of the aforementioned categories, resulting in a grand total of 600 palmar left-hand images for the 50 individuals that form part of the CIE data set. Note therefore that neither the right hand samples nor the wrist samples are considered within the context of this thesis.

5.4 Statistical performance measures

In this section we discuss a number of statistical performance measures employed for the purpose of quantifying the proficiency of the proposed identity verification systems. Recall that a trial constitutes a number of authentications made by a specific system. Such an authentication may have one of the following four outcomes:

1. A true positive (t^+) (when a *positive* sample is *accepted*);
2. A true negative (t^-) (when a *negative* sample is *rejected*);
3. A false positive (f^+) (when a *negative* sample is *accepted*); or
4. A false negative (f^-) (when a *positive* sample is *rejected*).

The first two of the aforementioned outcomes constitute correct decisions, while the last two constitute mistakes. The total number of positive and negative questioned samples within a specific trial are denoted by n^+ and n^- respectively. The total number of outcomes within a specific trial which are associated with points 1 to 4 listed above, are denoted by T^+ , T^- , F^+ and F^- respectively.

In this thesis we opt to employ the following 5 statistical performance measures for the purpose of quantifying the proficiency of the proposed systems:

$$\text{False rejection rate (FRR)} = \frac{F^-}{n^+} \quad (5.1)$$

$$\text{False acceptance rate (FAR)} = \frac{F^+}{n^-} \quad (5.2)$$

$$\text{Average error rate (AER)} = \frac{\text{FAR} + \text{FRR}}{2} \quad (5.3)$$

$$\text{Precision (PRES)} = \frac{1 - \text{FRR}}{(1 - \text{FRR}) + \text{FAR}} \quad (5.4)$$

$$\text{Recall (REC)} = \frac{1 - \text{FRR}}{(1 - \text{FRR}) + \text{FRR}} \quad (5.5)$$

Note that each of the 5 aforementioned statistical measures are within the range $[0, 1]$. The first 3 measures constitute error rates and should therefore be as close to zero as possible. The last 2 measures provide an indication of system accuracy and should therefore be as close to one as possible.

The remainder of this chapter is dedicated to simulations involving the *k-one* and *k-six* scenarios.

5.5 Experimental protocol: the *k-one* scenario

The *k-one* scenario simulates the real-world situation in which a certain company, that aims to employ a proposed authentication system, possesses *only one* positive sample per enrolled individual (client). The system is therefore compelled to use *said* sample as the *reference* sample for the corresponding individual during template matching (see Chapter 4). The experimental protocol for the *k-one* scenario furthermore has no other option but to employ a *global* decision threshold for *all* claimed identities within a certain trial. The estimation of an *individual-specific* decision threshold based on either the EER, FAR_{zero} or FRR_{zero} by considering an individual-specific optimisation set is not possible within this context, since such an approach requires *more* than one positive sample, as well as several negative samples for *every* client. Individual-specific *score normalisation* by (for example) calculating the z-score, followed by the estimation of an optimal *global* decision threshold (based on the *normalised* scores) for all enrolled individuals is not feasible either, since this approach *also* requires multiple positive samples per enrolled individual.

Whenever a score-based decision threshold is employed within the context of the *k-one* scenario, the utilised data set is partitioned into (1) an optimisation set and (2) an evaluation set (see Figure A.1). Whenever a rank-based decision threshold is employed, the utilised data set is however partitioned into (1) an optimisation set, (2) an evaluation set *and* (3) a holdout set (see Figure A.2). The purpose of the holdout set was explained in Section 5.1. Recall that each partition contains samples from *different* individuals.

The optimisation stage of a trial entails the authentication of *all* (labelled) optimisation samples by employing various *global* decision thresholds (within a suitable range) to the *raw* (unnormalised) scores or ranks. The *optimal* (global) decision threshold is subsequently *selected* so that it corresponds to either the EER, FAR_{zero} or FRR_{zero} . The selected optimal decision threshold is subsequently applied during system evaluation in order to authenticate a number of *unlabelled* evaluation samples associated with *other* individuals than those considered during system optimisation. This prevents biased results.

The set of performance measures obtained after the evaluation stage therefore indicates system proficiency when *no* individual-specific information is available for proficiency *enhancement* purposes². The number of permutations and folds associated with an experiment within the context of the *k-one* scenario is given by $N_P = 10$ and $N_F = 4$ respectively. The selection of the single *reference* sample is also cross-validated $N_R = 12$ times³ *within* a specific trial. The number of independent trials N_T therefore remains unchanged, and is given by $N_T = N_P N_F = 40$.

²Within the context of the *k-six* scenario, a number of proficiency enhancement strategies are proposed in Section 5.6.

³12 positive samples are acquired per individual in the data sets being considered.

The experimental protocol for the *k-one* scenario is diagrammatically conceptualised in Figures A.1 and Figure A.2 within the context of a score-based and rank-based classifier respectively.

5.6 Experimental protocol: the *k-six* scenario

The *k-six* scenario simulates the real-world situation in which a certain company, that aims to employ a proposed authentication system, possesses *six* positive samples per client. The employed system therefore has the opportunity to extract information from the *six* available positive samples in order to improve (enhance) system proficiency (accuracy). We investigate three so-called *proficiency enhancement strategies (PESs)* in this study.

Proficiency enhancement strategy 0 (PES0). PES0 involves the selection of the *most representative* of the *six* available samples for a specific individual as his/her *reference* sample (template) within the context of the *k-six* scenario. It is important to note that this stands in contrast to the *k-one* scenario, in which case the utilisation of the *only available* reference sample is cross-validated 12 times across the 12 different samples acquired per individual in the relevant data sets, after which an *average* performance is reported.

Within the context of PES0, the most representative sample associated with a specific individual may be *automatically* determined as follows: Each positive sample is (in turn) matched to the other five (available) positive samples by calculating the average (Euclidean or DTW-based) distance between their corresponding feature vectors. The sample that differs the *least* from the other five samples is deemed the most representative.

Note that PES0 *only* entails the selection of the most representative sample from the six available positive samples. Once selected, the remaining five positive samples are simply discarded and *not* employed for subsequent *individual-specific* optimisation purposes (as is the case for PES1 and PES2). Note that PES0 is *only* investigated within the context of a so-called *benchmark experiment* (benchmark system) that specifically involves the *k-six* scenario and a score-based classifier (see Section 5.7.1). PES0 is therefore compelled to employ a *global* decision threshold to the raw (unnormalised) scores.

Proficiency enhancement strategy 1 (PES1). *After* the templates have been selected for all enrolled individuals through the strategy employed for PES0, system proficiency may be *further* enhanced by applying appropriate individual-specific score normalisation by considering the remaining five positive (optimisation) samples for each client. An optimal *global* threshold is subsequently selected among the *normalised* scores.

Note that PES1 is *only* investigated within the context of a benchmark experiment that specifically involves the *k-six* scenario and a score-based clas-

sifier (see Section 5.7.1). For a detailed discussion on the employed z-score normalisation protocol which is specifically adapted to scenarios in which only a few positive optimisation samples (five in this case) are available per individual, the reader is referred to page 55 of Fick (2017).

The main disadvantages associated with implementing PES1 in this study involves the fact that (1) the small number of available positive (optimisation) samples per client (five) may result in an unreliable (statistical) estimate of the intra-class variance associated with each client, while (2) said intra-class variance may have already been adequately reduced through the normalisation techniques discussed in Chapter 4.

Proficiency enhancement strategy 2 (PES2). In this study we opt to extensively investigate PES2, which involves the estimation of an optimal tailor-made (individual-specific) decision threshold (based on either the EER, FAR_{zero} or FRR_{zero}) for each enrolled individual by considering both positive *and* negative individual-specific optimisation samples. Within this context, each trial therefore comprises of *individual-specific* optimisation and evaluation stages. Note that the disadvantages associated with PES1 are not applicable to PES2, since PES2 does not require an estimation of the intra-class variance. For this reason PES2 is expected to be more reliable.

The data partitioning protocol is (in this case) similar to the one outlined in Section 5.5. Recall that the purpose of the holdout-set within the context of a rank-based classifier has already been explained in Section 5.1. The optimisation set within the current context contains (1) *five* positive optimisation samples for each enrolled individual - recall that the most representative sample has been removed after selection through the strategy employed for PES0 - as well as (2) a set of negative samples associated with *other* individuals (so-called optimisation negatives). The evaluation set contains (1) the remaining *six* positive samples (i.e. different samples than those used for optimisation purposes) and (2) negative samples (evaluation negatives) associated with *other* individuals than those considered for optimisation purposes.

The individual-specific optimisation stage of a trial within the context of the *k-six* scenario therefore involves the iterative authentication of (1) the remaining *five* positive optimisation samples and (2) all of the negative optimisation samples. Said process is repeated for a suitable range of (rank-based or score-based) decision thresholds, after which an optimal *individual-specific* decision threshold is selected (based on either the EER, FAR_{zero} or FRR_{zero}).

The selected optimal threshold is subsequently employed in the (individual-specific) evaluation stage of the abovementioned trial for the purpose of authenticating (1) the remaining *six* positive *evaluation* samples and (2) all of the (negative evaluation) samples that do *not* belong to the claimed individual.

The evaluation stage of a trial within the current context outputs a set of performance measures for *each* individual (client). Said measures are subse-

quently averaged in order to obtain a set of performance measures across *all* enrolled (evaluation) individuals. The number of permutations and folds associated with an experiment within the context of the *k-six* scenario are given by $N_P = 10$ and $N_F = 4$ respectively. The selection of the *six* positive optimisation samples from the acquired 12 samples in the relevant data sets is also cross-validated $N_R = 12$ times *within* a specific trial. The number of independent trials N_T therefore remains unchanged and is given by $N_T = N_P N_F = 40$.

The experimental protocol for the *k-six* scenario (with PES2) is diagrammatically conceptualised in Figures A.3 and A.4 within the context of a score-based and rank-based classifier respectively.

5.7 Results

5.7.1 Preliminary discussion

Experimental categorisation. The parameters utilised in the experiments conducted in this study may be categorised into so-called *main* parameters and so-called *sub* parameters. The main parameters associated with a specific experiment (system) constitutes (1) the type of scenario simulated (*k-one* or *k-six*) and (2) the criterion by which an optimal decision threshold is selected (EER, FAR_{zero} or FRR_{zero}). The sub parameters include (1) the dissimilarity measure employed (Euclidean-based or DTW-based), (2) the type of classifier employed (rank-based or score-based) and (3) the number of opportunities granted for authentication during system *evaluation* (1, 3 or 6). Note that (by design) an individual is *only* granted one opportunity for authentication during the optimisation stage, and that the *bandwidth* associated with the DTW-based dissimilarity measure (see Section 4.3.2) is fixed at 10% of the dimension of a feature vector for all DTW-based experiments.

The purpose of the aforementioned categorisation is to facilitate the presentation of the results in groups that are based on the main parameters of each experiment. For example, the results presented in Section 5.7.2 constitute all experiments conducted within the context of the *k-one* scenario and the EER criterion. Note however that said categorisation is only applicable within the context of the Bosphorus data set, since only a selected number of experiments are conducted on the CIE data set (due to time constraints).

Benchmark experiments. The aforementioned experimental categorisation enables one to define a *benchmark* experiment for each main group. The benchmark system within each group (1) employs the EER for selecting the optimal decision threshold, (2) employs the Euclidean distance for quantifying the dissimilarity between two feature sets and (3) grants a questioned individual 3 opportunities for authentication. The purpose of each benchmark experiment is to obtain a *balanced* (reference) measure of the proficiency of

the benchmark system for a specific group of experiments. A benchmark experiment is deemed "balanced" in the sense that the utilisation of the EER for selecting the optimal decision threshold is neither strict (as in scenarios where FAR_{zero} is employed), nor lenient (as in scenarios where FRR_{zero} is employed). Granting a questioned individual three opportunities for authentication, instead of one (strict) or six (lenient), may also be deemed balanced within the current context. The benchmark results for *adapted* z-score normalisation (PES1) within the context of a score-based classifier and the *k-six* scenario are included in Table 5.11. The benchmark results for applying PES0 within the aforementioned context are also included in Table 5.11. In doing so, an estimate of the proficiency *gained* (if any) when employing PES1 or PES2 (instead of *only* PES0) may be obtained.

Ideal experiments. So-called *ideal* experiments are conducted in order to demonstrate the proficiency of the benchmark systems within the context of the *k-one* scenario when *only* the most representative sample for each enrolled individual is employed for reference purposes. This stands in contrast to the non-ideal (more realistic) experimental protocol for the *k-one* scenario (as outlined in Section 5.5), in which case the reference sample is cross-validated 12 times across all 12 acquired positive samples for an individual within the employed data sets. The rationale behind these *ideal* experiments is that any company, that decides to (1) employ a template-based authentication system and (2) acquire only *one* sample per prospective member during enrollment, is likely to ensure that the acquired sample is of *excellent quality*. However, it is in *reality* not possible to *select* the most representative sample within the context of the *k-one* scenario, since only one sample is available (per definition). However, since a *k-one* scenario is merely *simulated* in this thesis, we are not limited by the aforementioned constraint. Within the context of the *k-one* scenario, the most representative sample may therefore be selected from the 12 acquired positive samples (in the corresponding data set) associated with a specific individual, using the strategy outlined in Section 5.6 (PES0). In so doing, we *simulate* what a company is likely to do: The image acquisition protocol may be set up ideally (perfectly), or the most representative sample may be ideally (manually or automatically) selected from a group of candidates. The results for the ideal experiments are therefore based on the *assumption* that a company is able to acquire the *same* single high-quality reference sample for each individual as the one automatically obtained through the strategy outlined in Section 5.6 (PES0). The ideal experiments therefore simulate a scenario in which a high quality sample is acquired by the relevant company in *controlled* conditions. The (eleven) positive evaluation samples are of lesser quality and obtained in less controlled conditions at, for example, an ATM or POSM. The *ideal* results may therefore not be entirely unrealistic.

5.7.2 The *k-one* scenario in which EER is employed

An EER-based system may be labelled neither strict nor lenient, since it aims to select the optimal decision threshold associated with an equal number of errors (FAR = FRR) during system optimisation. The results for the *k-one* scenario in which EER is employed are presented in Tables 5.1, 5.2 and 5.3.

Table 5.1: Bosphorus data set: EER. Experimental results for the *k-one* scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	19.06	19.22	20.97	20.69
FRR	18.12	18.63	20.94	20.62
AER	18.59	18.92	20.95	20.66
PRES	81.12	80.89	79.04	79.33
REC	81.88	81.37	79.06	79.38

Table 5.2: Bosphorus data set: EER. Experimental results for the *k-one* scenario in which an individual is granted *three* opportunities for authentication. Note that the benchmark results are indicated in boldface and that the results for the *ideal* (benchmark) experiments are indicated in italics.

(%)	RANK			SCORE		
	Benchmark	Ideal	DTW	Benchmark	Ideal	DTW
FAR	24.05	<i>1.42</i>	24.06	26.35	<i>1.62</i>	26.19
FRR	5.43	<i>0.11</i>	2.49	8.36	<i>0.17</i>	4.56
AER	14.74	<i>0.77</i>	13.27	17.35	<i>0.89</i>	15.37
PRES	79.73	<i>98.58</i>	80.21	77.67	<i>98.4</i>	78.47
REC	94.57	<i>99.89</i>	97.51	91.64	<i>99.83</i>	95.44

Table 5.3: Bosphorus data set: EER. Experimental results for the *k-one* scenario in which an individual is granted *six* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	38.36	39.68	38.09	39.35
FRR	1.42	0.27	2.75	0.66
AER	19.89	19.98	20.42	20.01
PRES	71.99	71.54	71.86	71.63
REC	98.58	99.73	97.25	99.34

Whenever the EER criterion is employed during optimisation and a questioned individual is granted *only one* opportunity for authentication during evaluation, it is reasonable to expect that the error rates will be well-balanced ($FAR \approx FRR$). This is clearly illustrated in Table 5.1.

Increasing system leniency. The leniency of the system increases when more opportunities are granted to a questioned individual for authentication. This tendency is clearly indicated by the increase in FAR and decrease in FRR when the results in Table 5.1 are compared to those depicted in Tables 5.2 and 5.3.

Benchmark results. The AERs achieved by the rank-based and score-based benchmark systems are given by 14.74% and 17.35% respectively, which indicates that system proficiency is adequate, *even* in extreme cases where *only one* sample is available per enrolled individual, while said sample is not necessarily representative.

Ideal benchmark results. It is clear from Table 5.2 that the AERs achieved by the rank-based and score-based ideal benchmark systems are *significantly* smaller than those achieved by the corresponding benchmark systems (0.77% versus 14.74% and 0.89% versus 17.35%). This significant decrease in system proficiency clearly indicates the importance of selecting a reference sample of excellent quality for each enrolled client within the context of template-based authentication.

Rank-based versus score-based classification. The AERs achieved by the rank-based and score-based benchmark systems are given by 14.74% and 17.35% respectively. The higher proficiency associated with the rank-based classifier may be attributed to the fact that said classifier effectively ignores the relative differences between the successive scores in a ranked list.

Euclidean-based versus DTW-based distances. Recall that a DTW-based authentication system is (theoretically) more lenient than a Euclidean-based system by design, since it attempts to find a better correlation between the feature vectors being matched. For this reason, it is reasonable to expect a slight increase in the FAR, and a relatively larger decrease in the FRR, when the results associated with a DTW-based system is compared to those associated with a Euclidean-based system. This trend is clearly observed in Tables 5.2 and 5.3, but not in Table 5.1. The aforementioned changes may however be deemed insignificantly small, thereby rendering it irrelevant within the current context.

5.7.3 The k -one scenario in which FAR_{zero} is employed

A FAR_{zero} -based system may be considered to be very *strict*, in the sense that it aims to accept *no* negative samples. The results for the k -one scenario in which FAR_{zero} is employed are presented in Tables 5.4, 5.5 and 5.6.

Table 5.4: Bosphorus data set: FAR_{zero} . Experimental results for the k -one scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	0.33	0.26	0.0007	0.006
FRR	50.30	54.74	80.95	92.17
AER	25.31	27.50	40.47	46.09
PRES	99.34	99.44	99.99	99.92
REC	49.70	45.26	19.05	7.83

Table 5.5: Bosphorus data set: FAR_{zero} . Experimental results for the k -one scenario in which an individual is granted *three* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	0.27	0.14	0.0	0.0
FRR	30.10	35.26	76.83	100
AER	15.18	17.70	38.41	50.0
PRES	99.62	99.79	100	100
REC	69.91	64.74	23.18	0.0

Table 5.6: Bosphorus data set: FAR_{zero} . Experimental results for the k -one scenario in which an individual is granted *six* opportunities for authentication. The best results within the context of the k -one scenario (*excluding* those for the *ideal* scenario) are boxed.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	0.76	0.52	0.001	0.0
FRR	4.95	13.84	53.68	98.67
AER	2.85	7.18	26.84	49.33
PRES	99.21	99.39	99.99	100
REC	95.05	86.16	46.32	1.33

Whenever the FAR_{zero} criterion is employed during optimisation, it is reasonable to expect relatively large FRRs and small FARs during evaluation. This is clearly illustrated in Tables 5.4, 5.5 and 5.6.

Increasing system leniency. The aforementioned results also indicate an increase in system leniency when the number of opportunities for authentication is increased. Note that there is however an exceptional case in which the FAR slightly *decreases* when the number of opportunities for authentication is increased. This may simply be due to the fact that very few positive optimisation samples are available within the context of the relevant data sets being considered, whereby slight abnormalities may be witnessed in the corresponding results (as is the case for the FARs in Tables 5.4 and 5.5).

Rank-based versus score-based classification. It may be observed from the results presented in this section that a rank-based classifier is *significantly* superior to a score-based classifier within the current context. Note that a score-based classifier however achieves *smaller* FARs than a rank-based classifier across *all* experiments within the current context, while also achieving significantly *larger* FRRs (see Table 5.4 for example).

The most proficient system. The results that are boxed in Table 5.6 are achieved by employing the Euclidean-based dissimilarity measure, a rank-based classifier and the FAR_{zero} criterion, while granting an individual 6 opportunities for authentication. Although the aforementioned combination of sub-parameters may appear contradicting, the implementation of the FAR_{zero} criterion ensures that the FAR is kept sufficiently small, while the fact that 6 opportunities are granted to a questioned individual (for authentication) compensates for the possibility that a significant number of the 6 positive evaluation samples may be of poor quality. The AER for the most proficient system within the current context is given by 2.85%.

5.7.4 The *k-one* scenario in which FRR_{zero} is employed

A FRR_{zero} -based system may be considered to be very *lenient*, in the sense that it aims to accept *all* positive samples. The results for the *k-one* scenario in which FRR_{zero} is employed are presented in Tables 5.7, 5.8 and 5.9.

Table 5.7: Bosphorus data set: FRR_{zero} . Experimental results for the *k-one* scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	99.29	99.35	99.45	99.93
FRR	0.07	0.09	0.054	0.047
AER	49.68	49.72	49.75	49.99
PRES	50.16	50.14	50.12	50.01
REC	99.93	99.91	99.95	99.95

Table 5.8: Bosphorus data set: FRR_{zero} . Experimental results for the *k-one* scenario in which an individual is granted *three* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	99.91	99.97	99.87	99.99
FRR	0.0	0.0	0.0	0.0
AER	49.97	49.98	49.93	49.99
PRES	50.02	50.01	50.03	50.002
REC	100	100	100	100

Table 5.9: Bosphorus data set: FRR_{zero} . Experimental results for the *k-one* scenario in which an individual is granted *six* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	99.99	99.99	99.96	99.99
FRR	0.0	0.0	0.0	0.0
AER	49.99	49.99	49.98	49.99
PRES	50.002	50.002	50.01	50.001
REC	100	100	100	100

Whenever the FRR_{zero} criterion is employed during optimisation, it is reasonable to expect relatively large FARs and small FRRs during evaluation. This is clearly illustrated in Tables 5.7, 5.8 and 5.9.

Increasing system leniency. It is clear from the results presented in this section that the system is extremely lenient within the current context, even in cases where *only one* opportunity is granted for authentication. A slight increase in leniency is however still visible when the number of opportunities for authentication is increased.

Rank-based versus score-based classification. The results depicted in Tables 5.7, 5.8 and 5.9 do not clearly indicate whether or not a rank-based classifier is superior to a score-based classifier within the current context.

5.7.5 The *k-six* scenario in which EER is employed

Recall that the EER criterion may be considered to be neither strict nor lenient. The results for the *k-six* scenario in which EER is employed are presented in Tables 5.10, 5.11 and 5.12. Whenever the PES is not explicitly specified in the aforementioned tables it may be presumed that PES2 is being employed.

Table 5.10: Bosphorus data set: EER. Experimental results for the *k-six* scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	12.25	14.22	12.27	14.32
FRR	18.43	17.64	20.72	19.22
AER	15.34	15.93	16.5	16.77
PRES	86.94	85.28	86.6	84.94
REC	81.57	82.36	79.28	80.78

Table 5.11: Bosphorus data set: EER. Experimental results for the *k-six* scenario in which an individual is granted *three* opportunities for authentication. Note that the benchmark results are indicated in boldface. The results for PES0 and PES1 within the context of a score-based classifier are also included.

(%)	RANK		SCORE			
	Euclidean	DTW	Euclidean	DTW	PES1	PES0
FAR	16.02	18.78	16.14	18.95	21.72	22.9
FRR	1.08	0.98	1.86	1.13	1.08	3.29
AER	8.55	9.88	9.00	10.04	11.40	13.09
PRES	86.06	84.06	85.88	83.92	81.99	80.85
REC	98.92	99.02	98.14	98.87	98.92	96.71

Table 5.12: Bosphorus data set: EER. Experimental results for the *k-six* scenario in which an individual is granted *six* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	24.64	30.59	24.63	30.79
FRR	0.0	0.26	0.17	0.33
AER	12.32	15.42	12.39	15.56
PRES	80.23	76.53	80.21	76.39
REC	100	99.74	99.83	99.67

When the results in Tables 5.1 and 5.10 are compared, it is clear that the FRRs are similar in the majority of cases, while the FARs are *significantly* smaller in Table 5.10. The aforementioned smaller FARs may be attributed to the fact that individual-specific threshold optimisation appropriately cater for those individuals of which the samples are of relatively low quality by selecting a more lenient decision threshold. The system therefore tends to accept more negative samples within the context of the *k-one* scenario than is the case for the *k-six* scenario. The aforementioned gain in proficiency (as a result of employing PES2) is furthermore evident when the results depicted in Table 5.2 (or 5.3) are compared to those in Table 5.11 (or 5.12).

Proficiency gained when employing various PESs. The results presented in Table 5.11 firstly indicate that PES2 is superior to PES1. This may be due to the fact that intra-class variance had *already* been adequately reduced during feature extracting and matching (as explained in Section 5.6).

The aforementioned results also indicate that PES1 is superior to PES0. This is due to the fact that the remaining five positive (optimisation) samples for each client *are* relevant, and may therefore be sensibly employed for the purpose of enhancing system proficiency.

It is clear from the results presented in Tables 5.2 and 5.11 that PES0 (within the context of the *k-six* scenario) is superior to the corresponding *k-one* scenario in which the employed template is simply cross-validated 12 times. This furthermore indicates the importance of utilising an excellent quality reference sample within the context of template matching.

Rank-based versus score-based classification. The results presented in Section 5.7.5 *also* indicate that a rank-based classifier is superior to a score-based classifier with *no* exceptions. The improvement in FAR is however small in some cases (see Tables 5.10 and 5.12).

5.7.6 The k -six scenario in which FAR_{zero} is employed

Recall that a system that employs the FAR_{zero} criterion may be considered to be very *strict*, in the sense that it aims to reject *all* negative samples. The results for the k -six scenario in which FAR_{zero} is employed are presented in Tables 5.13, 5.14 and 5.15.

Table 5.13: Bosphorus data set: FAR_{zero} . Experimental results for the k -six scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	0.99	0.68	0.2	0.14
FRR	33.99	41.89	43.21	61.13
AER	17.49	21.29	21.71	30.63
PRES	98.53	98.84	99.64	99.65
REC	66.01	58.11	56.79	38.89

Table 5.14: Bosphorus data set: FAR_{zero} . Experimental results for the k -six scenario in which an individual is granted *three* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	1.02	0.63	0.17	0.05
FRR	9.2	14.69	15.72	40.46
AER	5.11	7.66	7.94	20.26
PRES	98.89	99.26	99.79	99.91
REC	90.8	85.31	84.28	59.54

Table 5.15: Bosphorus data set: FAR_{zero} . Experimental results for the k -six scenario in which an individual is granted *six* opportunities for authentication. The best results within the context of the k -six scenario are boxed.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	2.43	1.85	0.48	0.22
FRR	3.03	7.06	5.43	18.76
AER	2.73	4.45	2.95	9.49
PRES	97.55	98.05	99.43	99.73
REC	96.97	92.94	94.58	81.24

Recall that it is reasonable to expect relatively large FRRs and small FARs during evaluation whenever the FRR_{zero} criterion is employed. This is clearly illustrated in Tables 5.13, 5.14 and 5.15.

The most proficient system. The boxed results in Table 5.15 are achieved by employing the same specific combination of experimental parameters that also lead to the best results within the context of *k-one* scenario. The AER for the most proficient system within the current context is given by 2.73%.

5.7.7 The *k-six* scenario in which FRR_{zero} is employed

Recall that a system that employs the FRR_{zero} criterion may be considered to be very *lenient*, since it aims to accept *all* positive samples. The results for the *k-six* scenario in which FRR_{zero} is employed are presented in Tables 5.16, 5.17 and 5.18.

Table 5.16: Bosphorus data set: FRR_{zero} . Experimental results for the *k-six* scenario in which an individual is granted *only one* opportunity for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	21.4	31.54	21.56	32.19
FRR	14.72	13.31	17.06	15.29
AER	18.06	22.43	19.31	23.74
PRES	79.94	73.32	79.37	72.46
REC	85.28	86.69	82.94	84.71

Table 5.17: Bosphorus data set: FRR_{zero} . Experimental results for the *k-six* scenario in which an individual is granted *three* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	26.23	39.10	26.5	39.77
FRR	0.84	0.74	1.58	0.92
AER	13.53	19.92	14.04	20.34
PRES	79.08	71.74	78.78	71.36
REC	99.16	99.26	98.42	99.08

Table 5.18: Bosphorus data set: FRR_{zero} . Experimental results for the k -*six* scenario in which an individual is granted *six* opportunities for authentication.

(%)	RANK		SCORE	
	Euclidean	DTW	Euclidean	DTW
FAR	33.71	46.28	33.8	48.9
FRR	0.0	0.0	0.17	0.25
AER	16.85	23.14	16.98	24.57
PRES	74.79	68.36	74.71	67.1
REC	100	100	99.83	99.75

Recall that whenever the FRR_{zero} criterion is employed during optimisation, it is reasonable to expect relatively large FARs and small FRRs during evaluation. This is clearly illustrated in Tables 5.16, 5.17 and 5.18.

When the results presented in this section are compared to those presented in Section 5.7.4, it is clear that significantly smaller FARs are achievable within the context of the k -*six* scenario than is the case for the k -*one* scenario.

5.7.8 CIE data set

The results for a *selected* number of experiments conducted on the CIE data set are presented in Tables 5.19 and 5.20. Said experiments may be divided into (1) the benchmark experiments, (2) the *ideal* benchmark experiments and (3) experiments that employ the same experimental parameters for which the *best* results are achieved within the context of the Bosphorus data set. Recall that the aforementioned (best) parameters constitute (1) a rank-based classifier, (2) the FAR_{zero} criterion, (3) six opportunities for authentication and (4) the Euclidean-based dissimilarity measure (see Tables 5.6 and 5.15).

Table 5.19: CIE data set: k -*one* scenario. The results for the benchmark experiments, the ideal (benchmark) experiments, as well as for the system for which the **best** results are achieved when the Bosphorus data set is considered.

(%)	RANK			SCORE	
	Benchmark	Ideal	Best	Benchmark	Ideal
FAR	43.94	<i>3.10</i>	0.56	46.7	<i>3.28</i>
FRR	9.62	<i>0.16</i>	42.01	18.51	<i>0.52</i>
AER	26.78	<i>1.63</i>	21.29	32.60	<i>1.90</i>
PRES	67.29	<i>96.99</i>	99.04	63.57	<i>96.81</i>
REC	90.38	<i>99.84</i>	57.99	81.49	<i>99.48</i>

Table 5.20: CIE data set: k -six scenario. The results for the benchmark system and for the system for which the **best** results are achieved when the Bosphorus data set is employed.

(%)	RANK		SCORE
	Benchmark	Best	Benchmark
FAR	38.77	0.77	38.65
FRR	5.03	28.54	5.49
AER	21.9	14.66	22.07
PRES	71.01	98.93	70.98
REC	94.96	71.46	94.51

Rank-based versus score-based classification. It may be observed from the results depicted in Tables 5.19 and 5.20 that a rank-based classifier is also superior to a score-based classifier within the context of the CIE data set. For example, the AERs associated with the aforementioned classifiers within the context of the k -one scenario are given by 26.78% (rank-based) and 32.7% (score-based), while the corresponding AERs within the context of the k -six scenario are given by 21.9% (rank-based) and 22.07% (score-based).

Benchmark results. The benchmark results for the CIE data set are significantly worse than the corresponding results associated with the Bosphorus data set. This indicates that the effect of employing a relatively low quality reference sample is more severe within the context of the CIE data set. This may also be attributed to the fact that the CIE data set comprises of *palmar* hand vein images, which contains significantly more irrelevant information (like palmar lines) than is the case for the Bosphorus data set.

Ideal benchmark results. The ideal benchmark results associated with the CIE data set is comparable to the (exceptional) results achieved within the current context for the Bosphorus data set. This indicates that the system is highly proficient within the context of both data sets being considered, *provided* that an excellent quality reference sample is selected for each client.

Best results. The system is able to achieve similar (and even better) FARs when the most proficient system parameters within the context of the Bosphorus data set is applied to the CIE data set (compare Tables 5.6 and 5.15 to Tables 5.19 and 5.20). The corresponding FRRs are however significantly larger in the case of the CIE data set, which may also be attributed to the fact that the effect of a poor quality reference sample is much more severe within the context of the CIE (palmar hand vein) data set.

5.7.9 Experimental results for existing hand vein-based authentication systems

In order to place the proficiency of the systems proposed in this thesis into perspective, we briefly report on the proficiency of a number of existing hand vein-based authentication systems. One should however guard against drawing strong conclusions from such a comparison, since the relevant experimental protocols may differ *substantially*.

The experimental proficiency reported by Trabelsi *et al.* (2013) when all individuals in the Bosphorus data set are authenticated through an ANN is given by an *area under curve* (AUC) of 0.87 and an EER of approximately 13%.

The experimental proficiency reported by Krishnaveni *et al.* (2014) when all individuals (samples) in the Bosphorus data set are authenticated, by employing a dynamic score assignment scheme, is given by an EER of 1.26%. The so-called invariant *minutiae triangle* associated with a questioned hand vein sample (dynamically) determines the type of score assignment protocol employed.

Yuksel *et al.* (2011) estimated system proficiency by conducting a number of so-called *naive* and *informed* experiments on the Bosphorus data set. The main purpose of these experiments is to determine the effect of four *stress* exercises (see Section 5.2) on the structure of the hand veins associated with a specific individual. A total of five different *test* sets were utilised: four sets that contain samples acquired under the same stress exercise, as well as a fifth set that constitutes a combination of the first four sets. The authors subsequently conducted a number of *t-tests* for the purpose of comparing system proficiency when each of the five different test sets are employed. The proficiency of the various systems proposed by Yuksel *et al.* (2011) are evaluated by a number of statistical performance measures that include the EER and the so-called *identification rate*. Within the context of the naive tests, the authors reported an identification rate of 86.74% and an EER of 5.4%, while the best results within the context of the informed tests are given by an identification rate of 96.67% and an EER of 2.25%. They reported an even larger identification rate of 99.49% when certain feature matching and classification strategies are fused.

5.8 Concluding remarks

We are able to reach the following general conclusions based on the experiments conducted in this chapter:

1. A rank-based classifier is generally superior to a score-based classifier. A score-based classifier is however significantly more efficient.

2. The utilisation of a DTW-based distance instead of an Euclidean-based distance does not lead to a significant increase in system proficiency. The additional computational complexity is therefore not justifiable.
3. A tailor-made decision threshold (PES2) is superior to adapted z-score normalisation (PES1), which may be attributed to the fact that the statistical estimate of the intra-class variance is unreliable. PES1 is also superior to PES0, since PES0 does not utilise the five remaining positive optimisation samples for individual-specific optimisation purposes.
4. The utilisation of a relatively low quality template (instead of a high quality template) within the context of template-based authentication leads to a significant loss in the system's discriminative ability.
5. In scenarios where a template (reference sample) of *random* quality is considered for each client enrolled into the system, the best results achievable within the context of the Bosphorus data set are given by AERs of 2.85% and 2.73% within the context of the *k-one* and *k-six* scenarios respectively. Both of these AERs are achieved when a rank-based classifier and Euclidean-based distance measure are employed, in conjunction with a strict optimal decision threshold (based on the FAR_{zero} criterion), while an individual is granted six opportunities to present his/her hand for authentication purposes. Worse results are reported for the CIE data set, which indicates that, in the case of the CIE data set, the proficiency of the system is dependent on the acquisition of a high quality (representative) template (reference sample).
6. In ideal scenarios where a *single* template (reference sample) of *very high* quality is guaranteed for each client enrolled into the system, the best results achievable within the context of the Bosphorus and CIE data sets are comparable and given by AERs of 0.77% and 1.63% respectively. Both of these AERs are achieved when a rank-based classifier and Euclidean-based distance measure are employed, in conjunction with a balanced optimal decision threshold (based on the EER criterion), while an individual is granted three opportunities to present his/her hand for authentication purposes.

In order to gain some insight into typical (correct and erroneous) authentication examples, the reader is referred to the images depicted in Figure B.1. These authentication outcomes are the result of conducting the benchmark experiments within the context of the *k-one* scenario and a rank-based classifier.

Chapter 6

Conclusion and future work

6.1 Conclusion

A number of novel hand-vein based identity verification systems were developed during the course of this study. The hand veins are successfully isolated (segmented) within an acquired (NIR) hand vein image through a novel hand vein segmentation protocol that is based on morphological reconstruction. The *discrete Radon transform (DRT)* is utilised for the purpose of extracting a suitable feature set associated with the segmented hand veins. The dissimilarity between two feature sets is represented by either the average Euclidean distance or by the average dynamic time warping (DTW)-based distance between corresponding feature vectors. A score-based or rank-based classifier is subsequently employed for verification purposes.

The novel hand vein segmentation protocol developed in this study exploits the fact that hand veins may generally be associated with *dark* and *narrow* structures within an infra-red image. A hand vein *pixel* may additionally be expected to be near *strong edges*. The proposed segmentation protocol therefore utilises (1) the *black top-hat* transform to extract *dark* and *narrow* structures and (2) the *Laplacian of Gaussian* transform to extract pixels near *strong edges*. The *intersection* of the results of the two aforementioned transforms is calculated and subsequently utilised as a *seed image* in morphological reconstruction by dilation. The corresponding *mask image* is obtained through adaptive thresholding by *moving averages*. Said hand vein segmentation protocol is shown to be robust (in the sense that very little re-calibration is required) when applied to more than one hand vein data set. Said segmentation protocol is also shown to be proficient *despite* the presence of noise such as non-uniform illumination (local shadows) and artefacts (like hair, palmar grooves and jewellery).

The application of the DRT is followed by the necessary *normalisation* techniques so as to ensure translation, scale and rotation invariance across all feature sets associated with all enrolled individuals.

The proficiency of the proposed system is estimated by utilising (1) the Bosphorus (dorsal hand vein) data set and (2) the CIE (palmar hand vein) data set. System proficiency is demonstrated to be exceptional in *ideal* scenarios in which the selection of a single excellent quality reference sample is guaranteed for each enrolled individual (client). The system is also shown to perform adequately in *extreme* scenarios in which *only one* training sample is available for each client, and the selection of a high quality reference sample is *not* guaranteed. A significant increase in system proficiency is witnessed when various proficiency enhancement strategies (PESs) are employed within the context of the *k-six* scenario, that is a scenario in which six training samples are available for each client. A significant increase in system proficiency is also witnessed when various proficiency enhancement strategies (PESs) within the context of the *k-six* scenario, in which more than one training sample is available for each client, are employed.

The objectives listed in Section 1.3 have therefore been successfully achieved. In the following section we discuss possible avenues for future work within the context of the proposed authentication systems.

6.2 Future work

The systems proposed in this thesis are specifically designed to deal with a lack of training samples (only one in the extreme cases and maximally six) by employing suitable template matching strategies. In order to investigate the feasibility of *learning* (training) more sophisticated statistical models, like probabilistic graphical models (i.e. hidden Markov models, etc.) or artificial neural networks, the availability of only one (or even six) positive samples for each client enrolled into the system is simply not sufficient. The artificial generation of additional training samples therefore constitutes a legitimate avenue for future research. An individual-independent strategy through the utilisation of dissimilarity vectors (in dissimilarity space) instead of (conventional) feature vectors (in feature space) (see Swanepoel (2015)) may also substantially increase the number of available individual-independent training samples and therefore warrants further investigation.

Within the context of template matching, the experiments conducted in this research are also not nearly exhaustive, due to time constraints. For example, investigations into the utilisation of an arbitrary number of positive training samples and granting a questioned individual an arbitrary number of opportunities for authentication may have provided some additional insight.

Appendices

Appendix A

Experimental conceptualisations

Score-based classifier for the *k-one* scenario

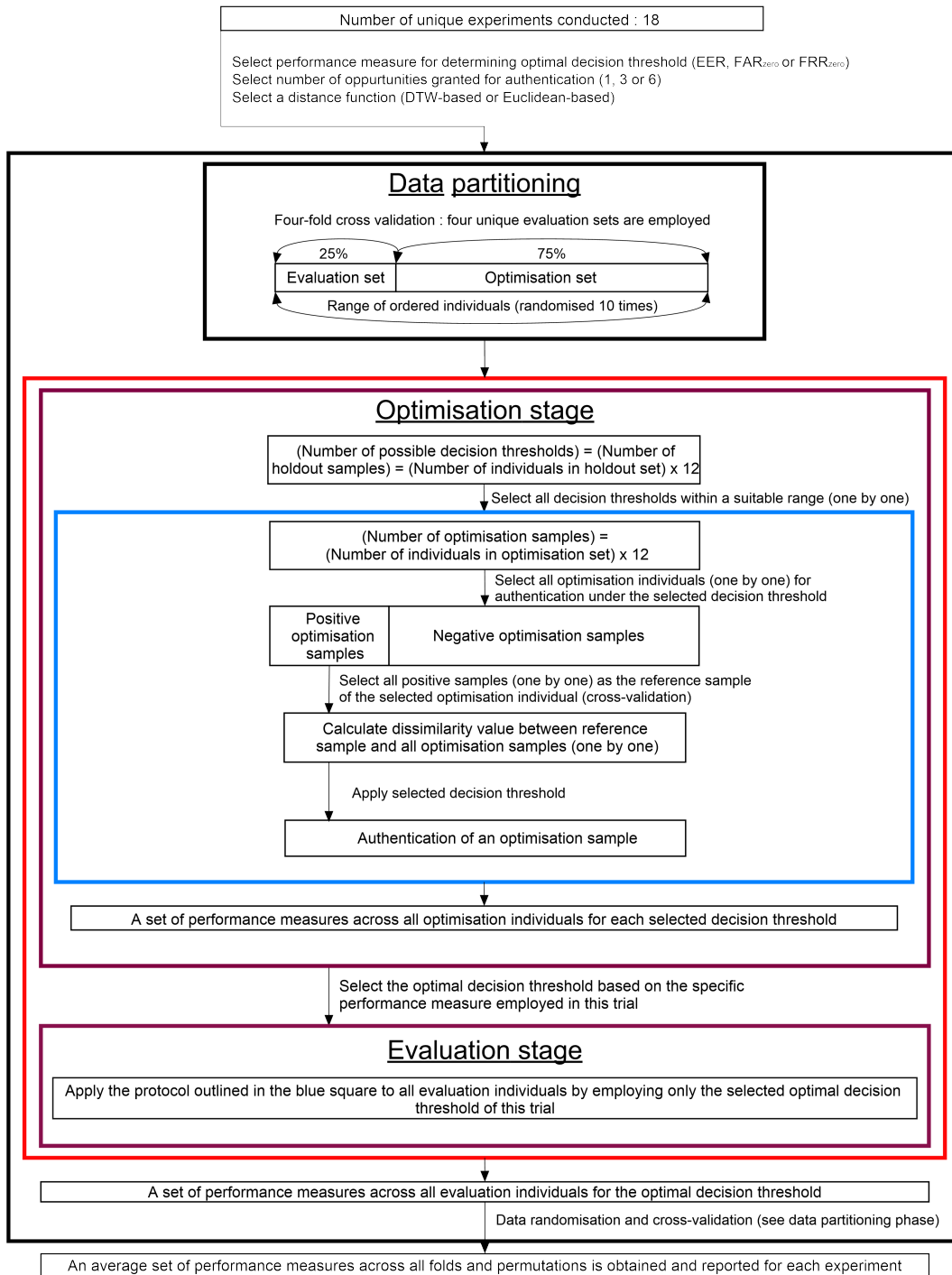


Figure A.1: Conceptualisation of the experimental protocol for the *k-one* scenario within the context of a score-based classifier.

Rank-based classifier for the *k-one* scenario

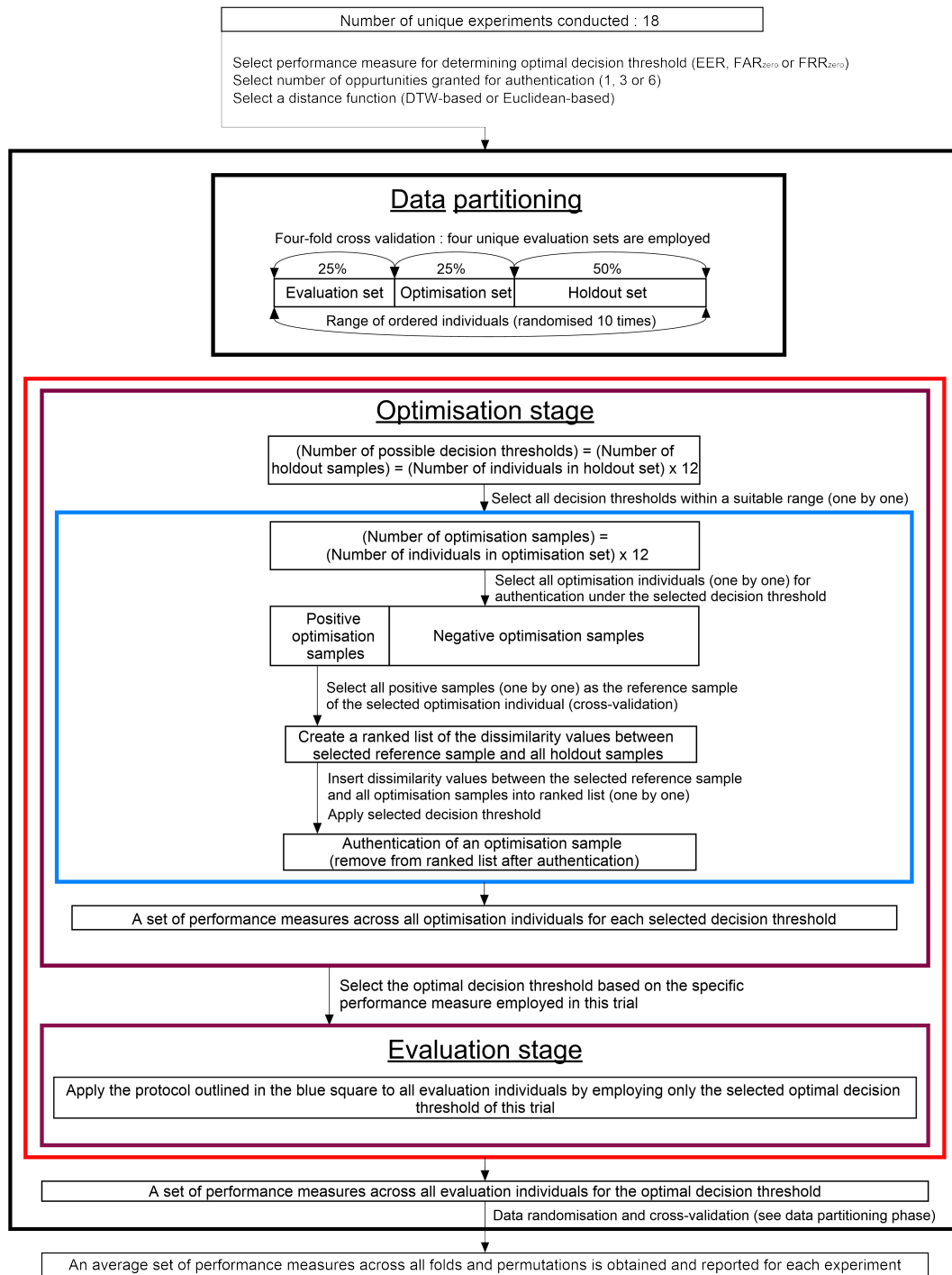


Figure A.2: Conceptualisation of the experimental protocol for the *k-one* scenario within the context of a rank-based classifier.

Score-based classifier for the *k-six* scenario

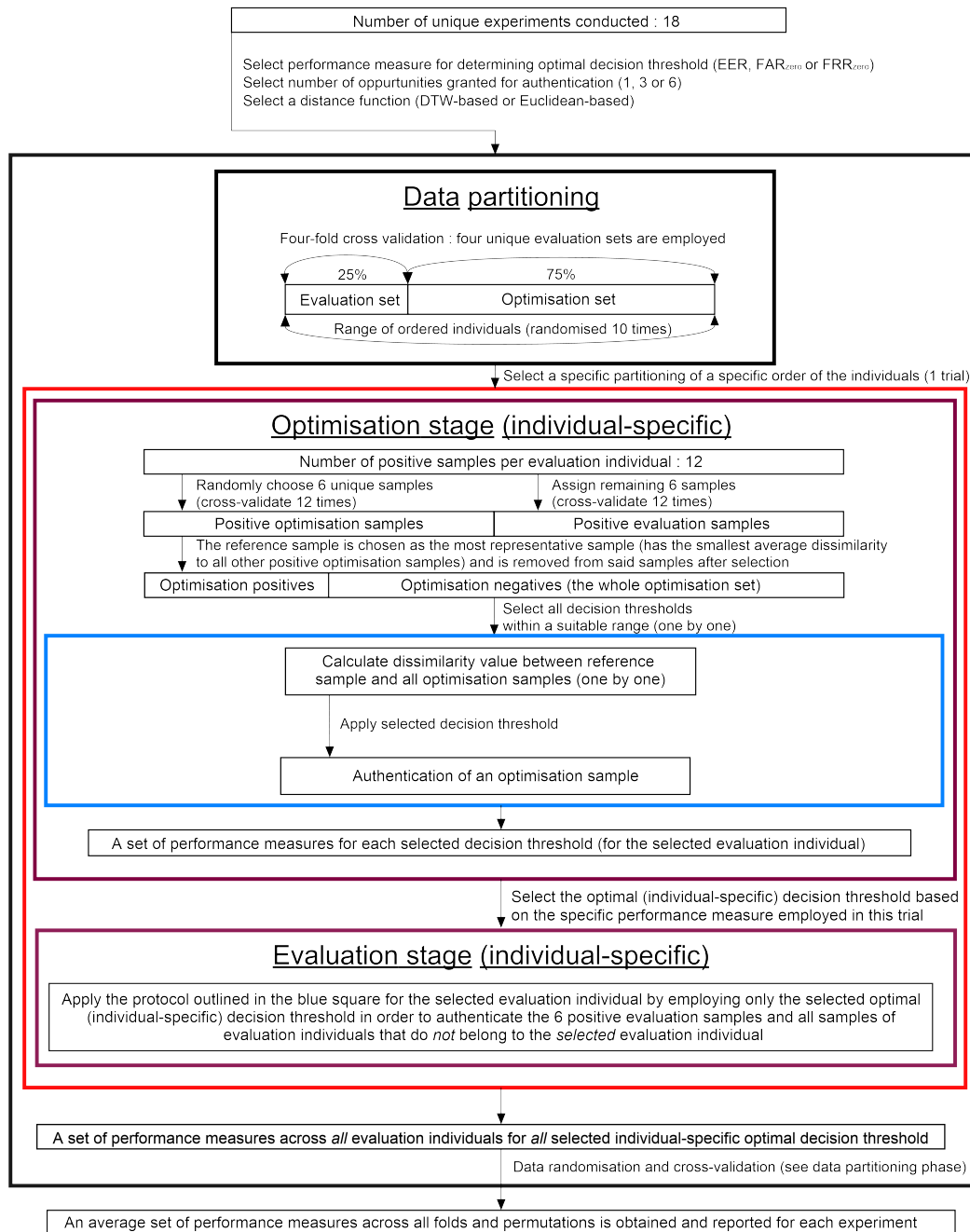


Figure A.3: Conceptualisation of the experimental protocol for the *k-six* scenario (PES2) within the context of a score-based classifier.

Rank-based classifier for the k -six scenario

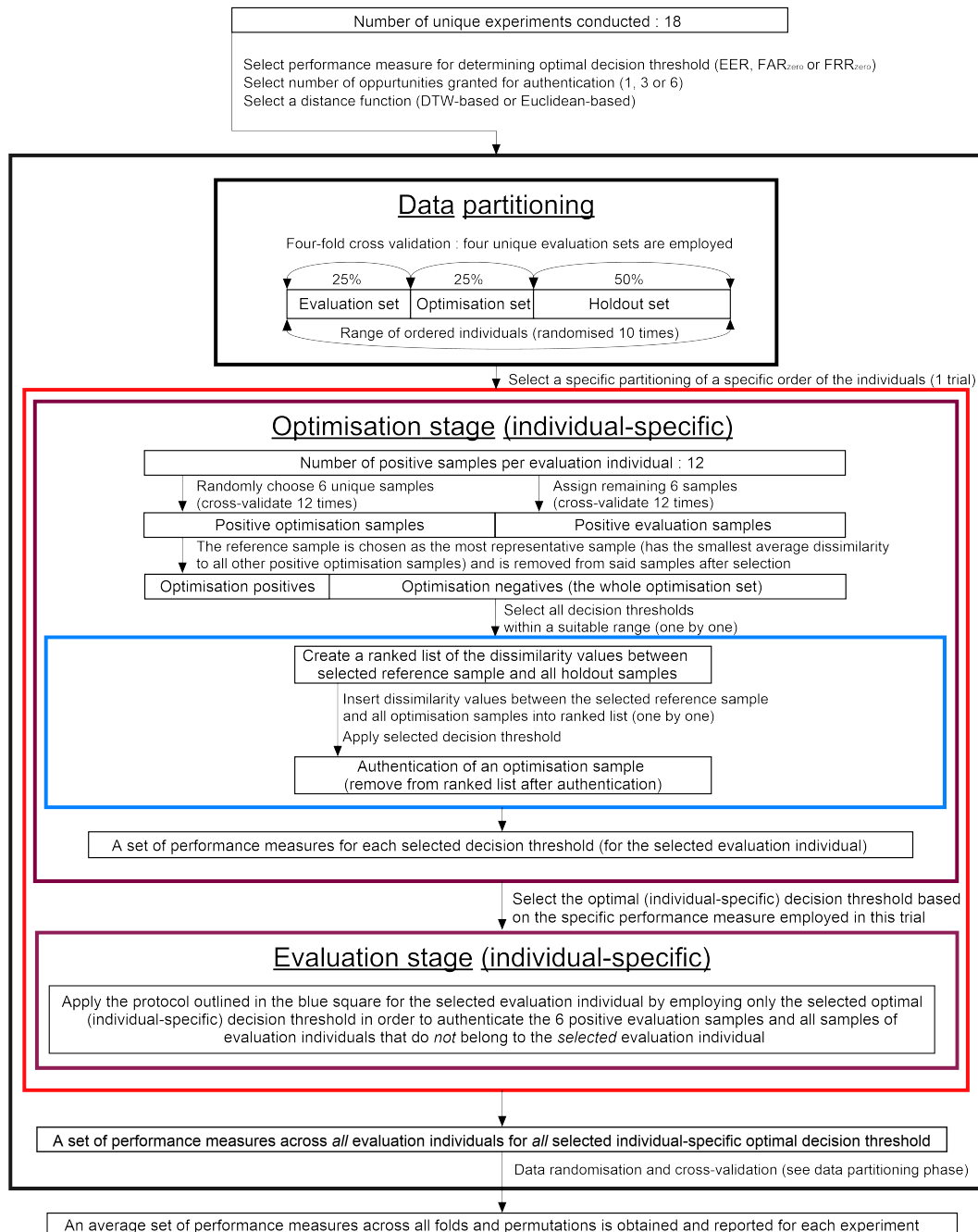


Figure A.4: Conceptualisation of the experimental protocol for the k -six scenario (PES2) within the context of a rank-based classifier.

Appendix B

Authentication examples

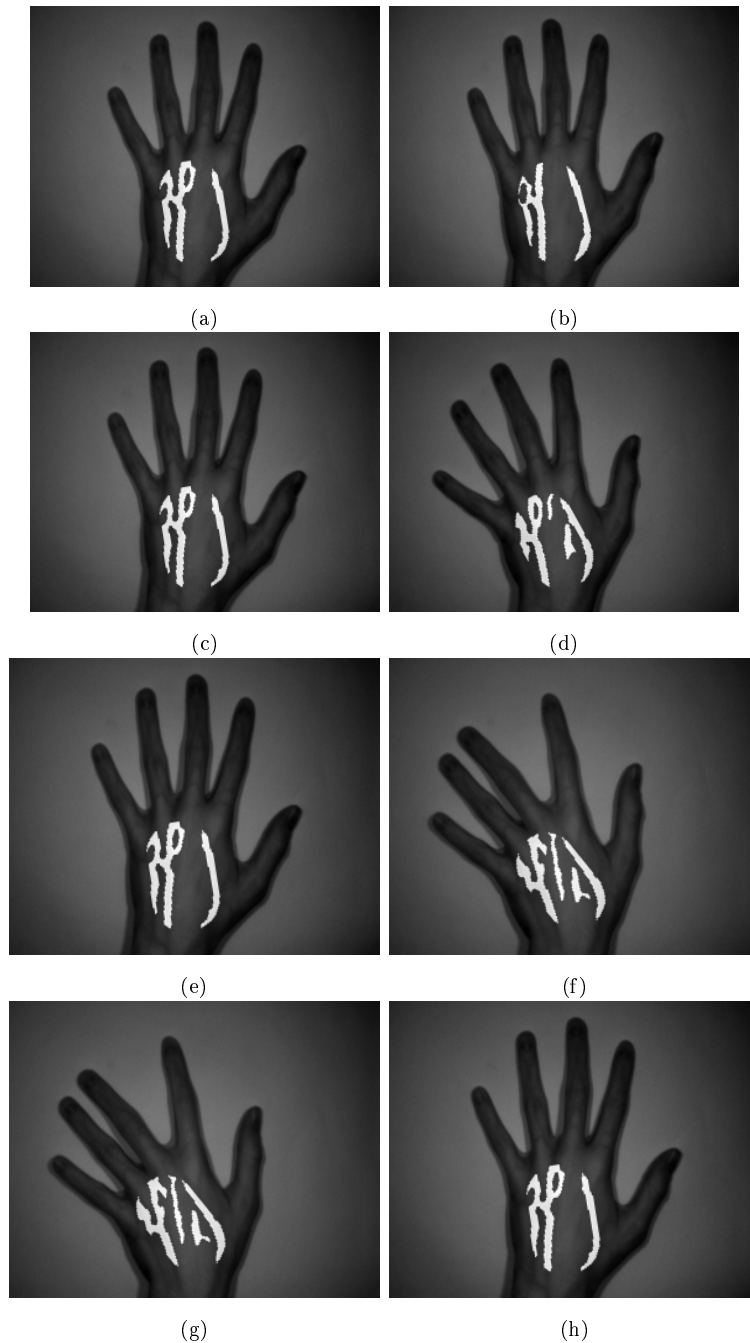


Figure B.1: (a) A relatively high quality template (reference sample) for a specific individual. (b) A *positive* evaluation sample that is *accepted* due to its similarity to the template in (a). (c) The image depicted in (a). (d) A *positive* evaluation sample that is *accepted* despite small differences in hand vein structure. (e) The image depicted in (a). (f) A *positive* evaluation sample *rejected* due to the significant structural differences of the hand veins. (g) A relatively low quality reference sample for the *same* individual as the one associated with the image depicted in (a). (h) The *positive* sample depicted in (b) is now *rejected* due to the low quality of the template depicted in (g).

List of References

- Ai, D., Yang, J., Fan, J., Zhao, Y., Song, X., Shen, J., Shao, L. and Wang, Y. (2016). Augmented reality based real-time subcutaneous vein imaging system. *Biomedical Optics Express*, vol. 7, no. 7.
- American Banking Association (2015). Deposit account fraud survey.
Available at: <https://www.aba.com/Products/Surveys/Pages/2015DepositAccount.aspx>
- Bogazici University (2010). Bosphorus hand vein database.
Available at: <http://bosphorus.ee.boun.edu.tr/hand/Home.aspx>
- Coetzer, J. (2005). *Off-line signature verification*. PhD, Stellenbosch Univ.
- Fick, C. (2017). *Modifying and generalising the Radon transform for improved curve-sensitive feature extraction*. MSc, Stellenbosch Univ.
- Gonzales, R. C., W.R.E. (2010). *Digital Image Processing*.
- Krishnaveni, N.V., Sivasankari, K. and Vijayan, V. (2014). Personal authentication using hand vein. *International Journal of Engineering Research and Technology (IJERT)*, vol. 3, no. 1, pp. 2333–2339.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66.
- Prasanna, R.D., Neelamegam, P., Sriram, S. and Nagarajan, R. (2012). Enhancement of vein patterns in hand image for biometric and biomedical application using various image enhancement techniques. *Procedia Engineering*, vol. 38, pp. 1174–1185.
- Rice, J. (1987 October 13). Apparatus for the identification of individuals. US Patent 4,699,149.
Available at: <https://www.google.com/patents/US4699149>
- Rossan, I., Heenay, M. and Khan, M. (2014). Impact of changing parameters when preprocessing dorsal hand vein pattern. *Procedia Computer Science*, vol. 32, pp. 513–520.
- Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165.

- Swanepoel, J. (2015). *Writer-independent handwritten signature verification*. PhD, Stellenbosch Univ.
- The Institute Of Control And Information Engineering (2013). Vein dataset. Available at: <http://biometrics.put.poznan.pl/vein-dataset/>
- Trabelsi, R.B., Masmoudi, A.D. and Masmoudi, D.S. (2013). A new multimodal biometric system based on finger vein and hand vein recognition. *International Journal of Engineering and Technology (IJET)*, vol. 5, no. 4.
- Wang, L., Leedham, G. and Cho, S. (2007). Infrared imaging of hand vein patterns for biometric purposes. *IET Computer Vision*, vol. 1, no. 3-4.
- Yuksel, A., Akarun, L. and Sankur, B. (2011). Hand vein biometry based on geometry and appearance methods. *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 398-406.