

Handwritten Signature Verification using Locally Optimized Distance-based Classification

by

Yaseen Moolla

Submitted in fulfilment of the academic
requirements for the degree of
Master of Science in Computer Science,
University of KwaZulu-Natal

Durban, South Africa, 2012

©Yaseen Moolla, 2012

As the candidate's supervisor I have/have not approved this dissertation for submission

Signed:

Name:

Date:

Abstract

Although handwritten signature verification has been extensively researched, it has not achieved optimum accuracy rate. Therefore, efficient and accurate signature verification techniques are required since signatures are still widely used as a means of personal verification. This research work presents efficient distance-based classification techniques as an alternative to supervised learning classification techniques (SLTs). Two different feature extraction techniques were used, namely the Enhanced Modified Direction Feature (EMDF) and the Local Directional Pattern feature (LDP). These were used to analyze the effect of using several different distance-based classification techniques. Among the classification techniques used, are the cosine similarity measure, Mahalanobis, Canberra, Manhattan, Euclidean, weighted Euclidean and fractional distances. Additionally, the novel weighted fractional distances, as well as locally optimized resampling of feature vector sizes were tested. The best accuracy was achieved through applying a combination of the weighted fractional distances and locally optimized resampling classification techniques to the Local Directional Pattern feature extraction. This combination of multiple distance-based classification techniques achieved accuracy rate of 89.2% when using the EMDF feature extraction technique, and 90.8% when using the LDP feature extraction technique. These results are comparable to those in literature, where the same feature extraction techniques were classified with SLTs. The best of the distance-based classification techniques were found to produce greater accuracy than the SLTs.

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

DECLARATION 1 - PLAGIARISM

I, YASEEN MOOLLA....., declare that

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a. Their words have been re-written but the general information attributed to them has been referenced
 - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed

.....

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE
DECLARATION 2 - PUBLICATIONS

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this dissertation

- Publication 1: **Handwritten Signature Verification using Weighted Fractional Distance Classification**
 - Status: Published in IEEE Xplore
 - Experimental Work: by Y. Moolla, with guidance from other authors
 - Writing: by Y. Moolla, with guidance and corrections from other authors
 - Appendix: A
- Publication 2: **Local Directional Pattern Based Signature Verification using Weighted Fractional Distance Classification**
 - Status: Published in ACM Digital Library
 - Experimental Work: by Y. Moolla, with guidance from other authors
 - Writing: by Y. Moolla, with guidance and corrections from other authors. In particular, S. Viriri contributed to “Abstract” and “Conclusion”; and J.R. Tapamo contributed to “Techniques and Methodology”.
 - Appendix: B
- Publication 3: **Offline Signature Verification using Locally Optimized Distance-based Classification**
 - Status: Submitted to South African Computer Journal
 - Experimental Work: by Y. Moolla, with guidance from other authors
 - Writing: by Y. Moolla, with guidance and corrections from other authors.
 - Appendix: C

Signed:


.....

Declaration Publications FHDR 22/05/08 Approved

Acknowledgements

I would like to thank

- CSIR, for their financial and academic assistance.
- My supervisors: Dr. S. Viriri, Prof. F.V. Nelwamondo and Prof. J.R. Tapamo. This dissertation would not have been possible without their guidance.
- The GPDS group, for the use of their signature database.

Dedication

This is dedicated to my family, for their never-ending encouragement and support.

Contents

1	General Introduction	1
1.1	Introduction	1
1.2	Motivation	1
1.3	Research Problem	2
1.4	Research Objectives	2
1.5	Contributions of the Dissertation	3
1.6	Dissertation Outline	3
2	Background and Literature Review	5
2.1	Introduction	5
2.2	Biometrics	5
2.2.1	History of Biometrics	5
2.2.2	Modalities	6
2.2.3	Verification and Recognition	7
2.3	Signature Biometric Modality	7
2.3.1	Online and Offline Formats	7
2.3.2	Advantages and Disadvantages	8
2.3.2.1	Disadvantages	8
2.3.2.2	Advantages	9
2.3.3	Signature Databases	9
2.3.3.1	GPDS	10
2.3.3.2	MCYT	10
2.3.3.3	SVC2004	10
2.3.3.4	Other signature databases	11
2.4	Feature Extraction	11
2.5	Biometric Performance Measures	12
2.5.1	False Rejection Rate	12

2.5.2	False Acceptance Rate	12
2.5.3	Equal Error Rate	13
2.5.4	ROC-Curve	13
2.5.5	Average Error Rate	13
2.6	Literature Survey of Offline Signature Classification Techniques	13
2.6.1	Supervised Learning Techniques (SLTs)	13
2.6.1.1	Hidden Markov Models (HMM)	14
2.6.1.2	Support Vector Machines (SVM)	14
2.6.1.3	Artificial Neural Networks (NN)	15
2.6.2	Distance-based Classification Techniques	16
2.6.2.1	Euclidean Distance	16
2.6.2.2	Mahalanobis Distance	17
2.6.2.3	Canberra Distance	18
2.6.2.4	Cosine Similarity Measure	18
2.6.2.5	Manhattan Distance	18
2.6.2.6	Weighted Euclidean Distance	18
2.6.2.7	Fractional Distances and the Concentration Phenomenon	18
2.7	Conclusion	19
3	Design and Methodology	20
3.1	Introduction	20
3.2	Design Overview	20
3.3	Preprocessing	21
3.3.1	Binarization	21
3.3.2	Thinning	22
3.3.3	Boundary Extraction	22
3.3.4	Dilation	23
3.4	Feature Extraction	25
3.4.1	Choosing the Feature extractions	25
3.4.2	Direction Feature	25
3.4.3	Modified Direction Feature	28
3.4.4	Ratio Feature	29
3.4.5	Maxima Feature	31
3.4.6	Energy Feature	32
3.4.7	Extended Modified Direction Feature	32

3.4.8	Local Direction Pattern (LDP)	34
3.4.9	Feature Vectors and Individually Optimized Resampling	35
3.5	Classification	38
3.5.1	Determining the Threshold	38
3.5.2	Euclidean Distance	38
3.5.3	Manhattan Distance	39
3.5.4	Fractional Distance	39
3.5.5	Ratio Distance	39
3.5.6	Mahalanobis Distance	40
3.5.7	Cosine Similarity Measure	41
3.5.8	Canberra Distance	42
3.5.9	Weighted Euclidean Distance	42
3.5.10	Weighted Manhattan Distance	43
3.5.11	Weighted Fractional Distance	43
3.6	Experimental Methodology	43
3.7	Conclusion	43
4	Results and Discussions	45
4.1	Introduction	45
4.2	Data Set	45
4.3	Threshold Calculation	45
4.4	Tests with Extended Modified Direction Feature	46
4.4.1	Euclidean Distance	47
4.4.2	Manhattan Distance	47
4.4.3	Mahalanobis Distance	48
4.4.4	Cosine Similarity Measure	49
4.4.5	Canberra Distance	49
4.4.6	Ratio Distance	50
4.4.7	Fractional Distance	50
4.4.8	Weighted Euclidean Distance	51
4.4.9	Weighted Manhattan Distance	53
4.4.10	Weighted Fractional Distance	53
4.4.11	Individually Optimized Resampling	54
4.5	Tests with Local Directional Pattern	55
4.5.1	Euclidean Distance	55

4.5.2	Manhattan Distance	56
4.5.3	Fractional Distance	57
4.5.4	Weighted Euclidean Distance	59
4.5.5	Weighted Manhattan Distance	60
4.5.6	Weighted Fractional Distance	61
4.5.7	Individually Optimized Resampling	62
4.5.7.1	Euclidean distance	63
4.5.7.2	Manhattan distance	63
4.5.7.3	Fractional distance	64
4.5.7.4	Weighted Euclidean distance	66
4.5.7.5	Weighted Manhattan distance	66
4.5.7.6	Weighted fractional distance	67
4.6	Literature Comparison	69
4.6.1	Literature Comparison of EMDF	69
4.6.2	Literature Comparison of LDP	70
4.7	Optimal Design Overview	71
4.8	Implementation Scenarios	72
4.9	Conclusion	72
5	Conclusion and Future Work	73
5.1	Summary of Work	73
5.2	Recommendations for Future Work	73
A	First Conference Paper	80
B	Second Conference Paper	87
C	First Journal Paper	93

List of Tables

3.1	MDF feature vector sizes	38
3.2	LDP feature vector sizes	38
4.1	The Euclidean and Manhattan distances investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF .	48
4.2	The Mahalanobis distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF	49
4.3	The cosine similarity measure, Canberra distance and Ratio distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF	50
4.4	The fractional distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF	51
4.5	The weighted Euclidean, weighted Manhattan and weighted fractional distances investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF	52
4.6	Individually optimized resampling investigated with the EMDF	55
4.7	The effect of different $split_H$ and $split_V$ on EER(%) using the Euclidean distance	56
4.8	The effect of different $split_H$ and $split_V$ on FARR(%) using Euclidean distances	56
4.9	The effect of different $split_H$ and $split_V$ on EER(%) using the Manhattan distance	57
4.10	The effect of different $split_H$ and $split_V$ on FARR(%) using Manhattan distances	57
4.11	The effect of different $split_H$ and $split_V$ on EER(%) using the fractional distance	59
4.12	The effect of different $split_H$ and $split_V$ on FARR(%) using fractional distance	59
4.13	The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Euclidean distance	60
4.14	The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Euclidean distances	60

4.15	The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Manhattan distance	61
4.16	The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Manhattan distances	61
4.17	The effect of different $split_H$ and $split_V$ on EER(%) using the weighted fractional distance	62
4.18	The effect of different $split_H$ and $split_V$ on FARR(%) using weighted fractional distance	62
4.19	The effect of implementing individually optimized resampling by choosing the best Euclidean distance EER from 4 feature vectors	63
4.20	The effect of implementing individually optimized resampling by choosing the best Euclidean distance EER from 9 feature vectors	63
4.21	The effect of implementing individually optimized resampling by choosing the best Manhattan distance EER from 4 feature vectors	64
4.22	The effect of implementing individually optimized resampling by choosing the best Manhattan distance EER from 9 feature vectors	64
4.23	The effect of implementing individually optimized resampling by choosing the best fractional distance EER from 4 feature vectors	65
4.24	The effect of implementing individually optimized resampling by choosing the best fractional distance EER from 9 feature vectors	65
4.25	The effect of implementing individually optimized resampling by choosing the best weighted Euclidean distance EER from 4 feature vectors	66
4.26	The effect of implementing individually optimized resampling by choosing the best weighted Euclidean distance EER from 9 feature vectors	66
4.27	The effect of implementing individually optimized resampling by choosing the best weighted Manhattan distance EER from 4 feature vectors	67
4.28	The effect of implementing individually optimized resampling by choosing the best weighted Manhattan distance EER from 9 feature vectors	67
4.29	The effect of implementing individually optimized resampling by choosing the best weighted fractional distance EER from 4 feature vectors	68
4.30	The effect of implementing individually optimized resampling by choosing the best weighted fractional distance EER from 9 feature vectors	68
4.31	A comparison of results from literature with the results obtained in this work	71

List of Figures

3.1	Overview of the design of a biometric verification system	21
3.2	A sample of a signature before thinning and direction feature extraction	22
3.3	A signature after the Zhang-Suen thinning algorithm has been applied .	22
3.4	A signature image before boundary extraction of the signature occurs. .	23
3.5	A signature image cropped to the bounding box size.	23
3.6	A standard black and white signature of a user before dilation preprocessing is applied	24
3.7	A standard black and white signature of a user after dilation preprocessing using a 3×3 OR mask	25
3.8	In depth demonstration of direction feature extraction	26
3.9	A signature after the direction feature has been applied with different colours for each type of line segment and for intersection pixels	27
3.10	The direction value that will be assigned to a current pixel based on it's directional relation to the previous pixel, p	27
3.11	Processing of the Direction Transition (DT) and Location Transition (LT) feature values in the Left-to-Right direction [12]	29
3.12	Resampling of DT and LT values in the left-to-right direction [12] . . .	30
3.13	A signature image cropped to the bounding box size	31
3.14	A depiction of the Maxima Feature extraction process after Direction Feature preprocessing. The distances between the longest vertical and horizontal segments is shown.	32
3.15	The 8 orientations of Kirsch Masks [21]. Each orientation is applied to a pixel and its 8 neighbours to calculate 8 mask values.	34
3.16	calculation of the LDP code [21] obtained by applying each of the 8 Kirsch masks	35
3.17	A sample LDP histogram showing the occurrences of each directional permutation from the image in Figure 3.7 with no splits	36
3.18	A dilated image with splits segmented by 3 horizontal splits and 4 vertical splits	36
4.1	A ROC-curve showing the point of intersection between the FRR and FARS curves (obtained from classification of the EMDF using the weighted fractional distance and individually optimized resampling)	46

4.2	EER in Relation to individual fractional distances for large and small feature vector sizes	58
4.3	The number of occurrences of each p-norm in the fractional distance with individually optimized resampling combinations [1:3] and [6:8]	65
4.4	The number of occurrences of each p-norm in the weighted fractional distance with individually optimized resampling combinations [1:3] and [6:8]	69
4.5	Overview of the design of a biometric verification system with specified optimal configurations	71

List of Abbreviations

DF	Direction feature
EMDF	Extended modified direction feature
HMM	Hidden Markov model
LDP	Local directional pattern
MDF	Modified direction feature
NN	Artificial neural network
RBF	Radial Basis Function
SLT	Supervised learning technique
SVM	Support vector machine

Chapter 1

General Introduction

1.1 Introduction

1.2 Motivation

Biometrics is the study of measuring and quantifying human characteristics for authentication or identification of individuals. Biometric modalities are regularly becoming an important aspect of automated electronic security systems. Successful biometric systems require methods and techniques that can authenticate or identify individuals with a high level of accuracy.

Biometric systems are used either for recognition or verification. Recognition entails the identification of a biometric trait, or set of traits, as belonging to a specific individual from a given set of individuals. Conversely, verification entails authenticating a claim that a biometric trait, or set of traits, belongs to a specific individual.

One of the most common and widely accepted biometric modality is the handwritten signature. It has been used for the manual verification of individuals for centuries. Due to its wide-spread use and acceptance, handwritten signatures are an ideal candidate for automated biometric verification systems. The two processes for capturing signatures are categorized as offline and online. For offline capture, a static image of a completed signature is recorded. For online capture, the creation of a signature is recorded as a function of time.

There are several cases in which online signatures can not be used. This includes automated authentication of bank cheques and legal documents. Additionally, equipment for the capture of offline signatures is cheaper, which will allow for greater adoption of automated signature verification, especially for small-to-medium businesses and in developing economies.

Due to security and privacy concerns, an automated offline signature verification system requires techniques that can verify authentic signatures and reject forged signatures with a high level of accuracy. However, offline signature verification systems have a low accuracy in comparison to other biometric modalities, such as fingerprints and irises.

Accurate offline signature verification is an open problem in the area of biometrics. This dissertation aims to improve verification accuracy through the use of distance-based

classification techniques. In this chapter, the importance of improving offline signature verification accuracy is highlighted; insufficient distance-based classification in offline signature verification is identified as the research problem; the aims and objectives are specified to approach solving the research problem; and how these objectives are met is summarized.

1.3 Research Problem

Offline signature verification systems have much potential for world wide usage, but are hindered by low accuracy rates. These low accuracy rates are due to the inherent randomness that is characteristic of behavioural biometrics. Static signatures also have much less information in comparison to online signatures, since the time dimension is not present.

There are many current research attempts to improve the accuracy of offline signature verification systems through the use of various techniques. Some of the research focuses on feature extraction techniques. These are designed to extract the most relevant and constant features of a signature while also attempting to reduce the extraction of features that are not stable or constant in an individuals signatures. Other research concentrates on using different classification techniques and finding the optimal classification technique for a particular feature set. Classification can be categorized into supervised learning techniques (SLTs) and distance-based measures. Most work with classification concentrates on using different SLTs, such as support vector machines, hidden Markov models and artificial neural networks. There is very little research into using different distance-based classification techniques, except for the Euclidean distance, which is the most commonly known distance measure in geometric space.

There are many other distance-based measures that could be used for classification. In comparison to SLTs, Distance-based measures have the potential for equal or more accurate classification; they are less complex and easier to implement; and their training and classification times may be faster. This work investigates the use of some of these techniques and applies them to various different feature extraction techniques to measure their classification accuracy.

1.4 Research Objectives

This work aims to improve the accuracy of offline signature verification systems by investigating various classification techniques. To achieve this, the following questions are investigated.

- Can distance-based classification techniques perform better than SLTs at verifying offline signatures?
- Which distance-based classification technique works best in signature verification systems?
- Does the classification accuracy translate across the use of multiple different feature extraction techniques?

1.5 Contributions of the Dissertation

The following contributions are made from this dissertation.

- Various different distance-based classification techniques are investigated.
 - The cosine similarity measure, Mahalanobis, Ratio, Canberra and various L^P -space distances are investigated as alternatives to SLTs in signature verification.
 - Fractional distances, in L^P -space, are investigated to improve verification accuracy over the Euclidean and Manhattan distances by combating the concentration phenomenon and non-Gaussian randomness within the feature vectors.
 - The weighted Euclidean distance is investigated to improve verification over the Euclidean distance by normalization of the feature vectors and adding statistical importance to the most stable features.
 - The weighted fractional distance, a novel classification technique obtained from the fusion of the weighted Euclidean distance and fractional distances, is presented. Its use is investigated to further improve the accuracy of offline signature verification systems.
- Multiple feature extraction techniques are used to verify the results of applying distance-based classification techniques to offline signature verification.
 - The effect of local features (Modified Direction Feature and Local Directional Pattern Feature) and global features (Energy, Ratio and Maxima features), and the effect of combining both types of features on distance-based classification techniques, are investigated.
 - Individually optimized resampling is investigated to improve the classification of local features.
- The best results from the distance-based classification are compared with results from literature to determine which of the distance-based classification techniques perform better than SLTs.

1.6 Dissertation Outline

The rest of this dissertation is ordered as follows.

- Section 2: reviews the literature and provides background of the topic.
- Section 3: describes the methodology and techniques employed in this research.
- Section 4: discusses the results obtained from applying the methodology and various techniques.
- Section 5: concludes the dissertation and discusses potential avenues for future work.

- Appendix: contains the conference papers and journal articles associated with this work.

Chapter 2

Background and Literature Review

2.1 Introduction

Offline signature verification is an ongoing research topic in the field of biometrics. In this chapter, a review of databases, feature extraction techniques, statistical performance measures and classification techniques are discussed in relation to the signature modality. Issues related to the classification of offline signatures will be highlighted.

2.2 Biometrics

Biometrics involves the measurement of one or more intrinsic physiological or behavioural human characteristics to verify or identify a person. It is often used as part of identity access management systems, for access control, recognition, identification and authentication of individuals [29].

2.2.1 History of Biometrics

Unique characteristics have been used to identify individuals for thousands of years, with records dating back to the Ancient Egyptian and Chinese civilizations [25]. In the 1800's biometrics were formalized and began being used in crime solving investigations. Alphonse Bertillion, Paris Prefecture of Police's chief of criminal identification division, developed and practiced the idea of using several head and body measurements to identify criminals [29]. At the same time, Scottish doctor Henry Faulds learned of the ancient Japanese practice of identifying the creator of pottery by embedded fingerprints. This lead him to publish an article on the forensics of fingerprints which were later adopted by Scotland Yard [15]. Later, Edward Henry, Azizul Haque and Hemchandra Bose, of the Bengal Police, together developed the Henry classification system to improve the identification time of fingerprints, which is the basis of many modern fingerprint identification systems. These are now used by criminal justice organizations the world over [63].

While fingerprints remain the biometric modality of choice for individual identification, many others, face recognition for example, have been researched over the past

few decades. Traditionally, a photograph of an individual was matched to the individual in real time, using manual human intervention. Systems are now being developed to automatically analyse an individual's face in three dimensional space in an attempt to improve the robustness and correctness of the technique [32].

Other examples of biometrics include handwritten signatures, retina blood vessel patterns, hand geometry and iris biometrics, with the iris being regarded as the best practical biometric modality for recognition and verification at the moment. There is also research into using DNA as a biometric modality but this is currently hampered by the low accuracy, slow speeds and high cost of hardware used to analyze DNA [65].

The most important factors to consider when choosing a biometric system are universality, distinctiveness, permanence, collectibility, overall performance of the analysis, public acceptability and possibility of circumvention. For practicality, a biometric system must be accurate, fast, non-resource intensive, harmless to users, be accepted by the target population and robust in terms of security [29].

2.2.2 Modalities

Biometric traits, or modalities, can be split into two broad categories, namely, physiological and behavioural. Physiological biometrics are intrinsic, constant traits of the human body. They are a measure of physiological and anatomical characteristics. These are constant, without much change throughout a person's life. Behavioural biometrics are manifestations of human activity, based on actions that are learned and regularly repeated. These actions may deviate or change over a period of time. [13]

The commonly known physiological biometric traits include fingerprints, palm prints, dental patterns, facial features, irises and retinas. Other new types of physiological biometrics include infrared thermograms of the human face; vein patterns of palms and fingers; the wrinkle patterns in knuckles; patterns in the nail bed of finger nails; skin spectroscopy; ear prints; vibrations of the inner ear; dental radiographs; and DNA.

Behavioural biometrics can be put into 5 different subcategories [70]:

1. Motor skills: This is the most common category and the most useful for automated recognition and verification of individuals. Modalities in this category include handwriting and signatures; voice; walking gait; movement of facial muscles and lips when smiling or talking; eye movement in response to stimulation.
2. Authorship: unique traits of individuals in the way they structure sentences, their choice of words, skills in artistry and programming style. These may be used to prove identity or authors during criminal investigations but are not very useful for automated authentication systems.
3. Direct Human-Computer Interactions (HCI): this is the quantification of skills exhibited by individuals during interaction with electronic devices. It can be further split into input device-based HCI, such as typing speed of key strokes and the movement of electronic mice; and software interaction-based HCI such as gaming strategies.
4. Indirect HCI: These are digital footprints left by users during regular interactions with electronic devices, such as network traffic; choices between GUI interaction and keyboard shortcuts; storage activity patterns and audit logs.

5. Purely behavioural: These are pure behavioural patterns created from the use of unique individual knowledge and skills during the execution of mentally strenuous exercises, such as car driving skills, behaviour under stress and credit card usage. The measurement of ECG of heartbeats, EEG of brain patterns and latencies in cognitive function also fall into this category.

Each biometric trait has its own set of advantages and possible uses. There is no known biometric that is applicable for all possible security uses. Therefore, there is a constant search for new biometric traits that can uniquely identify individuals [24].

Behavioural biometrics generally rely on muscular movement and therefore contain an inherent randomness which is not present in physiological biometrics. This randomness makes biometric matching a greater challenge with behavioural biometrics. To reduce the negative effects of this randomness, reference feature vectors are created through the averaging of multiple records of a biometric trait, for an individual.

2.2.3 Verification and Recognition

Biometric authentication is categorized into verification and recognition. Recognition, also called identification, is used to discern an individual of unknown identity. Verification is used to affirm an individual's claim to a specific identity.

2.3 Signature Biometric Modality

Signatures, which are one of the oldest used and most widely accepted biometric for identification and verification [73], are handwritten depictions of a person's name, nickname or other personal symbol. They are classified as a behavioural biometric trait, and are most often used for the verification of a signer's identity.

2.3.1 Online and Offline Formats

The capturing of signature information can be either online or offline. Offline recording is the capture of a completed static signature image. This can be via scanning of a signature on a page, photography or an offline style writing pad. Online signatures are a dynamic recording of the signature creation as a function of time, i.e. how much of a signature is created per unit of time. These can be recorded via video camera or recording pen, but are most often recorded with high quality writing pads. Depending on the technology and cost of equipment, online signatures can capture additional information such as the angle or pressure of the pen used to write the signature. Due to the additional information present in online signatures, and the difficulty of forging a signature at the exact same pace as the original signature writer, online signatures generally provide higher authentication accuracy in comparison to offline signatures. Online signatures also nullify the disadvantage of publicly visible signatures. However, there are some cases in which online signatures can not be used, such as with bank cheques. Additionally, small to medium businesses may not be able to afford the higher cost of online recording equipment. Thus, a system that uses financially more cost effective equipment

will allow smaller businesses access to better security solutions, both in developed and developing countries. In this work, an offline biometric signature verification system is investigated.

2.3.2 Advantages and Disadvantages

There are seven key aspects that need to be analyzed when weighing the advantages and disadvantages of a biometric trait. These aspects are, in no particular order: uniqueness, permanence, acceptability, collectability, universality, resistance to circumvention and performance [13]. The strengths and weaknesses of signatures are discussed below, in terms of each of these aspects.

2.3.2.1 Disadvantages

One of the greatest disadvantages of signatures is that they are very public, since they are visible throughout a person's legal documents. This makes it easy for potential forgers to obtain samples of a user's signature to practice copying. This is a greater disadvantage for offline signatures than for online signatures. The latter contains dynamic information which lowers the probability of successful security circumvention via forgeries. Users of a signature verification system must be educated on security measures and encouraged to create more complex signatures to protect against skilled forgeries.

While signatures are easily collectible, the use of low quality capturing equipment can distort the data and negatively affect the accuracy of the system. An investment in good quality devices is important to reduce the risk of distorted or noisy data. While most people are willing to use signatures as a form of personal identification, the illiterate and physically disabled provide a hurdle to the universal application of signature verification. The former group, illiterate individuals, is a shrinking minority. However, systems are required to cater for those who do not have full use of their hands and are thus incapable of writing signatures. A possible method to overcome these problems is to create multi-modal biometric systems which combine signatures with other biometrics.

Another disadvantage is that signatures are not infinitely permanent. They tend to change over a long period of time. To overcome this, systems often require that reference signatures, also called specimen signatures, be updated at regular intervals, which may require additional interaction with users and potential inconvenience. Signatures also have smaller variations every time they are written. This randomness is inherent to most behavioural biometrics. It is overcome to a large extent by creating reference feature vectors by averaging multiple authentic feature vectors. However, this requires additional user interaction when new individuals are added to the system so that a sufficient number of signatures can be stored and used for training and testing. Furthermore, the quality and accuracy of signatures, like many behavioural biometrics, can be negatively affected by a user's current emotional state. This means that if a user is distressed or in a hurry, their signature may be significantly different than when written in a calm state of mind. Signatures are also often written at different angles onto the page or other capturing medium. This requires many feature extraction techniques to adjust for the angular difference during preprocessing.

2.3.2.2 Advantages

The lack of permanence also has advantages. If a system's security is compromised or a user's signature is successfully forged, a new signature can be easily created, thus nullifying the security risks. Conversely, if the storage of physiological biometrics is compromised, it is a much more serious matter, since physiological biometric traits of an individual cannot be changed at will. The only potential recourse for the compromise of a physiological biometric system is to change the feature extraction techniques. A user can also have more than one signature in use concurrently, such that a different signature is used for authenticated access of different systems or even different areas of the same system.

Signatures also have a high degree of public acceptance and are easily collectible. Due to centuries of regular use, the general public is relatively comfortable with the recording of signatures and have a high degree of trust in the recognition, storage and use of signatures. Signatures are also less personal, not private and are non-invasive. In comparison, users may be uncomfortable with the digital recording of more personal physiological information such as their fingerprints and facial features. Trust and acceptance of signature systems can be further increased by systems that do not store the actual signature, but rather only store extracted information from the signatures. High quality capturing devices are inexpensive, especially for offline signatures. Low entry-level costs and public acceptance can promote fast adoption of signature verification systems in industry.

Signatures are also used throughout the world, by almost all people, with the exception of physically disabled and illiterate minorities. This makes a universal transition from paper-based to electronics-based signature systems a credible possibility. It also means that users do not require special training when signature-based biometric security systems are implemented.

Each person creates their own unique signature, which is most usually based on their name. Even if two users share the same name, their writing styles will be significantly different. The level of uniqueness is even higher for online systems, where dynamic information is stored. This is especially advantageous for signature recognition systems.

Due to the use of dynamic information, online signatures have very high performance accuracy. However, offline systems have lower, less accurate performance. Due to the lower cost of high quality offline equipment in comparison to online equipment, and due to instances where online signature cannot be used, developing good offline verification is an important research topic.

2.3.3 Signature Databases

Using pre-compiled, publicly available signature databases is preferred over compiling a private signature database. This is due to the high cost in terms of time and monetary expenses that are incurred in creating these databases, which require both authentic and skilled forgeries. Further, the use of widely accepted signature databases allows better comparison between the works of different researchers. The most commonly used signature dataset for offline signature testing are the GPDS, MCYT and SVC2004 datasets. The first contains purely offline information, while the other three contain both online and offline information.

2.3.3.1 GPDS

The Grupo de Procesador Digital de Senales (GPDS) signature database [66] is a collection of offline signatures that were captured by scanning signatures written onto paper. The GPDS300 corpus contains black and white signatures for 300 individuals, with 24 authentic signatures and 30 skilled forgeries per individual. This gives a total of 16200 signatures in the database, of which 7200 are authentic and 9000 are skilled forgeries. All signatures are written in Latin script. For the creation of skilled forgeries, the potential forgers were provided with authentic copies of user's signatures. They practiced copying these signatures until confident of their forgery skill. Only then were skilled forgeries added to the database. For each individual, 4 forgers were used. All authentic signatures for an individual were captured during a single sitting. For the testing of random forgeries, random authentic signatures of different individuals were chosen. A larger GPDS corpus of 960 individual's signatures (GPDS960) is commercially available, while smaller corpuses of 75 (GPDS75) and 160 (GPDS160) individuals are also publicly available.

2.3.3.2 MCYT

The MCYT multimodal biometric database [53] contains fingerprint, online signature information for 330 individuals. For the signature subcorpus, 25 authentic and 25 forgeries are captured for each individual, using a WACOM pen tablet. This gives a total of 16500 signatures in the database, of which 8250 are authentic and 8250 are skilled forgeries. Authentic signatures were captured in sets of 5. For skilled forgeries, 5 forgers were used, writing 5 forged signatures each. The forgers practiced the forgeries after observing offline (static) copies of the signatures. Although it is designed as an online signature database, the tablet also captures offline information that is used for testing of some offline signature verification systems.

2.3.3.3 SVC2004

The SVC2004 signature database [71] was created for the First International Signature Verification Competition, held in 2004. The complete database hold authentic and forged online signatures for 100 individuals, but only 40 of these were initially made available for earlier tests. Signatures were captured using the WACOM Intuos writing tablet. For each individual, 20 authentic signatures were captured over 2 sittings, i.e. 10 signatures per sitting. To allay privacy concerns, users designed new signatures and were allowed to practice these new signatures on the capturing device as much as they wanted before recording and storage to the database was performed. They were also allowed to remove instances of their signatures with which they were unsatisfied. Additionally, 20 skilled forgeries per individual were created, using at least 4 forgers. This gives a total of 4000 signatures in the database, of which 2000 are authentic and 2000 are skilled forgeries. The forgers observed the online creation of signatures before practicing and entering their forgeries into the database. The database contains some signatures using Latin characters and some using Chinese characters. This database contains fewer individuals and less information per individual in comparison to the GPDS and MCYT signature

databases, especially for those researchers that had access to the signatures of 40 individuals only. This brings into question the accuracy and precision of results obtained with this database, as larger datasets allow for more consistency in results.

2.3.3.4 Other signature databases

There are many other less commonly used signature databases such as: Dolfing's dataset [17] with signatures of 51 individuals from Germany; the SIGMA database [2] with 213 individuals from Malaysia; the CEDAR database, which has authentic and forged signatures for 55 individuals [51]; the Stellenbosch database [14] with 22 individuals from South Africa; the FUM-Persian Handwritten Signature Database (FUM-PHSDB) [52] with 20 individuals from Iran; an Indian database [60]; a Turkish signature database [54] with authentic signatures from 40 individuals; and the BioSecure DS2 database with European signatures of 330 individuals as part of a commercially available multimodal database [22]. Many of these databases are small, containing sets from a limited number of individuals. System accuracy tends to decrease as datasets become larger, as there is a greater chance of users with signatures that are easily forged. The effect of database size can be seen in the results of [21]. Larger datasets provide more precise results.

The GPDS is the most easily available, has among the largest number of signatures and is regularly used in literature. This makes it best suited for comparison between different literature works.

2.4 Feature Extraction

Feature extraction is split into 2 categories: local features and global features. Local features refer to feature extraction techniques that treat patterns holistically and provide in-depth information of an image by analyzing individual parts of a pattern. Conversely, global features refer to feature extraction techniques that extract information from an overview of the image. This makes them less affected by noise and variations in the image, but means that they extract less information. Most recent work utilizes the combination of multiple feature extraction techniques into a single feature vector. This yields higher accuracies than using just a single feature extraction technique.

In Armand et al. (2006) [5], Armand et al. (2007) [6] and Nguyen et al. (2007) [49] the Modified Direction Feature (MDF), which is a local feature, is combined with several global features. These features are the Ratio, Length, Centroid, Tri-surface, Sixfold-surface and Best Fit features. These are all combined to create the Enhanced MDF. Nguyen et al. (2007) [49] reports a best EER of 17.78% with an FARR of 0.16%. While the two earlier papers report higher accuracies, they also used much fewer signatures for testing. In Nguyen et al. (2009) [48], the MDF is combined with the Ratio, Energy and Maxima features to create the Extended MDF (EMDF) feature which provides a better result than the Enhanced MDF. The best EER for the EMDF was 17.25%, with an FARR of 0.08%. The Local Binary Pattern (LBP) and Local Directional Pattern are compared in Ferrer et al. [21]. The LDP performed better with an EER of 17.8% and FARR of 0.68%. Yilmaz et al. [72] combine the LBP with a histogram of gradients (HOG) along with a fusion of SVM classifiers to obtain an EER of 15.21%.

2.5 Biometric Performance Measures

The most important performance indicator of traditional biometric systems is the recognition accuracy which is expressed in terms of the False Rejection Rate (FRR) and False Acceptance Rate (FAR). These are also sometimes called the false negatives and false positives, respectively. In verification systems, it is important to achieve the lowest possible FRR and FAR to achieve the best recognition accuracy for a biometric system. These are expanded upon below.

2.5.1 False Rejection Rate

The False Rejection Rate (FRR) is a measure of the probability that the system will reject an authorized user. It can be expressed mathematically for a set of authentication attempts against a single individual as

$$FRR(n) = \frac{\text{unsuccessful authentication attempts of authorised individual } n}{\text{total authentication attempts of authorised individual } n} \times 100 \quad (2.1)$$

For improved precision and accuracy of results, a large set of signatures from different individuals must be used. Therefore, for N number of individuals, the equation becomes

$$FRR = \frac{1}{N} \sum_{n=1}^N FRR(n) \quad (2.2)$$

2.5.2 False Acceptance Rate

The False Acceptance Rate (FAR) is a measure of the probability that the system will accept an unauthorized user. It can be expressed mathematically for a set of forgery attempts against a single individual as

$$FAR(n) = \frac{\text{successful forgery attempts of unauthorised individual } n}{\text{total forgery attempts of unauthorised individual } n} \times 100 \quad (2.3)$$

For improved precision and accuracy of results, a large set of forged signatures against different individuals must be used. Therefore, for N number of individuals, the equation becomes

$$FAR = \frac{1}{N} \sum_{n=1}^N FAR(n) \quad (2.4)$$

FAR is often determined separately for skilled forgeries (FARS) and for random forgeries (FARR).

2.5.3 Equal Error Rate

The Equal Error Rate (EER) is the point where FRR and FARR are equal. This is used to gauge the accuracy of the system.

2.5.4 ROC-Curve

The receiver operating characteristic curve (ROC-curve) is often used in calibrating signature verification systems. It is a plot of FRR and FAR percentages at different thresholds. This allows for easy determination of the EER, where the FRR and FAR are equal. Additionally, by using a ROC-curve, the threshold can be dynamically adjusted in real world systems. For example, in a scenario of an individual with a large bank balance, the threshold can be adjusted to a point on the ROC-curve where the FAR is lower than the EER, and the FRR is higher than the EER. This will provide greater security against forgeries, in comparison to the EER [7].

2.5.5 Average Error Rate

The Average Error Rate (AER) is used in place of the EER when the system is not calibrated to have an equal FAR and FRR, such as in [7]. It is a calculation of the total error, along with a priori probabilities.

2.6 Literature Survey of Offline Signature Classification Techniques

Many techniques exist for the classification of signatures and other biometrics. They can be broadly categorized into supervised learning techniques (SLTs) and distance-based classification techniques. SLTs include neural networks [27], hidden Markov models [7], support vector machines [48] and fuzzy logic [31]. Linear techniques include Euclidean distance, Mahalanobis [35], Manhattan distance, weighted Euclidean distances [75] and fractional distances [68].

2.6.1 Supervised Learning Techniques (SLTs)

SLTs use complex mathematical and statistical models to predict the class into which a signature's feature vector belongs. The most commonly used SLTs are hidden Markov models (HMM), support vector machines (SVM), artificial neural networks (NN) and fuzzy logic. Recently, there is also much research in classifier fusion, where a multi-hypothesis approach combines an ensemble of different classifiers or classification of different feature extraction techniques.

2.6.1.1 Hidden Markov Models (HMM)

HMMs are a form of SLT. They are used to analyze the sequence of data points in a feature vector [3]. It was first documented by Baum and Petrie in 1966 [9].

Shakil et al. [59] compares the performance of several online and offline feature extraction techniques with HMM classification, using the SIGMA database of Malaysian signature [2]. The EER values for offline features were between 45 and 46%, which signifies a low accuracy. Coetzer et al. [14] applies HMM classification to discrete Radon transform, which is a global feature, and achieves an EER of 12.2% using skilled forgeries and 4.5% using random forgeries. Panton and Coetzer further improves the EER to 8.6% by using a fusion of HMM classifiers and adding local features to the feature vector [55]. These results are achieved using the small Dolfig and Stellenbosch datasets, since larger datasets were not yet publicly available. However, the results are among the best results obtained when using these specific databases. Ferrer et al. [20] compares the Euclidean distance, HMM and SVM, using geometric features based on Cartesian and polar coordinates. Signatures of 160 individuals from the GPDS database are used. This is a larger dataset than the Dolfig and Stellenbosch datasets, and larger datasets provide greater precision through the use of more datapoints. It is found that in [20], the best performance is for HMM, with an FRR of 14.1% and FAR of 12.3% for skilled forgeries. The SVM performs better than the Euclidean distance, but worse than the HMM.

Batista et al. [7] proposes a multi-hypothesis approach with a best AER of 7.79%. This multi-hypothesis approach is designed such that the most suitable HMM from several is dynamically chosen per individual, and user-specific codebooks are employed. A codebook is used to define discrete clusters of feature vectors, where each cluster is associated with a single individual. It is shown that a multi-hypothesis approach works better than a single-hypothesis approach, and that user-specific codebooks work better than codebooks with clusters for multiple users. User-specific codebooks also allow easier addition and removal of users from a database, since large portions of the database will not require re-training. However, this approach is computationally expensive, in comparison to single-hypothesis classification, due to the use of more than one classifier.

2.6.1.2 Support Vector Machines (SVM)

SVM classification is based on algorithms of statistical learning theory [3]. The technique was first documented by Cortes and Vapnik from Bell Labs in 1995 [16]. These are kernel based techniques for supervised learning. The most common kernels are the polynomial and radial basis function kernels. It can be used for both linear and non-linear approximations.

Yilmaz et al. [72] performs SVM classification on a combination of gradient-based and binary pattern-based features, extracted from the GPDS160 database, achieving a best AER of 15.41%. It is found that user-dependent, also called user-specific, classifiers work better than a globally applied classifier for all users. Low et al. [41] proposes a unique method where a signature is embedded in an image with SVM biometric watermarking (SVM-BW), so that the biometrics are secured until extracted from the image. Feature extraction is performed using discrete Radon transform and principle component analysis. Vargas et al. [67] proposes a system with least squares SVM classification on

features extracted from the Fourier transform of a signature image, using the GPDS100 database. An EER of 6.20% is obtained, which, is shown to perform better than similar techniques, using the same database. Kiani et al. [34] applied an SVM with RBF kernel to a combination of Local Radon Transform and Line Segment features. The results are compared with Panton and Coetzer [55], which applied an HMM to Local Radon Transform. The SVM worked better for the Persian database, but worse for the Stellenbosch database.

Nguyen et al. [50] compares squared Mahalanobis distance and Gaussian kernel SVM classification on a local gradient-based feature extraction. The AER for the SVM is 15.02%. The gradient distance with SVM classification performs better than the squared Mahalanobis distance, and better than previous tests with MDF feature extraction and SVM classification. Batista et al. [8] designs a multi-hypothesis approach where a further classification stage, using an ensemble of SVMs, is added to their HMM-based multi-hypothesis approach from [7]. When used individually, the ensemble of SVMs performs better than the HMMs. Once combined, an AER of 5.50% is achieved. However, the fusion of multiple classifiers negatively affects the speed efficiency of classification, since more processing is needed for training and classification. Additionally, fusing the best single-hypothesis classifiers into a multi-hypothesis approach leads to the best classification. Therefore, it is still essential to find the optimal single-hypothesis solutions before the successful implementation of a multi-hypothesis approach.

2.6.1.3 Artificial Neural Networks (NN)

NNs are mathematical models designed to mimic the behaviour of neurons in the human brain. It was first conceptualized by McCulloch and Pitts in 1943 [44] and first successfully implemented in software, by Widrow and Hoff of Stanford in 1960, to reduce phone line echoes [69]. NN also use kernels. The most common forms are back-propagation multi-layer perceptrons (BP-MLP) and radial basis function (RBF) kernels. NNs can be designed for temporal adaptation. This is a useful feature for signature authentication, since a user's signature changes over a period of time. There is much research involving NNs for offline signature classification. Some of the recent and notable works are discussed here.

Kovari et al. [36] uses NNs for classifying a feature vector comprised of both local and global features. Tests are performed using 40 individual's signature sets from the SVM database. While the database is small in size, the use of machine generated artificial forgeries is an interesting and novel approach. However, a poor EER of over 20% is achieved. Miskhat et al. [47] compared NN and SVM with a fusion of feature extraction techniques, on signatures from 30 individuals. An NN with a feed-forward back-propagation kernel produced better results than the SVM classification. However, no skilled forgeries were used in their tests. It is unknown if random forgeries were used during training and testing and the calculations for the accuracy measure are not specified.

Kumar et al. [38] uses and MLP NN to select the best subset of feature from a feature vector before classification via SVM. An EER of 11.59% is achieved. Mirzaei et al. [46] uses modular NNs to classify three different feature vectors. The results from these three classifiers are then fused, using the Mamdani fuzzy inference system. Using a private signature database comprising of authentic signatures from 30 individuals,

and no forgeries, a recognition accuracy of 96.6% was achieved. Sisodia et al. used an error back-propagation NN to classify signatures from 6 individuals and obtained promising results. Armand et al. (2007) [6] tested the Enhanced MDF with a fusion of NN classifiers. However, the tests were performed using just 44 signatures from the GPDS database. Multi-hypothesis fusion of NNs with other NNs, or with other SLTs, is a promising approach for improving accuracy, but is computationally expensive, since it requires each signature to be classified multiple times.

Nguyen et al. (2007) [49] compares NNs and SVMs with different kernels with the Enhanced MDF, using the GPDS160 database. An SVM with an RBF kernel produced their best EER of 17.78%. Nguyen et al. (2009) [48] compares NN and SVM classification with various kernels and the Extended MDF (EMDF) feature extraction, also using the GPDS160. The SVM provided the best EER of 17.25%. This is more reliable than earlier work [5, 6] where signatures of less than 45 individuals were used.

2.6.2 Distance-based Classification Techniques

Besides SLTs, the other classification techniques can be categorized as distance-based classification techniques. Distance measures are used to calculate the amount of difference, or distance, between two feature vectors: a reference feature vector and a sample feature vector. The reference feature vector is created by averaging several authentic feature vectors for an individual. This makes it a standard reference point to which other signatures' feature vectors, sample feature vectors, are compared. If the difference is below a chosen value, or threshold, the sample feature vector is regarded as authentic, else it is regarded as a forgery.

2.6.2.1 Euclidean Distance

The most well known distance-based measure is the Euclidean distance. It is attributed to the Ancient Greek mathematician Euclid, who developed Euclidean geometry. It was also previously known as the Pythagorean metric, as it is used to calculate the distance between points in a Cartesian plane, as part of the Pythagorean Theorem. It is a distance calculated in L^P -space, also called Lebesgue space, where the p-norm value is 2, or the L_2 distance. There are many works that have used the Euclidean distance for authentication. Some notable and recent works are discussed.

Shekar et al. [60] uses the Euclidean distance for the verification of feature vectors which are created using an Eigenvector-based feature extraction technique. Different sizes of feature vectors are tested, and a single globally applied feature vector size is eventually chosen. An EER of 14.33% is achieved when using 10 signatures for training and 14 for testing. A better EER of 8.78% is also achieved when using 15 signatures for training and 9 for testing, but it is noted that in compared works, the former configuration of training and testing samples were usually used. The former configuration is more commonly used since there are more signatures for testing, leading to greater precision of results. Rekik et al. [57] tests global and local feature extractions with Euclidean distance classification, and concludes that a fusion of local and global features performs better than using local or global features independently. Tests are performed on 50 individuals from BioSecure DS2 and on the GPDS104 database. The best EER

obtained with skilled forgeries is 14.2% and 4.8% using random forgeries. However, the fusion approach was not tested with the larger GPDS104 database. Ramachandra et al. [56] finds the smallest Euclidean distance between cross-validated graphs of signatures, using the Hungarian method [37]. Various feature vector sizes are tested on signatures from just 5 individuals, and an EER of 27.78% is achieved using skilled forgeries and 6.25% using random forgeries for two different vector sizes. In comparison with other literature, this is a poor accuracy obtained using a very small dataset. Majhi et al. [43] applies Euclidean distance classification to feature vectors based on geometric centres of several regions in an image. An FRR of 14.58%, FARR of 2.08% and FARS of 16.36% are obtained.

Kisku et al. [35] documents an interesting technique in which local and global feature extractions are performed, and the results of Euclidean distance, Mahalanobis distance and Gaussian empirical rule from these feature vectors are fused together using SVM. A private Indian signature database is used, with 9 authentic signatures and a single forged signature for each of 180 individuals. Individually, each classification technique performs very well with EER values below 10%, and combined, a best EER of 2.15% is achieved. This multi-hypothesis approach shows great promise, although tests with signatures from other global locations, and tests with more skilled forgeries, are required.

Ferrer et al. [20] performs a comparison between the Euclidean distance, HMM and SVM, using geometric features based on Cartesian and polar coordinates. The best Euclidean distance obtained was with FARS of 15.50% and FRR of 16.39%. However, the HMM and SVM performed better, as discussed in Section 2.6.1.

2.6.2.2 Mahalanobis Distance

The Mahalanobis distance was first documented in 1936 by Prasanta Mahalanobis [42]. It performs best with multivariate normal data distributions [30]. Some notable and recent works are discussed.

Fang et al. [19] performs one of the earliest classifications using the Mahalanobis distance, with a best EER of 19.1%. Nguyen et al. [50] compares squared Mahalanobis distance and SVM classification on a local gradient-based feature extraction. The AER for the squared Mahalanobis distance was 16.52%, but the SVM performed a little better, as discussed previously in Section 2.6.1.2. Sigari et al. [62] performs Mahalanobis distance classification on features extracted using Gabor wavelets. Verification tests are performed on 3 signature databases which are based on Iranian and Latin (South African and Turkish) scripts. EER values of 15.0%, 16.8% and 9.0% are obtained respectively. While these values are promising and performed across a wide range of scripts and cultures, it must be noted that each of these signature databases contained signatures of just between 20 and 40 individuals. The proposed system is also shown to outperform humans at discerning between authentic and forged signatures.

Kisku et al. [35] employs the use of the Mahalanobis distance in their combination of classifiers and previously described in 2.6.2.1. From the classifications without combination, the Mahalanobis distance performs better than the Euclidean distance and Gaussian classifier, for the chosen set of feature extraction techniques.

2.6.2.3 Canberra Distance

The Canberra distance was developed by Lance and Williams between 1966 and 1967 in Canberra, Australia [39]. It is the sum of the ratio set between points in two vectors. It is a less common distance measure, which, to the author's knowledge, was not previously used for signature verification. However, it is used in Shirdhonkar and Kokare [61] for offline signature retrieval, as a step towards signature recognition, with promising results.

2.6.2.4 Cosine Similarity Measure

The cosine similarity measures the similarity between two vectors by calculating the angle between each set of points. It is not a commonly used classification technique for offline signature verification. It is used in Impedovo et al. [26] to measure the suitability of local stability features. However, no information is reported on the accuracy of the stability features in classification of signatures. Thus, no accuracy information is available.

2.6.2.5 Manhattan Distance

The Manhattan distance is another p -norm distance in L^p -space, where $p = 1$. It can also be called the L_1 distance, or city block distance. While the Euclidean distance is the shortest distance between points, the Manhattan distance is a sum of the absolute difference of the points. It was first documented by Hermann Minkowski. Although similar to the Euclidean distance, it is not a commonly used technique for offline signature verification. It is used for offline signature recognition in Adamski and Saeed [1], as well as in Ismail et al. [27].

2.6.2.6 Weighted Euclidean Distance

The weighted Euclidean distance is a normalization of the Euclidean distance, where statistical weight or importance is given to each feature in the vector based on its standard deviation. The standard deviation is calculated from the signatures that make up the reference feature vector. Features with a lower standard deviation have smaller differences between multiple signatures of the same author. Therefore, these features are given greater weight in the weighted Euclidean distance classification.

It is not a commonly used classification technique. Zhu et al. [75] uses the weighted Euclidean distance for iris recognition in 2000. Kalera et al. [45] adapts the weighted Euclidean distance to a weighted k-nearest neighbour classification for offline signature recognition. However, the weighting function is a unique similarity measure rather than the standard deviation. A Bayesian classifier is used for the verification. Alizadeh et al. [4] uses the weighted Euclidean distance for online signature verification.

2.6.2.7 Fractional Distances and the Concentration Phenomenon

The fractional distance is another distance in L^p -space where the p -norm value, also called the Minowski norm exponent, is any fractional value less than 1. Francois and

Wertz [23] discusses the use of fractional distance as an alternative to the Euclidean distance to counteract the concentration phenomenon. This phenomenon is when large feature vectors cause the results of the Euclidean distance to concentrate, or cluster. This clustering of values, which is an intrinsic property of L^P -space distances, makes classification difficult for large feature vectors. Fractional distances generally produce less concentrated results than the Euclidean distance, which allows for better classification of datasets.

Francois et al. [23] states, “*Fractional norms are not always less concentrated than other norms. They seem, however, to be more relevant as a measure of similarity when the noise affecting the data is strongly non-Gaussian.*” Much of the noise generated by behavioral biometrics is due to random variations in human action, that may not follow a normal distribution. This makes fractional distances a viable investigative route of classifying offline handwritten signatures.

Fractional distances have not been used in offline signature verification before. However, they are used for online signature recognition in Vivaracho-Pascual et al. [68], and for face recognition in Espinosa-Duró et al. [18].

2.7 Conclusion

In this chapter, signatures as a biometric modality were discussed along with its advantages and disadvantages. Distinctions were specified between offline and online recording, as well as between recognition and verification. Existing signature databases, feature extraction techniques, performance measures and classification techniques were discussed. There is much research into the use of SLTs for the verification of signatures, but much less for distance-based techniques for classification. Some work exists on testing different feature vector sizes for optimal classification in specific cases, as well as choosing the local optimal classifier per individual. However, there is no work, to the author’s knowledge, on localized optimizations in terms of resampled feature vectors for offline signature verification. Then next chapter discusses the methodology and techniques used to obtain high accuracy through the use of distance-based classification techniques.

Chapter 3

Design and Methodology

3.1 Introduction

In this chapter, a discussion is provided on the roles of preprocessing, feature extraction and classification in an offline signature verification system. Each of the preprocessing techniques that are used and the feature extraction processes that follow are discussed. Several classification techniques are applied to the extracted feature vectors in order to discover the best possible distance-based classification technique. These classification techniques are described after the discussion on the feature extraction techniques.

3.2 Design Overview

Figure 3.1 shows the overview of the verification system. The three main steps in a biometric verification system are preprocessing, feature extraction and classification. Preprocessing prepares the signature image for the extraction of features which are then used to classify the signature as authentic or forged. The preprocessing techniques used in this work include binarization, bounding box extraction, thinning and dilation. For feature extraction, the Extended Modified Direction Feature (EMDF) and the Local Directional Pattern (LDP) features are used. Each feature extraction technique requires a different set of preprocessing techniques. For classification, distance-based classification techniques, such as the Euclidean, Manhattan and fractional distances are tested.

In the training phase, the reference feature vector is determined by averaging a subset of randomly chosen authentic feature vectors. Sample feature vectors are obtained from other authentic and forged signatures individually. The classifier is then trained to determine the threshold that provides the optimal accuracy for the system, using sample feature vectors of known classification as input. In the testing phase, further sample signatures are used, but the classifier will determine their classification independently based on the threshold obtained from the training phase. Accuracy is gauged based on the number of signatures that the classifier correctly accepts as authentic and correctly rejects as forgeries.

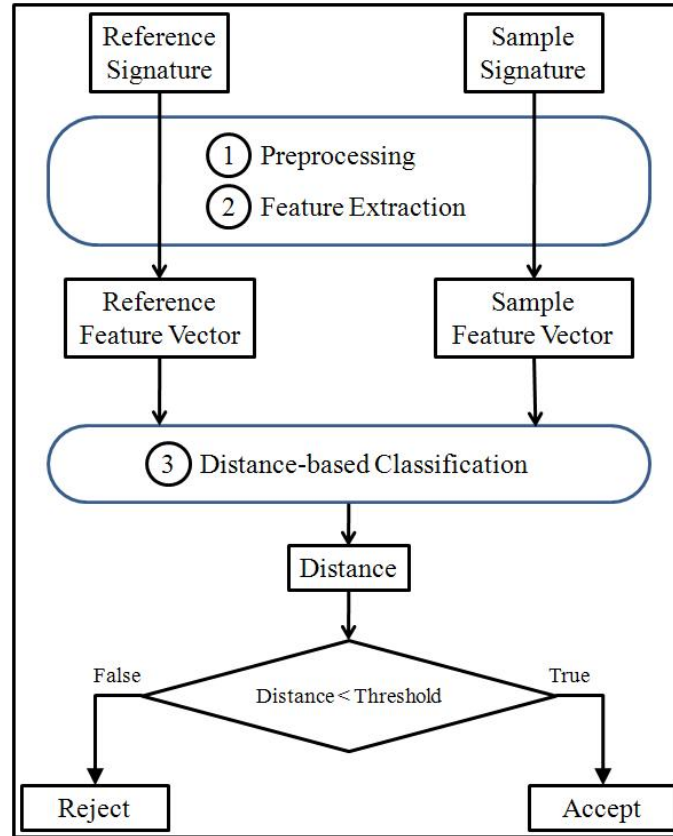


Figure 3.1: Overview of the design of a biometric verification system

3.3 Preprocessing

Preprocessing entails the transformations performed on an image to prepare it for feature extraction. Different preprocessing techniques are used as prerequisites for each of the feature extraction techniques.

3.3.1 Binarization

Binarization is the conversion of greyscale or colour images into simpler black and white images to allow faster processing and extraction of data.

Given a greyscale source image, $I(x, y)$, and threshold greyscale value, T , the binarized image, $B(x, y)$ can be determined as

$$B(i, j) = \begin{cases} fg, & \text{if } I(i, j) > T. \\ bg, & \text{otherwise.} \end{cases} \quad (3.1)$$

where fg is the value for all foreground pixels and bg is the value for all background pixels.

3.3.2 Thinning

Some of the feature extraction techniques used in this work require the image to be of a 1-pixel thickness before feature extraction is performed. To create a 1-pixel thick skeleton of the signature images, the Zhang-Suen thinning algorithm, described in Zhang and Suen [74], is used to allow a more direct comparison with literature results, which used this thinning algorithm as well. The Zhang-Suen is one of the most commonly used thinning algorithms for pattern recognition. However, it is designed for speed rather than accuracy. It may lose connectivity between segments. More accurate thinning algorithms exist, such as the Guo-Hall thinning algorithm, but these usually require much higher processing time [33] in comparison to Zhang-Suen. Multi-thread or multi-processor techniques are used to alleviate the constraints of computationally expensive thinning algorithms. Figure 3.2 and figure 3.3 show samples of a signature before and after thinning.



Figure 3.2: A sample of a signature before thinning and direction feature extraction

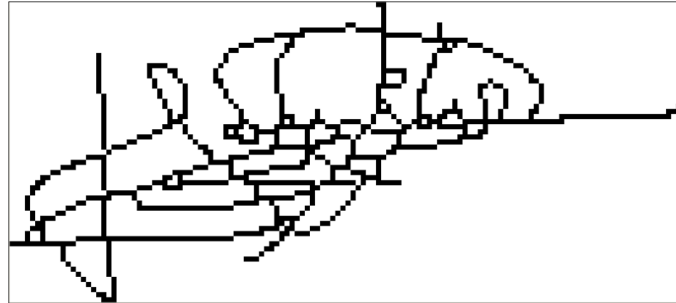


Figure 3.3: A signature after the Zhang-Suen thinning algorithm has been applied

3.3.3 Boundary Extraction

For uniformity between all signatures of an individual, the borders (bounding box) of signatures need to be detected and all background pixels outside of the area discarded [11]. The algorithm used for cropping the image is described in Algorithm 1 and an example of an image before and after cropping to the bounding box is shown in Figures 3.4 and 3.5 respectively.

Algorithm 1 Bounding box, I_{bound} , of a binarized image, I_{src}

Require: I_{src} , \triangleright Source image
Require: fgd , \triangleright The colour of foreground pixels
Ensure: I_{bound} , \triangleright Image transformed using bounding box

- 1: leftBorder = column in I_{src} with leftmost fgd
- 2: rightBorder = column in I_{src} with rightmost fgd
- 3: topBorder = row in I_{src} with topmost fgd
- 4: bottomBorder = row in I_{src} with bottommost fgd
- 5: $I_{bound}(0, 0, x, y) = I_{src}(\text{leftBorder}, \text{topBorder}, \text{rightBorder}, \text{bottomBorder})$

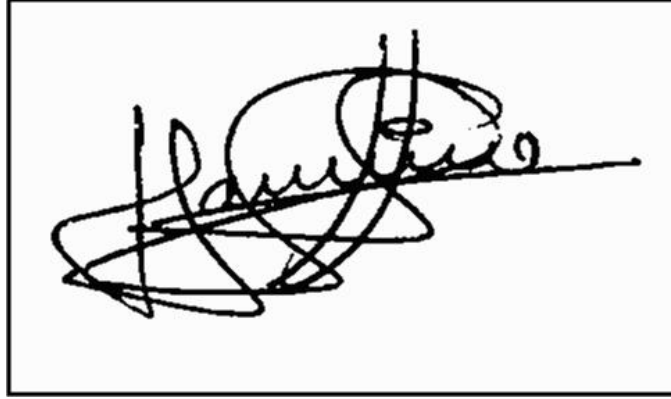


Figure 3.4: A signature image before boundary extraction of the signature occurs.

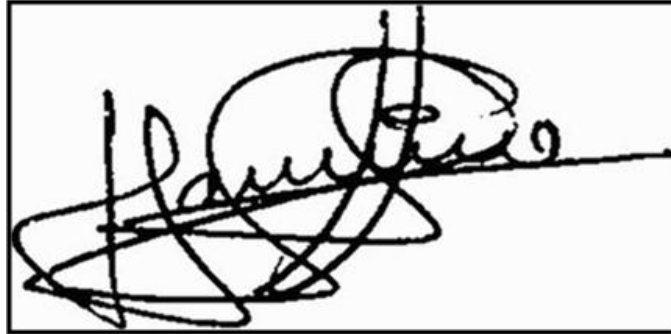


Figure 3.5: A signature image cropped to the bounding box size.

3.3.4 Dilation

For dilation a 3x3 pixel mask is applied to the image using the binary AND operation on a binarized image. This expands the foreground and emphasizes the contours of the image, which allows better extraction of some spacial structure features, such as the Local Directional Pattern [21].

Given a binarized image I and a mask, M , defined as

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Given each $I(x, y)$ we want to compute its transform by dilation, $I_D(x, y)$ with the mask M . It can be done by defining the function

$$\begin{aligned}
 FNCTOR(x, y) = & \left(\sum_{i=-\frac{n}{2}}^{-1} \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} I(x+i, y+j) \right) \\
 & + \left(\sum_{j=-\frac{n}{2}}^{-1} I(x, y+j) \right) + \left(\sum_{j=1}^{\frac{n}{2}} I(x, y+j) \right) \\
 & + \left(\sum_{i=\frac{n}{2}}^1 \sum_{j=-\frac{n}{2}}^{\frac{n}{2}} I(x+i, y+j) \right)
 \end{aligned} \tag{3.2}$$

where \sum and $+$ are boolean sums. In this case $n = 3$.

The dilated pixel, $I_D(x, y)$, in the dilated image, I_D , will be

$$I_D(x, y) = \begin{cases} 1 & \text{if } FNCTOR(x, y) = 1, \\ I(x, y) & \text{otherwise.} \end{cases} \tag{3.3}$$

Figures 3.6 and 3.7 show a signature before and after the dilation, respectively.

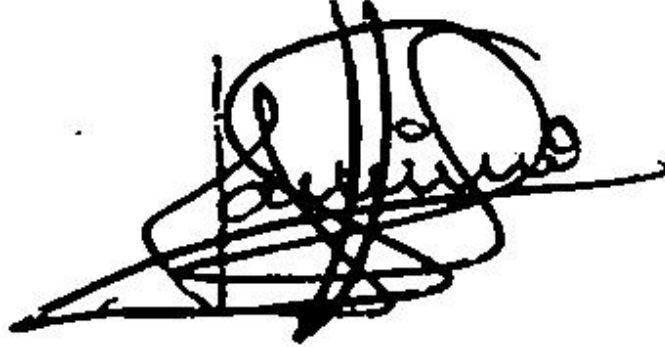


Figure 3.6: A standard black and white signature of a user before dilation preprocessing is applied

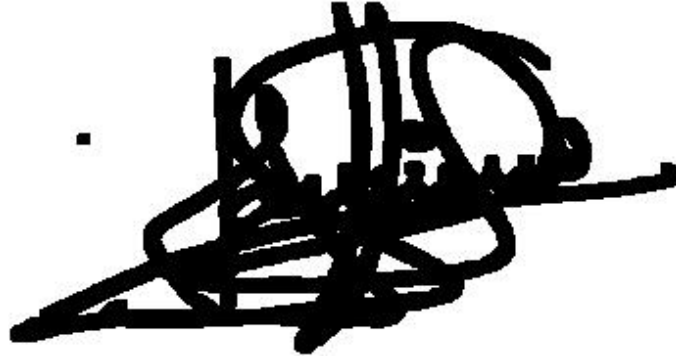


Figure 3.7: A standard black and white signature of a user after dilation preprocessing using a 3×3 OR mask

3.4 Feature Extraction

3.4.1 Choosing the Feature extractions

The Enhanced Modified Direction Feature (EMDF) and Local Directional Pattern (LDP) feature extraction techniques were chosen to test the distance-based classification techniques. When selecting feature extraction techniques several factors were taken into account.

- **Database:** Both techniques were previously tested using the GPDS signature database, which is also used in this work.
- **Offline verification:** The feature extraction techniques were designed for the verification of offline signatures.
- **Local feature extraction:** The feature vectors from both techniques were easily re-scalable by resampling due to the nature of the local feature extractions.
- **Classification:** Both feature extraction techniques were previously tested with SLTs. The MDF was tested with multiple types of NNs and SVMs, and the LDP was tested with an RBF SVM. This allowed for more direct comparison with the distance-based techniques employed in this work.

3.4.2 Direction Feature

The direction feature extraction technique was introduced by Blumenstein et. al. as part of a optical character recognition system for segmented handwriting [11]. It aims to extract the direction of each segment between intersections within a signature, i.e. whether the segment is horizontal, vertical, diagonal left or diagonal right. The steps are sequenced as follow:

1. Preprocessing: The image binarized, cropped to the bounding box area and then thinned to a 1-pixel thickness using the Zhang-Suen thinning algorithm [74].

2. Identification of intersections: Intersections are defined as any foreground pixel that has more than 2 neighbouring foreground pixels.
3. Distinguishing of line segments: Line segments are defined as a continuous connection of foreground pixels of at least 4 pixels in length parameterized by intersections or image borders.
4. Labelling of line segment: Each line segment is labelled with a direction based on the direction of the majority of its pixels, as shows in Figure 3.8, where red represents intersections; green, vertical segments; dark blue, horizontal segments; purple, 45° diagonal segments; light blue, 135° diagonal; and grey represents segments too short to determine a segment direction.

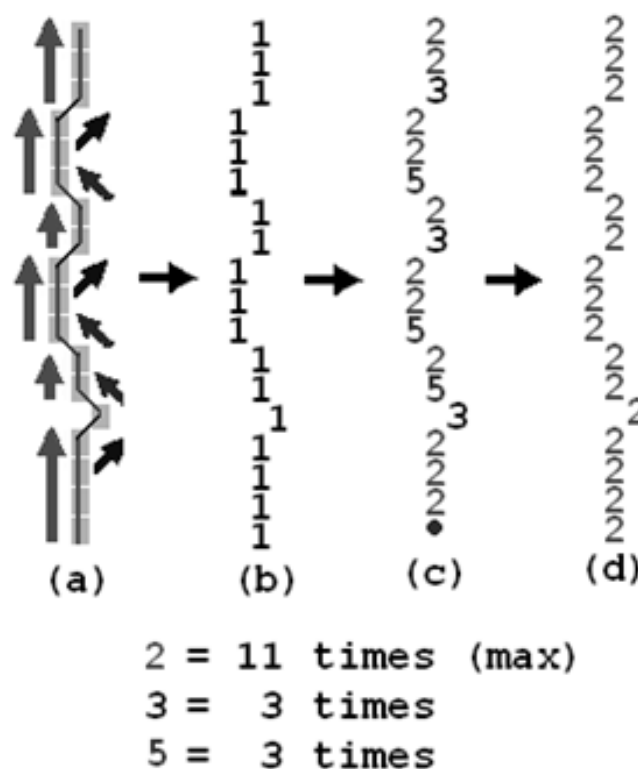


Figure 3.8: In depth demonstration of direction feature extraction

(a) represents the original line, (b) the binary line, (c) after each pixels direction was distinguished, (d) after the line segments direction as distinguished[11, 12]

A sample of a signature image after the direction feature extractions are applied is shown in figure 3.9. Intersections are marked in red, vertical segments in green, horizontal segments in dark blue, right diagonal segments in pink, left diagonal segments in light blue and short ignored segments in grey.

Figure 3.8 shows an example of part of the process for distinguishing the direction of line segments;

(a) shows the graphical representation of the line;

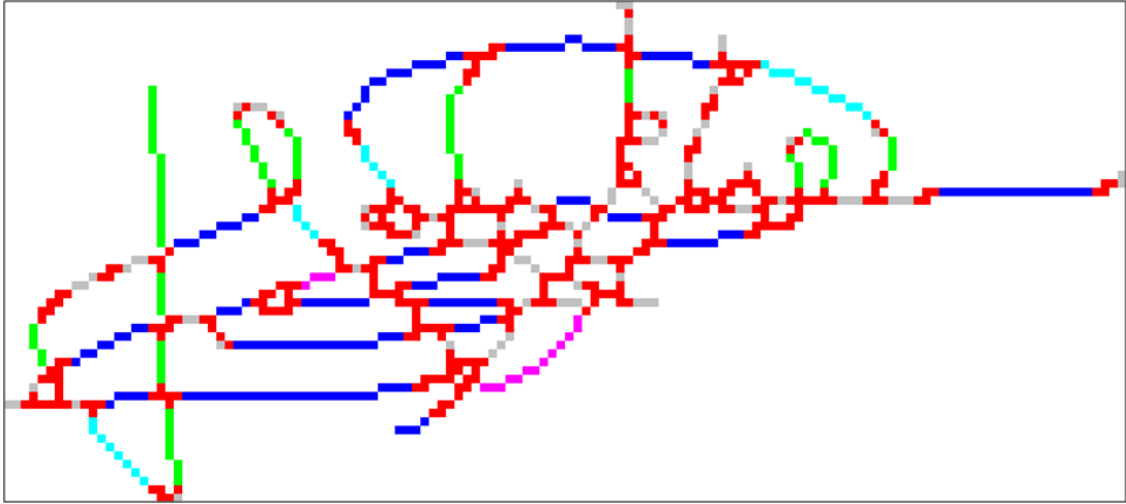


Figure 3.9: A signature after the direction feature has been applied with different colours for each type of line segment and for intersection pixels

- (b) shows the binarized data of a line segment with no direction information (1 for the foreground and 0 for the background, which has been omitted for simplicity).
- (c) each pixel is analyzed starting from the pixel of the bottom-most coordinates. The current pixel is given a different number based on it's position in relation to the previous pixel as shown in Figure 3.10.
- (d) Once each pixel is labelled with a direction number, the highest occurring direction number is assigned to all pixels in the line segment, if the segment is greater than 3 pixels in length. If the segment is 3 pixels or shorter in length, all pixels are assigned with a different number to signify they are to be disregarded.

5	2	3
4	<i>p</i>	4
3	2	5

Figure 3.10: The direction value that will be assigned to a current pixel based on it's directional relation to the previous pixel, ***p***

This yields 9 features, namely, the total length of each line direction set (4 features), the number of lines in each direction set (4 features) and the total number of intersection points (1 feature). Each feature is normalized to be within the range of 0 and 1000 as in Equation 3.4 where FV is the feature vector and max is the maximum value. This range is also used for the Modified Direction, Ratio, Maxima and Energy features. The uniform range is applied to allow better concatenation of these feature vectors when they are combined to form the Extended Modified Direction Feature.

$$FV_{normalized} = \frac{FV(i)}{max} \times 1000 \quad \forall i \in FV \quad (3.4)$$

3.4.3 Modified Direction Feature

The modified direction feature extraction technique [12, 10] results from a combination of the direction feature [11] and the transition feature. It has been used as part of optical character recognition systems. A series of steps are required:

1. Preprocessing: The direction feature extraction is performed so as to label each line segment with a direction.
2. Feature extraction: The image is parsed in four directions, namely, right to left, left to right, top to bottom and bottom to top. In each direction, the location transition (LT) and direction transition (DT) features are recorded. LT is the location of the transition of pixels from background to foreground and DT is the direction of the line segment at the point of transition. For the sake of uniformity, limits are placed to the maximum number of transitions, *max_transitions*, recorded in any given direction. This provided 8 arrays of features with sizes *max_transitions* \times (*image_height* or *image_width*). Figure 3.11 shows an example of processing of the DT and LT in the left-to-right direction.
3. Resampling: Window resampling is performed. This means that the height and width values of each array are resampled for normalization, or averaging, so that feature vectors for all signatures of an individual are uniform in size. Different numbers of (*rs_strips*) and *max_transitions* sizes are tested. Figure 3.12 shows the vertical resampling of the image as would have to be performed for (a) LT transition set and (b) DT transition set of left-to-right transitions. The 15 rows are split into 5 sets of 3 rows and each of these 3 rows are averaged into 1 row. Vertical resampling, as show in Figure 6, must be performed for left-to-right and right-to-left transitions and horizontal resampling must be performed for top-to-bottom and bottom-to-top transitions.

To formalize the window resampling, let m be the number of *maximum_transitions* and p be the number of pixels in height or width where each pixel now stores the LT or DT value for the row or column, so that the transitioned image will be $I_T(p, m)$. Let s be the number of pixels in one strip, as calculated in Equation 3.5. Let r be the number of *rs_strips* and v be the value of an element in a calculated *rs_strip*.

$$s = p/r \quad (3.5)$$

Then the resulting set of *rs_strip* values is

$$\begin{aligned} v_1 &= (x_1^1, x_1^2, \dots, x_1^r) \\ v_2 &= (x_2^1, x_2^2, \dots, x_2^r) \\ &\vdots \\ v_m &= (x_m^1, x_m^2, \dots, x_m^r) \end{aligned} \quad (3.6)$$

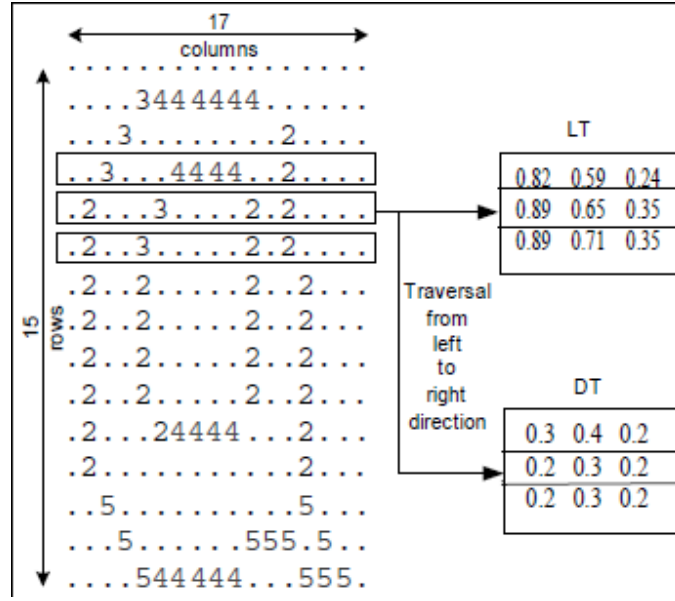


Figure 3.11: Processing of the Direction Transition (DT) and Location Transition (LT) feature values in the Left-to-Right direction [12]

Let v_j^i be a single calculated component where $1 \leq i \leq r$ and $1 \leq j \leq m$. Then the value v_j^i is calculated as

$$v_j^i = \frac{1}{s} \sum_{k=n \times s}^{k < (n \times (s+1)) - 1} I(k, j) \quad (3.7)$$

This provides 8 feature vectors ($4 \times \text{LT} + 4 \times \text{DT}$) with sizes $\text{max_transitions} \times \text{rs_strips}$ (of height or width). For simplification, in each case, an equal value for max_transitions and rs_strips is used. For example, if a maximum of 5 transitions are used, along with resampling of 5, the feature vector will have 200 components.

Since all values are in the range $[0 : 1]$, normalization of the feature vector values is

$$FV_{\text{normalized}} = FV(i) \times 1000 \quad \forall i \in FV \quad (3.8)$$

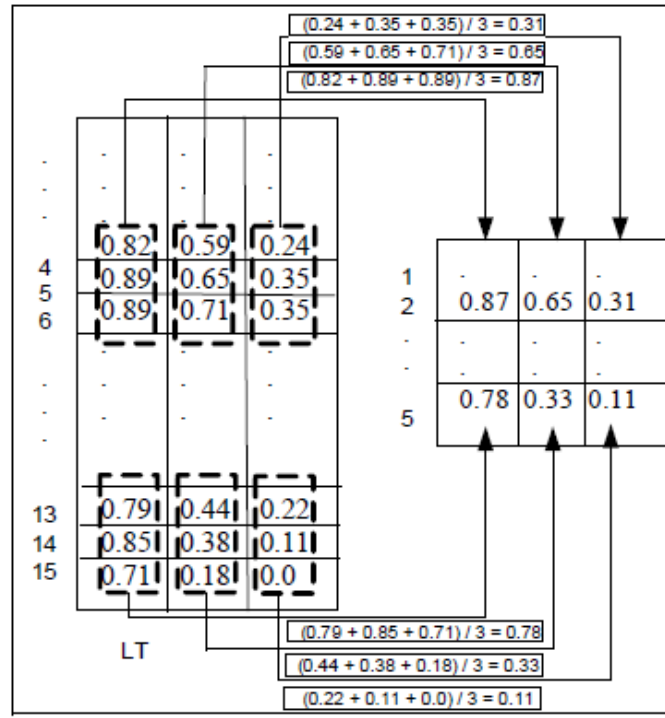
where FV is the feature vector.

3.4.4 Ratio Feature

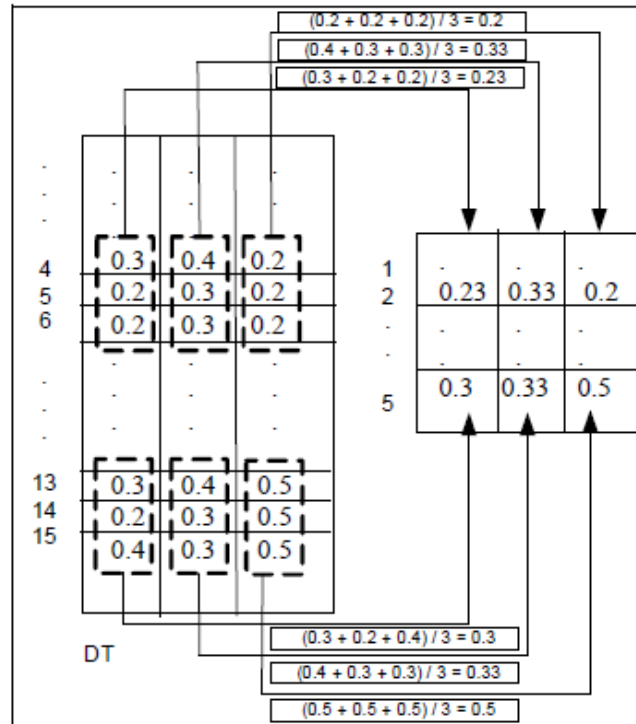
The Ratio feature is a measure of the ratio between the height and width of the image after bounding box preprocessing. Nguyen et. al. [48] tried two variations of the equation as shown in Equations 3.9 and 3.10.

The original equation used by Nguyen et. al. is

$$\text{Ratio}_1 = \frac{\arctan(\text{width}/\text{height})}{\pi/2} \quad (3.9)$$



(a)



(b)

Figure 3.12: Resampling of DT and LT values in the left-to-right direction [12]

where *width* and *height* are the dimensions of the signature image after cropping to the bounding box, as displayed in Figure 3.13.

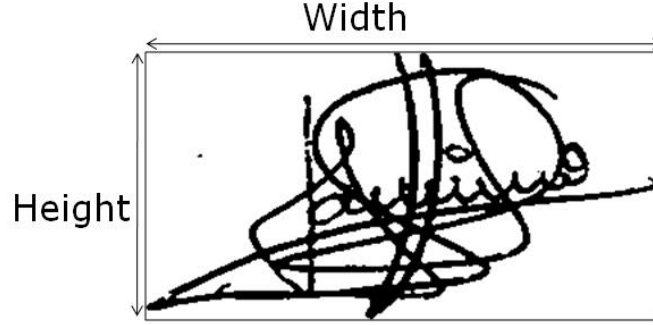


Figure 3.13: A signature image cropped to the bounding box size

In later works, a variation of the equation that provides better accuracy when training the system with supervised learning techniques is used. This variation is defined as

$$Ratio_2 = \frac{\min(width, height)}{\max(width, height)} \quad (3.10)$$

where $\min(width, height)$ and $\max(width, height)$ determine the smaller and larger between the width and height of the image, respectively. This returns a single feature, which is normalized as in Equation 3.8.

In this work, both variations are tested with distance-based classification techniques. The results are discussed in Chapter 4.

3.4.5 Maxima Feature

The Maxima feature, proposed by Nguyen et. al. [48], determines the distance between the two dominant vertical strokes as a fraction of the height and the distance between the two dominant horizontal strokes as a fraction of the width. The vertical maxima is defined as

$$Maxima_{vertical} = \left(\frac{|row_{max}^1 - row_{max}^2|}{height} \right) \quad (3.11)$$

and the horizontal maxima is defined as

$$Maxima_{horizontal} = \left(\frac{|col_{max}^1 - col_{max}^2|}{width} \right) \quad (3.12)$$

where row_{max}^1 and row_{max}^2 are the positions of the longest (or most dominant) and second longest horizontal segments and col_{max}^1 and col_{max}^2 are the positions of the longest and second longest vertical segments, respectively.

Two preprocessing variations of the Maxima feature are tested in this work.

1. In the first variation (Maxima_T), binarization, bounding box cropping and Zhang-Suen thinning [74] are performed. The segments were then identified by finding the longest perfectly straight lines of pixels in the horizontal and vertical directions.
2. In the second variation (Maxima_S), the Direction Feature is applied after the preprocessing steps mentioned directly above. The longest segments were then chosen from the average direction segments determined by the Direction Feature processing.

Figure 3.14 shows the identification of the two dominant horizontal segments (in blue) and two dominant vertical segments (in green) after the direction feature is applied. This returns 2 features, which are normalized as in Equation 3.8.

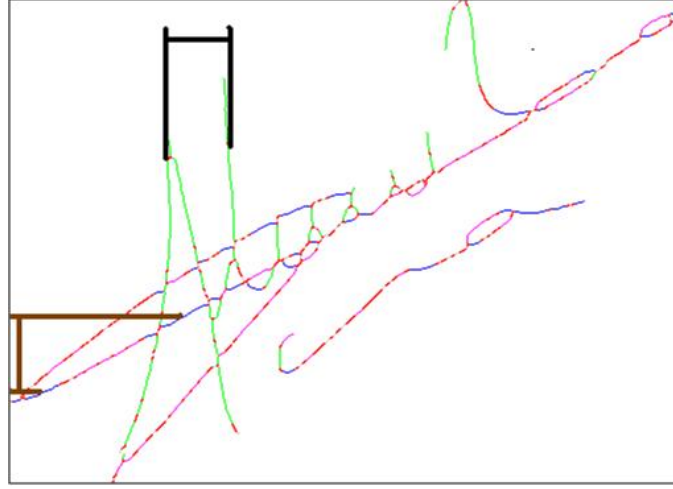


Figure 3.14: A depiction of the Maxima Feature extraction process after Direction Feature preprocessing. The distances between the longest vertical and horizontal segments is shown.

3.4.6 Energy Feature

The Energy feature, proposed by Nguyen et. al. [48], is a global measurement of the amount of energy used by an individual to write a signature. Binarization, bounding box cropping and thinning are required for preprocessing. After the preprocessing, the energy features are extracted as described in Algorithm 2. This returns 5 features, which are normalized as in Equation 3.8.

3.4.7 Extended Modified Direction Feature

The Extended Modified Direction Feature (EMDF) is a concatenation of the DF, MDF, Ratio, Maxima and Energy features as proposed by Nguyen et. al. [48]. It was found to improve the accuracy of classification since it is a combination of “local features” and “global features”. Local features refer to feature extraction techniques that treat patterns holistically and provide in-depth information of an image by analyzing individual parts of a pattern. The local features, in this work, are obtained from the MDF. Conversely,

Algorithm 2 Energy Feature extraction of an image, I

Require: I , \triangleright Source Image
Require: fgd , \triangleright The colour of foreground pixels
Require: sum_V , \triangleright Sum of vertical neighbouring pixels
Require: sum_H , \triangleright Sum of horizontal neighbouring pixels
Require: sum_{D45} , \triangleright Sum of 45° diagonal neighbouring pixels
Require: sum_{D135} , \triangleright Sum of 135° diagonal neighbouring pixels
Require: $height$, \triangleright Height of I
Require: $width$, \triangleright Width of I
Ensure: $energy_V$, \triangleright Vertical energy
Ensure: $energy_H$, \triangleright Horizontal energy
Ensure: $energyRatio_H$, \triangleright Ratio of horizontal to vertical energy
Ensure: $energyRatio_D$, \triangleright Ratio of 45° diagonal to 135° diagonal energy
Ensure: $energyRatio_G$, \triangleright Global ratio energy

- 1: **for** each pixel(x,y) **do**
- 2: **if** pixel(x,y+1)= fgd **then**
- 3: sum_V++
- 4: **end if**
- 5: **if** pixel(x,y-1)= fgd **then**
- 6: sum_V++
- 7: **end if**
- 8: **if** pixel(x+1,y)= fgd **then**
- 9: sum_H++
- 10: **end if**
- 11: **if** pixel(x-1,y)= fgd **then**
- 12: sum_H++
- 13: **end if**
- 14: **if** pixel(x+1,y+1)= fgd **then**
- 15: sum_H++ ; sum_V++ ; $sum_{D135}++$;
- 16: **end if**
- 17: **if** pixel(x-1,y-1)= fgd **then**
- 18: sum_H++ ; sum_V++ ; $sum_{D135}++$;
- 19: **end if**
- 20: **if** pixel(x+1,y-1)= fgd **then**
- 21: sum_H++ ; sum_V++ ; $sum_{D45}++$;
- 22: **end if**
- 23: **if** pixel(x-1,y+1)= fgd **then**
- 24: sum_H++ ; sum_V++ ; $sum_{D45}++$;
- 25: **end if**
- 26: **end for**
- 27: $energy_V = height / sum_V$
- 28: $energy_H = width / sum_H$
- 29: $energyRatio_H = \frac{\min(sum_V, sum_H)}{\max(sum_V, sum_H)}$
- 30: $energyRatio_D = \frac{\min(sum_{D45}, sum_{D135})}{\max(sum_{D45}, sum_{D135})}$
- 31: $energyRatio_G = \frac{\min(energyRatio_H, energyRatio_D)}{\max(energyRatio_H, energyRatio_D)}$

global features refer to feature extraction techniques that extract information from an overview of the image. This makes them less affected by noise and variations in the image, but means that they extract less information. The local features, in this work, are obtained from the Energy, Direction, Ratio and Maxima features. All the features are normalized into the range of [0:1000], as described previously, and then concatenated to create the Extended Modified Direction Feature.

3.4.8 Local Direction Pattern (LDP)

The Local Direction Pattern (LDP) was proposed by Jabid et. al. [28] for use in face recognition. It was later used for signature verification by Ferrer et. al. [21]. This technique utilizes the 8 orientations of Kirsch masks, as shown in Figure 3.15, to detect the presence of edges or corners and their orientations.

For preprocessing, the image is first binarized, then dilated and cropped to the bounding box area. For each pixel in the source image, $I_{src}(x, y)$, the LDP code values are calculated to transform it into a new image, $I_{LDP}(x, y)$, as shown in Figure 3.16. This is done by performing a convolution of Kirsch masks with the image at each pixel, so as to obtain the values of the 8 mask orientations, m_0, m_1, \dots, m_7 . The absolute mask values are then ranked in descending order and then binarized to create the code bits. For the mask value binarization, the three highest ranking mask values are set to 1 and the rest are set to 0. In other words, given the source image I_{src} , we will compute I_{LDP} , which is a transformed image using Algorithm 3.

$$\begin{array}{cccc}
 \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} & \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \\
 \text{East } M_0 & \text{North East } M_1 & \text{North } M_2 & \text{North West } M_3 \\
 \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} \\
 \text{West } M_4 & \text{South West } M_5 & \text{South } M_6 & \text{South East } M_7
 \end{array}$$

Figure 3.15: The 8 orientations of Kirsch Masks [21]. Each orientation is applied to a pixel and its 8 neighbours to calculate 8 mask values.

Following this, a histogram, H_{LDP} , is created from $I_{LDP}(x, y)$. However, since each 8-bit pixel has exactly three bits with the value 1, and 5 bits with the value 0, this allows for only 56 possibly permutations out of the standard 256 permutations that are possible from 8 bits. Therefore, the histogram will only account for these 56 possible values. Figure 3.17 shows a sample LDP histogram.

Further, it is possible to divide the image $I_{LDP}(x, y)$ into blocks, also called grids or zones, by splitting it into a specified number of parts vertically ($split_V$) and horizontally ($split_H$) and have a 56-value histogram for each block. The final feature vector, FV_{LDP} , is then obtained by concatenating all of these histograms, $FV = H_{LDP}^1 + H_{LDP}^2 + \dots + H_{LDP}^{split_V \times split_H}$. Figure 3.18 shows a sample of an image with 4 vertical splits and 3 horizontal splits. This is also regarded as a form of resampling, which may increase or decrease the feature vector size and the depth of information.

85	31	26
53	50	10
60	38	45

↓

Mask index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask value	-300	100	164	540	308	92	-508	-396
Rank	5	7	6	1	4	8	2	3
Code bit	0	0	0	1	0	0	1	1
LDP code	19							

Figure 3.16: calculation of the LDP code [21] obtained by applying each of the 8 Kirsch masks

Algorithm 3 Local Directional Pattern, I_{LDP} , calculation of an image, I_{src}

Require: I_{src} ,

▷ Source Image

Ensure: I_{LDP} ,

▷ Image Transformed
using Local Direc-
tional Pattern

```

1: for each pixel (x,y) do
2:   for i=0 do 7
3:     for k= - 1 do 1
4:       for l = -1 do 1
5:          $m_i = m_i + M_i(k + 1, l + 1) \times I_{src}(x + k, y + l)$ 
6:       end for
7:     end for
8:     Transform the three highest val-
       ues  $m_i$  into 1s and the rest into
       0s
9:   end for
10:   $powerof2 = 1$ 
11:   $I_{LDP}(x, y) = 0$ 
12:  for i = 0 do 7
13:     $I_{LDP}(x, y) = I_{LDP}(x, y) + m_i \times powerof2$ 
14:     $powerof2 = 2 \times powerof2$ 
15:  end for
16: end for

```

3.4.9 Feature Vectors and Individually Optimized Resampling

Resampling increases, or decreases, the amount of information extracted from an image, through resizing of the feature vector. This is a form of spatial normalization. Different resampling sizes results in difference accuracies due to changes in the amount of extracted information. It is expected that choosing the best resampling size per user will optimize the overall accuracy of the system. When a feature vector is resampled, its size is normalized to produce a uniform feature vector size, either globally for all signature sets or locally per individual set. Vivaracho-Pascual et. al. [68] tries several resampling

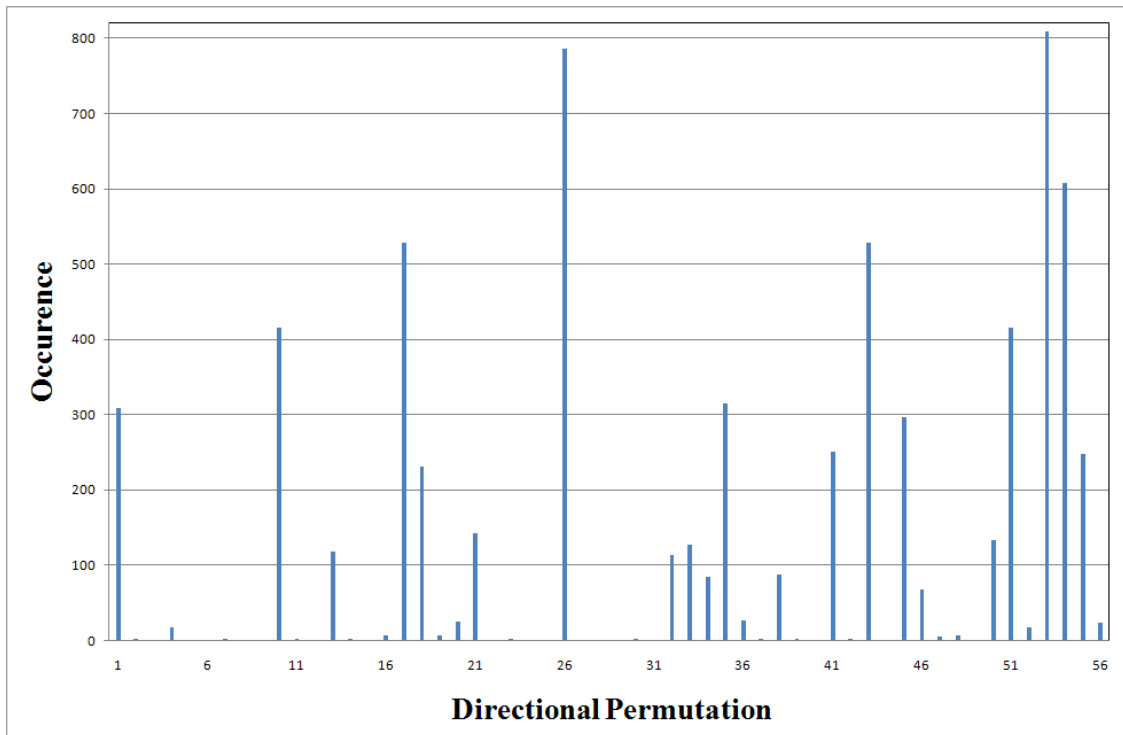


Figure 3.17: A sample LDP histogram showing the occurrences of each directional permutation from the image in Figure 3.7 with no splits



Figure 3.18: A dilated image with splits segmented by 3 horizontal splits and 4 vertical splits

sizes for their feature vectors, in research with online signatures. They note that there is no single resampling size that is optimal for all signatures. They further state that local optimization, also called individually optimized resampling, obtained by choosing the best feature vector size per individual, is a non-trivial approach and interesting for future study. However, for their work, they chose a global resampling size for all individuals. In this work, multiple feature vector sizes are analyzed so as to understand the effect of resampling. Resampling was performed on the MDF and on the LDP.

Since the resampling technique is usually tied closely to the feature extraction technique used, different resampling approaches were applied for each of the Modified Direction Feature (MDF) and Local Directional Pattern (LDP) feature extraction techniques. Below are brief outlines of the resampling that is performed.

The resampling technique used for the MDF is described in Sections 3.4.3. Different *rs_strips* and *max_transitions* sizes are used to change the size of the feature vector. The smallest possible is a size of 2. Tests with incrementally large sizes are used, until size 8. By this point, the resampling has little positive impact on the accuracy and the last feature vector size was negatively affecting processing time. It was decided to keep the maximum size of 8. In the previous studies of the MDF techniques [10, 12, 48] a *rs_strip* size of 5 was used, only. Blumstein et. al. [12] tests 2 different sizes of *max_transitions*, namely, 3 and 4. It was found that, when using a neural network with a radial basis function for classification, 3 transitions work better than 4 transitions in one case only. They therefore use 4 transitions only for the rest of their work. Feature vector sizes for MDF and EMDF respectively are calculated as

$$FV_{MDF} = 8 \times \text{max_transitions} \times \text{rs_strips} \quad (3.13)$$

and

$$FV_{EMDF} = 8 \times \text{max_transitions} \times \text{rs_strips} + FV_{DF} + FV_{Energy} + FV_{Ratio} + FV_{Maxima} \quad (3.14)$$

where

$$FV_{DF} + FV_{Energy} + FV_{Ratio} + FV_{Maxima} = 9 + 5 + 1 + 2 \quad (3.15)$$

The resampling technique used for LDP is described in Sections 3.4.8. In the works of Ferrer et. al. [21], the signatures are split into 4 blocks vertically and 3 blocks horizontally, giving a total of 12 blocks. In this work, images are tested with block splits between 1 and 8 in both the vertical, *split_V*, and horizontal, *split_H*, directions. This provides between 1 and 64 blocks per image. Since each histogram has a size of 56 and histograms are concatenated, the feature vector size for LDP feature extraction is calculated as

$$FV_{LDP} = 56 \times \text{split}_V \times \text{split}_H \quad (3.16)$$

The tested feature vector sizes for the MDF and EMDF are shown in Table 3.1 and the feature vector sizes for LDP are shown in Table 3.2.

Table 3.1: **MDF feature vector sizes** - The feature vector sizes with each different combination of *rs_strips* and *max_transitions*

<i>rs_strips</i>	<i>max_transitions</i>	MDF FV size	EMDF FV size
2	2	32	49
3	3	72	89
4	4	128	145
5	5	200	217
6	6	288	305
7	7	392	409
8	8	512	529

Table 3.2: **LDP feature vector sizes** - The feature vector sizes with each different combination of *split_H* and *split_V*

H \ V	1	2	3	4	5	6	7	8
1	56	112	168	224	280	336	392	448
2	112	224	336	448	560	672	784	896
3	168	336	504	672	840	1008	1176	1344
4	224	448	672	896	1120	1344	1568	1792
5	280	560	840	1120	1400	1680	1960	2240
6	336	672	1008	1344	1680	2016	2352	2688
7	392	784	1176	1568	1960	2352	2744	3136
8	448	896	1344	1792	2240	2688	3136	3584

3.5 Classification

3.5.1 Determining the Threshold

Authentic signatures are expected to have distance values below a certain threshold while forged signatures would have values above that threshold. This threshold is determined by finding the optimal Equal Error Rate (EER) during the training phase. Authentic signatures with distances above the threshold are regarded as false negatives and contribute to the False Rejection Rate (FRR) while forged signatures with distances below the threshold are regarded as false positives and contribute to the False Acceptance Rate (FAR). This is further split into the FAR for skilled forgeries (FARS) and for random forgeries (FARR). The threshold is chosen where the distance for the FRR and FARS are equal.

3.5.2 Euclidean Distance

One of the most common distance-based classification techniques for determining the accuracy of biometric systems is the calculation of the Euclidean distance between a reference vector (derived as a mean of several authentic signatures of an individual) and other feature vectors.

The equation for determining the Euclidean distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is defined as

$$||x - y||_p = \left(\sum_{i=1}^m |(x^i - y^i)|^p \right)^{1/p} \quad (3.17)$$

where $p = 2$. This makes the equation

$$||x - y||_2 = \left(\sum_{i=1}^m |(x^i - y^i)|^2 \right)^{1/2} \quad (3.18)$$

3.5.3 Manhattan Distance

The Manhattan distance, also called the City-block distance, is the distance between two points determined as the sum of the absolute difference of their respective coordinates. The equation for determining the Manhattan distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is computed as defined in Equation (3.17) where $p = 1$. Therefore, Equation (3.17) can be re-written as

$$||x - y||_1 = \sum_{i=1}^m |(x^i - y^i)| \quad (3.19)$$

3.5.4 Fractional Distance

A drawback of using Euclidean and other p -norm distances where $p \in \mathbb{N}_1$ is that as the vectors get larger, the distance values tend to cluster. This is called the concentration phenomenon. To overcome this limitation of distance-based classification, Vivaracho-Pascual et. al. [68] introduced the use of fractional p -norm distances in their work on online signature verification.

The equation for determining fractional p -norm distance between vectors x and y is computed as

$$\min(||x - y||_p) = (\sum |(x - y)|^p)^{1/p} \quad (3.20)$$

where $0.1 \leq p \leq 2.0$.

The optimal value of p is when the distance calculated using Equation (3.20) is at its minimum for all values of p within the given range.

3.5.5 Ratio Distance

The equation for determining the Ratio distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is computed as defined in equation (3.21).

$$||x - y|| = \left(\sum_{i=1}^m \frac{\min(|x^i|, |y^i|)}{\max(|x^i|, |y^i|)} \right) \quad (3.21)$$

where \min and \max determine the minimum and maximum values between the two input parameters, respectively.

3.5.6 Mahalanobis Distance

The Mahalanobis distance was first documented in 1936 by Prasanta Mahalanobis [42]. This description of the distance is based on the algorithm and equations of Kardi Teknomo [64].

Given the set of feature vectors X , with a the number of components of each vector, and b its cardinality. X is represented as a matrix

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_b^1 \\ x_1^2 & x_2^2 & \dots & x_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^a & x_2^a & \dots & x_b^a \end{bmatrix}$$

and Y another set of feature vectors, with a the number of components of each vector, and d its cardinality. Y is represented as a matrix

$$Y = \begin{bmatrix} y_1^1 & y_2^1 & \dots & y_d^1 \\ y_1^2 & y_2^2 & \dots & y_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ y_1^a & y_2^a & \dots & y_d^a \end{bmatrix}$$

The i^{th} vector in X is $x_i = (x_i^1, x_i^2, \dots, x_i^a)$

and i^{th} vector in Y is $y_i = (y_i^1, y_i^2, \dots, y_i^a)$

The Mahalanobis distance is computed as follows.

The mean values, $\overline{x^i}$ and $\overline{y^i}$, for i^{th} components in subsets X and Y respectively, are defined in Equations Equations 3.22 and 3.23, respectively.

$$\overline{x^i} = \left(\frac{1}{b} \sum_{j=1}^b x_j^i \right), \quad i = 1, 2, \dots, a \quad (3.22)$$

and

$$\overline{y^i} = \left(\frac{1}{d} \sum_{j=1}^d y_j^i \right), \quad i = 1, 2, \dots, a \quad (3.23)$$

We then obtain $MD = (MD^1, MD^2, \dots, MD^a)$, where the mean difference, MD^i , is determined using \bar{x}^i and \bar{y}^i .

$$MD^i = (\bar{x}^i - \bar{y}^i), \quad i = 1, 2, \dots, a \quad (3.24)$$

The centered data matrices, \tilde{X} and \tilde{Y} , are determined by centering the data on the mean

$$\tilde{x}_i^j = (x_i^j - \bar{x}^j), \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b \quad (3.25)$$

$$\tilde{y}_i^j = (y_i^j - \bar{y}^j), \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, d \quad (3.26)$$

The covariance matrices, $COV(X)$ and $COV(Y)$, are calculated using the centred data matrices.

$$COV(X) = \frac{1}{b} \tilde{X}^T \times \tilde{X} \quad (3.27)$$

$$COV(Y) = \frac{1}{d} \tilde{Y}^T \times \tilde{Y} \quad (3.28)$$

The, $a \times a$, pooled covariance matrix, $PCM(X, Y)$, is determined using a weighted average of covariance matrices $COV(X)$ and $COV(Y)$ of X and Y , as follows

$$PCM(X, Y) = \frac{b}{b+d} COV(X) + \frac{d}{b+d} COV(Y) \quad (3.29)$$

The final Mahalanobis distance, M , is then determined by using the mean difference and the pooled covariance matrix.

$$M = \sqrt{MD^T \times PCM(X, Y)^{-1} \times MD} \quad (3.30)$$

3.5.7 Cosine Similarity Measure

The Cosine Similarity Measure is another distance-based technique which determines the similarity between two vectors, $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$, as defined in Equation 3.31.

$$\|x - y\| = \left(\frac{\sum_{i=1}^m x^i y^i}{\sqrt{\sum_{i=1}^m (x^i)^2} \times \sqrt{\sum_{i=1}^m (y^i)^2}} \right) \quad (3.31)$$

3.5.8 Canberra Distance

The equation for determining the Canberra distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is computed as defined in equation (3.32).

$$||x - y|| = \left(\sum_{i=1}^m \frac{|(x^i - y^i)|}{|x^i| + |y^i|} \right) \quad (3.32)$$

3.5.9 Weighted Euclidean Distance

The weighted Euclidean distance measure is a technique used in iris verification by Zhu et al. [75] to improve the classification accuracy by adding weight, or statistical importance, to the most reliable features from the feature vector. Firstly, the standard deviation for the reference signatures is obtained.

Let the n reference signatures, containing m features, be

$$\begin{aligned} x_1 &= (x_1^1, x_1^2, \dots, x_1^m) \\ x_2 &= (x_2^1, x_2^2, \dots, x_2^m) \\ &\vdots \\ x_n &= (x_n^1, x_n^2, \dots, x_n^m) \end{aligned} \quad (3.33)$$

Let x_i^j be the j^{th} component of the i^{th} reference signature where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Then the mean of the j^{th} component the reference signatures, μ^j , is computed as

$$\mu^j = \frac{1}{n} \sum_{i=1}^n x_i^j \quad (3.34)$$

and their standard deviation σ^j is defined as

$$\sigma^j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^j - \mu^j)^2} \quad (3.35)$$

The weighted Euclidean distance between 2 reference signatures x_k and x_l , for $1 \leq k, l \leq n$ is

$$||x - y||_p = \left(\sum_{j=1}^m \frac{|x_k^j - x_l^j|^p}{\sigma^j} \right)^{1/p} \quad (3.36)$$

where $p = 2$.

The equation can be rewritten as

$$||x - y||_2 = \left(\sum_{j=1}^m \frac{|x_k^j - x_l^j|^2}{\sigma^j} \right)^{1/2} \quad (3.37)$$

3.5.10 Weighted Manhattan Distance

The Manhattan distance and weighted Euclidean distance can be combined to form the novel weighted Manhattan distance

$$||x - y||_1 = \sum_{j=1}^m \frac{|x^j - y^j|}{\sigma^j} \quad (3.38)$$

where $x = (x^1, x^2, \dots, x^m)$ and $y = (y^1, y^2, \dots, y^m)$

3.5.11 Weighted Fractional Distance

The fractional distances and weighted Euclidean distance can then be combined to form the novel weighted fractional distance

$$\min(||x - y||_p) = \left(\sum_{j=1}^m \frac{|x^j - y^j|^p}{\sigma^j} \right)^{1/p} \quad (3.39)$$

where $x = (x^1, x^2, \dots, x^m)$, $y = (y^1, y^2, \dots, y^m)$, and $0.1 \leq p \leq 2.0$

As with Equation 3.20, the optimal value of p is when the distance calculated using Equation 3.39 is at its minimum for all values of p within the given range.

3.6 Experimental Methodology

All of the classification techniques described in Section 3.5 were tested with the DF, Energy, Ratio, Maxima, MDF and EMDF features, using thresholding as described in Section 3.5.1 and individually optimized resampling as described in Section 3.4.9. The results for each of the classification techniques were compared and the most successful classification techniques were then tested with the LDP feature extraction to verify the effectiveness of the classification techniques.

The GPDS300 signature database [66] was used in this work. It contains signatures from 300 individuals, with 24 authentic and 30 forgeries per individual. This gives a total of 16200 signatures for training and testing. For each individual 10 authentic signatures were used to create the reference signature, while the other 14 authentic signatures and 30 skilled forgeries were used for training and testing. Additionally, for each individual, authentic signatures from each of the other 299 individuals were used as random forgeries for testing.

3.7 Conclusion

An overview of the signature verification system was described. After this, each of the three main components of this system were discussed in further detail, namely, pre-processing, feature extraction and classification. Different preprocessing techniques are

used for each feature extraction technique. The local feature extraction techniques used, are the MDF and LDP. The global techniques used, are the DF, Energy, Ratio and Maxima features. The global techniques are combined with the MDF, to form the EMDF. The combination of local and global features aids in reduction of noise in a feature vector, Individually Optimized Resampling, which chooses the optimal feature vector size per individual, is also used. Various classification techniques are described. Each will be used to classify the extracted feature vectors and the accuracy of each classification technique will be compared. In the next section, results of the classification tests are discussed. These tests involved analyzing the accuracy of the various distance-based classification techniques with the feature extraction techniques.

Chapter 4

Results and Discussions

4.1 Introduction

This chapter presents the results achieved, as well as a critical analysis of these results. The data set and the thresholding used in the analysis are discussed. Then the results of tests with the Extended Modified Direction Feature (EMDF) with various distance-based classification techniques are covered. Further tests with the most successful distance-based classification techniques are applied to the Local Directional Pattern (LDP) feature vectors and their results are discussed. A comparison with results from literature is then performed.

4.2 Data Set

For the analysis of the techniques, signatures from the Grupo de Procesado Digital de Senales (GPDS) signature database were used [66]. It is among the largest and most commonly used publicly available offline signature database. The database consists of black and white signatures of multiple individuals, with 24 authentic copies and 30 skilled forgeries for each individual. A set of 300 individuals (GPDS300) was obtained from this database and used for the experiments. This gives a total of 16200 signatures in the database used, of which 7200 are authentic and 9000 are skilled forgeries.

Each signature's file had a naming format of c-xxx-yy.bmp where xxx is the number of the signer in the range 001-300 for GPDS300, and the yy is the repetition number of the signature in the range 01-24.

10 authentic signatures are used to create the reference signature, the other 14 authentic signatures and the 30 skilled forgeries are used for the classification and verification. For random forgeries per individual, a single authentic signature from each of the other 299 individuals is used.

4.3 Threshold Calculation

There are two possible types of thresholds, a global threshold and a localized threshold. With a global threshold, a single threshold is used for analysis and decision making

with all signatures. With a localized threshold, a different threshold is used for analysis and decision making with each individual's set of signatures. Schafer and Viriri [58] reports that a locally applied threshold performs better than a globally applied threshold. Therefore, localized thresholding is used in this work.

Quantification of result accuracy is measured in terms of the False Rejection Rate (FRR); False Acceptance Rate (FAR) which is further broken down into FAR for skilled forgeries (FARS) and FAR for random forgeries (FARR); and the Equal Error Rate (ERR). The FRR is a measure of false negatives, which is the percentage of authentic signatures that are classified forgeries. Conversely, the FAR is a measure of false positives, which is the percentage of forged signatures that are classified as authentic. The EER is the point at which the FRR and FARS converge. Figure 4.3 shows an example of the ROC curve for obtaining the EER.

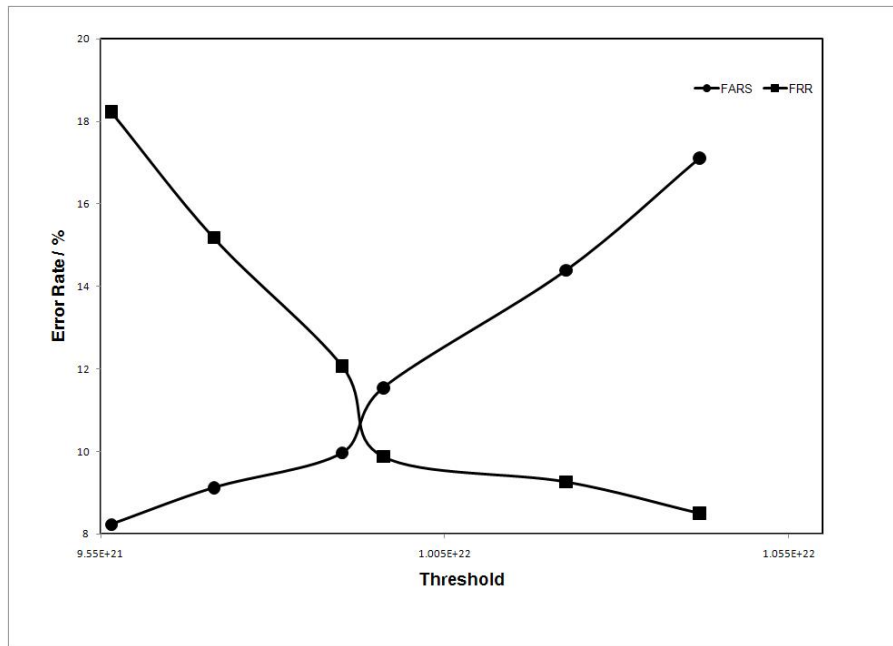


Figure 4.1: A ROC-curve showing the point of intersection between the FRR and FARS curves (obtained from classification of the EMDF using the weighted fractional distance and individually optimized resampling)

4.4 Tests with Extended Modified Direction Feature

Various distance-based classification techniques are tested with each of the global and local feature extractions separately. The global features are the Direction feature, Energy feature, and two variations each of the Maxima and Ratio features. The variations of the Maxima feature are $Maxima_T$, where preprocessing involved thinning; and $Maxima_{T_s}$, where preprocessing involves both thinning and segmentation. The variations of the Ratio feature are $Ratio_1$, as defined in Equation 3.9; and $Ratio_2$, as defined in Equation 3.10. The global features are the Modified direction feature with different feature vector sizes for differing levels of information extraction. Further, the most successful distance-based measures are then applied to the Extended MDF (EMDF), which is a concatenation

of the best global features and different feature vector sizes for the local features of the MDF. The feature vector sizes are varied through changing the resampling strip size and maximum transitions equally during the MDF feature extraction phase. In Tables 4.1 to 4.5, the variations in resampling are symbolizes with “rs x ”, where x is the resampling size used.

4.4.1 Euclidean Distance

Table 4.1 shows the results of Euclidean distance tests performed with various feature extraction techniques. The global feature extractions, when tested with the Euclidean distance, perform poorly in individual setting, with the best being the combination of energy and Ratio₁ features that provide an EER of 28.6% and FARR of 2.00%. While the DF provides a low EER of 6.61%, the FARR is worse than the ERR at 11.0%. The MDF feature extractions perform better with the best EER of 22.5% and corresponding lowest FARR of 1.10% occurring with a resampling of 7. While the EER improves for each increasing resampling size, this trend is broken by the MDF rs8 feature extraction. This is due to the concentration phenomenon [23], [68]. This phenomenon refers to the concentration of distances in p -norm space, also called L^P -space, as feature vectors become larger. It is generally most apparent will larger p -norms, such as the Euclidean distance, where the p -norm value is 2. The concentration of distances results in less accurate classification.

The EMDF performs much better as a whole in comparison to its constituent parts, which are the MDF and the global features. This is most pronounced in the 7.6% difference between the MDF and EMDF with a resampling strip size of 2 for both. Both the EER and FARR show improvement, confirming the report of Nguyen et. al. [48] that a combination of global and local feature extraction techniques enhances the classification. The best EER of 19.2% was obtained with EMDF with resampling of 3. The larger feature vectors perform worse, while some of the largest feature vectors perform worse than the smallest feature vector in terms of the EER. The poor performance of larger feature vectors can be attributed to the clustering of distances as described by the concentration phenomenon. This phenomenon is usually very apparent when the Euclidean distance is applied to large data sets.

4.4.2 Manhattan Distance

Table 4.1 shows the results of Manhattan distance tests performed with various feature extraction techniques. The global feature extractions, when tested with the Manhattan distance, perform poorly in individual setting, with the best being the combination of energy and Ratio₁ features that provide an EER of 27.1% and FARR of 1.50%. While the DF provides a low EER of 9.0%, the FARR is poor at 5.85%. The MDF feature extractions perform better with the best EER of 22.2% and corresponding FARR of 1.10% occurring with a resampling of 8. The EER improves for each increasing feature vector size. This suggests that the Manhattan distance is not affected by the concentration phenomenon as much as the Euclidean distance. While there is very little difference for the EER between the Euclidean and Manhattan distances for the MDF, the improvement of using the Manhattan distance is more pronounced when using the LDP feature extraction, discussed further down in this chapter.

Table 4.1: The Euclidean and Manhattan distances investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF

Feature Vectors	Euclidean Distance		Manhattan Distance	
	EER%	FARR%	EER%	FARR%
Direction Feature	6.61	11.0	9.0	5.85
Energy	29.7	8.29	27.8	0.00
Energy + Maxima _S	36.9	11.5	32.9	7.06
Energy + Maxima _T	34.1	9.74	30.6	6.21
Energy + Ratio ₁	28.6	2.00	27.1	1.50
Energy + Ratio ₂	29.6	5.85	27.9	4.53
MDF rs2	27.7	1.91	27.8	1.74
MDF rs3	25.5	1.39	25.4	1.25
MDF rs4	24.4	1.12	24.0	1.09
MDF rs5	23.3	1.26	23.1	1.02
MDF rs6	22.7	1.16	22.4	1.00
MDF rs7	22.5	1.10	22.3	1.06
MDF rs8	22.5	1.26	22.2	1.10
EMDF rs2	20.1	0.73	18.7	0.48
EMDF rs3	19.2	0.44	19.7	0.39
EMDF rs4	19.4	0.37	20.2	0.36
EMDF rs5	19.7	0.36	20.6	0.44
EMDF rs6	20.2	0.36	20.3	0.43
EMDF rs7	20.2	0.44	20.5	0.60
EMDF rs8	20.4	0.49	20.8	0.62

The EMDF performs much better as a whole in comparison to its constituent parts. The best EER in this group was 18.7% for the EMDF with a resampling of 2, with a corresponding FARR of 0.48%. This is better than the best EMDF with Euclidean distance by 0.5%, and is better than the best MDF with Manhattan distance by 3.5%. The FARR also improves through the concatenation of the global and local features.

The improvement trend is reversed for the EMDF, where the smallest feature vectors provide the best EER, while the largest MDF feature vectors provide the best EER in that group. This is due to combining the local and global features together, since the global features and smaller p -norm are both used to reduce the effect of non-Gaussian randomness within the feature vector. The global features have more impact in the smaller feature vectors. As the feature vectors become larger and their are a much greater number of data points, the impact of the small set of global features become insignificant.

4.4.3 Mahalanobis Distance

Table 4.2 shows the results of Mahalanobis distance tests performed with various feature extraction techniques. The Mahalanobis distance did not perform well, with the best EER of 37.2% and a corresponding FARR of 20.05%. In comparison with results in literature, where most EER results with skilled forgeries are below 20%, the Mahalanobis distance EER is too high. This high EER makes it a non-viable classification technique

candidate for the offline feature extraction techniques. It's poor performance is due to an asymmetric distribution of data in the feature vectors. The Mahalanobis distance performs best with multivariate normal data distributions [30].

Table 4.2: The Mahalanobis distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF

Feature Vectors	Mahalanobis Distance	
	EER%	FARR%
MDF rs2	41.7	23.39
MDF rs3	39.9	22.30
MDF rs4	38.9	21.56
MDF rs5	37.9	21.09
MDF rs6	37.4	20.66
MDF rs7	37.2	20.36
MDF rs8	37.2	20.05

4.4.4 Cosine Similarity Measure

Table 4.3 shows the results of the cosine similarity measure tests performed with various feature extraction techniques. The cosine similarity measure performed the worst from all of the distance-based classification techniques that were measured. The best EER for the global features was 56.7% for the combination of Energy and thinned Maxima feature extractions and the best for the local feature extractions was 61.6% when using resampling strip size of 2 with the MDF. This indicates that forgeries are accepted as authentic and authentic signatures are regarded as forgeries in over 60% of cases, in a best case scenario. The FARR performs even worse than the EER in all cases, where random forgeries are regarded as authentic in between 67% and 93% of case, depending on the feature extractions used. According to Liu et. al. [40], the cosine similarity measure works best for orthogonal data sets where each data point is independent of each other. Conversely, there is a high level of inter-dependence between features of a biometric feature vector and the feature vectors are not orthogonal.

4.4.5 Canberra Distance

Table 4.3 shows the results of the Canberra distance tests performed with various feature extraction techniques. The Canberra distance performed slightly better than the Mahalanobis distance in terms of the EER, and much better in terms of FARR. For the global feature extractions, the best EER was 25.5%, for a combination of Energy and Ratio₁ feature vectors, with a corresponding FARR of 1.03%. For the local feature vectors, the best was MDF with an resampling strip size of 6. This provides an EER of 26.1% and an FARR of 9.38%. The FARR is relatively low for small feature vectors but becomes very high as feature vectors become larger in size. This classification technique would not be good for LDP feature extraction technique, whose feature vectors are much larger in size. Additionally, even the low FARR values are high in comparison with the distances in L^p space, namely, the Euclidean, Manhattan and fractional distances.

Table 4.3: The cosine similarity measure, Canberra distance and Ratio distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF

Feature Vectors	Cosine Similarity		Canberra Distance		Ratio Distance	
	EER%	FARR%	EER%	FARR%	EER%	FARR%
Direction Feature	63.5	92.96	30.4	3.27	13.1	9.62
Energy	56.7	70.76	26.2	0.00	26.2	0.26
Energy+ Maxima _S	58.0	66.95	39.3	8.45	36.2	6.13
Energy+ Maxima _T	56.7	79.99	39.9	12.55	36.2	8.05
Energy+ Ratio ₁	63.0	65.02	25.5	1.03	25.6	1.09
Energy+ Ratio ₂	59.5	67.67	26.1	3.43	26.0	3.40
MDF rs2	61.6	87.26	29.8	2.29	29.7	2.35
MDF rs3	68.1	79.89	30.2	2.73	31.5	2.91
MDF rs4	69.9	78.53	27.9	3.12	32.8	3.16
MDF rs5	70.8	75.84	26.9	5.64	31.6	2.70
MDF rs6	71.5	73.37	26.1	9.38	31.2	2.35
MDF rs7	72.1	73.15	26.1	13.56	31.2	2.20
MDF rs8	72.5	73.34	26.2	17.69	30.6	1.83

4.4.6 Ratio Distance

Table 4.3 shows the results of the ratio distance tests performed with various feature extraction techniques. While the Ratio distance performs better than the Mahalanobis distance and cosine similarity measure, it also performs worse than the Canberra, Euclidean and Manhattan distances. Among the global feature extractions, only the DF performs well, with an EER of 13.1%. However, the corresponding FARR is very high at 9.62%, in comparison with other classification techniques applied to feature extraction techniques in this work. The best MDF, with an EER of 29.7%, is with a resampling of 2. This EER and its corresponding FARR of 2.35% are still too high. The high EER and FARR make it an inappropriate candidate for offline signature classification.

4.4.7 Fractional Distance

Table 4.4 shows the results of the fractional distance tests performed with various feature extraction techniques. Both the local MDF and global feature extractions, when tested with the fractional distance, performed better than all of the classification tests before it. This is due to locally optimizing the classification by choosing the best p -norm value per individual. Further, the fractional distances help to overcome the effect of the concentration phenomenon.

The best EER for the global features was 23.9% for both the energy feature and the concatenation of the energy and ratio₁ features. These provided a better EER than the Euclidean distance by 5.8% and 4.7% respectively. While the DF achieved an EER of 5.7%, the FARR was higher, at 10.78%. This is unacceptably high. The best EER for the local MDF features was 19.0%, which was achieved with the largest resampling size of 8.

The EMDF performs much better as a whole in comparison to its constituent parts. This is most pronounced in the 9.0% difference between the MDF and EMDF with resampling of 2 for both. Both the EER and FARR show improvement, confirming the report of Nguyen et. al. [48] that a combination of global and local feature extraction techniques enhances the classification. The best EER of 16.0% was obtained with EMDF with resampling of 2. The improvement trend is reversed for the EMDF, where the smallest feature vectors provide the best EER, while the largest MDF feature vectors provide the best EER in that group. This is attributed to an effect of combining the local and global features together, where the global features have more impact in the smaller feature vectors. As the feature vectors become larger, the impact of the small set of global features become insignificant. Additionally, the feature vectors are too small to utilize the full effect of the fractional distances in overcoming the concentration phenomenon.

Table 4.4: The fractional distance investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF

Feature Vectors	Fractional Distance	
	EER%	FARR%
Direction Feature	5.7	10.78
Energy	23.9	5.41
Energy + Maxima _S	26.2	6.00
Energy + Maxima _T	24.9	5.62
Energy + Ratio ₁	23.9	2.90
Energy + Ratio ₂	24.4	3.10
MDF rs2	25.0	1.96
MDF rs3	22.6	1.32
MDF rs4	21.5	1.07
MDF rs5	20.2	1.11
MDF rs6	19.6	0.98
MDF rs7	19.2	1.04
MDF rs8	19.0	1.16
EMDF rs2	16.0	0.60
EMDF rs3	16.5	0.45
EMDF rs4	17.1	0.42
EMDF rs5	17.3	0.38
EMDF rs6	17.4	0.40
EMDF rs7	17.5	0.52
EMDF rs8	17.6	0.68

4.4.8 Weighted Euclidean Distance

Table 4.5 shows the results of the weighted Euclidean distance tests performed with various feature extraction techniques. The weighted Euclidean distance is better than the Euclidean and Manhattan distances in all cases, and better than fractional distances in all the individual tests and some of the EMDF tests.

The best EER for the global features was 21.1% for the concatenation of the energy

Table 4.5: The weighted Euclidean, weighted Manhattan and weighted fractional distances investigated with individual global feature extraction techniques, the local MDF and concatenated EMDF

Feature Vectors	Weighted Euclidean		Weighted Manhattan		Weighted Fractional	
	EER%	FARR%	EER%	FARR%	EER%	FARR%
Direction Feature	4.0	19.50	7.0	7.62	3.9	7.89
Energy	21.9	1.69	23.0	1.81	20.6	2.24
Energy+ Maxima _S	21.8	1.60	23.1	1.82	20.4	2.35
Energy+ Maxima _T	21.1	1.89	22.3	2.02	19.6	2.45
Energy+ Ratio ₁	22.0	1.49	23.2	1.63	20.6	1.98
Energy+ Ratio ₂	22.1	1.46	23.2	1.70	20.6	2.04
MDF rs2	23.1	1.08	24.9	1.26	21.9	1.38
MDF rs3	20.3	0.56	22.0	0.72	19.1	0.81
MDF rs4	19.4	0.69	20.9	0.73	18.0	0.76
MDF rs5	20.0	1.16	20.2	0.97	17.7	0.90
MDF rs6	21.1	2.48	20.2	1.35	17.6	1.22
MDF rs7	21.5	4.11	20.2	2.02	17.5	1.46
MDF rs8	22.2	6.63	20.6	2.97	17.7	1.89
EMDF rs2	15.3	0.24	16.9	0.30	13.8	0.25
EMDF rs3	15.3	0.20	17.3	0.24	14.3	0.24
EMDF rs4	16.1	0.36	17.8	0.29	14.8	0.29
EMDF rs5	18.0	0.71	18.5	0.54	15.7	0.53
EMDF rs6	20.3	1.81	18.8	0.93	16.4	0.98
EMDF rs7	20.8	3.37	19.2	1.47	16.5	1.51
EMDF rs8	22.1	5.68	20.0	2.45	17.1	2.46

and Maxima_T features. This provided a better EER than the fractional distance for the same features by 3.8%. It also performed better than the best global feature vector from fractional distance classification by 1.8%. This feature vector was the concatenation of the energy and ratio₁ features which had the best EER for all of the previous global feature vector tests. While the DF achieved an EER of 4.0%, its FARR was higher, at 19.50%. This is unacceptably high. The best EER for the local MDF features was 19.4%, which was achieved with the resampling size of 4. This is slightly worse than the best MDF with fractional distance classification, of 19.0%, which was achieved with a resampling size of 8. It is also better than the best MDF with Euclidean distance classification by 3.1%.

The EMDF performs much better than any of its constituent parts perform individually. This is most pronounced in the 7.8% difference between the MDF and EMDF with resampling of 2 for both. Both the EER and FARR show improvement, confirming the report of Nguyen et. al. [48] that a combination of global and local feature extraction techniques enhances the classification. The best EER of 15.3% was obtained with EMDF with resampling of 3 and corresponding lowest FARR of 0.20%. As observed with the Euclidean, Manhattan and fractional distances, the improvement trend is reversed for the EMDF, where the smallest feature vectors provide the best EER, while the largest MDF feature vectors provide the best EER in that group. This is attributed to the combination

of the local and global features together.

While the EER shows great improvement, the FARR worsens with the using of the weighting by standard deviation. This is especially pronounced with the largest MDF and EMDF feature vectors. The reasoning is that because the weighting can not always differentiate between different sets of signatures, some features in random forgeries may be given extra weight if they are similar to authentic signatures. This is particularly visible with the largest feature vectors which will also may experience the effects of the concentration phenomenon as well.

4.4.9 Weighted Manhattan Distance

Table 4.5 shows the results of the weighed Manhattan distance tests performed with various feature extraction techniques. While the Manhattan distance performed better than the Euclidean distance, the weighted Manhattan distance performs worse than the Euclidean distance in an overall analysis. The weighed Manhattan distance performs better than the weighted Euclidean distance in the largest feature vectors only. The overall best weighted Manhattan distances in each group are still worse than those for the weighted Euclidean distance.

The best global feature vector is the concatenation of the energy and Maxima_T features with an EER of 22.3% and FARR of 2.02%. The corresponding EER and FARR for the weighted Euclidean distance is 21.1% and 1.89%. The best MDF EER with the weighed Manhattan distance is 20.2% whereas the best EER for the weighted Euclidean distance is 19.4%. The best EMDF EER with the weighed Manhattan distance is 16.9% whereas the best EER for the weighted Euclidean distance is 15.3%. While the EER for the weighted Manhattan distances is worse, the FARR improves for the larger feature vectors in comparison to the weighted Euclidean distance. This relative improvement for large feature vectors is due to smaller p -norm values experiencing less effect from the concentration phenomenon. As observed with the Euclidean, Manhattan, fractional and weighted Euclidean distances, the improvement trend is reversed for the EMDF, where the smallest feature vectors provide the best EER, while the largest MDF feature vectors provide the best EER in that group.

4.4.10 Weighted Fractional Distance

Table 4.5 shows the results of the weighted fractional distance tests performed with various feature extraction techniques. The weighted fractional distance performs better than all distance-based classification tests discussed previously.

The best global feature vector is the concatenation of the energy and Maxima_T features with an EER of 19.6% and FARR of 2.45%. This is better than the corresponding lowest weighted Euclidean distance for global features by 1.5%, although the FARR for the weighted fractional distance is higher. The best EER for the local MDF features was 17.5%, which was acheived with the resampling size of 7. This is better than the best MDF with weighted Euclidean distance classification, of 19.4%, which was achieved with a resampling size of 4. It is also better than the best MDF with fractional distance classification by 1.7% and better than the best MDF with Euclidean distance classification by 5.0%.

The EMDF performs much better than either the MDF or the local features individually, due to the combination of both. This is most pronounced in the 8.1% difference between the MDF and EMDF with resampling of 2 for both. While the FARR improves from the smaller feature vectors, it becomes worse for the largest feature vector, although it is still better than the FARR for the weighted Euclidean distance. This may be an effect of the weighting function. The best EER of 13.8% was obtained with EMDF with resampling of 2 and corresponding FARR of 0.25%.

As observed with the Euclidean, Manhattan, fractional, weighted Euclidean and weighted Manhattan distances, the improvement trend is reversed for the EMDF, where the smallest feature vectors provide the best EER, while the largest MDF feature vectors provide the best EER in that group.

4.4.11 Individually Optimized Resampling

Table 4.6 shows the results of individual optimization for different types of classification with the EMDF.

When dynamic individually optimized resampling was applied to the Euclidean distance an EER of 15.3% was obtained. This improves the the EER of the Euclidean distance applied to the EMDF without individually optimized resampling by up to 5.1% in comparison to the worst Euclidean distance EER, which was for a resampling size of 8. It is also better than the best Euclidean distance, which was far a resampling size of 3, by 3.9%.

With individually optimized resampling, the Manhattan distance becomes 15.4%. While this is slightly worse than the optimized EER for the Euclidean distance, it is still an improvement over the non-optimized Manhattan distance by up to 5.0%, in the case of the worse Manhattan distance resampling size, and by 3.3% in the case of the best Manhattan distance, which had a resampling size of 2.

The fractional distances, with individually optimized resampling, obtains an EER of 12.9%. This is better in comparison to the Euclidean distance with individually optimized resampling by 2.4%. It is also better than the best non-optimized fractional distance by 3.1% and better than the worse by 4.7%.

The weighted Euclidean distance with individually optimized resampling, obtains an EER of 11.7%. This is better than the EER of the unweighted Euclidean distance with optimized re-sampling by 3.6% and better than the best weighted Euclidean distance without resampling optimization by 3.6% as well. It is also better than the worst weighted Euclidean distance without resampling optimization by 10.4%. It is also better than individually optimized fractional distance by 1.2%.

With individual optimized resampling, the weighted Manhattan distance performs better than the unweighted Manhattan distance but worse than the weighted Euclidean distance and on par with the un weighted fractional distance. It performs better than the worst weighted Manhattan distance without resampling optimization by 7.1% and better than the best by 4.0%.

The final analysis for the EMDF was of the weighted fractional distances with individually optimized resampling, which combines the weighted Euclidean distance, fractional distance and resampling optimization concepts. It provided the best accuracy from

all of the individually optimized resampling tests. By combining these weighted fractional distances with individually optimized re-sampling, an EER of 10.8% was obtained. This is better than all of the previous combinations, including the fractional and weighted Euclidean distances. It is also better than the best weighted fractional distance without resampling optimization by 3.0% and better than the worst by 6.3%.

In all cases, the dynamic re-sampling combination provides a much higher accuracy than any individual static resampling size. Thus, the greatest accuracy was achieved by a combination of weighted distances, fractional distances and dynamic resampling with an EER of 10.8%. This is more accurate by almost 10% in comparison to any non-optimized resampling with a standard Euclidean distance measure.

Table 4.6: Individually optimized resampling investigated with the EMDF

Classification	EER%	FARR%
Euclidean	15.3	0.54
Manhattan	15.4	0.55
Fractional	12.9	0.57
Weighted Euclidean	11.7	0.71
Weighted Manhattan	12.9	0.69
Weighted Fractional	10.8	0.67

4.5 Tests with Local Directional Pattern

After the success of weighted fractional distances and individually optimized resampling with the EMDF feature extraction, further tests were performed with a different feature extraction technique, namely, the Local Directional Pattern. Tables 4.7 to 4.18 show the results of tests using different numbers of horizontal and vertical splits($split_H$ and $split_V$ respectively), with various different classification techniques. The feature vector sizes are determined by the $split_H \times split_V \times H$ where H is the length of the histogram, which is always 56 in the LDP extraction technique. Splits between 1 (no split) and 8 were tested. Further splits were omitted due to the very large vector size's adverse effect on processing speed. The same experimental setup was repeated for each of the Euclidean, Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances.

4.5.1 Euclidean Distance

Table 4.7 shows the EER results using the Euclidean distance calculations as defined in Equation 3.18. The EER improved with increasing feature vector sizes at first, since more data points allow for better classification. The smallest feature vector, from 1×1 splits, provided a high EER of 25.0%. This signifies a poor accuracy. As feature vector size increased, the EER improved to a best of 21.7% at splits 2×3 . However, due to the concentration phenomenon, which causes distance values to cluster, the EER worsened for feature vectors larger than this, while some of the largest feature vectors resulted in worse accuracy than the smallest feature vector. The worst EER of 26.0% was obtained with the largest feature vector which had splits of 8×8 .

A similar trend was observed with the FARR, in Table 4.8, obtained using the Euclidean distance calculations, with the best FARR in the same feature vector size vicinity as the best EER, and the worst FARR correlates with the worst EER. Once again this is attributed to the concentration phenomenon, which makes the differentiation between authentic and forged classes difficult with large feature vectors. The distance measures discussed below are used to counter the effects of the concentration phenomenon.

Table 4.7: The effect of different $split_H$ and $split_V$ on EER(%) using the Euclidean distance

$\begin{matrix} \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	25.0	23.9	22.6	23.0	22.6	22.8	23.0	23.0
2	23.8	22.3	21.7	22.0	22.0	22.3	22.4	22.6
3	24.0	22.7	21.9	21.9	22.2	22.3	22.6	22.8
4	24.1	23.0	22.5	22.5	22.8	23.0	23.1	23.5
5	24.3	23.7	23.1	23.2	23.5	23.6	23.9	24.3
6	24.8	23.9	23.4	23.7	24.0	24.2	24.5	25.1
7	25.2	24.4	24.2	24.1	24.5	25.0	25.3	25.4
8	26.0	25.2	24.7	25.0	25.2	25.6	25.7	26.0

Table 4.8: The effect of different $split_H$ and $split_V$ on FARR(%) using Euclidean distances

$\begin{matrix} \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	1.70	0.96	0.60	0.58	0.55	0.55	0.59	1.35
2	1.09	0.48	0.31	0.34	0.34	0.35	0.41	0.45
3	0.85	0.43	0.28	0.33	0.37	0.38	0.44	0.54
4	0.72	0.43	0.35	0.37	0.42	0.54	0.59	0.72
5	0.81	0.52	0.42	0.53	0.61	0.80	0.88	1.09
6	0.89	0.57	0.52	0.68	0.83	1.02	1.28	1.52
7	0.97	0.77	0.73	0.94	1.20	1.53	1.84	2.26
8	1.02	0.90	0.92	1.19	1.50	1.89	2.24	2.87

4.5.2 Manhattan Distance

Table 4.9 shows the EER results using the Manhattan distance calculations as defined in Equation 3.19. The best EER of 19.2% was with splits 3×3 , which gave a feature vector slightly larger than the Euclidean distance's best EER with 2×3 splitting. However, while the largest feature vectors did not provide the best EER, they still provided a better EER than the smallest feature vectors. This is in contrast to the Euclidean distance where the largest feature vectors resulted in a worse EER than the smallest feature vectors. This shows that p -norm distance measures other than the Euclidean distance can provide a better result when feature vectors are larger and information is greater.

The FARR, in Table 4.10, showed a similar trend, where the best FARR of 0.18% was also with splits of 3×3 . While the FARR of larger feature vectors increased in relation to

the best FARR, they were still better than the FARR of the smallest feature vector. This further advocates the use of p -norm distances other than the Euclidean distance.

Table 4.9: The effect of different $split_H$ and $split_V$ on EER(%) using the Manhattan distance

H \ V	1	2	3	4	5	6	7	8
1	23.3	21.6	20.4	20.3	20.2	20.1	20.2	20.2
2	21.7	20.2	19.3	19.7	19.4	19.7	19.8	20.0
3	21.5	20.0	19.2	19.4	19.3	19.4	19.5	19.7
4	21.4	19.9	19.6	19.5	19.8	19.5	19.6	20.1
5	21.6	20.2	19.6	19.8	19.9	20.1	20.1	20.3
6	21.5	20.6	20.0	20.3	20.2	20.4	20.6	20.8
7	21.9	20.6	19.8	20.1	20.5	20.5	20.8	21.0
8	22.0	21.0	20.2	20.5	20.7	20.9	20.9	21.2

Table 4.10: The effect of different $split_H$ and $split_V$ on FARR(%) using Manhattan distances

H \ V	1	2	3	4	5	6	7	8
1	1.39	0.77	0.46	0.43	0.39	0.43	0.37	0.89
2	0.79	0.34	0.24	0.24	0.23	0.23	0.28	0.30
3	0.54	0.22	0.18	0.20	0.23	0.24	0.29	0.32
4	0.45	0.22	0.22	0.23	0.26	0.32	0.36	0.44
5	0.47	0.25	0.24	0.27	0.37	0.45	0.54	0.60
6	0.45	0.25	0.30	0.34	0.44	0.55	0.64	0.79
7	0.47	0.30	0.31	0.43	0.55	0.71	0.93	1.08
8	0.50	0.35	0.38	0.50	0.67	0.87	1.10	1.33

4.5.3 Fractional Distance

Figure 4.2 shows the effect of globally applied fractional distances on small and large feature vectors. The small feature vector that was used in Figure 4.2 was obtained from feature extraction with no splitting (1 block), while the large feature vector was obtained from feature extraction with 7×7 splitting (49 blocks). While fractional distances provide an improvement over the Euclidean distance with the small feature vector, the effect is much more pronounced with the large feature vector. The best EER obtained with the small feature vector was with $p = 0.4$. This correlates with the observation of Vivaracho-Pascual et al. [68], where the best EER was also obtained with $p = 0.4$. Accuracy then sharply worsens and $p = 0.1$ gave a worse EER than the Euclidean distance. Conversely, when the feature vector is very large, the Euclidean distance provided a slightly worse EER in comparison to the small feature vector, but the EER greatly increased when $p < 2.0$. The best EER for the large feature vector occurred with $p = 0.1$, which was 5.8% better than the best globally applied fractional distance for the smaller feature vector. This discrepancy with Vivaracho-Pascual et al. is due to their use of insufficiently

large feature vectors. Additionally, the best EER when the fractional distance was applied to the large feature vector was 8.8% better than the associated Euclidean distance and 11.2% better than the worst fractional distance with the small feature vector. This proves that the fractional distances do counter the concentration effect that occurs especially with large feature vectors.

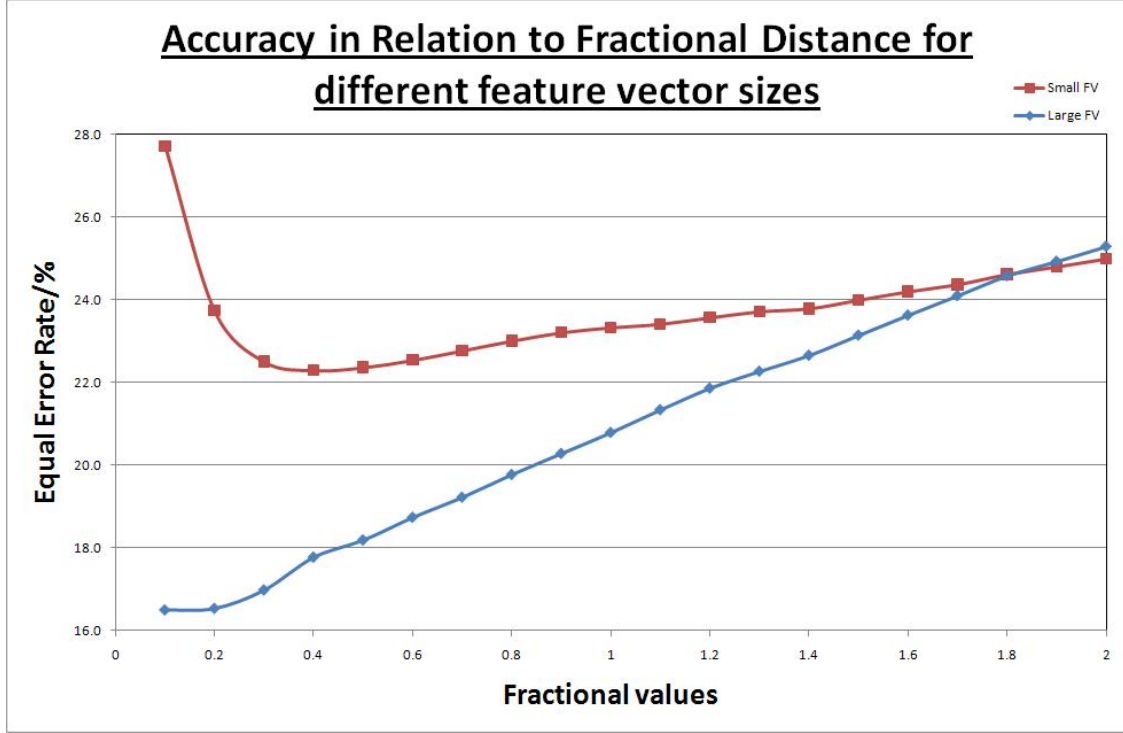


Figure 4.2: EER in Relation to individual fractional distances for large and small feature vector sizes

The fractional distances were then applied with local optimization. Table 4.11 shows the EER results using the locally optimized fractional distance calculations as defined in Equation 3.20. This means that the best fractional distance within the range $0.1 \leq p \leq 2.0$ was chosen per individual, since some fractional distances work better than others for different individuals. The best EER of 14.7% was with one of the largest feature vectors, 7×8 . This EER was better than the EER of smallest feature vector, using fractional distances, by 4.2% and it was better than the best Euclidean distance by 7.0%. This further shows that the fractional distances provide a solution to the concentration phenomenon that occurs with the Euclidean distance, particularly in the case of offline signature verification. There was a trend of improvement as feature vectors become larger. The fractional distances reached their best improvement in overcoming the concentration phenomenon between splits 7×7 and 8×8 , which correlated with feature vector sizes between 2744 features and 3584 features. There was also an improvement over globally applied fractional distance, by 1.8%.

The FARR, in Table 4.12, showed a slightly different trend, where the best FARR did not correlate with the best EER. This is due to choosing the best p -norm value locally for each individual, rather than a globally used p -norm as was the case of the Euclidean and Manhattan distances. Due to this localized optimization of the fractional distances,

larger p -norm values were used for the best EER in some cases, resulting in a worsening of the concentration phenomenon for larger feature vectors. However, the FARR for fractional distances was still better than the FARR for Euclidean and Manhattan distances, especially in the case of larger feature vectors.

Table 4.11: The effect of different $split_H$ and $split_V$ on EER(%) using the fractional distance

H \ V	1	2	3	4	5	6	7	8
1	18.9	17.7	17.0	16.9	16.7	16.6	16.6	16.7
2	17.9	15.0	16.2	16.1	16.1	15.9	16.0	16.0
3	17.5	16.4	15.6	15.7	15.4	15.2	15.2	15.5
4	17.5	16.2	15.9	15.6	15.4	15.3	15.3	15.2
5	17.5	16.3	15.3	15.4	15.3	15.2	15.0	15.2
6	17.7	16.1	15.7	15.4	15.2	15.0	14.9	14.9
7	17.4	16.2	15.4	15.3	15.0	14.8	14.7	14.7
8	17.7	16.5	15.5	15.3	15.3	14.8	14.7	14.7

Table 4.12: The effect of different $split_H$ and $split_V$ on FARR(%) using fractional distance

H \ V	1	2	3	4	5	6	7	8
1	1.59	0.95	0.53	0.50	0.50	0.48	0.45	0.52
2	1.02	0.48	0.40	0.37	0.38	0.44	0.51	0.52
3	0.70	0.33	0.32	0.30	0.35	0.32	0.42	0.43
4	0.57	0.31	0.37	0.39	0.34	0.40	0.45	0.51
5	0.59	0.34	0.39	0.31	0.45	0.47	0.54	0.57
6	0.54	0.36	0.44	0.44	0.45	0.51	0.54	0.73
7	0.52	0.42	0.41	0.47	0.51	0.66	0.77	0.78
8	0.53	0.54	0.50	0.51	0.60	0.70	0.76	0.89

4.5.4 Weighted Euclidean Distance

Table 4.13 shows the EER results obtained when using the weighted Euclidean distance calculations, as defined in Equation 3.37. The best EER of 14.5% was with splits 4×5 , which gave a feature vector that was larger than the feature vector that was obtained with the best unweighted Euclidean and Manhattan distances. The weighted Euclidean distance resulted in a better EER than the Euclidean and Manhattan distances, and of an almost equal EER in comparison with the fractional distance. The low EER is due to extra weight, or importance, that was given to the most reliable features in the feature vector, which had the smallest intra-class difference per individual. This weighting also counteracted the clustering of the concentration phenomenon to a small extent. This was emphasized further since the largest feature vectors had a lower EER than the smallest feature vectors.

In contrast, the FARR, in Table 4.14, showed a different trend, where the FARR worsened as the feature vectors became larger. This was because the weighting function can not always differentiate between different sets of signatures and some features in random forgeries were given extra weight if the feature was similar to authentic signatures. This was particularly visible with the largest feature vectors which also experienced the effects of the concentration phenomenon.

Table 4.13: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Euclidean distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	18.6	17.9	16.8	16.5	16.1	16.0	16.0	16.0
2	18.4	16.5	15.8	15.7	15.4	15.3	15.2	15.2
3	17.3	16.0	15.5	15.3	14.6	14.8	14.6	14.9
4	17.4	15.7	15.2	15.1	14.5	14.8	14.8	15.0
5	17.3	15.6	14.7	14.8	14.6	14.6	14.7	15.1
6	17.0	15.7	14.8	15.0	14.7	14.8	14.7	14.8
7	16.9	15.5	14.8	14.9	15.0	14.7	14.9	15.1
8	17.2	15.8	15.1	15.3	15.2	14.9	15.1	15.3

Table 4.14: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Euclidean distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	0.98	0.65	0.50	0.48	0.60	0.64	0.71	1.00
2	0.69	0.51	0.69	0.96	1.08	1.44	1.59	1.85
3	0.55	0.62	0.84	1.04	1.32	1.48	1.53	1.80
4	0.54	0.68	0.93	1.26	1.46	1.69	1.84	1.94
5	0.68	0.87	1.33	1.44	1.79	1.86	2.10	2.22
6	0.77	1.14	1.51	1.66	1.93	2.05	2.17	2.42
7	0.81	1.39	1.54	1.82	1.95	2.19	2.37	2.66
8	0.91	1.35	1.70	1.90	2.08	2.39	2.50	2.92

4.5.5 Weighted Manhattan Distance

Table 4.15 shows the EER results using the weighted Manhattan distance calculations as defined in Equation 3.38. The best EER of 14.3% was obtained with splits 7×7 , which gave a feature vector that was larger than those which were obtained with the best unweighted Euclidean, Manhattan and weighted Euclidean distances. The best weighted Manhattan distance feature vector was of equal size in comparison to the fractional distance. The larger feature vector size provided a better EER due to a combination of a smaller p -norm distance and the weighting by standard deviation, both of which counteracted the concentration phenomenon. Due to this combination of weighting and a smaller p -norm, the weighted Manhattan distance provided a better EER than was obtained with the Euclidean, Manhattan, fractional and weighted Euclidean distances.

The FARR, in Table 4.16, showed a similar trend to the FARR of the weighted Euclidean distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a smaller p -norm value, the largest feature vectors had a smaller FARR in comparison with the largest feature vectors of the weighed Euclidean distance.

Table 4.15: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Manhattan distance

H \ V	1	2	3	4	5	6	7	8
1	19.4	18.6	17.3	17.3	17.0	16.6	16.7	16.7
2	19.3	17.2	16.6	16.3	15.9	15.6	15.5	15.6
3	18.3	16.5	15.8	15.5	15.1	15.0	14.7	14.7
4	18.1	16.2	15.4	15.3	14.8	14.8	14.5	14.7
5	18.0	16.2	15.0	14.9	14.8	14.7	14.7	14.7
6	17.7	16.1	15.1	15.1	14.9	14.6	14.5	14.7
7	17.4	16.0	15.0	15.0	14.5	14.3	14.3	14.4
8	18.0	16.3	15.2	15.1	14.7	14.6	14.5	14.6

Table 4.16: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Manhattan distances

H \ V	1	2	3	4	5	6	7	8
1	0.91	0.56	0.36	0.35	0.31	0.36	0.38	0.50
2	0.63	0.32	0.37	0.48	0.52	0.59	0.70	0.86
3	0.44	0.32	0.36	0.44	0.50	0.66	0.74	0.85
4	0.36	0.32	0.44	0.53	0.62	0.74	0.81	0.95
5	0.39	0.42	0.60	0.67	0.91	0.96	1.11	1.26
6	0.42	0.46	0.70	0.80	1.01	1.16	1.29	1.50
7	0.43	0.53	0.77	0.96	1.02	1.31	1.52	1.71
8	0.47	0.59	0.82	1.01	1.23	1.44	1.70	1.94

4.5.6 Weighted Fractional Distance

Following the success of the weighted Manhattan distance, the weighted fractional distances were tested. The weighted fractional distance combines the weighting function of the weighted Euclidean distance with the fractional distances. Table 4.17 shows the EER results using the weighted fractional distance calculations as defined in Equation 3.39. This provided the best EER of 12.2%, at splits 7×7 , which is a better EER than obtained from any of the classification techniques tested on the LDP so far. Once again, the use of fractional distances and weighting allowed greater accuracy with larger feature vectors. By combining two techniques that are individually better than the Euclidean distance, an overall much better accuracy was achieved. The worst EER for the weighted fractional distance was 16.83%, which used no splitting of the image. This was nearly 5% better than the best results for the standard Euclidean distance. Conversely, the best EER was

9.5% better than the best recorded Euclidean distance, and 2.5% and 2.7% better than the fractional and weighted Euclidean distances respectively.

The FARR, in Table 4.18, showed a similar trend to the FARR of the weighted Euclidean distance and the weighted Manhattan distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a wide range of p -norm values, the FARR values are better than the FARR for the Euclidean distance, but slightly worse than for the Manhattan distance.

Table 4.17: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted fractional distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	16.8	16.1	15.1	14.8	14.5	14.4	14.2	14.6
2	16.7	15.0	14.2	14.0	13.7	13.4	13.3	13.4
3	15.7	14.3	13.6	13.2	12.7	12.6	12.5	12.6
4	15.8	14.0	13.3	13.0	12.5	12.7	12.4	12.5
5	15.3	14.1	12.8	12.7	12.5	12.3	12.3	12.5
6	15.1	13.9	12.9	12.7	12.6	12.5	12.3	12.4
7	15.0	13.8	12.9	12.7	12.5	12.3	12.2	12.4
8	15.4	13.9	13.0	13.0	12.6	12.5	12.5	12.6

Table 4.18: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted fractional distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	1.08	0.70	0.47	0.42	0.52	0.57	0.57	0.72
2	0.65	0.48	0.54	0.84	0.81	0.93	1.22	1.55
3	0.54	0.42	0.66	0.75	0.75	0.81	1.09	1.33
4	0.56	0.53	0.23	0.91	1.00	1.11	1.14	1.34
5	0.63	0.79	1.00	1.02	1.20	1.28	1.28	1.61
6	0.68	0.76	1.16	1.08	1.34	1.36	1.45	1.67
7	0.68	0.88	1.18	1.36	1.44	1.63	1.64	1.78
8	0.66	1.04	1.21	1.42	1.53	1.84	1.84	2.11

4.5.7 Individually Optimized Resampling

Tables 4.19 to 4.30 show the test results for individually, or locally, optimized resampling performed in conjunction with each of the above 6 discussed distance-based classification techniques, namely, the Euclidean, Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances. Optimizations from 4 split and 9 split combinations were tested. These are symbolized in the following tables with the [x:y] notation. For example, [1:2] represents the 4 combinations of 1×1 , 1×2 , 2×1 and 2×2 ; while [1:3] represents the 9 combinations of 1×1 , 1×2 , 1×3 , 2×1 , 2×2 , 2×3 , 3×1 , 3×2 and 3×3 . In each case, the 9 combinations for individual optimizations performed better than the 4 combinations.

4.5.7.1 Euclidean distance

The best EER for 4-combination individually optimized resampling on the Euclidean distance was 18.4% from the [2:3] combination, in Table 4.19, and the best for the 9-combination was 16.5% from the [1:3] combination, in Table 4.20. The optimal combinations of feature vector sizes correlated with the lowest single Euclidean distance EER, of 21.7%, from splits 2×3 when no individual optimization was performed. Both combinations provided better EER than non-optimized Euclidean and Manhattan distances. The FARR values also improved through the individually optimized resampling. This shows a correlation between individuals with a low EER and a low FARR.

Table 4.19: The effect of implementing individually optimized resampling by choosing the best Euclidean distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	19.0	0.97
[2:3]	18.4	0.39
[3:4]	18.8	0.29
[4:5]	20.0	0.47
[5:6]	21.0	0.71
[6:7]	22.1	1.20
[7:8]	23.2	2.00

Table 4.20: The effect of implementing individually optimized resampling by choosing the best Euclidean distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	16.5	0.67
[2:4]	17.1	0.31
[3:5]	17.7	0.34
[4:6]	19.0	0.55
[5:7]	20.1	0.80
[6:8]	21.4	1.36

4.5.7.2 Manhattan distance

The best EER for 4-combination individual optimization on the Manhattan distance was 16.2% from the [2:3] combination, in Table 4.21, and the best for the 9-combination was 14.3% from the [1:3] combination, in Table 4.22. The optimal combinations of feature vector sizes correlated with the lowest single Manhattan distance EER, of 19.2%, from splits 3×3 when no individual optimization was performed. The 4-combination provided better EER than non-optimized Euclidean and Manhattan distances and the locally optimized Euclidean distance. Conversely, the 9-combination provided an EER better than or equal to all of the non-locally optimized distance-based measures, except for the weighted fractional distance whose best EER was 12.2%. The FARR values also improved through the individually optimized resampling.

Table 4.21: The effect of implementing individually optimized resampling by choosing the best Manhattan distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	17.1	0.74
[2:3]	16.2	0.24
[3:4]	16.4	0.18
[4:5]	17.1	0.24
[5:6]	17.7	0.40
[6:7]	18.4	0.65
[7:8]	19.0	0.96

Table 4.22: The effect of implementing individually optimized resampling by choosing the best Manhattan distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	14.3	0.49
[2:4]	14.9	0.21
[3:5]	15.3	0.21
[4:6]	16.3	0.30
[5:7]	17.1	0.45
[6:8]	17.8	0.69

4.5.7.3 Fractional distance

The worst EER for 4-combination individual optimization on the fractional distances was 12.7% from the [6:7] combination, in Table 4.23. Conversely, the best for the 9-combination was 11.3% from the [1:3] combination and worst was 11.8% in the [6:8] combination, in Table 4.24. The 4-combination provided better EER than all but the non-optimized weighted fractional distance. Conversely, the 9-combination provided an EER better than all measures tested before it, including the the weighted fractional distance, whose best EER was 12.2%. The FARR values also showed an improvement through the individually optimized resampling.

While the optimal combinations of feature vector sizes correlated with the lowest single fractional distance EER in the 4-combination, this is not the case with the 9-combination, even though the best and worst EER are very close, with only a 0.5% difference between them. This discrepancy is explained by Figure 4.3. In the worse case scenario, with the combination [6:8], the majority occurring p -norm was 0.1. This was larger than the second highest occurrence ($p = 0.2$) by over 40%. Conversely, for the best case scenario of [1:3], the 5 highest occurring distances were all within a 10% range of each other, and were the 5 smallest p -norm values. This allowed a higher accuracy, different p -norm values were occasionally better for difference individuals among smaller feature vectors. However, with the largest feature vectors, the p -norm value of 0.1 outperformed all others. A combination of smaller feature vectors with more variable p -norm values can therefore perform slightly better than a large feature vector with a single dominant p -norm value.

Table 4.23: The effect of implementing individually optimized resampling by choosing the best fractional distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	13.6	0.80
[2:3]	13.1	0.29
[3:4]	12.9	0.26
[4:5]	13.0	0.31
[5:6]	12.9	0.36
[6:7]	12.7	0.50
[7:8]	12.8	0.77

Table 4.24: The effect of implementing individually optimized resampling by choosing the best fractional distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	11.3	0.57
[2:4]	11.7	0.31
[3:5]	11.7	0.29
[4:6]	11.8	0.36
[5:7]	11.9	0.45
[6:8]	11.8	0.62

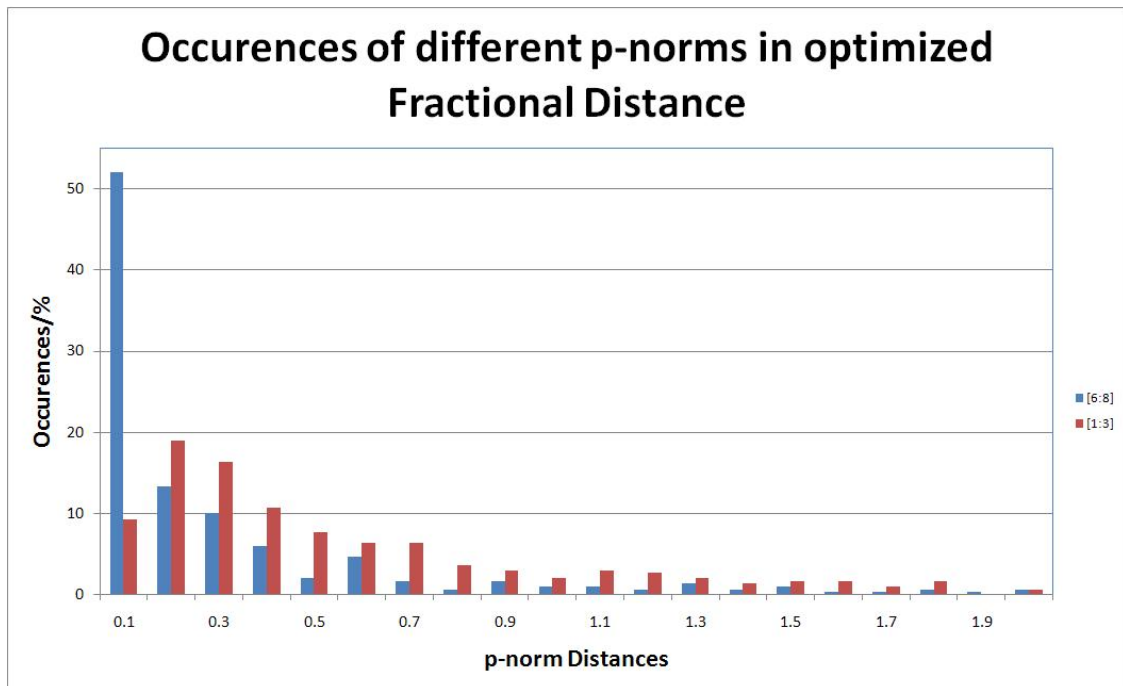


Figure 4.3: The number of occurrences of each p-norm in the fractional distance with individually optimized resampling combinations [1:3] and [6:8]

4.5.7.4 Weighted Euclidean distance

The best EER for 4-combination individual optimization on the weighted Euclidean distance was 12.1% from the [4:5] combination, in Table 4.25, and the best for the 9-combination was 10.7% from the [3:5] combination, in Table 4.26. The optimal combinations of feature vector sizes correlated with the lowest single weighted Euclidean distance EER, of 14.5%, from splits 4×5 where no individual optimized resampling was performed. Both the 4-combination and 9-combination provided better EER than previous distance-based classification measures. This includes the best weighted fractional distance of 12.2% without individually optimized resampling and all of the previously tested individually optimized resamplings of the LDP. The FARR values correlated closely with those for the weighted Euclidean distance without individually optimized resampling.

Table 4.25: The effect of implementing individually optimized resampling by choosing the best weighted Euclidean distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	13.5	0.70
[2:3]	12.5	0.60
[3:4]	12.2	0.93
[4:5]	12.1	1.46
[5:6]	12.2	1.73
[6:7]	12.6	2.10
[7:8]	13.1	2.38

Table 4.26: The effect of implementing individually optimized resampling by choosing the best weighted Euclidean distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	11.0	0.70
[2:4]	10.8	0.72
[3:5]	10.7	1.18
[4:6]	11.1	1.59
[5:7]	11.2	1.88
[6:8]	11.8	2.20

4.5.7.5 Weighted Manhattan distance

The best EER for 4-combination individual optimization on the weighted Manhattan distance was 12.6% from the [4:5] combination, in Table 4.27, and the best for the 9-combination was 11.4% from the [3:5] combination, in Table 4.28. The optimal combinations of feature vector sizes correlated with the lowest single weighted Manhattan distance EER, of 14.5%, from splits 7×7 where no individual optimized resampling was

performed. Both the 4-combination and 9-combination provided better EER than previous distance-based classification measures, where no individually optimized resampling was performed. However, the individually optimized weighted Manhattan distance performed worse than the individually optimized weighted Euclidean distance for both the 4- and 9-combinations. This is because the standard deviation used for the weighting equation was calculated in p -norm space of 2 while the Manhattan distance was calculated in p -norm space of 1. The FARR values correlated closely with those for the weighted Manhattan distance without individually optimized resampling.

Table 4.27: The effect of implementing individually optimized resampling by choosing the best weighted Manhattan distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	14.0	0.54
[2:3]	13.3	0.31
[3:4]	12.8	0.38
[4:5]	12.6	0.60
[5:6]	12.8	0.93
[6:7]	12.7	1.18
[7:8]	12.9	1.64

Table 4.28: The effect of implementing individually optimized resampling by choosing the best weighted Manhattan distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	11.5	0.50
[2:4]	11.8	0.34
[3:5]	11.4	0.48
[4:6]	11.6	0.68
[5:7]	11.9	1.06
[6:8]	12.0	1.27

4.5.7.6 Weighted fractional distance

The best EER for 4-combination individual optimization on the weighted fractional distance was 10.3% from the [4:5] combination, in Table 4.29, and the best for the 9-combination was 9.25% from the [3:5] combination, in Table 4.30. The best EER for both the 4- and 9-combinations of the individually optimized weighted fractional distance performed better than all other test before it, while the 9-combination performed better than the 4-combination. This high accuracy is due to choosing both the best p -norm and best split size per individual, i.e. locally optimized classification.

The optimal split sizes for the individually optimized fractional distance did not correlate with the optimal split sizes for the non-locally optimized weighted fractional distance, although there was a mere difference of 0.75% between the best and worst EER.

Table 4.29: The effect of implementing individually optimized resampling by choosing the best weighted fractional distance EER from 4 feature vectors

Split Range	EER/%	FARR/%
[1:2]	12.0	0.69
[2:3]	11.2	0.52
[3:4]	10.7	0.66
[4:5]	10.3	0.95
[5:6]	10.5	1.19
[6:7]	10.6	1.37
[7:8]	10.8	1.60

Table 4.30: The effect of implementing individually optimized resampling by choosing the best weighted fractional distance EER from 9 feature vectors

Split Range	EER/%	FARR/%
[1:3]	9.65	0.60
[2:4]	9.66	0.49
[3:5]	9.25	0.80
[4:6]	9.50	0.97
[5:7]	9.60	1.18
[6:8]	10.0	1.36

An analysis of Figure 4.4 showed that the pattern for the highest EER, from combination [6:8], was similar to that for the highest EER for the individually optimized fractional distance, in Figure 4.3. In both cases, the smallest p -norm size of 0.1 had the greatest percentage of occurrences by far. Similarly, the best EER from combination [3:5] of individually optimized weighted fractional distance had a comparable pattern with the best EER of [1:3] for the individually optimized fractional distance. In both cases, the first 4 p -norm distances were among the largest and within a close range. Additionally, the greatest occurrence was still for the p -norm value of 0.1. The smaller feature vectors in combination [1:3] provided an EER in between the highest and lowest. In Figure 4.4, it was seen that the greatest occurrence was for p -norm of 0.6 rather than 0.1. The latter provided much lower EER values with larger feature vector sizes. The larger p -norm of 0.6 was more effect with smaller feature vectors, in comparison with 0.1. However, 0.1 performed much better than 0.6 in comparison with larger feature vectors.

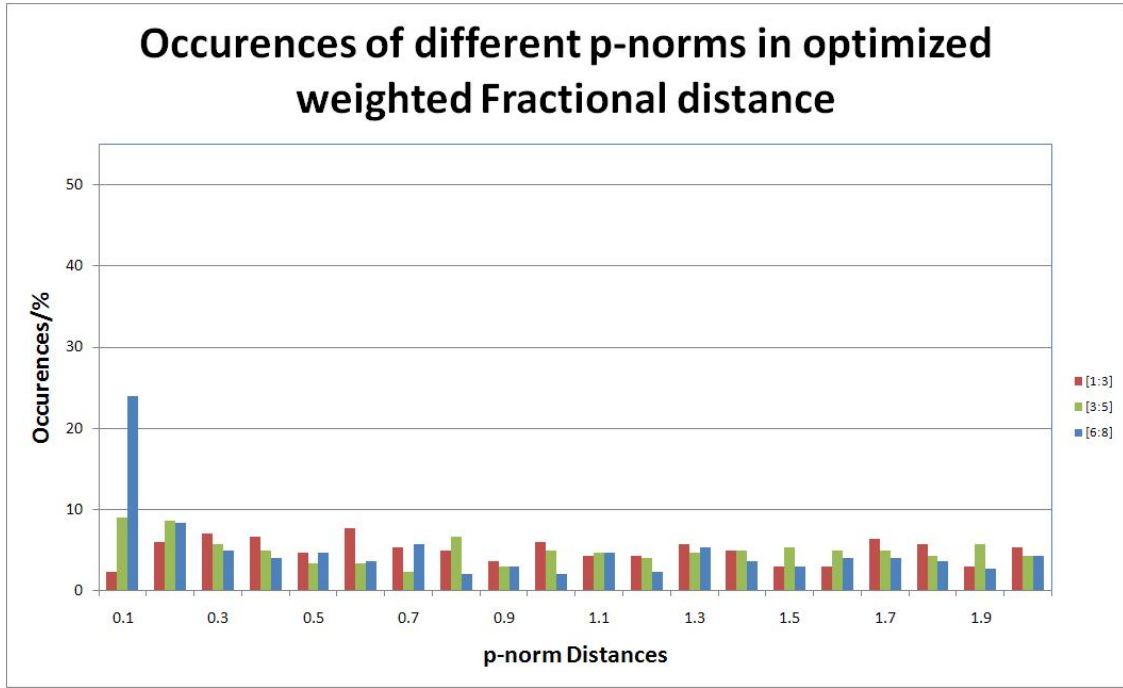


Figure 4.4: The number of occurrences of each p-norm in the weighted fractional distance with individually optimized resampling combinations [1:3] and [6:8]

4.6 Literature Comparison

4.6.1 Literature Comparison of EMDF

The work of Nguyen et al. (2009) [48] was chosen for the comparison and benchmarking of the EMDF. They performed many different configurations to find the optimal supervised learning technique (SLT) classification technique for the chosen feature vectors. Their results were obtained through tests with: multilayer perceptrons, also called artificial neural networks, with both radial basis function (RBF) and back propagation kernels; support vector machines (SVM) with linear, polynomial and RBF kernels; and different configurations of authentic and forged signatures for the training and testing phase. According to [12], for all feature extractions, the *rs_strip* size was 5 and the *max_transitions* were 3 or 4. The smaller feature vector size with *max_transition* of 3 provided better accuracy with a RBF neural network kernel for classification. It is assumed that a *max_transition* of 3 is used for all of their future work. Their best results were found in [48], where an SVM with an RBF kernel was used to obtain an EER of 17.25% and FARR of 0.08%. Signatures of 160 GPDS individuals were used [66].

In this work, the same feature extraction technique was implemented and tested with multiple distance-based classification techniques, including the Euclidean, fractional, weighted Euclidean and weighted fractional distances. Signatures of 300 GPDS individuals were used. A larger range of *rs_strip* and *max_transition* sizes were tested. This range was from 2 to 8, where both sizes were kept equal. While the best Euclidean and Manhattan distance performances for the EMDF was worse than the literature result, at 19.2% and 18.7% respectively, the fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances all performed much better, at 16.0%, 15.3%, 16.9% and

13.8%. It was seen that the weighted fractional distance performed better than the SVM with an RBF kernel for the classification of the EMDF. Further, individually optimized resampling was performed, where the best resampling (*rs_strip* and *max_transition*) sizes per individual were chosen. The best individually optimized resampling was for the weighted fractional distance, with an EER of 10.8% and corresponding FARR of 0.67%. The obtained EER was better than the best results in the literature by 6.5% and both systems obtained an FARR of below 1%.

4.6.2 Literature Comparison of LDP

Ferrer et al. [21] tested the Local Binary Pattern (LBP) and Local Directional Pattern (LDP), separately, using several data sets. Their results are used for the comparison and benchmarking of LDP results that were obtained in this work. These sets were 75 individuals from the MCYT database [53], and 75, 300 and 960 users from the GPDS database [66]. While the EER with the data sets using 75 individuals was low, these results can not be used for comparison, since the small size of the data set brings the precision and accuracy of results into question. A better comparison was the results obtained with the data sets containing 300 and 960 individuals. Classification was performed with a Least Squares SVM (LS-SVM) with an RBF kernel. The signatures were split into 12 blocks for the feature extraction, consisting of 4 vertical splits and 3 horizontal splits, and an overlap of 60%. From these, the best EER was 17.8% using 300 individuals of the GPDS database, with a corresponding FARR of 0.68%.

In this work, the same feature extraction technique was implemented and tested with multiple distance-based classification techniques, including the Euclidean, fractional, weighted Euclidean and weighted fractional distances. Signatures of 300 GPDS individuals were used, which was also the same database used for the best results in [21]. A larger range of splits sizes were tested. This range was from 1 to 8 splits in both the vertical and horizontal directions. While the best Euclidean and Manhattan distance performances for the EMDF were worse than the literature result, at 21.7% and 19.2% respectively, the fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances all performed much better, at 14.7%, 14.5%, 14.3% and 12.2%. It was seen that the weighted fractional distance performed better than the LS-SVM with an RBF kernel for the classification of the LDP. Further, individually optimized resampling was performed, where the best resampled block sizes per individual were chosen from a possibility of either 4 or 9. The best individually optimized resampling was for the weighted fractional distance, with an EER of 9.25% and corresponding FARR of 0.80%. The obtained EER is better than the best results in the literature by 8.6% and both systems obtained an FARR of below 1%. The literature comparison with the best results is summarized in Table 4.31.

Table 4.31: A comparison of results from literature with the results obtained in this work

Feature Extraction	Classification	Database	EER/%	FARR/%
EMDF [48]	SVM, RBF kernel	GPDS160	17.25	0.08
LDP [21]	LS-SVM, RBF kernel	GPDS300	17.8	0.68
EMDF (current)	weighted fractional distance, with individually optimized resampling	GPDS300	10.8	0.67
LDP (current)	weighted fractional distance, with individually optimized resampling	GPDS300	9.25	0.80

4.7 Optimal Design Overview

Figure 4.5 shows the detailed configuration of the verification system, which was briefly described in Figure 3.1, that achieved the best results. For preprocessing the images were binarized, dilated and then cropped to the bounding box. For feature extraction, LDP features performed better than the EMDF features. For classification, a combination of the weighted fractional distances and individually optimized resampling, which are distance-based classification techniques, provided the best accuracy.

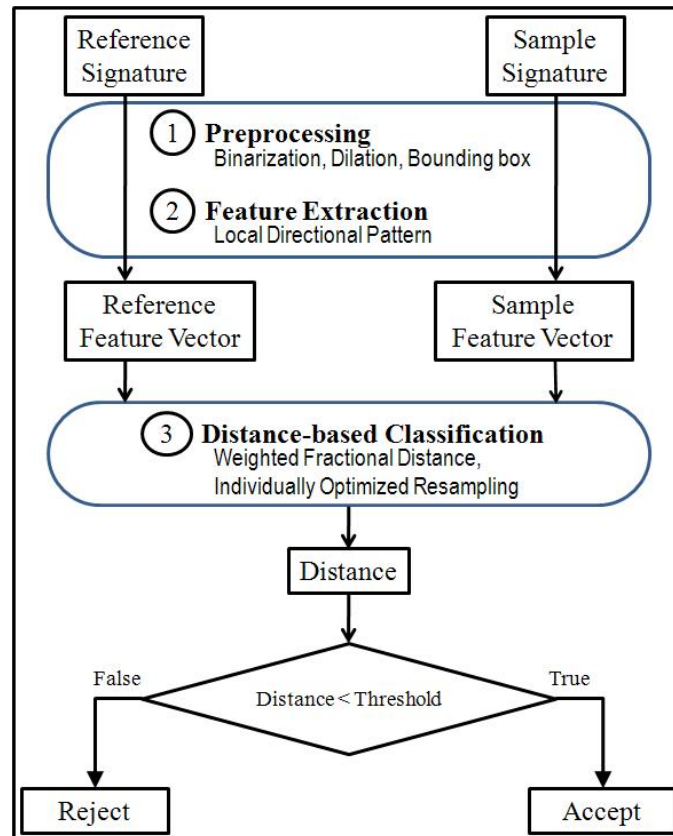


Figure 4.5: Overview of the design of a biometric verification system with specified optimal configurations

4.8 Implementation Scenarios

The most likely scenarios for the use of an offline signature verification system are to verify the identity of signers on bank cheques and to assist questioned document examiners in forensic science laboratories. In both cases, known authentic signatures will be captured using a scanner device. These will be used to generate the reference feature vector and to train the system. The signature from a new cheque, or a questioned document will also be scanned into the system, and the extracted feature vector will be compared with the reference feature vector. This automated technique can improve productivity. Since automated signature verification systems are more accurate than humans, implementation will increase security and reduce the occurrences of identity theft and forged documents.

4.9 Conclusion

Results were obtained and analyzed for distance-based classification with the EMDF and LDP feature extraction techniques. The weighted fractional distance with individually optimized resampling produced the best EER in this work, and the LDP performed better than the EMDF. The distance-based classification technique also produced better results than supervised learning techniques in literature when applied to the sample feature extraction techniques. In the next chapter, the dissertation will be concluded and possible extensions will be discussed.

Chapter 5

Conclusion and Future Work

5.1 Summary of Work

In this dissertation, a framework for the verification of offline signatures, using distance-based classification techniques was presented. Firstly, a detailed review of literature in the area was provided. Then methods and techniques employed were described. Results were discussed, with comparisons between different distance-based classification techniques, the effects of using different feature extraction techniques with the distance-based classification and a comparison with results from literature.

It was found that the cosine similarity measure, Mahalanobis, Canberra and Ratio distances performed poorly during classification. The Euclidean, Manhattan and fractional distances, which are all p -norm distances, performed much better. Weighted p -norm distances and individually optimized resampling, in conjunction with p -norm distances also performed well. The best EER of 9.25% was obtained with the Local Directional Pattern feature extraction technique, using individually optimized weighted fractional distances for classification. This result is better than literature results using the same feature extraction technique, and SLTs for classification.

5.2 Recommendations for Future Work

Limitations and suggestions for improvement of this offline signature verification system are:

- **Effect of time:** By nature biometric traits change over a long period of time. Counteracting this limitation is an ongoing research area.
- **Multi-modal:** fusion of signature biometrics with other biometric modalities could be considered to improve the security of systems.
- **Databases:** The proposed system was tested with the GPDS database, which comprises signatures of Spanish individuals. It would be interesting to observe the effects of signatures based on other language character sets, such as Chinese, Japanese and Arabic.

List of References

- [1] M. Adamski and K. Saeed. Signature image recognition by shape context image matching. *Journal of Medical Informatics and Technologies*, 11:89–95, 2007. 18
- [2] S. M. S. Ahmad, A. Shakil, A. R. Ahmad, M. Agil, M. Balbed, and R. M. Anwar. Sigma-a malaysian signatures database. In *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, pages 919–920. IEEE, 2008. 11, 14
- [3] Y. M. Al-Omari, S. N. H. S. Abdullah, and K. Omar. State-of-the-art in offline signature verification system. In *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on*, volume 1, pages 59–64. IEEE, 2011. 14
- [4] A. Alizadeh, T. Alizadeh, and Z. Daei. Optimal threshold selection for online verification of signature. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, pages 17–19, 2010. 18
- [5] S. Armand, M. Blumenstein, and V. Muthukkumarasamy. Off-line signature verification using the enhanced modified direction feature and neural-based classification. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 684–691. IEEE, 2006. 11, 16
- [6] S. Armand, M. Blumenstein, and V. Muthukkumarasamy. Off-line signature verification using an enhanced modified direction feature with single and multi-classifier approaches. *Computational Intelligence Magazine, IEEE*, 2(2):18–25, 2007. 11, 16
- [7] L. Batista, E. Granger, and R. Sabourin. Improving performance of hmm-based off-line signature verification systems through a multi-hypothesis approach. *International journal on document analysis and recognition*, 13(1):33–47, 2010. 13, 14, 15
- [8] L. Batista, E. Granger, and R. Sabourin. A multi-classifier system for off-line signature verification based on dissimilarity representation. *Multiple Classifier Systems*, pages 264–273, 2010. 15
- [9] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966. 14
- [10] M. Blumstein and X. Y. Liu. Experimental analysis of the modified direction feature for cursive character recognition. *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*, (0-7695-2187-8/04), 2004. 28, 37

- [11] M. Blumstein, B. Verma, and H. Basli. A novel feature extraction technique for the recognition of segmented handwritten characters. In *International Conference on Document Analysis and Recognition*, pages 137–141, 2003. 22, 25, 26, 28
- [12] M. Blumstein, B. Verma, and X. Y. Liu. A modified direction feature for cursive character recognition. *IEEE Computer Society*, (0-7803-8359-1/04), 2004. xiv, 26, 28, 29, 30, 37, 69
- [13] P. Campisi, R. L. Carter, C. W. Crooks, V. Govindaraju, W. Hamilton, J. Hurt, A. A. Ross, C. J. Tilton, and Jr. D. M. Waymire. *Module 2 Biometric Modalities in Certified Biometrics Professional (CBP)*. IEEE, 2010. 6, 8
- [14] J. Coetzer, B. M. Herbst, and J. A. Du Preez. Offline signature verification using the discrete radon transform and a hidden markov model. *EURASIP Journal on Applied Signal Processing*, 2004:559–571, 2004. 11, 14
- [15] S. Cole. History of fingerprint pattern recognition. In *Automatic Fingerprint Recognition Systems*, pages 1–25. Springer New York, 2004. 5
- [16] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 14
- [17] J. G. A. Dolfig, E. H. L. Aarts, and J. J. G. M. van Oosterhout. On-line signature verification with hidden Markov models. In *International Conference on Pattern Recognition*, volume 2, 1998. 11
- [18] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska. A new face database simultaneously acquired in visible, near-infrared and thermal spectrums. *Cognitive Computation*, pages 1–17, 2012. 19
- [19] B. Fang, C. H. Leung, Y. Y. Tang, K. W. Tse, P. C. K. Kwok, and Y. K. Wong. Off-line signature verification by the tracking of feature and stroke positions. *Pattern recognition*, 36(1):91–101, 2003. 17
- [20] M. A. Ferrer, J. B. Alonso, and C. M. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):993–997, 2005. 14, 17
- [21] M. A. Ferrer, F. Vargas, C. M. Traviesto, and J. B. Alonso. Signature verification using local directional pattern. In *International Carnahan Conference on Security Technology*, pages 336–340. IEEE, Oct 2010. xiv, 11, 23, 34, 35, 37, 70, 71
- [22] “Association BioSecure” (BioSecure Foundation). <http://biosecure.it-sudparis.eu>. Accessed 06/11/2012. 11
- [23] D. Francois, V. Wertz, and M. Verleysen. The concentration of fractional distances. *Knowledge and Data Engineering, IEEE Transactions on*, 19(7):873–886, 2007. 19, 47
- [24] G. Goudelis, A. Tefas, and I. Pitas. Emerging biometric modalities: a survey. *Journal on Multimodal User Interfaces*, 2:217–235, 2008. 7
- [25] S. Huopio. Biometric identification. *Eye*, 3:1, 1998. 5

- [26] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla, and C. A. Trullo. Analysis of stability in static signatures using cosine similarity. In *Proceedings of the International Conference on Frontiers in Handwriting Recognition*, pages 231–235. IEEE, 2012. 18
- [27] I. A. Ismail, M. A. Ramadan, T. S. El-Danaf, and A. H. Samak. An efficient offline signature identification method based on fourier descriptor and chain codes. *IJC-SNS International Journal of Computer Science and Network Security*, 10(5):29–35, 2010. 13, 18
- [28] T. Jabid, H. Kabir, and O. Chae. Gender classification using local directional pattern (LDP). In IEEE Computer Society, editor, *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR’10, pages 2162–2165, 2010. 34
- [29] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14:4–20, 2004. 5, 6
- [30] N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto. A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:258–262, 1999. 17, 49
- [31] H. B. Kekre, V. A. Bharadi, S. Gupta, A. A. Ambardekar, and V. B. Kulkarni. Off-line signature recognition using morphological pixel variance analysis. In *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*, ICWET ’10, pages 3–10. ACM, 2010. 13
- [32] E. J. C. Kelkboom, B. G okberk, T. A. M. Kevenaer, and A. H. M. Akkermans. “3D face”: Biometric template protection for 3D face recognition. In *In Proceedings of Second International Conference on Biometrics*, pages 566-573, 2007. 6
- [33] N. P. Khanyile, J. R. Tapamo, and E. Dube. A comparative study of fingerprint thinning algorithms. 2011. 22
- [34] V. Kiani, R. Pourreza Shahri, and H. R. Pourreza. Offline signature verification using local radon transform and support vector machines. *International Journal of Image Processing*, 3, 2009. 15
- [35] D. R. Kisku, A. Rattani, P. Gupta, and J. K. Sing. Offline signature verification using geometric and orientation features with multiple experts fusion. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, volume 5, pages 269–272. IEEE, 2011. 13, 17
- [36] B. Kovari, A. Horvath, B. Toth, H. Charaf, L. Perlovsky, D. D. Dionysiou, L. A. Zadeh, M. M. Kostic, C. Gonzalez-Concepcion, H. Jaberg, et al. Local feature based off-line signature verification using neural network classifiers. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, number 11. WSEAS, 2009. 15
- [37] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 2006. 17

- [38] R. Kumar, J. D. Sharma, and B. Chanda. Writer-independent off-line signature verification using surroundedness feature. *Pattern Recognition Letters*, 2011. 15
- [39] G. N. Lance and W. T. Williams. Mixed-data classificatory programs i - agglomerative systems. *Australian Computer Journal*, 1(1):15–20, 1967. 18
- [40] N. Liu, B. Zhang, J. Yan, Q. Yang, S. Yan, Z. Chen, F. Bai, and W. Y. Ma. Learning similarity measures in non-orthogonal space. In *International Conference on Information and Knowledge Management*, pages 334–341, 2004. 49
- [41] C. Y. Low, A. Beng-Jin Teoh, and C. Tee. Support vector machines (svm)-based biometric watermarking for offline handwritten signature. In *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on*, pages 2095–2100. IEEE, 2008. 14
- [42] P. C. Mahalanobis. On the generalized distance in statistics. In *Proceedings of the National Institute of Science (India)*, volume 12, pages 49–55. National Institute of Science, 1936. 17, 40
- [43] B. Majhi, Y. S. Reddy, and D. P. Babu. Novel features for off-line signature verification. *International Journal of Computers, Communications & Control*, 1(1):17–24, 2006. 17
- [44] W. S. McCulloch and W. Pitts. Artificial neural network. *Bull. Math. Biophys*, 5:115–133, 1943. 15
- [45] K. K. Meenakshi, S. Srihari, and A. Xu. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(07):1339–1360, 2004. 18
- [46] O. Mirzaei, H. Irani, and H. R. Pourreza. Offline signature recognition using modular neural networks with fuzzy response integration. In *Proceedings of 2011 1st International Conference on Network and Electronics Engineering (ICNEE 2011)*, 2011. 15
- [47] S. F. Miskhat, M. Ridwan, E. Chowdhury, S. Rahman, and M. A. Amin. Profound impact of artificial neural networks and gaussian svm kernel on distinctive feature set for offline signature verification. In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on*, pages 940–945. IEEE, 2012. 15
- [48] V. Nguyen, M. Blumenstein, and G. Leedham. Global features for the off-line signature verification problem. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, pages 1300–1304. IEEE Computer Society, 2009. 11, 13, 16, 29, 31, 32, 37, 47, 51, 52, 69, 71
- [49] V. Nguyen, M. Blumenstein, V. Muthukkumarasamy, and G. Leedham. Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines. In *Ninth International Conference on Document Analysis and Recognition*, pages 734 –738. IEEE Computer Society, 2007. 11, 16

- [50] V. Nguyen, Y. Kawazoe, T. Wakabayashi, U. Pal, and M. Blumenstein. Performance analysis of the gradient feature and the modified direction feature for off-line signature verification. In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, pages 303–307. IEEE, 2010. 15, 17
- [51] Center of Excellence for Document Analysis and Recognition. Cedar signature database. <http://www.cedar.buffalo.edu/NIJ/publications.html>. Accessed 06/11/2012. 11
- [52] Ferdowsi University of Mashhad. FUM-PHSDB: The FUM-Persian handwritten signature database. <http://mvlab.um.ac.ir/index.php?module=htmlpages&func=display&pid=11>, 2006. Accessed 11/11/2012. 11
- [53] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro. Mcyt baseline corpus: A bimodal biometric database. In *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, volume 150, pages 395–401, December 2003. 10, 70
- [54] E. Özgündüz, T. Şentürk, and M. E. Karşılıgil. Off-line signature verification and recognition by support vector machine. In *European Signal Processing Conference. Turkey*, 2005. 11
- [55] M. S. Panton. *Off-line signature verification using ensembles of local Radon transform-based HMMs*. PhD thesis, Stellenbosch University, 2011. 14, 15
- [56] A. C. Ramachandra, K. Pavithra, K. Yashasvini, K. B. Raja, K. R. Venugopal, and L. M. Patnaik. Offline signature authentication using cross-validated graph matching. In *Proceedings of the 2nd Bangalore Annual Compute Conference*, page 7. ACM, 2009. 17
- [57] Y. Rekik, N. Houmani, M. A. El Yacoubi, S. Garcia-Salicetti, and B. Dorizzi. A comparison of feature extraction approaches for offline signature verification. In *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*, pages 1–6. IEEE, 2011. 16
- [58] B. Schafer and S. Viriri. An off-line signature verification system. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on*, pages 95–100. IEEE, 2009. 46
- [59] A. Shakil, S. M. S. Ahmad, R. B. M. Anwar, and M. A. M. Balbed. Analysis of the effect of different features’ performance on hidden markov modeling based online and offline signature verification systems. In *Computing: Techniques and Applications, 2008. DICTA’08. Digital Image*, pages 572–577. IEEE, 2008. 14
- [60] B. H. Shekar and R. K. Bharathi. Eigen-signature: A robust and an efficient off-line signature verification algorithm. In *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, pages 134–138. IEEE, 2011. 11, 16
- [61] M. S. Shirdhonkar and M. B. Kokare. Off-line handwritten signature retrieval. In *Proceedings of the CUBE International Information Technology Conference*, pages 290–293. ACM, 2012. 18

- [62] M. H. Sigari, M. R. Pourshahabi, and H. R. Pourreza. Offline handwritten signature identification and verification using multi-resolution gabor wavelet. *International Journal of Biometric and Bioinformatics*, 5, 2011. 17
- [63] G. S. Sodhi and J. Kaur. The forgotten indian pioneers of fingerprint science. *Current Science*, 88(1):185–191, 2005. 5
- [64] K. Teknomo. Similarity measurement: Mahalanobis distance. <http://people.revoledu.com/kardi/tutorial/Similarity/MahalanobisDistance.html>, 2006. Accessed 06/08/2012. 40
- [65] P. Tuly, B. Scorić, and T. Kevenaar, editors. *Security with Noisy Data: Private Biometrics Secure Key Storage and Anti-Counterfeiting*. Springer-Verlag, 2007. 6
- [66] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso. Off-line handwritten signature gpd960 corpus. In *Ninth International Conference on Document Analysis and Recognition*, pages 764 – 768. IEEE Computer Society, 2007. 10, 43, 45, 69, 70
- [67] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso. Offline signature verification based on pseudo-cepstral coefficients. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 126–130. IEEE, 2009. 14
- [68] C. Vivaracho-Pascual, M. Faundez-Zanuy, and J. M. Pascual. An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances. *Pattern Recognition*, 42(1):183 – 193, 2009. 13, 19, 35, 39, 47, 57
- [69] B. Widrow, M. E. Hoff, et al. Adaptive switching circuits. 1960. 15
- [70] R. V. Yampolskiy and V. Govindaraju. Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, 1, 2008. 6
- [71] D. Y. Yeung, H. Chang, Y. Xiong, S. E. George, R. S. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First International Signature Verification Competition. In *International Conference on Biometric Authentication*, pages 16–22, 2004. 10
- [72] M. B. Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov. Offline signature verification using classifier combination of HOG and LBP features. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, 2011. 11, 14
- [73] D. Zhang and A. K. Jain, editors. *First International Conference on Biometric Authentication*, volume 3072 of *Lecture Notes in Computer Science*, Hong Kong, July 2004. Springer Verlag. 7
- [74] T. Y. Zhang and C. Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, pages 236–239, March 1984. 22, 25, 32
- [75] Y. Zhu, T. Tan, and Y. Wang. Biometric personal identification based on iris patterns. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 801–804. IEEE, 2000. 13, 18, 42

Appendix A

Conference paper accepted for and presented at IEEE ICSPCC '12

- Conference Name: 2012 International Conference on Signal Processing, Communications and Computing
- Conference Location: Hong Kong
- Conference Date: 12 - 15 August 2012
- Conference Publisher: IEEE

HANDWRITTEN SIGNATURE VERIFICATION USING WEIGHTED FRACTIONAL DISTANCE CLASSIFICATION

Y. Moolla^{1,2}, S. Viriri¹, F.V. Nelwamondo², J.R. Tapamo³

¹ School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal
Durban, South Africa
{viriris; 205500527}@ukzn.ac.za

² Modelling and Digital Science Unit
Council for Scientific and Industrial Research
Pretoria, South Africa
fnelwamondo@csir.co.za

³ School of Engineering
University of KwaZulu-Natal
Durban, South Africa
tapamoj@ukzn.ac.za

ABSTRACT

Signatures are one of the behavioural biometric traits, which are widely used as a means of personal verification. Therefore, they require efficient and accurate methods of authenticating users. The use of a single distance-based classification technique normally results in a lower accuracy compared to supervised learning techniques. This paper investigates the use of a combination of multiple distance-based classification techniques, namely individually optimized re-sampling, weighted Euclidean distance, fractional distance and weighted fractional distance. Results are compared to a similar system that uses support vector machines. It is shown that competitive levels of accuracy can be obtained using distance-based classification. The best accuracy obtained is 89.2%.

Index Terms— Pattern Recognition; Biometrics, Handwritten Signatures, Optimized Re-sampling, Weighted Euclidean Distance, Fractional Distance

1. INTRODUCTION

Biometrics is the use of one or more intrinsic physical or behavioural human characteristics to verify or identify a person. Biometric traits should be unique, universal, long lasting, collectible, commonly accepted, difficult to falsely duplicate and identifiable efficiently and accurately by machine. Examples of physiological biometric traits include fingerprints, DNA,

hand and palm geometry and iris recognition, whereas examples of behavioural biometric (behaviometric) traits include voice recognition, writing patterns and signatures [1].

The basic process of automated biometrics verification involves capturing the biometric traits onto a machine and then using biometric feature extraction algorithms to create a digital representation template of the trait. For authentication of an individual, the system creates a biometric template from newly captured data and compares the two templates [1].

A signature, which is one of the oldest used and most widely accepted biometric for identification and verification [2], is a handwritten depiction of a persons name, nickname or other personal symbol. It is classified as a behavioural biometric trait.

There are two ways to capture signatures: online and offline. Offline signatures are static images while online signatures are dynamic and capture the progress of signature writing as a function of time. Since online signatures hold a greater amount of information, they intrinsically allow greater accuracy than a static image. However, there are still many systems that require the improved accuracy of offline signatures. For instance, online signatures are not available for bank cheques or credit cards, and accurate offline signature verification is essential. The electronic writing pads for the capture of offline signatures are also much more cost effective than that for online signatures. Thus, the availability of competitively accurate offline signature verification could improve security measures for businesses in poorer emerging

markets.

There are many different techniques for classifying signatures and other biometrics. They can be broadly categorized into supervised learning techniques (SLTs) and distance-based classification techniques. SLTs include neural networks [3], hidden Markov models [4], support vector machines [5] and fuzzy logic [6]. Distance-based techniques include Euclidean distance, Mahalanobis [7], Manhattan distance, weighted Euclidean distances [8] and fractional distances [9].

SLTs in general provide a greater accuracy than basic distance-based techniques. This paper aims to combine several distance-based techniques to gain accuracy comparable with SLTs. The weighted Euclidean distance and fractional distance classification techniques are investigated. The two are then combined to create a novel weighted fractional distance classification.

Individually optimized re-sampling space normalization of the feature vector is also investigated to further improve the overall accuracy of the system. The feature extraction techniques from Nguyen et. al. [5] are used, namely, the Direction, Modified Direction, Energy, Ratio and Maxima features. These features were re-sampled to a set static size so that all feature vectors were of equal length. Vivaracho-Pascual et. al. [9] tested several different re-sampled sizes for a different online feature extraction, but settled on a single re-sampled size for all signatures. However, they suggest optimization of the re-sampled size for each user as an interesting study. This paper aims to investigate the effect of individually optimized re-sampling on the modified direction feature from [5]. Results achieved are compared to those obtained with a supervised classification technique using the same feature extraction techniques.

Feature extraction and classification techniques are described in section 2. Results obtained are discussed in section 3 and the conclusion is given in section 4.

2. METHODS AND TECHNIQUES

Verification of signatures is a multi-step process. Firstly, preprocessing is performed to clean the original image and remove noise and other unwanted data and prepare the next step. The next step is feature extraction, which entails extracting essential information. The third step is training, where feature vectors from known authentic and forged signatures are compared for the calibration of the classification technique. The fourth step, classification, is where the system must accurately and independently determine whether signatures are authentic or forged.

In our context, preprocessing requires binarization, finding the bounding box and thinning of the signature image. The Zhang-Suen thinning algorithm [10] is used.

Features used in this paper are the one used in [5], namely, direction, modified direction, energy, ratio and maxima features.

In the training and classification phases, dynamic re-sampling to normalize the feature vector [11], Euclidean distance, weighted Euclidean distance [8], fractional distances [9], and weighted fractional distances are tested.

Feature extraction techniques used are described in section 2.1, and classification techniques are described in section 2.2.

2.1. Feature Extraction

2.1.1. Direction Feature

The direction feature extraction technique extracts the direction of each segment between intersections within a signature, i.e. whether the segment is horizontal, vertical, diagonal left or diagonal right. It yields 9 biometric features, namely, the total length of each line direction set (4 features), the number of lines in each direction set (4 features) and the total number of intersection points (1 feature). Detailed information on direction feature can be found in [11].

2.1.2. Modified Direction Feature

The modified direction feature extraction technique [11, 12] is the result of a combination of the direction feature [13] and the transition feature. It has been used as part of optical character recognition systems. A series of steps are required:

Firstly, the direction feature extraction is performed so as to label each line segment with a direction.

Then, the image is parsed in four directions, namely, right to left, left to right, top to bottom and bottom to top. In each direction, the location transition (LT) and direction transition (DT) features are recorded. LT is the location of the transition of pixels from background to foreground and DT is the direction of the line segment at the point of transition. For the sake of uniformity, limits are placed to the maximum number of transitions, *max_transitions*, recorded in any given direction. This provided 8 arrays of features with sizes *max_transitions* \times (*image.height* or *image.width*).

Finally, window re-sampling is performed. This means that the height and width values of each array are re-sampled for normalization, or averaging, so that feature vectors for all signatures of an individual are uniform in size. Different numbers of re-sampling strips (*rs_strips*) are tested.

To formalize the window re-sampling, let *m* be the number of *maximum_ttransitions* and *p* be the number of pixels in height or width where each pixel now stores the LT or DT value for the row or column, so that the transitioned image will be $I_T(p, m)$. Let *s* be the number of pixels in one strip, as calculated in Equation (1). Let *r* be the number of *rs_strips* and *v* be the value of an element in a calculated *rs_strip*.

$$s = p/r \quad (1)$$

Then the resulting set of *rs_strip* values is

$$\begin{aligned}
v_1 &= (x_1^1, x_1^2, \dots, x_1^r) \\
v_2 &= (x_2^1, x_2^2, \dots, x_2^r) \\
&\vdots \\
v_m &= (x_m^1, x_m^2, \dots, x_m^r)
\end{aligned} \tag{2}$$

Let v_j^i be a single calculated component where $1 \leq i \leq r$ and $1 \leq j \leq m$. Then the value v_j^i is calculated as

$$v_j^i = \frac{1}{s} \sum_{k=n \times s}^{k < (n \times (s+1)) - 1} I(k, j) \tag{3}$$

This provides 8 feature vectors ($4 \times \text{LT} + 4 \times \text{DT}$) with sizes *max.transitions* \times *rs.strips* (of height or width). Direction features are added to this feature vector. For example, if a maximum of 4 transitions are used, along with re-sampling of 5, the feature vector will have 169 components.

More detailed information on the modified direction feature can be read in [11].

2.1.3. Other Feature Extractions

The Energy, Ratio and Maxima feature extractions were used as described by Nguyen et. al. [5].

2.2. Classification

2.2.1. Euclidean distance and thresholds

One of the most common distance-based classification techniques for determining the accuracy of biometric systems is the calculation of the Euclidean distance between a reference vector (derived as a mean of several authentic signatures of an individual) and other feature vectors.

Authentic signatures are expected to have Euclidean distance values below a certain threshold while forged signatures would have values above that threshold. Authentic signatures with distances above the threshold are regarded as false negatives and contribute to the False Rejection Rate (FRR) while forged signatures with distances below the threshold are regarded as false positives and contribute to the False Acceptance Rate (FAR). This is further split into the FAR for skilled forgeries (FARS) and for random forgeries (FARR). The threshold is chosen where the distance for the FRR and FARS are equal. This rate is also called the Equal Error Rate (ERR).

The equation for determining the Euclidean distance between vectors x and y is computed as defined in Equation (4).

$$\|x - y\|_p = (\sum |(x - y)|^p)^{1/p} \tag{4}$$

where $p = 2$

2.2.2. Weighted Euclidean distance

The weighted Euclidean distance measure is a technique adapted from [8] to improve the classification accuracy by adding weight, or statistical importance, to the most reliable features from the feature vector. Firstly, the standard deviation for the reference signatures is obtained.

Let the n reference signatures be

$$\begin{aligned}
x_1 &= (x_1^1, x_1^2, \dots, x_1^m) \\
x_2 &= (x_2^1, x_2^2, \dots, x_2^m) \\
&\vdots \\
x_n &= (x_n^1, x_n^2, \dots, x_n^m)
\end{aligned} \tag{5}$$

Let x_i^j be the j^{th} component of the i^{th} reference signature where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Then the mean of the j^{th} component of the reference signatures, μ^j , is computed as in Equation (6)

$$\mu^j = \frac{1}{n} \sum_{i=0}^{i < n} x_i^j \tag{6}$$

and their standard deviation σ^j is defined in Equation (7) as

$$\sigma^j = \sqrt{\frac{1}{n} \sum_{i=0}^{i < n} (x_i^j - \mu^j)^2} \tag{7}$$

The weighted Euclidean distance can then be calculated using the standard deviation as

$$\|x - y\|_p = \left(\sum_{j=0}^{j < m} \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \tag{8}$$

where $p = 2$.

2.2.3. Fractional distances

A drawback of using Euclidean and other p -norm distances where $p \in \mathbb{N}_1$ is that as the vectors get larger, the distance values tend to cluster. This is called the concentration phenomenon. To overcome this limitation of distance-based classification, Vivaracho-Pascual et. al. [9] introduced the use of fractional p -norm distances.

The equation for determining fractional p -norm distance between vectors x and y is computed as defined in equation (9)

$$\min(\|x - y\|_p) = (\sum |(x - y)|^p)^{1/p} \tag{9}$$

where $0.1 \leq p \leq 2.0$.

The optimal value of p is when the distance calculated using Equation (9) is at its minimum for all values of p within the given range.

2.2.4. Weighted fractional distances

The fractional distances and weighted Euclidean distance can then be combined to form the weighted fractional distance as defined in equation (10)

$$\min(\|x - y\|_p) = \left(\sum_{j=0}^{j < m} \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \quad (10)$$

where $0.1 \leq p \leq 2.0$

Just like with Equation (9), the optimal value of p is when the distance calculated using Equation (10) is at its minimum for all values of p within the given range.

2.2.5. Individually optimized re-sampling

Re-sampling of the feature vector allows it to be re-sized. This is a form of spatial normalization. Different re-sampling sizes results in changing accuracies. By choosing the best re-sampling size per user, it is expected that the overall accuracy of the system may be optimized.

The re-sampling method is described in Section 2.1.2. Different *rs_strips* sizes were used to change the size of the feature vector. The smallest possible was an *rs_strip* size of 2. Tests with incrementally large sizes were used, until size 8. By this point, the re-sampling was having little positive impact on the accuracy and the last feature vector size was negatively affecting processing time. It was decided to keep the maximum *rs_strip* size of 8.

3. RESULTS AND DISCUSSION

3.1. Data Set

Experiments were performed using signatures from the Grupo de Procesado Digital de Senales (GPDS) signature database [14]. The database consists of black and white signatures of 300 different individuals, with 24 authentic copies and 30 skilled forgeries for each individual.

10 authentic signatures are used to create the reference signature, the other 14 authentic signatures and the 30 skilled forgeries are used for the classification and verification. For random forgeries per individual, a single authentic signature from each of the other 299 individuals is used.

3.2. Results

Quantification of result accuracy is measured in terms of the False Rejection Rate (FRR); False Acceptance Rate (FAR) which is further broken down into FAR for skilled forgeries (FARS) and FAR for random forgeries (FARR); and the Equal Error Rate (ERR), which is the point at which the FRR and FARS converge. Fig. 1 shows an example of the ROC curve for obtaining the EER.

Table 1. The effect of different re-sampling sizes using the Euclidean distance

<i>rs_strips</i>	vector size	FRR /%	FARS /%	FARR /%
2	49	20.11	20.11	0.7268
3	89	19.23	19.23	0.4436
4	145	19.41	19.41	0.3667
5	217	19.73	19.73	0.3633
6	305	20.17	20.17	0.3567
7	409	20.16	20.16	0.4391
8	527	20.35	20.35	0.4882
min(2:8)	mixed	15.32	15.32	0.5417

Table 2. The effect of different re-sampling sizes using the weighted Euclidean distance

<i>rs_strips</i>	vector size	FRR /%	FARS /%	FARR /%
2	49	15.19	15.19	0.2429
3	89	15.31	15.31	0.1973
4	145	16.14	16.14	0.3567
5	217	18.03	18.03	0.7101
6	305	20.15	20.15	1.8138
7	409	20.82	20.82	3.367
8	527	21.97	21.97	5.680
min(2:8)	mixed	11.72	11.72	0.7056

Table 3. The effect of different re-sampling sizes using fractional distances

<i>rs_strips</i>	vector size	FRR /%	FARS /%	FARR /%
2	49	15.99	15.99	0.5986
3	89	16.52	16.52	0.4525
4	145	17.07	17.07	0.4213
5	217	17.29	17.29	0.3767
6	305	17.39	17.39	0.4035
7	409	17.54	17.54	0.5217
8	527	17.61	17.61	0.6788
min(2:8)	mixed	12.85	12.85	0.5718

Table 1 shows the results of tests using different re-sampling sizes, without the weighting function in the Euclidean distance calculations. The resampling sizes are deter-

Table 4. The effect of different re-sampling sizes using weighted fractional distances

<i>rs_strips</i>	vector size	FRR /%	FARS /%	FARR /%
2	49	13.82	13.82	0.2541
3	89	14.31	14.31	0.2352
4	145	14.81	14.81	0.2942
5	217	15.71	15.71	0.5295
6	305	16.35	16.35	0.9821
7	409	16.52	16.52	1.512
8	527	17.09	17.09	2.457
min(2:8)	mixed	10.76	10.76	0.6655

mined by the *rs_strip* size in the MDF extraction technique. Sizes between 2 and 8 are tested. In the final row, results for individually optimized dynamic re-sampling are shown, i.e. for each individual, the best re-sampling between 2 and 8, is chosen. The worst EER was 20.35% for re-sampling size 8, while the best EER of 15.32% was for the dynamic re-sampling. Using dynamic re-sampling improves the EER over static re-sampling by up to over 5%.

This experiment is repeated using the weighted Euclidean distance, the fractional distance and weighted fractional distance functions. These results are shown in Tables 2, 3 and 4 respectively. In each case, the individually optimized re-sampling improved the accuracy of the classification.

An analysis of the static re-sampling results shows that the weighted Euclidean distance provides a great improvement over the traditional Euclidean distance when feature vectors are small in size, but it begins to have an adverse effect when the feature vectors become too large. By combining the weighted Euclidean distance with individually optimized resampling, the EER becomes 11.72%. This is better than the 15.32% from the combination of unweighted Euclidean distance with optimized re-sampling.

The unweighted fractional distances have slightly worse results compared to weighted Euclidean distance for smaller re-sampling sizes, but still better than the unweighted Euclidean distance. With larger re-sampling sizes, when the feature vectors become larger, the effect of fractional distances becomes more pronounced and provides better results than both weighted and unweighted Euclidean distances. By combining this with individually optimized re-sampling, the EER obtained is 12.85%. This is better in comparison to unweighted Euclidean distance, but worst in comparison to weighted Euclidean distance.

The final analysis was of the weighted fractional distances, which combines the weighted Euclidean distance and fractional distance measures. This provides the best accuracy

for all static re-sampling sizes. By combining these weighted fractional distances with individually optimized re-sampling, an EER of 10.76% was obtained. This is better than all of the previous combinations.

In all cases, the dynamic re-sampling combination provides a much higher accuracy than any individual static re-sampling size. Thus, the greatest accuracy was achieved by a combination of weighted distances, fractional distances and dynamic re-sampling with an EER of 10.76%. This is more accurate by almost 10% in comparison to any non-optimized resampling with a standard Euclidean distance measure.

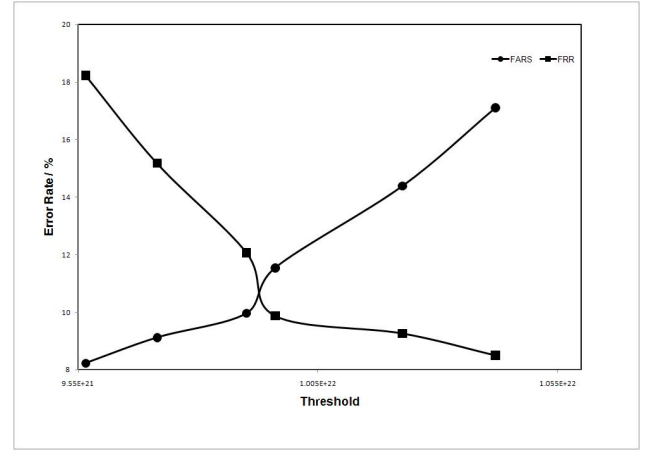


Fig. 1. Calculation of EER on applying the weighted fractional distances with individually optimized resampling

3.3. Literature Comparison

The work of Nguyen et. al. [5], [15] was chosen for the comparison in this paper. Their work is well documented and experimentally sound. They performed many different configurations to find the optimal supervised learning technique (SLT) classification technique for the chosen feature vectors. Tests were performed with multilayer perceptrons with both radial basis function (RBF) and back propagation kernels; support vector machines (SVM) with linear, polynomial and RBF kernels; and different configurations of authentic and forged signatures for the training and testing phase. Their best results were found in [5], where an SVM with an RBF kernel were used to obtain an EER of 17.25% and FARR of 0.08%. In this work, we obtained a best EER of 10.76% with an FARR of 0.67% using locally optimized distance-based classification techniques instead. The obtained EER is better than the best results in the literature by 6.5%. While both systems obtained an FARR of below 1%, the better FARR by Nguyen et. al. can be attributed to their training of the SVM using random forgeries which was not done in our system.

While further tests need to be performed, especially with regards to the training set used, and different feature extrac-

tions from literature, the results are promising.

4. CONCLUSION

Using the Modified Direction, Energy, Ratio and Maxima feature extraction techniques; and a combination of weighted fractional distance and individually optimized re-sampling classification techniques, an improved Equal Error Rate (EER) of 10.76% is obtained against skilled forgeries. This is compared to literature results where supervised learning techniques were applied to the same feature extraction techniques, which obtained an EER of 17.8%.

The weighted fractional distances technique works better as a whole than either of its individual constituent parts, namely, the weighted Euclidean distance and fraction distances. Individually optimizing the re-sampling of feature vectors also allows for improvement of overall accuracy.

It can be concluded that in some instances of signature verification, a combination of distance-based classification techniques can to be more accurate than supervised learning techniques.

5. REFERENCES

- [1] A.K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.
- [2] "Biometric authentication," in *International Conference on Biometric Authentication*, D. Zhang and A. K. Jain, Eds., 2004.
- [3] I.A. Ismail, M.A. Ramadan, T.S. El-Danaf, and A.H. Samak, "An efficient off line signature identification method based on fourier descriptor and chain codes," *International Journal of Computer Science and Network Security*, vol. 10, no. 5, pp. 29–35, 2010.
- [4] L. Batista, E. Granger, and R. Sabourin, "Improving performance of HMM-based off-line signature verification systems through a multi-hypothesis approach," *International Journal on Document Analysis and Recognition*, vol. 13, no. 1, pp. 33–47, 2010.
- [5] V. Nguyen, M. Blumenstein, and G. Leedham, "Global features for the off-line signature verification problem," in *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*. 2009, pp. 1300–1304, IEEE Computer Society.
- [6] H.B. Kekre, V.A. Bharadi, S. Gupta, A.A. Ambardekar, and V.B. Kulkarni, "Off-line signature recognition using morphological pixel variance analysis," in *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*. 2010, pp. 3–10, ACM.
- [7] D.R. Kisku, P. Gupta, and J.K. Sing, "Offline signature identification by fusion of multiple classifiers using statistical learning theory," *Computing Research Repository*, vol. abs/1003.5865, no. 1, 2010.
- [8] L. Masek, *Recognition of Human Iris Patterns for Biometric Identification*, Ph.D. thesis, School of Computer Science and Software Engineering, University of Western Australia, 2003.
- [9] C. Vivaracho-Pascual, M. Faundez-Zanuy, and J.M. Pascual, "An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances," *Pattern Recognition*, vol. 42, no. 1, pp. 183 – 193, 2009.
- [10] T.Y. Zhang and C.Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications ACM*, vol. 27, no. 3, pp. 236–239, 1984.
- [11] M. Blumenstein and X.Y. Liu, "A modified direction feature for cursive character recognition," in *Proceedings of the International Joint Conference on Neural Networks*. 2004, pp. 2983–2987, IEEE Computer Society.
- [12] X.Y. Liu and M. Blumenstein, "Experimental analysis of the modified direction feature for cursive character recognition," in *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*. 2004, pp. 353–358, IEEE Computer Society.
- [13] M. Blumenstein, B. Verma, and H. Basli, "A novel feature extraction technique for the recognition of segmented handwritten characters," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*. 2003, pp. 137–142, IEEE Computer Society.
- [14] J. F. Vargas, M. A. Ferrer, C. M. Travieso, and J. B. Alonso, "Off-line handwritten signature gpds-960 corpus," in *Ninth International Conference on Document Analysis and Recognition*. 2007, pp. 764 – 768, IEEE Computer Society.
- [15] V. Nguyen, M. Blumenstein, V. Muthukkumarasamy, and G. Leedham, "Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines," in *Ninth International Conference on Document Analysis and Recognition*. 2007, pp. 734 – 738, IEEE Computer Society.

Appendix B

Conference paper accepted for and presented at ACM SAICSIT '12

- Conference Name: South African Institute for Computer Scientists and Information Technologists 2012
- Conference Location: Centurion, South Africa
- Conference Date: 1 - 3 October 2012
- Conference Publisher: ACM

Local Directional Pattern Based Signature Verification using Weighted Fractional Distance Classification

Yaseen Moolla
University of KwaZulu-Natal
Durban, South Africa
Council for Scientific and
Industrial Research
Pretoria, South Africa
205500527@stu.ukzn.ac.za

Serestina Viriri
University of KwaZulu-Natal
Durban, South Africa
viriris@ukzn.ac.za

Jules R. Tapamo
University of KwaZulu-Natal
Durban, South Africa
tapamoj@ukzn.ac.za

Fulufhelo V.
Nelwamondo
Council for Scientific and
Industrial Research
Pretoria, South Africa
fnelwamondo@csir.co.za

ABSTRACT

Although handwritten signature verification has been extensively researched, it has not achieved optimum accuracy rate yet. Therefore, efficient and accurate signature verification techniques are required since signatures are still widely used as a means of personal verification. This paper presents an alternative efficient classification technique to supervised learning classification techniques. The signature features are extracted using the Local Directional Pattern algorithm, and classified using a combination of multiple distance-based techniques: *weighted Euclidean distance*, *fractional distance* and *weighted fractional distance*. This combination of multiple distance-based classification techniques achieved accuracy rate of 87.8%, which is comparable to a similar system that used Support Vector Machines, a supervised learning technique. Therefore, competitive levels of accuracy can be obtained using distance-based classification.

Categories and Subject Descriptors

I.5.3 [Pattern Recognition]: Clustering—*Similarity measures*; I.4.7 [Image Processing]: Feature Measurement—*Feature representation*

General Terms

Pattern Recognition; Biometrics

Keywords

Handwritten Signatures; Weighted Euclidean Distance; Fractional Distance

1. INTRODUCTION

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAICSIT '12, October 01-03 2012, Pretoria, South Africa
Copyright 2012 ACM 978-1-4503-1308-7/12/10\$15.00.

Biometrics is the use of one or more intrinsic physical or behavioural human characteristics to verify or identify a person. Biometric traits should be unique, universal, long lasting, collectible, commonly accepted, difficult to falsely duplicate and identifiable efficiently and accurately by machine. Examples of physiological biometric traits include fingerprints, DNA, hand and palm geometry and iris recognition, whereas examples of behavioural biometric (behaviometric) traits include voice recognition, writing patterns and signatures [6].

The basic process of automated biometrics verification involves capturing the biometric traits onto a machine and then using biometric feature extraction algorithms to create a digital representation template of the trait. For authentication of an individual, the system creates a biometric template from newly captured data and compares the two templates [6].

Signatures, which are one of the oldest used and most widely accepted biometric for identification and verification [13], are handwritten depictions of a person's name, nickname or other personal symbol. They are classified as a behavioural biometric trait.

There are two ways to capture signatures: online and offline. Offline signatures are static images while online signatures are dynamic and capture the progress of signature writing as a factor of time. Since online signatures hold a greater amount of information, they intrinsically allow greater accuracy than a static image. However, there are still many systems that require the improved accuracy of offline signatures. For instance, online signatures are not available for bank cheques or credit cards, and accurate offline signature verification is essential. The electronic writing pads for the capture of offline signatures are also much more cost effective than that for online signatures. Thus, the availability of competitively accurate offline signature verification could improve security measures for businesses in poorer emerging markets.

Many techniques exist for the classification of signatures and other biometrics. They can be broadly categorized into supervised learning techniques (SLTs) and distance-based classification techniques. SLTs include neural networks [4], hidden Markov models [1], support vector ma-

chines [11] and fuzzy logic [8]. Linear techniques include Euclidean distance, Mahalanobis [9], Manhattan distance, weighted Euclidean distances [10] and fractional distances [12].

SLTs, in general, provide a greater accuracy than basic distance-based classification techniques. This paper aims to combine several distance-based techniques to gain accuracy comparable with SLTs. The weighted Euclidean distance and fractional distance classification techniques are investigated. The two are then combined to create a novel weighted fractional distance classification. This paper aims to investigate this novel classification technique. The results are compared to those of a SLT that classified signatures using the same feature extraction techniques.

The feature extraction and classification techniques are described in section 2. The results obtained are discussed in section 3 and the conclusion is given in section 4.

2. TECHNIQUES AND METHODOLOGY

The first step in image processing based signature verification is to capture the signature image. The likelihood of this image containing noise is high. The next important step is then to preprocess this image, in order to remove noise and unwanted data, while preserving essential information for the final decision. Features are thereafter extracted and represented in a certain way; in this article, they are represented as vectors. A training is then carried out, where feature vectors from known authentic and forged signatures are compared for the calibration of the classification technique. The last step, classification, is where the system must accurately and independently determine whether signatures are authentic or forged.

Preprocessing includes smoothing, binarization, dilation and finding the bounding box of the signature image.

For feature extraction, Local Directional Pattern (LDP) is used. LDP is a gray-scale texture based feature method that characterizes the spatial structure; it was first introduced by Jabid *et al*[5], and used to classify gender; Kabir *et al*[7] later used one of its variant for facial expression recognition. More recently Ferrer *et. al* [3] performed signature verification with LDP.

In this paper we evaluate Local Directional Pattern performance in signature verification when combining multiple similarity measures, namely Euclidean distance, weighted Euclidean distance [10], fractional distances [12], and weighted fractional distances.

The feature extraction techniques used are described in section 2.1, and the classification techniques are described in section 2.2.

2.1 Feature Extraction

Feature extraction is performed using Local Directional Pattern[5]. This technique utilizes the 8 orientations of Kirsch masks, as shown in Figure 1, to detect the presence of edges or corners and their orientations. Values of the 8 mask orientations, m_0, m_1, \dots, m_7 , are obtained by performing a convolution of Kirsch masks with the image at each pixel, followed by a binarization. In other words, given the source image I_{src} , we will compute I_{LDP} , which is a transformed image using Algorithm 1.

In Figure 2, an example of Local Directional Pattern transformation of a source image $I_{src}(x, y)$ into a new image $I_{LDP}(x, y)$ is shown.

A histogram, H_{LDP} , can then be created from the image $I_{LDP}(x, y)$. However, since each 8-bit pixel has exactly three bits with the value 1 and 5 bits with the value 0,

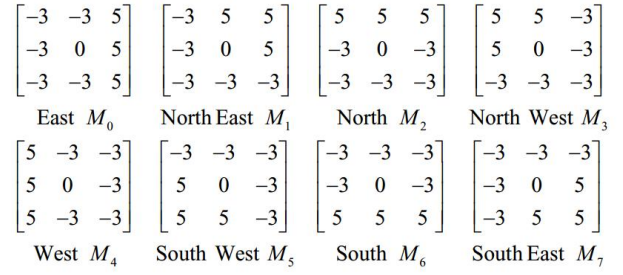


Figure 1: The 8 orientations of Kirsch Masks

this allows for only 56 possible permutations out of the usual 256. Therefore, the histogram will only account for these 56 possible values.

Further, it is possible to divide the image $I_{LDP}(x, y)$ into blocks by splitting it a specified number of parts vertically ($split_V$) and horizontally ($split_H$) and have a 56-value histogram for each block. The final feature vector, FV_{LDP} , is then obtained by concatenating all of these histograms, $FV = H_{LDP}^1 + H_{LDP}^2 + \dots + H_{LDP}^{split_V \times split_H}$.

Algorithm 1 Local Directional Pattern, I_{LDP} , calculation of an image, I_{src}

Require: I_{src} , \triangleright Source Image
Ensure: I_{LDP} , \triangleright Image Transformed using Local Directional Pattern

```

1: for each pixel (x,y) do
2:   for i=0 to 7 do
3:     for k= - 1 to 1 do
4:       for l = -1 to 1 do
5:          $m_i = m_i + M_i(k + 1, l + 1) \times I_{src}(x + k, y + l)$ 
6:       end for
7:     end for
8:     Transform the three highest values  $m_i$  into
       1s and the rest into 0s
9:   end for
10:   $powerof2 = 1$ 
11:   $I_{LDP}(x, y) = 0$ 
12:  for i = 0 to 7 do
13:     $I_{LDP}(x, y) = I_{LDP}(x, y) + m_i \times powerof2$ 
14:     $powerof2 = 2 \times powerof2$ 
15:  end for
16: end for
```

2.2 Classification

2.2.1 Euclidean distance and thresholds

One of the most common distance-based classification techniques for determining the accuracy of biometric systems is the calculation of the Euclidean distance between a reference vector (derived as a mean of several authentic signatures of an individual) and other feature vectors.

Authentic signatures are expected to have Euclidean distance values below a certain threshold while forged signatures would have values above that threshold. Authentic signatures with distances above the threshold are regarded as false negatives and contribute to the False Rejection Rate (FRR) while forged signatures with distances below the threshold are regarded as false positives and contribute to the False Acceptance Rate (FAR).

The equation for determining the Euclidean distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$

	85	32	26					
	53	50	10					
	60	38	45					
	↓							
Mask index	m ₇	m ₆	m ₅	m ₄	m ₃	m ₂	m ₁	m ₀
Mask value	-303	97	161	537	313	97	503	-399
Rank	5	7	6	1	4	8	2	3
Code bit	0	0	0	1	0	0	1	1
LDP code	19							

Figure 2: Calculation of the LDP code [3]

is computed as defined in equation (1).

$$\|x - y\|_p = (\sum_{i=1}^m |(x^i - y^i)|^p)^{1/p} \quad (1)$$

where $p = 2$.

2.2.2 Weighted Euclidean distance

The weighted Euclidean distance measure is a technique adapted from [10] to improve the classification accuracy by adding weight. In our case we add weight that has statistical importance, related to the most reliable features from the feature vector.

Given a set of signatures $S = \{sig_1, sig_2, \dots, sig_n\}$, represented by their feature vectors spaces as

$$sig_j = (sig_j^1, sig_j^2, \dots, sig_j^m)$$

where $j = 1, 2, \dots, n$, and sig_i^j is the i^{th} component of the j^{th} signature, with $i = 1, 2, \dots, m$.

We will compute the standard deviation of each component of reference signatures as follows:

The mean μ^i is computed as

$$\mu^i = \frac{1}{n} \sum_{l=1}^n sig_l^i \quad (2)$$

and the standard deviation σ^i is defined as

$$\sigma^i = \sqrt{\frac{1}{n} \sum_{l=1}^n (sig_l^i - \mu^i)^2} \quad (3)$$

The weighted Euclidean distance between two signature $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ can then be calculated using the standard deviation as shown in Equation 4

$$\|x - y\|_p = \left(\sum_{j=1}^m \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \quad (4)$$

where $p = 2$.

2.2.3 Fractional distances

A drawback of using Euclidean and other p -norm distances where $p \in \mathbb{N}_1$ is that as the vectors get larger, the distance values tend to cluster[2]. This is called the concentration phenomenon. To overcome this limitation of distance-based distance classification, Vivaracho-Pascual et. al. [12] introduced the use of fractional p -norm distances.

The equation for determining fractional p -norm distance between vectors x and y is computed as defined in equation (1) where $0.1 \leq p \leq 2.0$.

2.2.4 Weighted fractional distances

The fractional distances and weighted Euclidean distance can then be combined to form the weighted fractional distance as defined in equation (4) where $0.1 \leq p \leq 2.0$. The optimal value of p is calculated experimentally for each user.

3. RESULTS AND DISCUSSION

3.1 Data Set

The Grupo de Procesado Digital de Senales (GPDS) signature database is used in the analysis of the techniques. The database consists of authentic and forged signature sets for 300 different individuals. There are 24 authentic signatures and 30 skilled forgeries per each individual. All signatures are stored in black and white.

10 authentic signatures are used to create the reference signature, the other 14 authentic signatures and the 30 skilled forgeries are used for the classification and verification. For testing random forgeries with each individual, a single authentic signature from each of the other 299 individuals is used.

3.2 Thresholding

Quantification of result accuracy is measured in terms of the False Rejection Rate (FRR); False Acceptance Rate (FAR) which is further broken down into FAR for skilled forgeries (FARS) and FAR for random forgeries (FARR); and the Equal Error Rate (ERR), which is the point at which the FRR and FARS are equal.

The reference feature vector is created by calculating the mean of each feature in the feature vectors of 10 authentic signatures. The distance between the reference feature vector and each of the remaining 14 authentic signatures is determined using one of the Euclidean, fractional, weighted Euclidean or weighted fractional distance equations. A threshold distance value is determined. All distances below the threshold are regarded as authentic, while all distances about are regarded as forgeries. The FRR is calculated from this. Likewise, the FARS and FARR are obtained from the distance calculated between the reference feature vector and the skilled and random forged signatures respectively.

The optimal threshold distance value is determined locally for each individual as the value that produces the EER. This is the threshold where the FRR and FARS are equal. In Fig. 3, this is represented as the convergence point of the FRR and FARS on the ROC curve. The FARR is then determined using this threshold value.

3.3 Results

Tests are performed using the distance-based classification techniques. These techniques are the Euclidean, fractional, weighted Euclidean and weighted fractional distances. For each classification technique, multiple vertical (*split_V*) and horizontal (*split_H*) split sizes are tested. The number of splits are in the range of 1, i.e. no splits, to 8, providing a total of 49 different combinations of splits which were applied globally to all individuals. This is done to find the best possible global feature vector size per classification technique and investigate the effect of the weighted distances and fractional distances on the concentration phenomenon. The feature vector size is determined by *split_H* \times *split_V* \times H where H is the length of the histogram, which is always 56 in the LDP extraction technique. For the best classification technique with the

Table 1: The effect of different $split_H$ and $split_V$ on EER(%) using the Euclidean distance

H \ V	1	2	3	4	5	6	7
1	25.0	23.9	22.6	23.0	22.6	22.8	23.0
2	23.8	22.3	21.7	22.0	22.0	22.3	22.4
3	24.0	22.7	21.9	21.9	22.2	22.3	22.6
4	24.1	23.0	22.5	22.5	22.8	23.0	23.1
5	24.3	23.7	23.1	23.2	23.5	23.6	23.9
6	24.8	24.0	23.4	23.7	24.0	24.2	24.5
7	25.2	24.4	24.2	24.1	24.5	25.0	25.3

Table 2: The effect of different $split_H$ and $split_V$ on EER(%) using fractional distances

H \ V	1	2	3	4	5	6	7
1	18.9	17.7	17.0	16.9	16.7	16.6	16.6
2	18.0	16.8	16.2	16.1	16.1	15.9	16.0
3	17.5	16.5	15.6	15.7	15.4	15.2	15.2
4	17.5	16.2	15.9	15.6	15.4	15.3	15.3
5	17.5	16.3	15.3	15.4	15.3	15.2	15.0
6	17.7	16.1	15.7	15.4	15.2	15.0	14.9
7	17.4	16.2	15.4	15.3	15.0	14.8	14.7

Table 3: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Euclidean distance

H \ V	1	2	3	4	5	6	7
1	18.6	17.9	16.8	16.5	16.1	16.0	16.0
2	18.4	16.5	15.8	15.7	15.4	15.3	15.2
3	17.3	16.0	15.5	15.3	14.6	14.8	14.6
4	17.4	15.7	15.2	15.1	14.5	14.8	14.8
5	17.3	15.6	14.7	14.8	14.6	14.6	14.7
6	17.0	15.7	14.8	15.0	14.7	14.8	14.7
7	16.9	15.5	14.8	14.9	15.0	14.7	15.0

Table 4: The effect of different $split_H$ and $split_V$ on EER(%) using weighted fractional distances

H \ V	1	2	3	4	5	6	7
1	16.8	16.1	15.1	14.8	14.5	14.4	14.2
2	16.7	14.1	14.2	14.0	13.7	13.4	13.3
3	15.7	14.3	13.6	13.2	12.7	12.6	12.5
4	15.8	14.0	13.3	13.0	12.5	12.7	12.4
5	15.3	14.1	12.8	12.7	12.5	12.3	12.3
6	15.1	13.9	12.9	12.7	12.6	12.5	12.3
7	15.0	13.8	12.9	12.7	12.5	12.3	12.2

lowest EER, the FARR is also shown.

The first analysis is performed with the Euclidean distance and the results are shown in Table 1. The worst

Table 5: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted fractional distances

H \ V	1	2	3	4	5	6	7
1	1.08	0.70	0.47	0.43	0.52	0.57	0.57
2	0.65	0.48	0.54	0.84	0.81	0.93	1.22
3	0.54	0.42	0.66	0.75	0.75	0.81	1.09
4	0.56	0.53	0.23	0.91	1.00	1.11	1.14
5	0.63	0.79	1.00	1.02	1.20	1.28	1.28
6	0.68	0.76	1.16	1.08	1.34	1.36	1.45
7	0.68	0.88	1.18	1.36	1.44	1.63	1.64

EER when using the Euclidean distance was found to occur with the largest feature vector size, i.e. $7 \text{ vertical} \times 7 \text{ horizontal}$ splits. This EER was 25.3%. Conversely, the best EER occurred with one of the smallest feature vector sizes, i.e. 3×2 splits. This EER was 21.66%. The accuracy trend with standard Euclidean distance measuring was that the EER was poor for the smallest feature vectors, it then improved as the feature vectors became larger and more detailed, but then worsened once the feature vectors became too large. This reduction in accuracy confirms that the concentration phenomenon occurs with large feature vectors because the Euclidean distance causes the distance values to cluster. The weighted Euclidean distance and fractional distance discussed below are used to counter this phenomenon.

The second analysis is performed with the fractional distance and the results are shown in Table 2. The classification technique produced a higher accuracy in comparison with the Euclidean distance. The worst EER occurred with the smallest feature vector, i.e. 18.9% with 1×1 splits. This worst EER is still better than the best EER of the Euclidean distance, which was 21.7%. As the feature vectors increase in size, the accuracy of this classification technique gradually improves. The best EER was achieved with the largest feature vector, i.e. 14.7% with 7×7 splits. This shows that the fractional distances overcome the concentration phenomenon that adversely affects larger feature vectors when the Euclidean distance is applied. The best EER of the fractional distance is better than the best Euclidean distance EER by almost 7%. Larger feature vectors were tested, but have been excluded due to subsequent decreasing accuracies and lack of space.

The third analysis is performed with the weighted Euclidean distance and the results are shown in Table 3. This classification technique also produced a higher accuracy in comparison to the Euclidean distance and while also producing a slightly higher accuracy than the fractional distance. Once again, the worst EER occurred with the smallest feature vector, i.e. 18.6% with 1×1 splits. The best EER occurred among one of the largest feature vectors, i.e. 14.5% with 5×4 splits. This is better than both the Euclidean and fractional distances. While the weighted Euclidean distance does not counter the concentration phenomenon completely, it is effective enough to produce a good accuracy.

The fourth and final analysis is performed with the weighted fractional distance, which is a combination of the weighted and fractional distance techniques. The results are shown in Table 4. This combined technique

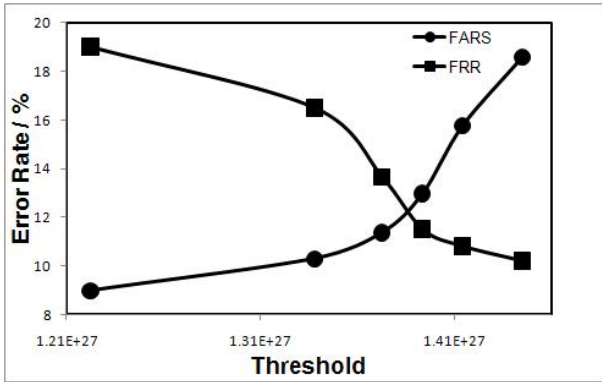


Figure 3: Calculation of EER on applying the weighted fractional distances with 7x7 splitting

provides the best accuracy from all four of the compared techniques. Just like with the fractional and weighted Euclidean distances, the worst EER occurred with the smallest feature vector, i.e. 16.8% with 1×1 splits. Just like with the fractional distance, the best EER was obtained with the largest feature vector, i.e. 12.2% with 7×7 splits. This EER is the best from all the compared techniques. It provides a 9.5% increase in accuracy in comparison with the best Euclidean distance EER. The FARR values for the weighted fractional distance are shown in Table 5. Conversely, it can be seen that increased feature vector sizes cause the undesired effect of an increase in the FARR.

The greatest accuracy was achieved by a combination of weighted and fractional distances with an EER of 12.2%. This is more accurate by almost 10% in comparison to the best standard Euclidean distance measure.

3.4 Literature Comparison

The results in this paper are compared to those of Ferrer et. al. [3], who used the same feature extraction technique, the Local Directional Pattern; and the same signature database, GPDS300. However, they used a different classification technique, namely, Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel. By using these techniques, they obtained an EER of 17.8% and FARR of 0.68%. In comparison, we obtained a best EER of 12.2% with an FARR of 1.64% by using weighted fractional distance classification instead.

While both systems obtained an FARR of below 2%, the better FARR was obtained by Ferrer et. al. This can be attributed to the use of random forgeries in the training of their classifier, which was not performed in our system.

While further tests need to be performed, especially with regards to the training set used, and different feature extractions from literature, the results are promising.

4. CONCLUSION

An alternative efficient handwritten signature classification technique has been presented. The proposed technique uses a combination of multiple distance-based classification techniques: *weighted Euclidean distance*, *fractional distance* and *weighted fractional distance*, to verify the Local Directional Pattern-based signature features. Experimental results show that the proposed approach obtains an Equal Error Rate (EER) of 12.2% on skilled forgeries. This is comparable to literature results where supervised learning techniques were applied to the same feature extraction techniques, which obtained an EER of 17.8%.

Therefore, a combination of multiple distance-based classification techniques is an alternative signature classification technique with an accuracy rate of 87.8%.

5. REFERENCES

- [1] L. Batista, E. Granger, and R. Sabourin. Improving performance of HMM-based off-line signature verification systems through a multi-hypothesis approach. *International Journal on Document Analysis and Recognition*, 13:33–47, March 2010.
- [2] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is “nearest neighbor” meaningful? In *Int. Conf. on Database Theory*, pages 217–235, 1999.
- [3] M. Ferrer, F. Vargas, C. Traviesto, and J. Alonso. Signature verification using local directional pattern. In *International Carnahan Conference on Security Technology*, pages 336–340. IEEE, Oct 2010.
- [4] I. Ismail, M. Ramadan, T. S. El-Danaf, and A. Samak. An efficient off line signature identification method based on fourier descriptor and chain codes. *International Journal of Computer Science and Network Security*, 10(5):29–35, 2010.
- [5] T. Jabid, H. Kabir, and O. Chae. Gender classification using local directional pattern (LDP). In I. C. Society, editor, *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR’10, pages 2162–2165, 2010.
- [6] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14:4–20, 2004.
- [7] H. Kabir, T. Jabid, and O. Chae. Local directional pattern variance(LDPv): A roust feature descriptor for facial expression recognition. *The International Arab Journal of Information Tecnology*, 9(4):382–391, July 2012.
- [8] H. B. Kekre, V. A. Bharadi, S. Gupta, A. A. Ambardekar, and V. B. Kulkarni. Off-line signature recognition using morphological pixel variance analysis. In *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*, ICWET ’10, pages 3–10. ACM, 2010.
- [9] D. Kisku, P. Gupta, and J. Sing. Offline signature identification by fusion of multiple classifiers using statistical learning theory. *Computing Research Repository*, abs/1003.5865(1), 2010.
- [10] L. Masek. *Recognition of Human Iris Patterns for Biometric Identification*. PhD thesis, School of Computer Science and Software Engineering, University of Western Australia, 2003.
- [11] V. Nguyen, M. Blumenstein, and G. Leedham. Global features for the off-line signature verification problem. In *10th International Conference on Document Analysis and Recognition*, pages 1300–1304, July 2009.
- [12] C. Vivaracho-Pascual, M. Faundez-Zanuy, and J. M. Pascual. An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances. *Pattern Recognition*, 42(1):183 – 193, 2009.
- [13] D. Zhang and A. K. Jain, editors. *First International Conference on Biometric Authentication*, volume 3072 of *Lecture Notes in Computer Science*, Hong Kong, July 2004. Springer Verlag.

Appendix C

Journal paper submitted to South African Computer Journal

- Journal Name: South African Computer Journal
- Submission date: 4 December 2012
- Authors: Y. Moolla, S. Viriri, F.V. Nelwamondo, J.R. Tapamo

Offline Signature Verification using Locally Optimized Distance-based Classification

ABSTRACT

Although handwritten signature verification has been extensively researched, it has not achieved an optimal classification accuracy rate. Therefore, efficient and accurate signature verification techniques are required since signatures are still widely used as a means of personal verification. This research work presents efficient distance-based classification techniques as an alternative to supervised learning classification techniques (SLTs). The Local Directional Pattern (LDP) feature extraction technique was used to analyze the effect of using several different distance-based classification techniques. The classification techniques tested, are the Euclidean, Manhattan, Fractional, weighted Euclidean, weighted Manhattan, weighted fractional distances and individually optimized resampling of feature vector sizes. The best accuracy, of 90.8%, was achieved through applying a combination of the weighted fractional distances and locally optimized resampling classification techniques to the Local Directional Pattern feature extraction. These results are compared with results from literature, where the same feature extraction technique was classified with SLTs. The distance-based classification was found to produce greater accuracy than the SLTs.

KEYWORDS: Biometrics, pattern recognition, distance-based classification

1 INTRODUCTION

Biometrics is the measure of human characteristics for authentication or identification of individuals. Biometric modalities are regularly becoming an important aspect of automated electronic security systems. For such a system to be successful, it requires methods and techniques that produce high accuracy.

These systems are used either for recognition or verification. Recognition entails the identification of a biometric trait, or set of traits, as belonging to a specific individual from a given set of individuals. Conversely, verification entails authenticating a claim that a biometric trait, or set of traits, belongs to a specific individual.

One of the most common and widely accepted biometric modality is the handwritten signature. It has been used for verification since before the advent of electronic machines. Due to its wide-spread use and acceptance, handwritten signatures are an ideal candidate for automated biometric verification systems. The two methods of capture for signatures are categorized as offline and online. For offline capture, a static image of a completed signature is recorded. For online capture, the creation of a signature is recorded as a function of time.

Offline signature verification systems have much potential for world wide usage and there are several cases in which only offline signatures can be used. This includes automated authentication of bank cheques and legal documents. Additionally, equipment for the capture of offline signatures is cheaper, which will allow for greater adoption of automated signature verification, especially for small-to-medium businesses and in developing economies.

A high accuracy is required for the acceptance of such a system, but offline signature verification systems are hindered by low accuracy rates, having not yet reached an acceptable level of accuracy. These low accuracy rates are due to the inherent randomness that is characteristic of behavioural biometrics. Static signatures also have much less information in comparison to online signatures, since the time dimension is not present.

There is much current research that attempts various techniques to improve the accuracy of offline signature verification systems. Some of the research focuses on novel, new feature extraction techniques. These are designed to extract the most relevant and constant features of a signature while also attempting to reduce the extraction of features that are not stable or constant in an individuals signatures.

Other research concentrates on using different classification techniques and finding the optimal classification technique for a particular feature set. Classification can be categorized into supervised learning techniques (SLTs) and distance-based measures. Most offline classification emphasizes the use of different SLTs, such as support vector machines, hidden Markov models and artificial neural networks. There is very little research into using different distance-based classification techniques, except for the Euclidean distance, which is the most commonly known distance measure in geometric space.

There are many other distance-based measures that could be used for classification. This work improves the accuracy of offline signature verification systems by investigating the use of some of these techniques, specifically weighted and fractional distances in L^P -space, and applies them to various different fea-

ture extraction techniques. The results obtained are compared with those from literature. In particular, the distance-based classification techniques are compared with SLTs with the use of the same feature extraction techniques.

1.1 Literature Survey

Many techniques exist for the classification of signatures and other biometrics. They can be broadly categorized into supervised learning techniques (SLTs) and distance-based classification techniques. SLTs include neural networks, hidden Markov models (HMM) [1], support vector machines (SVM) [2] and fuzzy logic. Linear techniques include Euclidean distance, Mahalanobis, Manhattan distance, weighted Euclidean distances [3] and fractional distances [4]. Some of the notable, recent works are discussed below.

Kovari et al. [5] uses artificial neural networks for classifying a feature vector comprised of both local and global features, and achieves an EER of just over 20%. Coetzer et al. [6] applies HMM classification to discrete Radon transform, which is a global feature, and achieved an EER of 12.2% using skilled forgeries and 4.5% using random forgeries. Panton and Coetzer further improves the EER to 8.6% by using a fusion of HMM classifiers and adding local features to the feature vector [7]. Yilmaz et al. [8] performs SVM classification on a combination of gradient-based and binary pattern-based features. It is found that user-dependent, also called user-specific, classifier worked better than a globally applied classifier for all users. Thy achieved a best AER of 15.41%. Vargas et al. [9] proposes a system with least squares SVM classification on features extracted from the Fourier transform of a signature image. An EER of 6.20% is obtained, which, by literature comparison, performs better than similar techniques, using the same database. Batista et al. [10] uses the statistical fusion of HMM classifiers via a multi-hypothesis approach and user specific codebooks. An AER of 7.79% is achieved. The best recent results for SLTs is by using multi-hypothesis fusion approaches. Ferrer et al. [11] performs a comparison between the Euclidean distance, HMM and SVM, using geometric features based on Cartesian and polar coordinates. It is found that, for the feature extractions used, the best performance was for HMM, with and FRR of 14.1% and FAR of 12.3%, using skilled forgeries. The SVM performed better than the Euclidean distance, but worse than the HMM. Kisku et al. [12] develops another multi-hypothesis approach where Euclidean distance, Mahalanobis distance and Gaussian empirical rule results are fused together using SVM. A private signature database, comprising of 9 authentic signatures and a single forged signature were collected for each of 180 individuals. Individually, each classification technique performs very well with EER values below 10%, and combined, a best EER of 2.15% is achieved. While multi-hypothesis techniques can greatly improve accuracy, they also require much more processing for training and testing,

since multiple classifiers are used in place of a single classifier.

The Mahalanobis distance was first documented in 1936 by Prasanta Mahalanobis [13]. It performs best with multivariate normal data distributions [14]. Fang et al. [15] reports one of the earliest classifications using the Mahalanobis distance. A best EER of 19.1% was achieved. Nguyen et al. [16] also compares squared Mahalanobis distance and Gaussian kernel SVM classification on a local gradient-based feature extraction. The AER for the SVM is 15.02%. The gradient distance with SVM classification performs better than the squared Mahalanobis distance, and better than their previous tests with MDF feature extraction and SVM classification. Sigari et al. [17] performs Mahalanobis distance classification on features extracted using Gabor wavelets. Verification tests are performed on 3 relatively small, yet diverse, signature databases. EER values of 15.0%, 16.8% and 9.0% were obtained. The Mahalanobis distance performs best with multivariate normal data distributions [14]. This limits its applicability to feature extraction techniques.

The most well known distance-based measure is the Euclidean distance. There are many works that have used the Euclidean distance for authentication due to its simplicity of implementation. Shekar et al. [18] uses the Euclidean distance for the verification of feature vectors created using an Eigenvector-based feature extraction technique. Different sizes of feature vectors were tested, and a single globally applied feature vector size was eventually chosen. An EER of 14.33% is achieved when using 10 signatures for training and 14 for testing. A better EER of 8.78% is achieved when using 15 signatures for training and 9 for testing, but it is noted that in compared works, the former configuration of training and testing samples were used. Rekik et al. [19] also tests global and local feature extractions with Euclidean classification. A fusion of local and global features is found to perform better than using local or global features independently. The best EER with skilled forgeries is 11.0% using signatures from 75 different individuals. Ramachandra et al. [20] finds the smallest Euclidean distance between cross-validated graphs of signatures, using the Hungarian method [21]. Various feature vector sizes are tested and an EER of 27.78% is achieved using skilled forgeries and 6.25% using random forgeries for two different vector sizes. The Euclidean distance accuracy is limited by the concentration of distances of large feature vectors.

The weighted Euclidean distance is not a commonly used classification technique. Zhu et al. [3] uses the weighted Euclidean distance for iris recognition in 2000. Alizadeh et al. [22] uses the weighted Euclidean distance for online signature verification. This is a promising technique that can improve the accuracy of distance-based classification.

Fractional distances are another uncommon, yet promising, classification technique. They have not been used in offline signature verification before. How-

ever, they have been used for online signature recognition in Vivaracho-Pascual et al. [4], and for face recognition in Espinosa-Dur et al. [23].

In this work, the concept of weighting by standard deviation is applied to the fractional distance to produce a novel distance-based classification technique, called the weighted fractional distance. Additionally, locally optimized thresholding is optimized by choosing the optimal feature vector length per individual.

2 TECHNIQUES AND METHODOLOGY

2.1 Design Overview

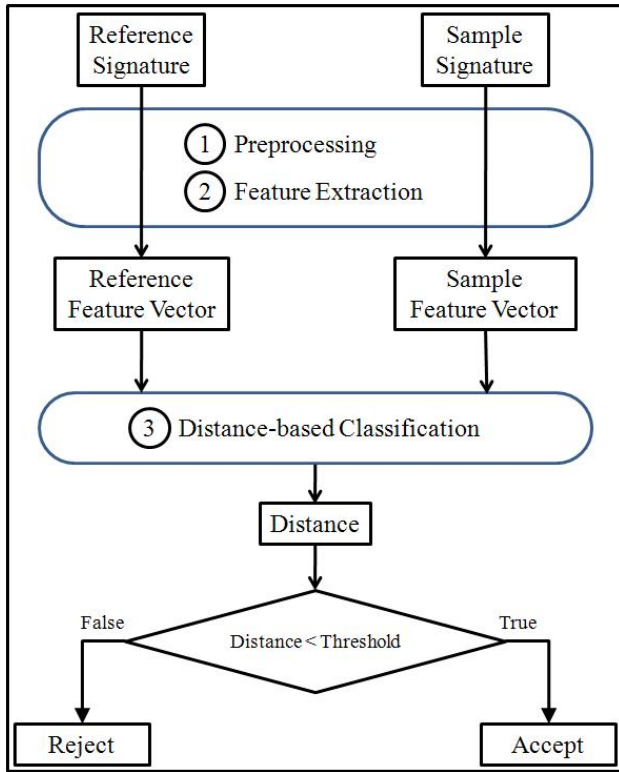


Figure 1: Overview of the design of a biometric verification system

Figure 1 shows the overview of the verification system. The three main steps are preprocessing, feature extraction and classification. Preprocessing prepares the signature image for the extraction of features which are then used to classify the signature as authentic or forged. For feature extraction, the Local Directional Pattern (LDP) features are used. Before feature extraction, the required preprocessing steps are binarization, dilation and bounding box extraction. For classification, distance-based classification techniques, such as the Euclidean, Manhattan and fractional distances are tested.

In the training phase, the reference feature vector is determined by averaging a subset of randomly chosen authentic feature vectors. Sample feature vectors are obtained from other authentic and forged signatures individually. The classifier is then trained to determine the threshold that provides the optimal ac-

curacy for the system, using sample feature vectors of known classification as input. In the testing phase, further sample signatures are used, but the classifier will determine their classification independently based on the threshold obtained from the training phase. Accuracy is gauged based on the number of signatures that the classifier correctly accepts as authentic and correctly rejects as forgeries.

2.2 Feature Extraction

2.3 Local Directional Pattern

The Local Directional Pattern (LDP) [24] [25] is a gray-scale texture based feature method that characterizes the spatial structure of an image. It utilizes the 8 orientations of Kirsch masks, as shown in Figure 2, to detect the presence of edges or corners and their orientations. Values of the 8 mask orientations, m_0, m_1, \dots, m_7 , are obtained by performing a convolution of Kirsch masks with the image at each pixel, followed by a binarization. In other words, given the source image I_{src} , we will compute I_{LDP} , which is a transformed image using Algorithm 1.

Algorithm 1 Local Directional Pattern, I_{LDP} , calculation of an image, I_{src}

Require: I_{src} , \triangleright Source Image
Ensure: I_{LDP} , \triangleright Image Transformed using Local Directional Pattern

```

1: for each pixel (x,y) do
2:   for i=0 to 7 do
3:     for k= - 1 to 1 do
4:       for l = -1 to 1 do
5:          $m_i = m_i + M_i(k+1, l+1) \times I_{src}(x+k, y+l)$ 
6:       end for
7:     end for
8:     Transform the three highest values  $m_i$ 
       into 1s and the rest into 0s
9:   end for
10:   $powerof2 = 1$ 
11:   $I_{LDP}(x,y) = 0$ 
12:  for i = 0 to 7 do
13:     $I_{LDP}(x,y) = I_{LDP}(x,y) + m_i \times powerof2$ 
14:     $powerof2 = 2 \times powerof2$ 
15:  end for
16: end for
  
```

In Figure 3, an example of Local Directional Pattern transformation of a source image $I_{src}(x, y)$ into a new image $I_{LDP}(x, y)$ is shown.

A histogram, H_{LDP} , can then be created from the image $I_{LDP}(x, y)$. However, since each 8-bit pixel has exactly three bits with the value 1 and 5 bits with the value 0, this allows for only 56 possible permutations out of the usual 256. Therefore, the histogram will only account for these 56 possible values. A sample histogram derived from a signature image is shown in Figure 4.

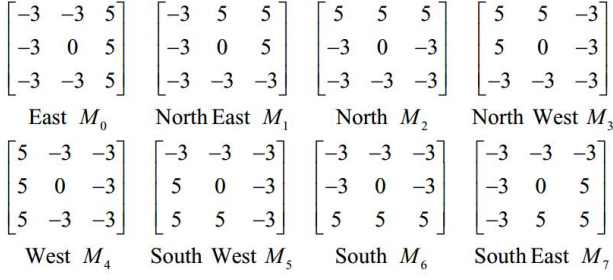


Figure 2: The 8 orientations of Kirsch Masks [25]. Each orientation is applied to a pixel and its 8 neighbours to calculate 8 mask values.

	85	31	26					
	53	50	10					
	60	38	45					
	↓							
Mask index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask value	-300	100	164	540	308	92	-508	-396
Rank	5	7	6	1	4	8	2	3
Code bit	0	0	0	1	0	0	1	1
LDP code	19							

Figure 3: calculation of the LDP code [25] obtained by applying each of the 8 Kirsch masks

2.4 Feature Vectors and Resampling

Further, it is possible to divide the image $I_{LDP}(x, y)$ into blocks by splitting it a specified number of parts vertically ($split_V$) and horizontally ($split_H$) and have a 56-value histogram for each block. The final feature vector, FV_{LDP} , is then obtained by concatenating all of these histograms, $FV = H_{LDP}^1 + H_{LDP}^2 + \dots + H_{LDP}^{split_V \times split_H}$. A sample signature after binarization, dilation, cropping and splitting is shown in Figure 5.

In this work, multiple feature vector sizes are analyzed so as to understand the effect of resampling on the LDP. Feature vectors for between 1 and 8 vertical splits, $split_V$, and horizontal splits, $split_H$, are tested. Since each histogram has a size of 56 and they are concatenated, the feature vector size for LDP feature extraction is calculated as

$$FV_{LDP} = 56 \times split_V \times split_H \quad (1)$$

Resampling will be discussed further in the Section 2.5.6 which discusses Individual Optimized Resampling.

2.5 Classification

2.5.1 Determining the Threshold

Authentic signatures are expected to have distance values below a certain threshold while forged signatures would have values above that threshold. This threshold is determined by finding the optimal Equal Error Rate (EER) during the training phase. Authentic signatures with distances above the threshold are

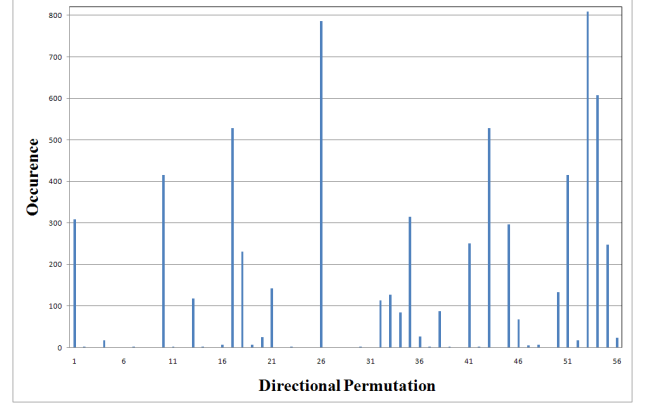


Figure 4: A sample LDP histogram showing the occurrences of each directional permutation from the image in Figure 5 with no splits



Figure 5: A dilated image with splits segmented by 3 horizontal splits and 4 vertical splits

regarded as false negatives and contribute to the False Rejection Rate (FRR) while forged signatures with distances below the threshold are regarded as false positives and contribute to the False Acceptance Rate (FAR). This is further split into the FAR for skilled forgeries (FARS) and for random forgeries (FARR). The threshold is chosen where the distance for the FRR and FARS are equal. This rate is also called the Equal Error Rate (ERR).

2.5.2 Euclidean Distance

One of the most common distance-based classification techniques for determining the accuracy of biometric systems is the calculation of the Euclidean distance between a reference vector (derived as a mean of several authentic signatures of an individual) and other feature vectors. The Euclidean distance is used to calculate the distance between points in a Cartesian plane. It is a distance calculated in L^p -space, also called Lebesgue space, where the p -norm value is 2, or the L_2 distance.

The equation for determining the Euclidean distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is defined as

$$\|x - y\|_p = \left(\sum_{i=1}^m |(x^i - y^i)|^p \right)^{1/p} \quad (2)$$

where $p = 2$. This makes the equation

$$\|x - y\|_2 = \left(\sum_{i=1}^m |(x^i - y^i)|^2 \right)^{1/2} \quad (3)$$

2.5.3 Manhattan Distance

The Manhattan distance, also called the City-block distance, is the distance between two points determined as the sum of the absolute difference of their respective coordinates. The equation for determining the Manhattan distance between vectors $x = (x^i)_{i=1,2,\dots,m}$ and $y = (y^i)_{i=1,2,\dots,m}$ is computed as defined in Equation (2) where $p = 1$. Therefore, Equation (2) can be re-written as

$$\|x - y\|_1 = \sum_{i=1}^m |(x^i - y^i)| \quad (4)$$

2.5.4 Fractional Distance

The fractional distance is another distance in L^P -space where the p -norm value, also called the Minkowski norm exponent, is any fractional value less than 1. Francois and Wertz [26] discusses the use of fractional distance as an alternative to the Euclidean distance to counteract the concentration phenomenon. This phenomenon is when large feature vectors cause the results of the Euclidean distance to concentrate, or cluster. This clustering of values, which is an intrinsic property of L^P -space distances, makes classification difficult for large feature vectors. Fractional distances generally produce less concentrated results than the Euclidean distance, which allows for better classification of data sets.

Francois et al. states “Fractional norms are not always less concentrated than other norms. They seem, however, to be more relevant as a measure of similarity when the noise affecting the data is strongly non-Gaussian.” [26] Much of the noise generated by behavioral biometrics is due to random variations in human action, that may not follow a normal distribution. This makes fractional distances a viable investigative route of classifying offline handwritten signatures.

The equation for determining fractional p -norm distance between vectors x and y is computed as

$$\min(\|x - y\|_p) = (\sum |(x - y)|^p)^{1/p} \quad (5)$$

where $0.1 \leq p \leq 2.0$.

The optimal value of p is when the distance calculated using Equation (5) is at its minimum for all values of p within the given range.

2.5.5 Weighted Distances in L^P -space

The weighted Euclidean distance measure is a technique adapted from [3] to improve the classification accuracy by adding weight, or statistical importance, to the most reliable features from the feature vector. Firstly, the standard deviation for the reference signatures is obtained.

Let the n reference signatures be

$$\begin{aligned} x_1 &= (x_1^1, x_1^2, \dots, x_1^m) \\ x_2 &= (x_2^1, x_2^2, \dots, x_2^m) \\ &\vdots \\ x_n &= (x_n^1, x_n^2, \dots, x_n^m) \end{aligned} \quad (6)$$

Let x_i^j be the j^{th} component of the i^{th} reference signature where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Then the mean of the j^{th} component the reference signatures, μ^j , is computed as

$$\mu^j = \frac{1}{n} \sum_{i=1}^n x_i^j \quad (7)$$

and their standard deviation σ^j is defined as

$$\sigma^j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^j - \mu^j)^2} \quad (8)$$

The weighted Euclidean distance is then calculated using the standard deviation

$$\|x - y\|_p = \left(\sum_{j=1}^m \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \quad (9)$$

where $p = 2$.

The equation can be rewritten as

$$\|x - y\|_2 = \left(\sum_{j=1}^m \frac{|(x^j - y^j)|^2}{\sigma^j} \right)^{1/2} \quad (10)$$

Further, the Manhattan distance and weighted Euclidean distance are combined to form the weighted Manhattan distance

$$\min(\|x - y\|_1) = \sum_{j=1}^m \frac{|(x^j - y^j)|}{\sigma^j} \quad (11)$$

The fractional distances and weighted Euclidean distance are also combined to form the weighted fractional distance

$$\min(\|x - y\|_p) = \left(\sum_{j=1}^m \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \quad (12)$$

where $0.1 \leq p \leq 2.0$

As with Equation 5, the optimal value of p is when the distance calculated using Equation 12 is at its minimum for all values of p within the given range.

2.5.6 Individually Optimized Resampling

Resampling of the feature vector allows it to be resized. This is a form of spatial normalization. Different resampling sizes results in changing accuracies. It is expected that choosing the best resampling size per user will optimize the overall accuracy of the system. When a feature vector is resampled, its size is

normalized to produce a uniform feature vector size, either globally for all signature sets or locally per individual set. Vivaracho-Pascual et. al. [4] try several resampling sizes for their feature vectors, in research with online signatures. They note that there is no single resampling size that is optimal for all signatures. They further state that local optimization, obtained by choosing the best feature vector size per individual, is a non-trivial approach and interesting for future study. However, for their work, they chose a global resampling size for all individuals.

The resampling method used for LDP is described in Sections 2.3 and 2.4. In the works of Ferrer et. al. [25], the signatures are split into 4 blocks vertically and 3 blocks horizontally, giving a total of 12 blocks. It is unclear if other split combinations were tested. In their work, each block overlapped by 60% and feature vectors were transformed using the discrete cosine transform before classification with a support vector machine. In this work, images are tested with block splits between 1 and 8 in both the vertical and horizontal directions. This provides block numbers between 1 and 64 blocks per image.

3 RESULTS AND DISCUSSION

Tables 1 to 12 show the results of tests on the LDP using different numbers of horizontal and vertical splits ($split_H$ and $split_V$ respectively), with various different classification techniques. The feature vector sizes are determined by the $split_H \times split_V \times H$ where H is the length of the histogram, which is always 56 in the LDP extraction technique. Splits between 1 (no split) and 8 are tested. Further splits were omitted due to the very large vector size's adverse effect on processing speed. The same experiment setup is repeated for each of the Euclidean, Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances.

3.1 Euclidean Distance

Table 1 shows the EER results using the Euclidean distance calculations as defined in Equation 3. The EER improved with increasing feature vector sizes at first, since more data points allows better classification. The smallest feature vector, from 1×1 splits, provided a high EER of 25.0%. This signifies a poor accuracy. As feature vector size increased, the EER improved to a best of 21.7% at splits 2×3 . However, due to the concentration phenomenon, which causes distance values to cluster, the EER worsened for feature vectors larger than this, while some of the largest feature vectors resulted in worse accuracy than the smallest feature vector. The worst EER of 26.0% was obtained with the largest feature vector which had splits of 8×8 .

A similar trend is observable with the FARR, in Table 2, obtained using the Euclidean distance calculations, with the best FARR in the same feature vector size vicinity as the best EER, and the worst

FARR correlates with the worst EER. Once again this can be attributed to the concentration phenomenon, which makes the differentiation between authentic and forged classes difficult with large feature vectors. The distance measures discussed below are used to counter the effects of the concentration phenomenon.

Table 1: The effect of different $split_H$ and $split_V$ on EER(%) using the Euclidean distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	25.0	23.9	22.6	23.0	22.6	22.8	23.0	23.0
2	23.8	22.3	21.7	22.0	22.0	22.3	22.4	22.6
3	24.0	22.7	21.9	21.9	22.2	22.3	22.6	22.8
4	24.1	23.0	22.5	22.5	22.8	23.0	23.1	23.5
5	24.3	23.7	23.1	23.2	23.5	23.6	23.9	24.3
6	24.8	23.9	23.4	23.7	24.0	24.2	24.5	25.1
7	25.2	24.4	24.2	24.1	24.5	25.0	25.3	25.4
8	26.0	25.2	24.7	25.0	25.2	25.6	25.7	26.0

Table 2: The effect of different $split_H$ and $split_V$ on FARR(%) using Euclidean distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	1.70	0.96	0.60	0.58	0.55	0.55	0.59	1.35
2	1.09	0.48	0.31	0.34	0.34	0.35	0.41	0.45
3	0.85	0.43	0.28	0.33	0.37	0.38	0.44	0.54
4	0.72	0.43	0.35	0.37	0.42	0.54	0.59	0.72
5	0.81	0.52	0.42	0.53	0.61	0.80	0.88	1.09
6	0.89	0.57	0.52	0.68	0.83	1.02	1.28	1.52
7	0.97	0.77	0.73	0.94	1.20	1.53	1.84	2.26
8	1.02	0.90	0.92	1.19	1.50	1.89	2.24	2.87

3.2 Manhattan Distance

Table 3 shows the EER results using the Manhattan distance calculations as defined in Equation 4. The best EER of 19.2% is with splits 3×3 , which gives a feature vector slightly larger than with the Euclidean distance. However, while the largest feature vectors do not provide the best EER, they still provide a better EER than the smallest feature vectors. This is in contrast to the Euclidean distance where the largest feature vectors resulted in a worse EER than the smallest feature vectors. This shows that p -norm distance measures other than the Euclidean distance can provide a better result when feature vectors are larger and information is greater.

The FARR, in Table 4 shows a similar trend, where the best FARR of 0.18% is also with splits of 3×3 . While the FARR increase for larger feature vector sizes, it is still better than for the smallest feature vector. This further advocates the use of p -norm distances other than the Euclidean distance.

The fractional distances discussed next are used to counter the concentration phenomenon even further.

Table 3: The effect of different $split_H$ and $split_V$ on EER(%) using the Manhattan distance

$\begin{matrix} \backslash \\ H \end{matrix} \begin{matrix} V \\ \end{matrix}$	1	2	3	4	5	6	7	8
1	23.3	21.6	20.4	20.3	20.2	20.1	20.2	20.2
2	21.7	20.2	19.3	19.7	19.4	19.7	19.8	20.0
3	21.5	20.0	19.2	19.4	19.3	19.4	19.5	19.7
4	21.4	19.9	19.6	19.5	19.8	19.5	19.6	20.1
5	21.6	20.2	19.6	19.8	19.9	20.1	20.1	20.3
6	21.5	20.6	20.0	20.3	20.2	20.4	20.6	20.8
7	21.9	20.6	19.8	20.1	20.5	20.5	20.8	21.0
8	22.0	21.0	20.2	20.5	20.7	20.9	20.9	21.2

Table 4: The effect of different $split_H$ and $split_V$ on FARR(%) using Manhattan distances

$\begin{matrix} \backslash \\ H \end{matrix} \begin{matrix} V \\ \end{matrix}$	1	2	3	4	5	6	7	8
1	1.39	0.77	0.46	0.43	0.39	0.43	0.37	0.89
2	0.79	0.34	0.24	0.24	0.23	0.23	0.28	0.30
3	0.54	0.22	0.18	0.20	0.23	0.24	0.29	0.32
4	0.45	0.22	0.22	0.23	0.26	0.32	0.36	0.44
5	0.47	0.25	0.24	0.27	0.37	0.45	0.54	0.60
6	0.45	0.25	0.30	0.34	0.44	0.55	0.64	0.79
7	0.47	0.30	0.31	0.43	0.55	0.71	0.93	1.08
8	0.50	0.35	0.38	0.50	0.67	0.87	1.10	1.33

3.3 Fractional Distance

Figure 6 shows the effect of globally applied fractional distances on small and large feature vectors. The small feature vector was obtained from images with no split (1 block), while the large feature vector was obtained from images with 7×7 splits (49 blocks). While fractional distances provide an improvement over the Euclidean distance with the small feature vector, the effect is much more pronounced with the large feature vector. With the small feature vector, the best EER was obtained with $p = 0.4$. This correlates with the observation of Vivaracho-Pascual et. al. [4] Accuracy then sharply decreases and $p = 0.1$ gives a worse EER than the Euclidean distance. Conversely, when the feature vector is very large, the Euclidean distance provides a slightly worse EER in comparison to the small feature vector, but the EER greatly increases when $p < 2.0$. The best EER was found to then be with $p = 0.1$, which is 5.8% better than the best globally applied fractional distance for the smaller feature vector. This discrepancy with Vivaracho-Pascual et. al. is most likely due to their use of insufficiently large feature vectors. Additionally, the best EER for the fractional distance with the large feature vector was 8.8% better than the associated Euclidean distance and 11.2% better than the worse fractional distance with the small feature vector. This proves that the fractional distances do counter the concentration effect that occurs especially with large feature vectors.

The fractional distances were then applied with local optimization. Table 5 shows the EER results using the locally optimized fractional distance calculations as defined in Equation 5. This means that the best

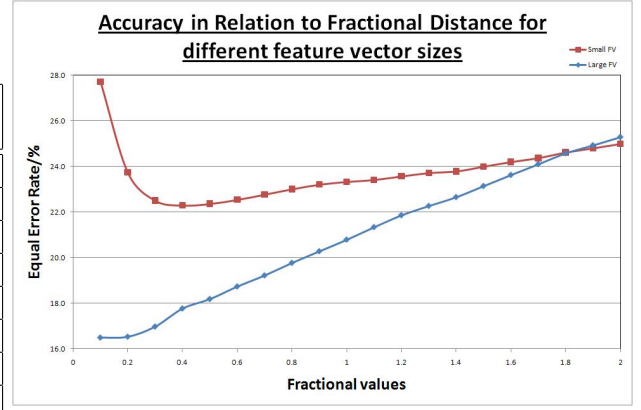


Figure 6: EER in Relation to individual fractional distances for large and small feature vector sizes

Table 5: The effect of different $split_H$ and $split_V$ on EER(%) using the fractional distance

$\begin{matrix} \backslash \\ H \end{matrix} \begin{matrix} V \\ \end{matrix}$	1	2	3	4	5	6	7	8
1	18.9	17.7	17.0	16.9	16.7	16.6	16.6	16.7
2	17.9	15.0	16.2	16.1	16.1	15.9	16.0	16.0
3	17.5	16.4	15.6	15.7	15.4	15.2	15.2	15.5
4	17.5	16.2	15.9	15.6	15.4	15.3	15.3	15.2
5	17.5	16.3	15.3	15.4	15.3	15.2	15.0	15.2
6	17.7	16.1	15.7	15.4	15.2	15.0	14.9	14.9
7	17.4	16.2	15.4	15.3	15.0	14.8	14.7	14.7
8	17.7	16.5	15.5	15.3	15.3	14.8	14.7	14.7

fractional distance within the range $0.1 \leq p \leq 2.0$ was chosen per individual, since some fractional distances work better than others for different individuals. The best EER of 14.7% is with one of the largest feature vectors, 7×8 . This is smaller than the smallest feature vector using fractional distances by 4.2% and it is better than the best Euclidean distance by 7.0%. This further shows that the fractional distances provide a solution to the concentration phenomenon that occurs with the Euclidean distance. There is a trend of improvement as feature vectors become larger. The fractional distances appears to reach their best improvement in overcoming the concentration phenomenon between splits 7×7 and 8×8 , which correlates with feature vector sizes between 2744 features and 3584 features. There is also an improvement over globally applied fractional distance, by 1.8%.

The FARR, in Table 6 shows a slightly different trend, where the best FARR does not correlate with the best EER. This is most likely due to choosing the best p -norm value locally for each individual, rather than a globally used p -norm as in the case of the Euclidean and Manhattan distances. Due to this localized optimization of the fractional distances, in some cases, larger p -norm values were used for the best EER, resulting in worsening of the concentration phenomenon for larger feature vectors. However, the FARR for fractional distances is still better than that for Euclidean and Manhattan distances, especially in the case of larger feature vectors.

Table 6: The effect of different $split_H$ and $split_V$ on FARR(%) using fractional distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	1.59	0.95	0.53	0.50	0.50	0.48	0.45	0.52
2	1.02	0.48	0.40	0.37	0.38	0.44	0.51	0.52
3	0.70	0.33	0.32	0.30	0.35	0.32	0.42	0.43
4	0.57	0.31	0.37	0.39	0.34	0.40	0.45	0.51
5	0.59	0.34	0.39	0.31	0.45	0.47	0.54	0.57
6	0.54	0.36	0.44	0.44	0.45	0.51	0.54	0.73
7	0.52	0.42	0.41	0.47	0.51	0.66	0.77	0.78
8	0.53	0.54	0.50	0.51	0.60	0.70	0.76	0.89

3.4 Weighted Euclidean Distance

Table 7 shows the EER results using the weighted Euclidean distance calculations as defined in Equation 10. The best EER of 14.5% is with splits 4×5 , which gives a feature vector larger than with the best unweighted Euclidean or Manhattan distance. This is of a better EER than the Euclidean and Manhattan distances, and of an almost equal EER in comparison with the fractional distance. The low EER is due to extra weight, or importance, being given to the most reliable features in the feature vector, which have the smallest intra-class difference per individual. This weighting also counteracts the clustering of the concentration phenomenon to a small extent. This is emphasized further by the largest feature vectors having a lower EER than the smallest feature vectors.

In contrast, the FARR in Table 8 shows a different trend, where the FARR worsens as the feature vectors become larger. This is because the weighting can not always differentiate between different sets of signatures and some features in random forgeries may be given extra weight if they are similar to authentic signatures. This is particularly visible with the largest feature vectors which will also experience the effects of the concentration phenomenon as well.

Table 7: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Euclidean distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	18.6	17.9	16.8	16.5	16.1	16.0	16.0	16.0
2	18.4	16.5	15.8	15.7	15.4	15.3	15.2	15.2
3	17.3	16.0	15.5	15.3	14.6	14.8	14.6	14.9
4	17.4	15.7	15.2	15.1	14.5	14.8	14.8	15.0
5	17.3	15.6	14.7	14.8	14.6	14.6	14.7	15.1
6	17.0	15.7	14.8	15.0	14.7	14.8	14.7	14.8
7	16.9	15.5	14.8	14.9	15.0	14.7	14.9	15.1
8	17.2	15.8	15.1	15.3	15.2	14.9	15.1	15.3

3.5 Weighted Manhattan Distance

Table 9 shows the EER results using the weighted Euclidean distance calculations as defined in Equation 11. The best EER of 14.3% is with splits 7×7 , which gives a feature vector larger than with the best

Table 8: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Euclidean distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	0.98	0.65	0.50	0.48	0.60	0.64	0.71	1.00
2	0.69	0.51	0.69	0.96	1.08	1.44	1.59	1.85
3	0.55	0.62	0.84	1.04	1.32	1.48	1.53	1.80
4	0.54	0.68	0.93	1.26	1.46	1.69	1.84	1.94
5	0.68	0.87	1.33	1.44	1.79	1.86	2.10	2.22
6	0.77	1.14	1.51	1.66	1.93	2.05	2.17	2.42
7	0.81	1.39	1.54	1.82	1.95	2.19	2.37	2.66
8	0.91	1.35	1.70	1.90	2.08	2.39	2.50	2.92

unweighted Euclidean, Manhattan and weighted Euclidean distances. It is of equal size in comparison to the fractional distance. The larger feature vector size provided a better EER due to a combination of a smaller p -norm distance and the weighting by standard deviation, which counteracted the concentration phenomenon in tandem. Due to this combination, it is also of a better EER than all of the distances tested with the LDP before it.

The FARR, in Table 8, shows a similar trend to the FARR of the weighted Euclidean distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a smaller p -norm value, the largest feature vectors have a smaller FARR in comparison with the largest feature vector sizes of the weighed Euclidean distance.

Table 9: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted Manhattan distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	19.4	18.6	17.3	17.3	17.0	16.6	16.7	16.7
2	19.3	17.2	16.6	16.3	15.9	15.6	15.5	15.6
3	18.3	16.5	15.8	15.5	15.1	15.0	14.7	14.7
4	18.1	16.2	15.4	15.3	14.8	14.8	14.5	14.7
5	18.0	16.2	15.0	14.9	14.8	14.7	14.7	14.7
6	17.7	16.1	15.1	15.1	14.9	14.6	14.5	14.7
7	17.4	16.0	15.0	15.0	14.5	14.3	14.3	14.4
8	18.0	16.3	15.2	15.1	14.7	14.6	14.5	14.6

Table 10: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted Manhattan distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	0.91	0.56	0.36	0.35	0.31	0.36	0.38	0.50
2	0.63	0.32	0.37	0.48	0.52	0.59	0.70	0.86
3	0.44	0.32	0.36	0.44	0.50	0.66	0.74	0.85
4	0.36	0.32	0.44	0.53	0.62	0.74	0.81	0.95
5	0.39	0.42	0.60	0.67	0.91	0.96	1.11	1.26
6	0.42	0.46	0.70	0.80	1.01	1.16	1.29	1.50
7	0.43	0.53	0.77	0.96	1.02	1.31	1.52	1.71
8	0.47	0.59	0.82	1.01	1.23	1.44	1.70	1.94

3.6 Weighted Fractional Distance

Following the success of the weighted Manhattan distance, the weighted fractional distances, which combines the weighted Euclidean distance, are tested. Table 11 shows the EER results using the weighted fractional distance calculations as defined in Equation 12. At splits 7×7 , This provides the best EER of 12.2% which is better than all of the classification techniques tested so far. Once again, the use of fractional distances and weighting allow greater accuracy with larger feature vectors. By combining two techniques that are individually better than the Euclidean distance, an overall much better accuracy was achieved. The worst EER for the weighted fractional distance was 16.83% using no splitting of the image. This is nearly 5% better than the best results for the standard Euclidean distance. Conversely, the best EER is 9.5% better than the best recorded Euclidean distance, and 2.5% and 2.7% better than the fractional and weighted Euclidean distances respectively.

The FARR, in Table 12, shows a similar trend to the FARR of the weighted Euclidean distance and the weighted Manhattan distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a wide range of p -norm values, the FARR values are better than the FARR for the Euclidean distance, but slightly worse than for the Manhattan distance.

Table 11: The effect of different $split_H$ and $split_V$ on EER(%) using the weighted fractional distance

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	16.8	16.1	15.1	14.8	14.5	14.4	14.2	14.6
2	16.7	15.0	14.2	14.0	13.7	13.4	13.3	13.4
3	15.7	14.3	13.6	13.2	12.7	12.6	12.5	12.6
4	15.8	14.0	13.3	13.0	12.5	12.7	12.4	12.5
5	15.3	14.1	12.8	12.7	12.5	12.3	12.3	12.5
6	15.1	13.9	12.9	12.7	12.6	12.5	12.3	12.4
7	15.0	13.8	12.9	12.7	12.5	12.3	12.2	12.4
8	15.4	13.9	13.0	13.0	12.6	12.5	12.5	12.6

Table 12: The effect of different $split_H$ and $split_V$ on FARR(%) using weighted fractional distances

$\begin{matrix} \text{H} \backslash \text{V} \\ \text{H} \end{matrix}$	1	2	3	4	5	6	7	8
1	1.08	0.70	0.47	0.42	0.52	0.57	0.57	0.72
2	0.65	0.48	0.54	0.84	0.81	0.93	1.22	1.55
3	0.54	0.42	0.66	0.75	0.75	0.81	1.09	1.33
4	0.56	0.53	0.23	0.91	1.00	1.11	1.14	1.34
5	0.63	0.79	1.00	1.02	1.20	1.28	1.28	1.61
6	0.68	0.76	1.16	1.08	1.34	1.36	1.45	1.67
7	0.68	0.88	1.18	1.36	1.44	1.63	1.64	1.78
8	0.66	1.04	1.21	1.42	1.53	1.84	1.84	2.11

3.7 Individually Optimized Resampling

Tables 13 to 14 show the test results for individually, or locally, optimized resampling performed in conjunction with each of the above six discussed distance-based classification techniques, namely, the Euclidean, Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances. Optimizations from 4 split and 9 split combinations were tested. It was found that 9 splits yield better results than 4 splits. These are symbolized in the following tables with the $[x:y]$ notation. For example, $[1:3]$ represents the 9 combinations of 1×1 , 1×2 , 1×3 , 2×1 , 2×2 , 2×3 , 3×1 , 3×2 and 3×3 .

3.7.1 Individually Optimized Euclidean Distance

Table 13 shows the results of tests with the p -norm distances and individually optimized resampling and the Table 14 shows those for weighted p -norm distances. The best EER for the Euclidean distance with individually optimized resampling was 16.5% from the $[1:3]$ combination. The optimal combinations of feature vector sizes correlate with the lowest single Euclidean distance EER, of 21.7%, from splits 2×3 when no individual optimization was performed. The combination provides a better EER than non-optimized Euclidean and Manhattan distances. The FARR values also improved through the individually optimized resampling. This suggests a correlation between individuals with a low EER and a low FARR.

3.7.2 Individually Optimized Manhattan Distance

The best EER for individual optimization on the Manhattan distance was 14.3% from the $[1:3]$ combination. The optimal combination of feature vector sizes correlate with the lowest single Manhattan distance EER, of 19.2%, from splits 3×3 when no individual optimization was performed. The 9-combination provides an EER better than or equal to all of the non-locally optimized distance-based measures, except for the weighted fractional distance whose best EER is 12.2%. The FARR values also improved through the individually optimized resampling.

3.7.3 Individually Optimized Fractional Distance

The best EER for individually optimized fractional distances was 11.3% from the $[1:3]$ combination and worst was 11.8% in the $[6:8]$ combination. The 9-combination provides an EER better than all measures tested before it, including the the weighted fractional distance whose best EER is 12.2%. The FARR values also showed an improvement through the individually optimized resampling.

The optimal combinations of feature vector sizes do not correlate with the lowest single fractional distance, even though the best and worst EER are very close, with only a 0.5% difference between them. This discrepancy may be explained by Figure 7. In the worse case scenario, with the combination $[6:8]$, the majority occurring p -norm in 0.1. This is larger than

Split Range	Euclidean		Manhattan		Fractional	
	EER	FARR	EER	FARR	EER	FARR
[1:3]	16.5	0.67	14.3	0.49	11.3	0.57
[2:4]	17.1	0.31	14.9	0.21	11.7	0.31
[3:5]	17.7	0.34	15.3	0.21	11.7	0.29
[4:6]	19.0	0.55	16.3	0.30	11.8	0.36
[5:7]	20.1	0.80	17.1	0.45	11.9	0.45
[6:8]	21.4	1.36	17.8	0.69	11.8	0.62

Table 13: The effect of implementing individually optimized resampling by choosing the best EER from 9 feature vectors for each of the p -norm distances

Split Range	Weighted Euclidean		Weighted Manhattan		Weighted Fractional	
	EER	FARR	EER	FARR	EER	FARR
[1:3]	11.0	0.70	11.5	0.50	9.65	0.60
[2:4]	10.8	0.72	11.8	0.34	9.66	0.49
[3:5]	10.7	1.18	11.4	0.48	9.25	0.80
[4:6]	11.1	1.59	11.6	0.68	9.50	0.97
[5:7]	11.2	1.88	11.9	1.06	9.60	1.18
[6:8]	11.8	2.20	12.0	1.27	10.0	1.36

Table 14: The effect of implementing individually optimized resampling by choosing the best EER from 9 feature vectors for each of the weighted p -norm distances

the second highest, 0.2, occurrence by over 40%. Conversely, for the best case scenario of [1:3], the 5 highest occurring distances are all within a 10% range of each other, and are the 5 smallest p -norm values. This allows a higher accuracy, since sometimes, among smaller feature vectors, different p -norm values are better for difference individuals. However, with the largest feature vectors, the p -norm value of 0.1 outperforms all others. A combination of smaller feature vectors with more variable p -norm values can therefore perform slightly better than a large feature vector with a single dominant p -norm value.

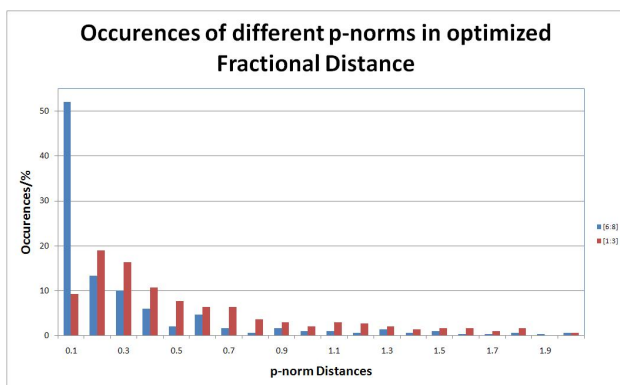


Figure 7: The number of occurrences of each p -norm in the fractional distance with individually optimized resampling combinations [1:3] and [6:8]

3.7.4 Individually Optimized Weighted Euclidean Distance

The best EER for individually optimized weighted Euclidean distance was 10.7% from the [3:5] combination. The optimal combination of feature vector sizes corre-

late with the lowest single weighted Euclidean distance EER, of 14.5%, from splits 4×5 when no individual optimized resampling was performed. This provides better EER than previous distance-based classification measures. This includes better performance than the best weighted fractional distance of 12.2% without individually optimized resampling and all of the previously tested individually optimized resamplings of the LDP. The FARR values correlate closely with those for the weighted Euclidean distance without individually optimized resampling.

3.7.5 Individually Optimized Weighted Manhattan Distance

The best EER for individually optimized weighted Manhattan distance was 11.4% from the [3:5] combination. The optimal combination of feature vector sizes correlate with the lowest single weighted Manhattan distance EER, of 14.5%, from splits 7×7 when no individual optimized resampling was performed. This provides better EER than previous distance-based classification measures, where no individually optimized resampling was performed. However, the individually optimized weighted Manhattan distance performs worse than the individually optimized weighted Euclidean distance. This may be because the standard deviation used in the weighting equation is calculated in L^P -space of 2 while the Manhattan distance is calculated in p -norm space of 1. The FARR values correlate closely with those for the weighted Manhattan distance without individually optimized resampling.

3.7.6 Individually Optimized Weighted Fractional Distance

The best EER for the individually optimized weighted fractional distance performed better than all other test before. This good performance is due to choosing both the best p -norm and best split size per individual, i.e. locally optimized classification. The optimal split sizes for the individually optimized fractional distance do not correlate with the optimal split sizes for the non-locally optimized weighted fractional distance, although there is a mere difference of 0.75% between the best and worst EER. An analysis of Figure 8, shows that the pattern for the highest EER, from combination [6:8], is similar to that for the highest EER for the individually optimized fractional distance in Figure 7. In both cases the smallest p -norm size of 0.1 has the greatest percentage of occurrences by far. Similarly, the best EER, from combination [3:5], individually optimized fractional distance has a comparable pattern with the best EER of [1:3] for the individually optimized fractional distance. In both cases, the first 4 p -norm distances are among the largest and within a close range. Additionally, the greatest occurrence is still for the p -norm value of 0.1. The smaller feature vectors in combination [1:3] provide an EER in between the highest and lowest. In Figure 8, it can be seen that the greatest occurrence was for p -norm of 0.6 rather than 0.1. The latter provides much lower EER values with larger feature vector sizes. The larger p -norm of 0.6 is more effect with smaller feature vectors, in comparison with 0.1. However, 0.1 performs much better than 0.6 in comparison with larger feature vectors.

An analysis of the weighted Euclidean, and the individually optimized weighted Euclidean, weighted Manhattan and weighted fractional distances shows that all of them provide their best EER in the feature vector range for splits 3×3 and 5×5 . This suggests a strong involvement of the standard deviation function that is used in the weighting. In all cases, the standard deviation was calculated in p -norm space of 2, whereas the Manhattan and fractional distances were calculates with smaller p -norm distances.

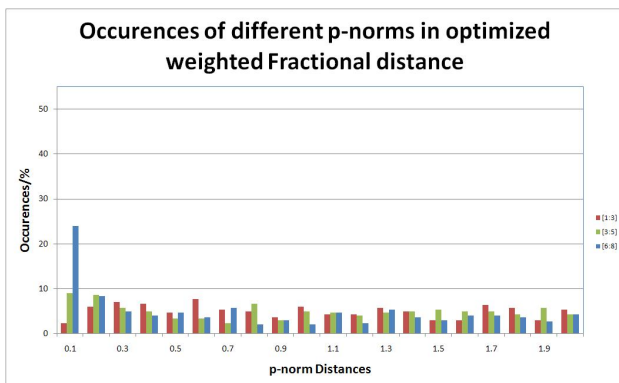


Figure 8: The number of occurrences of each p -norm in the weighted fractional distance with individually optimized resampling combinations [1:3] and [6:8]

3.8 Literature Comparison

Ferrer et. al. [25] tested the Local Binary Pattern (LBP) and Local Directional Pattern (LDP) using several data sets, separately. These sets were 75 individuals from the MCYT database [27], and 75, 300 and 960 users from the GPDS database [28]. While EER with the data sets using 75 individuals is low, these results can not be used for comparison, since the small size of the data set bring the precision and accuracy of results into question. A better comparison are the results obtained with the data sets containing 300 and 960 individuals. Classification was performed with an Least Squares SVM (LS-SVM) with an RBF kernel. The signatures were split into 12 blocks for the feature extraction, consisting of 4 vertical splits and 3 horizontal splits, and an overlap of 60%. From these, the best EER was 17.8% using 300 individuals of the GPDS database, with a corresponding FARR of 0.68%.

In this work, the same feature extraction technique was implemented and tested with multiple distance-based classification techniques, including the Euclidean, fractional, weighted Euclidean and weighted fractional distances. Signatures of 300 GPDS individuals were used, which was also the same database used for the best results in [25]. A larger range of splits sizes were tested. This range was from 1 to 8 splits in both the vertical and horizontal directions. While the best Euclidean and Manhattan distance performances for the EMDF was worse than the literature result, at 21.7% and 19.2% respectively, the fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances all performed much better, at 14.7%, 14.5%, 14.3% and 12.2%. It can be seen that the weighted fractional distance performs better than the LS-SVM with an RBF kernel for the classification of the LDP. Further, individually optimized resampling was performed, where the best resampled block sizes per individual were chosen from a possibility of 9. The best individually optimized resampling was for the weighted fractional distance, with an EER of 9.25% and corresponding FARR of 0.80%. The obtained EER is better than the best results in the literature by 8.6%, and both systems obtained an FARR of below 1%.

4 CONCLUSION AND FUTURE WORK

To improve the accuracy of offline signature verification, distance-based classification techniques were tested with application to LDP feature extraction technique.

The distances in L^P -space were tested. The Manhattan distance and fractional distances performed better than the Euclidean distance. The smaller p -norm distances performed better because they are less affected by the concentration phenomenon. The weighted distances in L^P -space also performed well. Greater weight is given to the most stable and reliable features through use of standard deviation. It also smooths out randomness from features and thus

improves classification accuracy. By combining the weighting and fractional distances, the weighted fractional distance was created. This novel distance measure performed best.

Additionally, many different feature vector sizes were tested. The sizes were varied through resampling. Fractional distances generally work better with larger feature vectors. Due to the uniqueness of each individual's signatures, the optimal feature vector size is not the same for all individuals. Therefore, individually optimized resampling was used to choose the optimal feature vector size per user. The best results were obtained when the weighted fractional distances were combined with individually optimized resampling. This combination of multiple distance-based classification techniques achieved accuracy rate of 90.8%.

Classification with the fractional distances, all weighted distances in L^P -space and individually optimized resampling performed better than literature results, where SLTs, namely SVMs and NNs, were used for classification. Therefore, distance-based classification techniques provide a viable alternative to SLTs for the verification of offline signatures.

Future work includes testing the system with foreign signature databases, such as Chinese and Persian; and fusing the signature modality with other biometric modalities to create a multi-modal biometric verification system.

REFERENCES

- [1] L. Baum and T. Petrie. "Statistical inference for probabilistic functions of finite state Markov chains". *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.
- [2] C. Cortes and V. Vapnik. "Support-vector networks". *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [3] Y. Zhu, T. Tan and Y. Wang. "Biometric personal identification based on iris patterns". In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, pp. 801–804. IEEE, 2000.
- [4] C. Vivaracho-Pascual, M. Faundez-Zanuy and J. Pascual. "An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances". *Pattern Recognition*, vol. 42, no. 1, pp. 183 – 193, 2009.
- [5] B. Kovari, A. Horvath, B. Toth, H. Charaf, L. Perlovsky, D. Dionysiou, L. Zadeh, M. Kostic, C. Gonzalez-Concepcion, H. Jaberg et al. "Local feature based off-line signature verification using neural network classifiers". In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, 11. WSEAS, 2009.
- [6] J. Coetzer, B. Herbst and J. Du Preez. "Offline signature verification using the discrete radon transform and a hidden Markov model". *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 559–571, 2004.
- [7] M. Panton. *Off-line signature verification using ensembles of local Radon transform-based HMMs*. Ph.D. thesis, Stellenbosch: Stellenbosch University, 2011.
- [8] M. Yilmaz, B. Yanikoglu, C. Tirkaz and A. Kholmatov. "Offline signature verification using classifier combination of HOG and LBP features". In *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–7. IEEE, 2011.
- [9] J. Vargas, M. Ferrer, C. Travieso and J. Alonso. "Offline signature verification based on pseudo-cepstral coefficients". In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pp. 126–130. IEEE, 2009.
- [10] L. Batista, E. Granger and R. Sabourin. "Improving performance of HMM-based off-line signature verification systems through a multi-hypothesis approach". *International journal on document analysis and recognition*, vol. 13, no. 1, pp. 33–47, 2010.
- [11] M. Ferrer, J. Alonso and C. Travieso. "Offline geometric parameters for automatic signature verification using fixed-point arithmetic". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 993–997, 2005.
- [12] D. Kisku, A. Rattani, P. Gupta and J. Sing. "Offline signature verification using geometric and orientation features with multiple experts fusion". In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, vol. 5, pp. 269–272. IEEE, 2011.
- [13] P. Mahalanobis. "On the generalized distance in statistics". In *Proceedings of the National Institute of Science (India)*, vol. 12, pp. 49–55. National Institute of Science, 1936.
- [14] N. Kato, M. Suzuki, S. Omachi, H. Aso and Y. Nemoto. "A Handwritten Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 258–262, 1999.
- [15] B. Fang, C. Leung, Y. Tang, K. Tse, P. Kwok and Y. Wong. "Off-line signature verification by the tracking of feature and stroke positions". *Pattern recognition*, vol. 36, no. 1, pp. 91–101, 2003.
- [16] V. Nguyen, Y. Kawazoe, T. Wakabayashi, U. Pal and M. Blumenstein. "Performance analysis of the gradient feature and the modified direction feature for off-line signature verification". In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, pp. 303–307. IEEE, 2010.
- [17] M. Sigari, M. Pourshahabi and H. Pourreza. "Offline Handwritten Signature Identification and Verification Using Multi-Resolution Gabor Wavelet". *International Journal of Biometric and Bioinformatics*, vol. 5, 2011.
- [18] B. Shekar and R. Bharathi. "Eigen-signature: A robust and an efficient offline signature verification algorithm". In *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, pp. 134–138. IEEE, 2011.
- [19] Y. Rekik, N. Houmani, M. El Yacoubi, S. Garcia-Salicetti and B. Dorizzi. "A comparison of feature extraction approaches for offline signature verification". In *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*, pp. 1–6. IEEE, 2011.

- [20] A. Ramachandra, K. Pavithra, K. Yashasvini, K. Raja, K. Venugopal and L. Patnaik. “Offline signature authentication using cross-validated graph matching”. In *Proceedings of the 2nd Bangalore Annual Compute Conference*, p. 7. ACM, 2009.
- [21] H. Kuhn. “The Hungarian method for the assignment problem”. *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 2006.
- [22] A. Alizadeh, T. Alizadeh and Z. Daei. “Optimal Threshold Selection for Online Verification of Signature”. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 17–19. 2010.
- [23] V. Espinosa-Duró, M. Faundez-Zanuy and J. Mekyska. “A New Face Database Simultaneously Acquired in Visible, Near-Infrared and Thermal Spectrums”. *Cognitive Computation*, pp. 1–17, 2012.
- [24] T. Jabid, H. Kabir and O. Chae. “Gender Classification using Local Directional Pattern (LDP)”. In I. C. Society (editor), *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR’10, pp. 2162–2165. 2010.
- [25] M. Ferrer, F. Vargas, C. Traviesto and J. Alonso. “Signature verification using local directional pattern”. In *International Carnahan Conference on Security Technology*, pp. 336–340. IEEE, Oct 2010.
- [26] D. Francois, V. Wertz and M. Verleysen. “The concentration of fractional distances”. *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, no. 7, pp. 873–886, 2007.
- [27] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero and Q. I. Moro. “MCYT baseline corpus: A bimodal biometric database”. In *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, vol. 150, pp. 395–401. December 2003.
- [28] J. Vargas, M. Ferrer, C. Travieso and J. Alonso. “Off-line handwritten signature GPDS-960 corpus”. In *Ninth International Conference on Document Analysis and Recognition*, pp. 764 – 768. IEEE Computer Society, 2007.