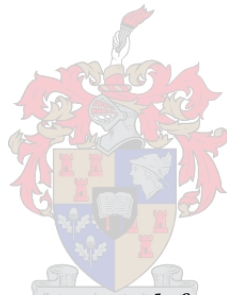


Writer-independent Handwritten Signature Verification

by

Jacques Philip Swanepoel



*Dissertation presented for the degree of
Doctor of Philosophy in Applied Mathematics
in the Faculty of Science at Stellenbosch University*

Supervisor: Dr Johannes Coetzer

December 2015

Declaration

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

Date:

Copyright © 2015 Stellenbosch University
All rights reserved.

Abstract

In this dissertation we present a novel strategy for automatic handwritten signature verification. The proposed framework employs a *writer-independent* approach to signature modelling and is therefore capable of authenticating questioned signatures claimed to belong to *any* writer, provided that at least one authentic sample of said writer's signature is available for comparison. We investigate both the traditional *off-line* scenario (where an existing pen-on-paper signature is extracted from a digitised document) as well as the increasingly popular *on-line* scenario (where the signature data are automatically recorded during the signing event by means of specialised electronic hardware). The utilised off-line feature extraction technique involves the calculation of several *discrete Radon transform* (DRT) based projections, whilst on-line signatures are represented in feature space by several spatial and temporal function features. In order to facilitate writer-independent signature analysis, these feature sets are subsequently converted into a dissimilarity-based representation by means of a suitable dichotomy transformation. The classification techniques utilised for signature modelling and verification include *quadratic discriminant analysis* (QDA) and *support vector machines* (SVMs). The major contributions of this study include two novel techniques aimed towards the construction of a robust writer-independent signature model. The first, a *dynamic time warping* (DTW) based dichotomy transformation for off-line signature representation, is able to compensate for reasonable intra-class variability by non-linearly aligning DRT-based projections prior to matching. The second, a writer-specific dissimilarity normalisation strategy, improves inter-class separability in dissimilarity space by considering only strictly relevant dissimilarity statistics when normalising the dissimilarity vectors belonging to a specific individual. This normalisation strategy is generic in the sense that it is equally applicable to both off-line and on-line signature model construction. The systems developed in this study are specifically aimed towards *skilled forgery detection*. System proficiency estimation is conducted using a rigorous experimental protocol. Several large signature corpora are considered. In both the off-line and on-line scenarios, the proposed SVM-based system outperforms the proposed QDA-based system. We also show that the systems proposed in this study outperform most existing systems that were evaluated on the same data sets. More importantly, when compared to state-of-the-art techniques currently employed in the literature, we show that the incorporation of the novel techniques proposed in this study consistently results in a *statistically significant* improvement in system proficiency.

Opsomming

In hierdie verhandeling stel ons 'n nuwe strategie vir outomatiese handtekening-verifikasie voor. Die voorgestelde raamwerk gebruik 'n *skrywer-onafhanklike* benadering tot handtekening-modellering en is dus in staat om bevragekende handtekeninge, wat aan *enige* skrywer behoort, te bekragtig, op voorwaarde dat minstens een outentieke voorbeeld vir vergelykingsdoeleindes beskikbaar is. Ons ondersoek die tradisionele *statische* geval (waarin 'n bestaande pen-op-papier handtekening vanuit 'n versyferde dokument onttrek word), asook die toenemend gewilde *dinamiese* geval (waarin handtekeningdata outomaties tydens ondertekening m.b.v. gespesialiseerde elektroniese hardeware bekom word). Die statiese kenmerk-onttrekkingstegniek behels die berekening van verskeie *diskrete Radon-transform* (DRT) projeksies, terwyl dinamiese handtekeninge deur verskeie ruimtelike en temporele funksie-kenmerke in die kenmerkruimte voorgestel word. Ten einde skrywer-onafhanklike handtekening-ontleding te bewerkstellig, word hierdie kenmerkstelle na 'n verskil-gebaseerde voorstelling d.m.v. 'n geskikte digotomie-transformasie omgeskakel. Die klassifikasietegnieke, wat vir handtekening-modellering en -verifikasie gebruik word, sluit *kwadratiese diskriminant-analise* (KDA) en *steunvektormasjiene* (SVM) in. Die hoofbydraes van hierdie studie sluit twee nuwe tegnieke, wat op die bou van 'n robuuste skrywer-onafhanklike handtekeningmodel gerig is, in. Die eerste, 'n *dinamiese tydsverbuiging* digotomie-transformasie vir statiese handtekening-voorstelling, is in staat om vir redelike intra-klas variasie te kompenseer, deur die DRT-projeksies voor vergelyking nie-lineêr te belyn. Die tweede, 'n *skrywer-spesifieke verskil-normaliseringstrategie*, is in staat om inter-klas skeibaarheid in die verskilruimte te verbeter deur slegs streng relevante statistieke tydens die normalisering van verskil-vektore te beskou. Die normaliseringstrategie is generies van aard in die sin dat dit ewe veel van toepassing op beide statiese en dinamiese handtekening-modelkonstruksie is. Die stelsels wat in hierdie studie ontwikkel is, is spesifiek op die opsporing van *hoë-kwaliteit vervalsings* gerig. Stelselvaardigheid-afskatting word met behulp van 'n omvattende eksperimentele protokol bewerkstellig. Verskeie groot handtekening-datastelle is oorweeg. In beide die statiese en dinamiese gevalle vaar die voorgestelde SVM-gebaseerde stelsel beter as die voorgestelde KDA-gebaseerde stelsel. Ons toon ook aan dat die stelsels wat in hierdie studie ontwikkel is, die meeste bestaande stelsels wat op dieselfde datastelle geëvalueer is, oortref. Dit is selfs meer belangrik om daarop te let dat, wanneer hierdie stelsels met bestaande tegnieke in die literatuur vergelyk word, ons aantoon dat die gebruik van die nuwe tegnieke, soos in hierdie studie voorgestel, konsekwent tot 'n *statisties beduidende* verbetering in stelselvaardigheid lei.

Acknowledgements

I would like to express my most sincere gratitude toward each of the following persons and/or institutions for their role in the successful completion of this study:

- My supervisor, Dr Johannes Coetzer, for his invaluable insights and eagerness to constantly explore new ideas. The work presented in this dissertation represents the culmination of almost nine years of research that Hanno and I started with a (now seemingly rudimentary) Honour's project in handwritten character recognition. My PhD candidature marks our third postgraduate research project together and words therefore cannot fully describe Hanno's immeasurable contribution towards my development as an academic. I am proud to call him my colleague, my mentor and my friend.
- Stellenbosch University, for their financial assistance.
- My family and friends, for their moral support.
- My parents ("old" and "new"), Derek, Sandra, Harriet, Leon and Lisbé, for their love and support.
- My late grandmother, Ethel (or "Gran"), for her love and encouragement.
- My fiancée, Elizna, for her endless patience, unconditional love and unwavering support.

Contents

Declaration	i
Abstract	ii
Opsomming	iii
Acknowledgements	iv
Contents	v
List of Figures	viii
List of Tables	xiv
List of Symbols	xvi
List of Acronyms	xvii
1 Introduction	1
1.1 Motivation	1
1.2 Key concepts	4
1.2.1 Personal identification systems	4
1.2.2 Handwritten signatures	6
1.2.3 Forgery categorisation	7
1.2.4 Signature acquisition	9
1.2.5 Pattern recognition	10
1.2.6 Writer-independent signature modelling	16
1.2.7 Performance metrics	17
1.3 Objectives	20
1.4 Overview of this study	20
1.4.1 System design	20
1.4.2 Experimental results	26
1.5 Contribution of this study	26
1.6 Outline	27

<i>CONTENTS</i>	vi
2 Literature Study	29
2.1 Introduction	29
2.2 Off-line signature verification	30
2.2.1 Writer-dependent systems	30
2.2.2 Writer-independent systems	34
2.2.3 Hybrid systems	35
2.3 On-line signature verification	36
2.3.1 Writer-dependent systems	36
2.3.2 Writer-independent systems	39
2.4 Concluding remarks	40
3 Signature Representation	41
3.1 Introduction	41
3.2 Off-line signatures	42
3.2.1 Signature acquisition and pre-processing	42
3.2.2 Feature extraction and normalisation	45
3.3 On-line signatures	48
3.3.1 Signature acquisition and pre-processing	48
3.3.2 Feature extraction	51
3.4 Dichotomy transformation	52
3.5 Concluding remarks	56
4 Classifiers	58
4.1 Introduction	58
4.2 Overview	58
4.3 Discriminant analysis	60
4.3.1 Overview	60
4.3.2 The curse of dimensionality	63
4.4 Support vector machines	64
4.4.1 Overview	64
4.4.2 SVM parameter selection	67
4.5 Concluding remarks	70
5 Signature Modelling and Verification	72
5.1 Introduction	72
5.2 Signature modelling	72
5.2.1 Outlier removal	73
5.2.2 Dissimilarity normalisation	75
5.2.3 Model training	79
5.3 Verification	80
5.4 Concluding remarks	82

<i>CONTENTS</i>	vii
6 Experiments	84
6.1 Introduction	84
6.2 Experimental setup	84
6.2.1 Data	85
6.2.2 Protocol	89
6.3 Results	91
6.3.1 Off-line verification systems	91
6.3.2 On-line verification systems	95
6.4 Comparison with previous work	96
6.4.1 Off-line verification systems	97
6.4.2 On-line verification systems	99
6.5 Contribution of this study	101
6.5.1 A DTW-based dichotomy transformation for writer-independent off-line signature representation	107
6.5.2 A writer-specific dissimilarity normalisation strategy for writer-independent handwritten signature modelling	109
6.5.3 A writer-independent off-line signature verification system that is both novel and proficient	110
6.5.4 A writer-independent on-line signature verification system that is both novel and proficient	112
6.6 Concluding remarks	112
7 Conclusion and Future Work	115
7.1 Conclusion	115
7.2 Future work	116
7.2.1 Improved image processing for off-line signature representation	116
7.2.2 Feature weighted signature modelling	117
7.2.3 Confidence weighted score fusion	119
7.2.4 Other applications	121
Bibliography	123
A Dynamic Time Warping	133
A.1 Algorithm	133

List of Figures

1.1	Categorisation of selected personal identification systems. Categories relevant to this study are emphasised in bold. The attributes associated with the design of a biometric system are sub-categorised in Figure 1.2.	5
1.2	Categorisation of selected biometric attributes. Categories relevant to this study are emphasised in bold.	6
1.3	Examples of historic documents that were authorised by means of handwritten signatures. (a) The Declaration of Independence of the United States of America, as signed by 56 delegates of the US Congress on July 4 th 1776. (b) A document signed by members of the Greek National Assembly in Athens (after the revolution on September 3 rd 1843) that drafted a new constitution and made Greece a constitutional monarchy.	7
1.4	Categorisation of forged handwritten signatures. Categories relevant to this study are emphasised in bold.	8
1.5	A typical example of (a) a genuine signature, as well as attempts at its reproduction in the form of a (b) random, (c) amateur skilled and (d) professional skilled forgery.	9
1.6	Schematic representation of the pattern recognition process. The additional data processing stages (indicated with dashed lines/borders) are considered optional, but are often included in order to improve system performance. . .	11
1.7	Categorisation of selected feature extraction techniques associated with <i>off-line</i> signature verification. Categories relevant to this study are emphasised in bold.	12
1.8	Categorisation of selected feature extraction techniques associated with <i>on-line</i> signature verification. Categories relevant to this study are emphasised in bold. Note that the arrow notation indicates that velocity and acceleration may be <i>derived from</i> position data.	13
1.9	Categorisation of selected classification techniques associated with handwritten signature verification. Categories relevant to this study are emphasised in bold.	14

1.10	(a) Idealised confidence distributions for the positive class G and the negative class F . Note that the degree of overlap between $P(G s)$ and $P(F s)$ is proportional to the quality of the forgeries considered. (b) Class membership prediction by means of confidence thresholding, where a confidence score above (or below) the chosen threshold τ is considered indicative of a genuine (or forged) signature. It should be clear that, in a scenario where $P(G s)$ and $P(F s)$ are non-separable (as is often the case with <i>skilled</i> forgery detection), <i>any</i> value for τ inevitably results in a number of erroneous class membership predictions (i.e. <i>misclassifications</i>). Several suitable methods commonly utilised to quantify system performance are discussed in Section 1.2.7.	15
1.11	Conceptual comparison of the (a) <i>writer-dependent</i> and (b) <i>writer-independent</i> approaches to signature representation in feature space and dissimilarity space respectively, when three writers are considered. Positive and negative instances are indicated with “+” and “-” respectively.	17
1.12	Conceptualisation of the system performance metrics considered in this study. (a) The FRR and FAR as functions of the verification threshold τ . Also indicated is the EER of the underlying continuous classifier. (b) Performance evaluation of a hypothetical classifier in ROC space. The continuous classifier C is depicted by a ROC curve, whilst each threshold-specific discrete classifier $C(\tau)$ is depicted by a single point in ROC space. Also indicated are the optimal α -based and EER-based discrete classifiers, denoted by $C(\tau_\alpha)$ and $C(\tau_{\text{EER}})$ respectively.	19
1.13	Schematic representation of a typical system developed in this study. Detailed schematics of the signature representation, signature modelling and signature verification processes are presented in Figures 1.14–1.17.	21
1.14	Schematic representation of the <i>off-line</i> signature representation process utilised in this study.	22
1.15	Schematic representation of the <i>on-line</i> signature representation process utilised in this study.	23
1.16	Schematic representation of the signature modelling process utilised in this study.	24
1.17	Schematic representation of the signature verification process utilised in this study.	25
3.1	Example of a grey-level intensity image that contains an off-line signature sample.	42
3.2	Image binarisation and noise reduction. Images obtained after (a) Otsu’s binarisation method and (b) the median filter are successively applied to the image depicted in Figure 3.1. The dashed borders indicate the boundary of the original grey-level image. Note that the median filter successfully removes several traces of impulse noise (encircled in red) and also <i>partially</i> repairs selected pen stroke segments that were decimated during the preceding binarisation stage.	44

3.3	Signature segmentation. (a) Grey-level image of a signature sample surrounded by traces of ink residue and (b) its resulting binary representation obtained after noise reduction. Although the pre-processing stages effectively dealt with most of the document degradation, a small portion of noise is still present near the image boundary (encircled in red). (c)-(d) Signature segmentation results, where the solid border indicates the sub-image extracted using (c) the traditional bounding box method and (d) the proposed method of discarding all zero-valued rows and/or columns. This example clearly illustrates the shortcomings of the bounding box method, as well as the ability of the proposed segmentation technique to minimise the adverse affects associated with sub-optimal noise reduction.	45
3.4	Geometric interpretation of the Radon transform. Each projection $R_\theta(f(x, y))$ constitutes the line integral of $f(x, y)$ parallel to the y_θ -axis.	46
3.5	Conceptualisation of the method used to calculate the discrete Radon transform for a specific angle θ with $\alpha_{ij} \approx 0.6$. This implies that the j^{th} beam overlaps approximately 60% of the i^{th} pixel, as indicated by the dark grey shaded region.	47
3.6	Example of (a) a binarised signature image and (b) its resulting DRT-based representation for $T = 180$. Each column of the DRT constitutes a feature vector $\bar{\mathbf{R}}_\theta$, that is the <i>normalised</i> projection profile associated with a specific projection angle θ	48
3.7	Conceptualisation of the measurements recorded during on-line signature acquisition at time i . These signature descriptors may include the pen stroke coordinates (x_i, y_i) , the pen pressure p_i , as well as the pen orientation, as described in terms of the azimuth angle θ_i and the altitude angle φ_i	49
3.8	Translation and scale normalisation. (a) Superposition of the pen position features \mathbf{x} and \mathbf{y} associated with three signature samples belonging to the same writer. Since their positions and sizes differ significantly, these samples are not yet fit for direct comparison. (b) Superposition of the normalised pen position features $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, that is the features depicted in (a) following successful translation and scale normalisation.	50
3.9	Conceptual comparison of the feature correspondences considered during dissimilarity vector construction when either (a) the Euclidean distance or (b) a DTW-algorithm is utilised. Note that, unlike the Euclidean distance-based approach, a DTW-algorithm is able to detect (and subsequently compensate for) non-linearly misaligned features, thereby producing a considerably more reliable measure of dissimilarity between the feature vectors submitted for comparison.	54

3.10	Conceptualisation of the dynamic time warping algorithm considered in this study. The algorithm identifies similar elements contained in the reference vector \mathbf{x}_k and questioned vector \mathbf{x}_q and subsequently constructs an optimal path between said vectors, based on these feature similarities. The resulting distance measure is calculated between elements matched according to the optimal path, as opposed to simply using corresponding elements. The <i>bandwidth</i> β restricts the search space and is used to regulate both the flexibility and computational requirements of the alignment process. Note that when $\beta = 0$, this algorithm is equivalent to utilising the Euclidean distance.	55
4.1	Distribution of hypothetical positive and negative training samples in a two-dimensional feature space.	60
4.2	Implementation of a linear discriminant. (a) Multivariate Gaussian PDFs estimated from the samples depicted in Figure 4.1, using class-specific means and a <i>pooled</i> covariance estimate. (b) The resulting linear decision boundary $f_{\text{LDA}}(\mathbf{x}) = 0$. Note that, although the positive and negative classes are clearly linearly separable, several negative samples are misclassified.	62
4.3	Implementation of a quadratic discriminant. (a) Multivariate Gaussian PDFs estimated from the samples depicted in Figure 4.1, using class-specific means and <i>class-specific</i> covariance estimates. (b) The resulting quadratic decision boundary $f_{\text{QDA}}(\mathbf{x}) = 0$. Note that, unlike the linear discriminant depicted in Figure 4.2 (b), the quadratic discriminant is able to successfully separate the positive and negative classes.	63
4.4	Conceptualisation of the maximally separating hyperplane and its associated margin in a hypothetical two-dimensional feature space for (a) linearly separable and (b) linearly non-separable training data.	65
4.5	Conceptualisation of the kernel trick. (a) Positive and negative training samples in a hypothetical two-dimensional <i>feature space</i> . There clearly exists no linear decision boundary (i.e. a straight line) capable of separating the two classes. (b) Non-linear mapping of the training samples depicted in (a) into a hypothetical three-dimensional <i>kernel space</i> , wherein it becomes possible to obtain a separating hyperplane. The inverse mapping of the hyperplane indicated in (b) corresponds to the dashed line indicated in (a), which successfully separates the two classes in feature space.	66
4.6	Implementation of a soft-margin SVM. Decision boundaries obtained from the training samples depicted in Figure 4.1 when (a) a linear kernel and (b) an RBF kernel are utilised. Since the training samples are linearly separable in feature space, both the linear and RBF kernels yield boundaries that are able to successfully separate the two classes. Note that the orientation (and also the curvature in the case of the RBF kernel) of each boundary is determined by the value specified for the internal parameter C (and also γ), which is further discussed in Section 4.4.2.	67

4.7	Significance of the regularisation parameter C . Optimal decision boundaries for the separation of two classes in a hypothetical two-dimensional feature space when (a) relatively large and (b) relatively small values for C are specified. Although the boundary in (a) reduces the number of margin errors on the training set, the boundary in (b) provides a more sensible overall separation of the two classes and is therefore expected to prove superior in the classification of future questioned samples.	69
4.8	Significance of the RBF kernel width parameter γ . Decision boundaries obtained from the training samples depicted in Figure 4.1 when (a) $\gamma = 10000$ and (b) $\gamma = 0.1$. The boundary in (a) is over-smoothed to such an extent that it resembles the boundary obtained when employing a linear kernel (see Figure 4.6 (a)), whilst the excessive non-linear flexibility of the boundary in (b) clearly over-fits the training set – this is easily confirmed by observing that each of the training samples is also identified as a support vector i.e. <i>every</i> training sample contributes towards the optimal solution.	70
5.1	(a) Typical representation of negative samples <i>superimposed</i> onto positive samples in dissimilarity space for $T = 2$, $N = 40$ and $K = 10$. (b) Outliers detected (and subsequently removed) by the IOR algorithm.	74
5.2	(a) The conventional logistic function. (b) The dissimilarity normalisation function utilised by the systems developed in this study for $c = 5$	76
5.3	The <i>global</i> dissimilarity normalisation strategy. (a) Dissimilarity values representative of positive and negative samples obtained from five different writers. Also included is the critical value $z = \mu + \sigma$, as obtained from <i>all</i> the samples. (b) Normalised dissimilarity values obtained using the <i>global</i> normalisation function $\eta(z, \mu + \sigma)$, as well as the mapped critical value $\eta(\mu + \sigma, \mu + \sigma) = 0.5$. Note that, although the original dissimilarity values have been successfully rescaled, no improvement is observed in terms of class separability.	77
5.4	The <i>writer-specific</i> dissimilarity normalisation strategy. (a) Dissimilarity values as also depicted in Figure 5.3 (a). Also included are the writer-specific critical values $z = \mu^{(\omega)} + \sigma^{(\omega)}$ for $\omega = 1, 2, \dots, 5$, as obtained from the samples belonging to writer ω only. (b) Normalised dissimilarity values obtained using the <i>writer-specific</i> normalisation function $\eta(z, \mu^{(\omega)} + \sigma^{(\omega)})$ for each writer separately, as well as their associated mapped critical values $\eta(\mu^{(\omega)} + \sigma^{(\omega)}, \mu^{(\omega)} + \sigma^{(\omega)}) = 0.5$. Note that this strategy not only successfully rescales the original dissimilarity values, but also significantly improves overall class separability when compared to the results illustrated in Figure 5.3 (b).	78
5.5	Comparison of (a) the <i>global</i> dissimilarity normalisation strategy and (b) the <i>writer-specific</i> dissimilarity normalisation strategy, when applied to the <i>retained samples</i> depicted in Figure 5.1 (b).	79

6.1	Selected samples from Dolfing’s data set. Note that all the signatures depicted here have the same uniform stroke width. Since the original images have been rescaled for improved representation, the aforementioned property is not always clearly visible.	86
6.2	Examples of typical signature images contained in MCYT-75.	88
6.3	Examples of partial signature images contained in MCYT-75. These incomplete samples, a result of image <i>cropping</i> , were most likely obtained as a result of writers who signed outside a designated signing area.	88
6.4	Example of an incorrectly extracted signature image contained in MCYT-75. The dark region visible on the right is in all likelihood a result of incorrect page positioning during the digitisation process.	88
6.5	Schematic representation of the experimental protocol utilised to estimate system performance for a <i>single</i> run within a <i>single</i> trial.	90
6.6	Average AUCs achieved by the QDS and SVMS, when evaluated on (a) Dolfing’s data set and (b) MCYT-75, as a function of the projection angle set size T	95
6.7	Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on Dolfing’s data set for configurations C1–C4.	103
6.8	Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on MCYT-75 for configurations C1–C4.	103
6.9	Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on the Philips database for configurations C2 and C4.	104
6.10	Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on MCYT-100 for configurations C2 and C4.	104
7.1	Average feature weights reported in Swanepoel and Coetzer (2014), as determined using (a) the F -score and (b) the linear support vector weighting methods. The feature indices correspond to the columns of the feature set considered, that is $\mathbf{X} = [\mathbf{p}, \mathbf{x}, \mathbf{y}, \dot{\mathbf{x}}, \dot{\mathbf{y}}, \ddot{\mathbf{x}}, \ddot{\mathbf{y}}, \boldsymbol{\theta}_x, \boldsymbol{\theta}_y]$. Although both strategies clearly identify the vertical position \mathbf{y} , horizontal velocity $\dot{\mathbf{x}}$ and vertical velocity $\dot{\mathbf{y}}$ (indices 3–5) as highly discriminative, it should be noted that substantially different weights are associated with the horizontal position \mathbf{x} and the vertical acceleration $\ddot{\mathbf{y}}$ (indices 2 and 7).	118
7.2	Confidence weighting function candidates. It should be clear that these functions may intuitively be associated with a (a) conservative, (b) neutral, and (c) liberal approach to weight assignment, respectively.	121

List of Tables

3.1	The initial set of <i>five</i> function features in \mathfrak{R}^d , that is the signature data captured during the acquisition process after appropriate normalisation.	51
3.2	The final set of <i>eighteen</i> function features. Each vector $\mathbf{f} \in \mathfrak{R}^d$ is obtained from the initial feature set listed in Table 3.1.	52
3.3	Interpretation of the feature set dimensions associated with off-line and on-line signatures.	53
6.1	Summary of the data set partitioning considered in this study, including the number of writers in the data set (Ω), the number of writers in the training set (Ω_T), the number of writers in the evaluation set (Ω_E), the number of genuine samples per writer (N_ω^+), as well as the number of forged samples per writer (N_ω^-).	90
6.2	The set of experimental parameter values considered for off-line system evaluation.	91
6.3	Average AUCs and EERs achieved by the QDS when evaluated on Dolfing's data set. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.	93
6.4	Average AUCs and EERs achieved by the SVMs when evaluated on Dolfing's data set. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.	93
6.5	Average AUCs and EERs achieved by the QDS when evaluated on MCYT-75. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.	94
6.6	Average AUCs and EERs achieved by the SVMs when evaluated on MCYT-75. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.	94
6.7	Average AUCs and EERs achieved by the QDS and SVMs when evaluated on the Philips database.	96
6.8	Average AUCs and EERs achieved by the QDS and SVMs when evaluated on MCYT-100.	96
6.9	Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on Dolfing's data set, with those reported in this study.	97

LIST OF TABLES

xv

6.10	Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on MCYT-75, with those reported in this study.	98
6.11	Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on the Philips database, with those reported in this study.	100
6.12	Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on MCYT-100, with those reported in this study.	101
6.13	The set of system configurations considered for system re-evaluation.	102
6.14	Performance gradients (and corresponding test statistics) obtained between configurations C2 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.	108
6.15	Performance gradients (and corresponding test statistics) obtained between configurations C2 and C1 when the QDS and SVMS are evaluated on MCYT-75.	108
6.16	Performance gradients (and corresponding test statistics) obtained between configurations C3 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.	111
6.17	Performance gradients (and corresponding test statistics) obtained between configurations C3 and C1 when the QDS and SVMS are evaluated on MCYT-75.	111
6.18	Performance gradients (and corresponding test statistics) obtained between configurations C4 and C2 when the QDS and SVMS are evaluated on the Philips database.	112
6.19	Performance gradients (and corresponding test statistics) obtained between configurations C4 and C2 when the QDS and SVMS are evaluated on MCYT-100.	112
6.20	Performance gradients (and corresponding test statistics) obtained between configurations C4 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.	113
6.21	Performance gradients (and corresponding test statistics) obtained between configurations C4 and C1 when the QDS and SVMS are evaluated on MCYT-75.	113

List of Symbols

General	
ω	Writer
G	Positive class
F	Negative class
\cup	Concatenation operator
'	Transpose operator
$\mu(\mathbf{f})$	Mean of vector \mathbf{f}
$\sigma(\mathbf{f})$	Standard deviation of vector \mathbf{f}
Signature Representation	
\mathbf{x}	Feature vector
$\dot{\mathbf{x}}$	First derivative of \mathbf{x}
$\ddot{\mathbf{x}}$	Second derivative of \mathbf{x}
\mathbf{X}	Feature set
\mathbf{z}	Dissimilarity vector
$\bar{\mathbf{z}}$	Normalised dissimilarity vector
\mathbf{Z}	Dissimilarity set
d	Feature vector dimension
T	Feature set length / Dissimilarity vector dimension
Classifiers	
$\boldsymbol{\mu}$	Class mean vector
$\boldsymbol{\Sigma}$	Class covariance matrix
$ \boldsymbol{\Sigma} $	Determinant of $\boldsymbol{\Sigma}$
$\boldsymbol{\Sigma}^{-1}$	Inverse of $\boldsymbol{\Sigma}$
\mathbf{w}	Support vector machine weight vector
b	Support vector machine bias
C	Support vector machine regularisation parameter
γ	Radial basis function kernel width
Signature Modelling & Verification	
λ	Signature model
K	Reference set size
s	Confidence score
τ	Confidence threshold
y	Class label

List of Acronyms

AER	Average Error Rate
AMF	Adaptive Median Filter
AUC	Area Under Curve
DRT	Discrete Radon Transform
DTW	Dynamic Time Warping
EER	Equal Error Rate
FAR	False Acceptance Rate
FPR	False Positive Rate
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
IOR	Iterative Outlier Removal
LBP	Local Binary Pattern
(L/Q)DA	(Linear/Quadratic) Discriminant Analysis
MLP	Multilayer Perceptron
PDF	Probability Density Function
QDS	QDA-based System
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
R-SVM	RBF-kernel SVM
SMO	Sequential Minimal Optimisation
SVM	Support Vector Machine
SVMS	R-SVM-based System
TPR	True Positive Rate

Chapter 1

Introduction

“He who seeks for methods without having a definite problem in mind, seeks in the most part in vain.”

- David Hilbert (1862–1943)

1.1 Motivation

The purpose of this study is to develop a *proficient writer-independent handwritten signature verification system*, that is a system that automatically classifies a questioned handwritten signature sample as being either authentic or fraudulent. The proposed system should be proficient in the sense that it produces as few erroneous classifications as possible. The system should also be writer-independent in the sense that it is able to construct a single, universal signature model and subsequently use said model to authenticate the signature of *any* potential writer.

The work presented in this study is mainly inspired by, although not at all limited to, the process of transaction authentication within a banking environment. Despite the rise in popularity of alternative payment channels, including account purchases and online shopping, banking institutions are still faced with a prodigious number of signed cheques and card receipts that require processing on a daily basis.

The sole purpose of transaction authentication is to prevent fraudulent transaction documents (such as cheques bearing forged signatures) from being accepted as valid and consequently facilitating payment by an unauthorised party. The acceptance of fraudulent cheques, for example, constitutes a major concern for banking institutions across the globe, since the economic losses associated with this type of fraud accumulate to an alarming annual amount¹.

¹Due to the sensitive nature thereof, it is very difficult to obtain comprehensive and accurate data regarding global financial losses due to cheque fraud. However, according to Business Day (2009), it is estimated that the annual financial losses in the United States of America, as a result of cheque fraud, increased from \$12.6bn to \$20bn during the period 1996–2009. Also, the Cheque & Credit Clearing Company (2014) reported that annual losses associated with cheque fraud in the United Kingdom averaged approximately £13.2m during the period 2009–2013.

Continuing with the example of cheque fraud, it is reported in Business Day (2009) that many countries are following the current European trend of *decreased* cheque usage. In South Africa, for example, the use of cheques has reportedly declined by as much as 20% annually in recent years. However, this article also notes that this trend is in stark contrast with recent tendencies in other countries such as the United States of America, where *increased* cheque usage has resulted in *increased* losses due to cheque fraud. Furthermore, a decrease in cheque usage is typically associated with the increased use of alternative payment methods, such as cheque/credit/account cards – many of which still require transaction authentication by means of a handwritten signature. In essence, the problem of inadequate forgery detection is therefore not avoided, but merely shifted to a different medium.

Within the context of a banking environment, the task of document authentication is typically performed by an unskilled² *human* operator. Due to the fact that these operators are vastly outnumbered by the documents requiring their attention, a relatively small number of cheques and card receipts are eventually submitted for manual verification. Although an *automatic* signature verification system is likely to function effectively without any human intervention, it is worth mentioning that this study does not necessarily advocate the utilisation of such a system as an *alternative* to employing human verifiers. Instead, we aim to *enhance* the existing human-centric signature verification environment as outlined below.

Machines as an aid to human operators

According to an experiment conducted by Coetzer *et al.* (2006), the average human operator takes approximately 3.5–4.7 seconds to authenticate a handwritten signature. In addition, the efficiency and reliability of a human verifier's performance may be adversely affected by external factors such as fatigue, illness or boredom. In contrast, an efficient automatic verification system outputs real-time decisions and does not suffer from any of the physical constraints (and consequent inconsistencies in performance) associated with a human operator.

In order to accommodate the above-mentioned human limitations in the banking sector, only those transactions that exceed a specific monetary value are presented for manual authentication. This *threshold amount* is usually determined by the relevant banking institution, in such a manner that a manageable workload is ensured for its employees, thereby maintaining a reasonable payment processing period for its clients. Unfortunately, this practice is also common knowledge amongst forgers, who are able to circumvent detection by simply submitting cheques of a sufficiently small monetary value.

It should be clear that the introduction of one or more machine-based verifiers into the banking environment will result in a dramatic increase in document processing resources and may therefore potentially provide an effective solution to the current shortcomings associated with a threshold amount-based approach. Any transaction that exceeds the

²Human verifiers in the banking sector are generally considered “unskilled” in the sense that they rarely possess professional training in forensic handwriting analysis.

threshold amount may, for instance, still be authenticated by a human operator, whilst those that fall below this amount are submitted for automated verification. Furthermore, the threshold amount may be adjusted on an ad hoc basis, in order to prevent processing backlogs or to alleviate the workload of the human workforce (should one or more of the operators become unavailable).

Human-machine collaboration

The strategy outlined above is based on the assumption that human verifiers and automatic verification systems operate on a mutually exclusive basis. This does not necessarily have to be case. In a recent study by Coetzer *et al.* (2012), where the decisions of a workforce of human verifiers are *combined* with the threshold-specific decisions of an automatic verification system, it is shown that the expected performance of the resulting *hybrid* verifiers exceeds that of both the unaided humans *and* the unaided machine.

The aforementioned protocol for human-machine collaboration holds another important advantage – it facilitates the *dynamic* selection of the most appropriate hybrid verifier, based on the expected cost associated with misclassification. The proposed classifier selection process is shown to reduce the average expected cost associated with the erroneous acceptance/rejection of transactions, regardless of their monetary value.

The ability to perform *cost-sensitive* transaction authentication may prove a particularly attractive prospect to financial institutions such as retail banks, where a trade-off has to be made between security concerns and customer satisfaction. For instance, the acceptance of a fraudulent transaction with an exceptionally large monetary value may result in an unacceptable financial loss for the banking institution. In order to avoid this scenario, a banking institution may be persuaded to instruct its staff to maintain a bias towards rejecting transactions. However, the repeated rejection of legitimate transactions, regardless of their monetary value, may eventually prove such an inconvenience to the client, that he/she may consider terminating his/her account(s) and seek the services of a competing institution. From the bank's perspective, this would of course also be considered unacceptable.

The introduction of a proficient automatic verification system, into a banking environment, is therefore expected to dramatically reduce the number of undetected fraudulent transactions, whilst the subsequent reduction in monetary losses undoubtedly warrants an investigation into the feasibility of such a system. In this study, we perform such an investigation.

Finally, it is worth mentioning that the motivations outlined thus far specifically pertains to the traditional problem of *off-line* signature verification. In the case of *on-line* signature verification (see Section 1.2.4), where the authentication process requires *no* human intervention, the motivation is more straightforward – the objective is simply to develop a system that is as accurate as possible, whilst the number of genuine signature samples required during writer enrolment is minimised.

1.2 Key concepts

The development of a proficient automatic handwritten signature verification system requires the successful exploitation of the synergy between various concepts throughout the biological, social and mathematical sciences. In this section, we present brief discussions on several of the fundamental concepts involved in such a development.

1.2.1 Personal identification systems

In a society where wealth is often difficult to acquire, it is of paramount importance for individuals to protect their accumulated assets. The desire to ensure that only authorised individuals should be granted access to certain resources gave rise to the development of *personal identification systems*. These security systems are designed to authenticate the claimed identity of a person, based on properties supposedly unique to said individual, in order to determine whether or not he/she should be granted access to specific resources.

A personal identification system may be categorised as being either *knowledge-based*, *possession-based* or *biometric*, as illustrated in Figure 1.1. In order to determine the most appropriate type of identification system for any given scenario, several factors demand consideration, since each type of system has its own set of advantages and limitations. Generally, the most important factor is the trade-off between the level of security provided and the practical/financial feasibility of system deployment.

A *knowledge-based* system grants its users access to the desired resource only upon the presentation of predetermined *information*, usually a password or access code. The deployment of such a system is relatively straightforward, since its only design requirements are the mechanisms necessary to store and match information. This ease of implementation has resulted in the widespread deployment of knowledge-based systems. The utilisation of user names and associated passwords is probably the most well-known modern example of this type of system. However, the major drawback associated with such a system is the fact that its security is solely dependent on the assumption that *all* (and *only*) authorised individuals possess the knowledge required to gain access to the protected resource. This is of course an unrealistic assumption. Passwords and access codes are easily (and therefore often) forgotten, which prevents authorised individuals from gaining access to the desired resource. Also, unless a concerted effort is made to protect the information in question (by means of e.g. an advanced encryption standard), tokens such as passwords and access codes may be intercepted by unauthorised individuals, which consequently nullifies the security system.

In order to prevent such an information-based security breach, a *possession-based* identification system requires the presentation of a *physical* token, usually a key or access card. This type of system can also be deployed with relative ease, since the hardware required is generally uncomplicated and inexpensive. The most well-known example of a possession-based identification system is undoubtedly the use of a unique physical lock with its corresponding key. The vulnerabilities of a possession-based system is, however, fairly similar to that of its knowledge-based counterpart. Tokens such as keys or access

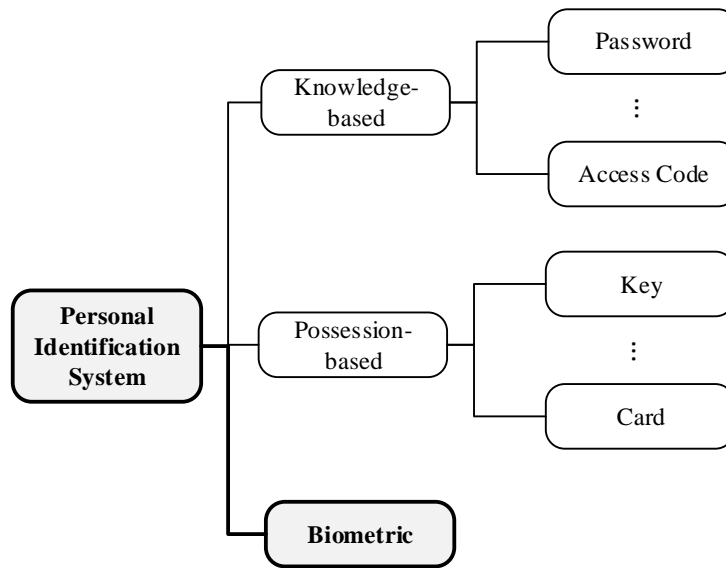


Figure 1.1: Categorisation of selected personal identification systems. Categories relevant to this study are emphasised in bold. The attributes associated with the design of a biometric system are sub-categorised in Figure 1.2.

cards may be lost or stolen (or duplicated with relative ease). Nevertheless, in order to circumvent a possession-based system, an unauthorised individual requires *direct* access to the physical token. This type of system therefore eradicates the possibility of a remote security breach.

It should be clear from the above discussions that the primary weakness of both knowledge-based and possession-based systems is the fact that the entity used for identity verification is not an attribute of the actual individual, but rather a token deemed *representative* of said individual. In order to avoid this pitfall, a *biometric* identification system strictly requires the physical presence of the authorised *individual* and performs ad hoc identity verification by considering an attribute measured directly from his/her person. This attribute may either be a measurable property of the person's physique (i.e. a *physiological* attribute) or a gestural feature that was developed/trained over time and therefore deemed unique to the individual in question (i.e. a *behavioural* attribute). Selected examples of physical and behavioural biometric attributes are presented in Figure 1.2.

It is a well-known fact that biometric identification systems based on physiological attributes are significantly more accurate than those based on behavioural traits and therefore provide a much higher level of security. However, due to the specialised hardware required to deploy a physiological biometric authentication system, the use of such systems is often deemed too impractical for widespread use. Furthermore, the invasive nature of many physiological feature acquisition devices (e.g. retinal scanners) often proves unpopular amongst

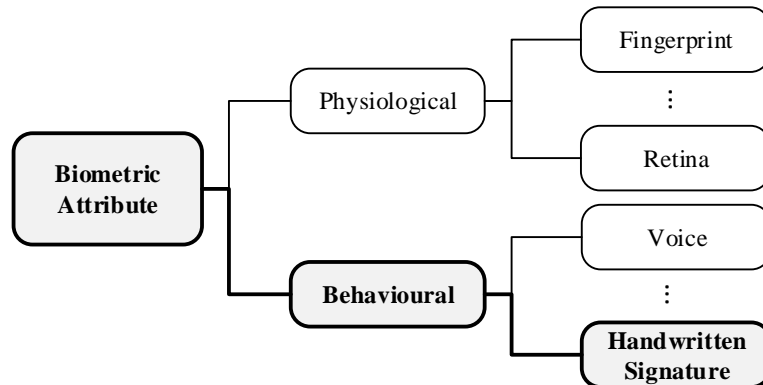


Figure 1.2: Categorisation of selected biometric attributes. Categories relevant to this study are emphasised in bold.

the general public. In contrast, behavioural attributes may be easily acquired with a non-invasive device – a property that makes the use of behavioural biometric authentication systems a particularly attractive prospect.

In this study we develop biometric authentication systems that are based on a behavioural attribute. The specific attribute considered here is the *handwritten signature*, which is discussed in the following section.

1.2.2 Handwritten signatures

The use of handwritten signatures (henceforth referred to only as *signatures*), as a means of identity verification, has long been one of the most widely implemented biometric authentication techniques the world over. Since signature production constitutes the result of a highly complex and well-trained neuromuscular process, as explained in e.g. Nguyen (2012), it is widely believed that signatures are able to reflect personal idiosyncrasies that are unique amongst individuals. This notion of an inherent connection between a person’s identity and his/her signature is supported by e.g. Schmidt (1994), where it is explained that an individual’s signature is typically composed of stroke sequences much unlike those used in ordinary handwriting and, as a result of sustained repetition over a prolonged period of time, tends to evolve towards a consistent, unique pattern.

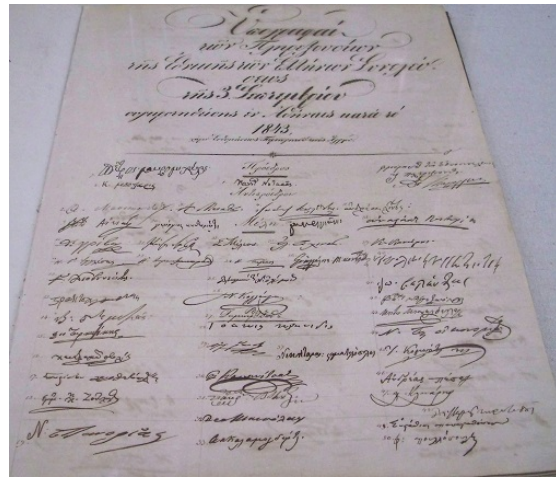
Even in our present day and age, where regular quantum leaps in the field of information technology strive towards ultimately delivering a so-called *paperless society*, signatures remain a socially and legally accepted proof of identification and consent. This universally accepted convention has stood for centuries, as illustrated by the examples of historic documents presented in Figure 1.3.

The above-mentioned convention is especially important within the financial sector, where transactions and other legal contracts are often subject to future disputes. In order

CHAPTER 1. INTRODUCTION



(a)



(b)

Figure 1.3: Examples of historic documents that were authorised by means of handwritten signatures. (a) The Declaration of Independence of the United States of America, as signed by 56 delegates of the US Congress on July 4th 1776. (b) A document signed by members of the Greek National Assembly in Athens (after the revolution on September 3rd 1843) that drafted a new constitution and made Greece a constitutional monarchy.

to resolve such a dispute, one is required to verify two key aspects. Firstly, that each party involved had given their consent at the time of the agreement and, secondly, that the consenting individuals do in fact represent the parties involved. As mentioned earlier, signatures are legally deemed representative of both identity and consent, and are therefore considered ideal for the endorsement of official documentation. For this reason, a legal document is generally only considered valid once it has been signed personally by each party (or by an officially designated proxy).

1.2.3 Forgery categorisation

The major drawback associated with biometric authentication based on handwritten signatures, as is the case with any behavioural biometric, is that the opportunity is created for an imposter to perpetrate identity fraud by producing imitations (i.e. forgeries) of the genuine article with relative ease. As illustrated in Figure 1.4, forged signatures may generally be categorised, in increasing order of quality, as either *random*, *casual* or *skilled*. Typical examples of these different types of forgeries are provided in Figure 1.5.

A random forgery (see Figure 1.5 (b)) is produced without *any* prior knowledge regarding a genuine signature. This type of forgery typically bears no resemblance to a genuine

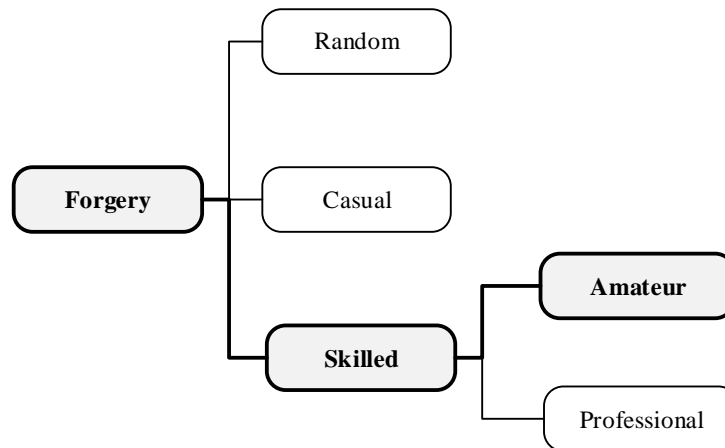


Figure 1.4: Categorisation of forged handwritten signatures. Categories relevant to this study are emphasised in bold.

signature and can be easily detected by a human verifier. For experimental purposes, genuine signatures belonging to writers *other* than the writer in question are often used to represent random forgeries.

Casual forgeries are the result of arbitrary attempts at signature reproduction, given *only* the name of the signature's owner. This type of forgery may therefore constitute a reasonable semantic reproduction of a genuine signature, but typically differs significantly from an authentic sample in terms of its stylistic properties. In most cases, casual forgeries can also be easily detected by human verifiers.

Only when a forger has unrestricted access to at least one genuine signature sample, as well as ample time to practise its reproduction, may a skilled forgery be produced. The forger therefore has the opportunity to not only produce a forgery similar in design, but also mimic the inherent stylistic properties of a genuine signature. In the case of skilled forgeries, a further distinction can be made between *amateur* (see Figure 1.5 (c)) and *professional* (see Figure 1.5 (d)) forgeries, where a professional forger also possesses a degree of forensic expertise. Skilled forgeries, especially those produced by professional forgers, are undoubtedly the most difficult to detect. Since an untrained human verifier (e.g. a typical bank official) is unlikely to detect such a forgery, the efforts of a forensic document examiner, or a sophisticated automatic verification system, is generally required. It is not reasonable to expect that a typical banking institution would employ forensic experts.

Apart from the *traditional* forgery types described above, a growing interest has recently developed in the detection of so-called *disguised* forgeries. Unlike the name suggests, this type of signature is actually not a forgery at all, but rather a deliberate attempt by the legitimate signer to create reasonable doubt regarding the authenticity of his/her signature at a later stage. The signer could, for instance, use his/her non-dominant hand to produce



Figure 1.5: A typical example of (a) a genuine signature, as well as attempts at its reproduction in the form of a (b) random, (c) amateur skilled and (d) professional skilled forgery.

the signature. The signing event is therefore considered legitimate, although future analysis of the signature may suggest a fraudulent attempt.

The systems developed in this study are specifically aimed towards the detection of amateur skilled forgeries. It should be clear, however, that any system developed for the detection of skilled forgeries should also be able to easily detect casual and random forgeries. These systems will not necessarily be proficient in the detection of professional or disguised forgeries.

1.2.4 Signature acquisition

In order to facilitate an effective machine-based analysis, signature data must first be captured using one of two fundamentally different modalities, namely *off-line* or *on-line* acquisition.

The acquisition of off-line signature data involves the digitisation, usually by means of a flatbed scanner, of an existing *pen-on-paper* sample. This process yields a static image, wherein each pixel is indicative either of the foreground (i.e. a pen stroke) or the background (i.e. the document). Unfortunately, since an off-line signature may be produced using an arbitrary writing implement on an arbitrary writing surface, a clear distinction between foreground and background is not always achievable. Furthermore, the digitisation process itself may introduce *noise* into the resulting image, that is the occurrence of actual background pixels that are erroneously digitised as apparent foreground pixels (or vice versa). A proficient off-line signature verification system therefore typically employs several image processing techniques, specifically aimed at addressing these potential im-

pediments. Finally, it should be clear that, since a static signature image is completely devoid of temporal information, the analysis of an off-line signature sample is strictly limited to the interpretation of spatial information, as indicated by the apparent pen stroke pixels.

In contrast, an on-line signature is produced using specialised hardware, such as a digitising pen and tablet (or similar electronic device). A major advantage associated with the use of an on-line signature acquisition platform is the fact that the data capturing process is completed accurately and automatically, during the signing event, by the device itself. Depending on the specific device utilised (several suitable signature acquisition devices are described in e.g. Brockly *et al.* (2014)), the data acquired during such a signing event is typically related to several different descriptors, such as the pen stroke coordinates, pen pressure and pen angle – each as a function of time. The information contained in this initial set of descriptors may be exploited even further through the calculation of additional temporal descriptors relating to e.g. pen trajectory, velocity and acceleration. Furthermore, unlike its off-line counterpart, the quality of on-line signature data is not adversely affected by factors such as poor contrast or background noise. Ultimately, the wealth of information contained in an on-line signature sample makes such instances significantly more difficult to imitate, since the forger needs to reproduce not only the spatial properties of the genuine article, but also several writer-specific stylistic idiosyncrasies. For this reason, the proficiency of a typical on-line signature verification system is expected to be significantly superior to that of a typical off-line system.

In this study we are concerned with the generalised concept of handwritten signature verification and therefore do not limit our investigation to a specific type of signature data. Consequently, we aim to develop proficient systems for *both* the off-line *and* on-line scenarios and show that the design of these fundamentally different systems differ *only* in terms of the respective techniques employed for the conversion of the input signal (i.e. static images or time series data) into a representation suitable for machine-based analysis.

1.2.5 Pattern recognition

The human brain performs numerous tasks relating to pattern recognition during every conscious moment. Our innate ability to recognise and interpret patterns (e.g. faces, voices, gestures, etc.) governs nearly every concept that allows modern society to function. For example, without the pattern recognition process, the concept of communication would be inconceivable. Furthermore, it should come as no surprise that this ability to (practically instantaneously) recognise a specific scenario, based on past experiences, inherently forms the basis of any type of intelligent decision-making.

From a machine-based perspective, the pattern recognition process involves the automatic classification of a questioned entity, or *pattern*, as belonging to a specific category of predefined entity types, or *pattern classes*, based on properties that are considered historically representative of the classes in question. A comprehensive review of this process, as conceptualised in Figure 1.6, may be found in e.g. Jain *et al.* (2000).

The pattern recognition process is generally subdivided into two key phases, namely

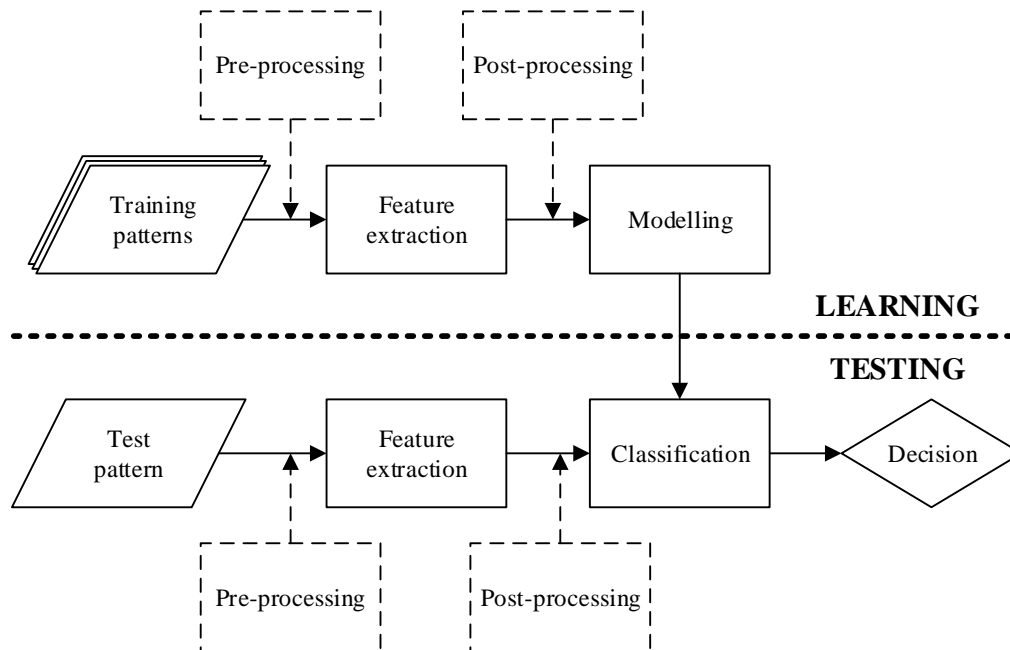


Figure 1.6: Schematic representation of the pattern recognition process. The additional data processing stages (indicated with dashed lines/borders) are considered optional, but are often included in order to improve system performance.

learning and *testing*. We now present a generalised discussion of the aforementioned phases, whilst the specific design elements relating to the systems developed in this study are discussed in Section 1.4.1.

Learning

The learning phase is concerned with the construction of a mechanism that is able to interpret and analyse historic patterns (typically of *known* origin), in such a way that it should be equally adept at a similar interpretation and analysis of future patterns (of *unknown* origin). As seen in Figure 1.6, the learning phase is characterised by two primary processes, namely *feature extraction* and *modelling*.

Feature extraction. From a mathematical perspective, a pattern (e.g. a signature) may be represented by an *observation sequence*, or *feature set* $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$. A feature set is constructed from a collection of T , d -dimensional *observations*, or *feature vectors* $\mathbf{x}_i = \{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(d)}\}$. Each feature vector element $x_i^{(j)}$ constitutes an arbitrary measurable quantity, or *feature*. The process of converting a pattern from its original state into a suitable representation in *feature space*, that is the multidimensional Cartesian space

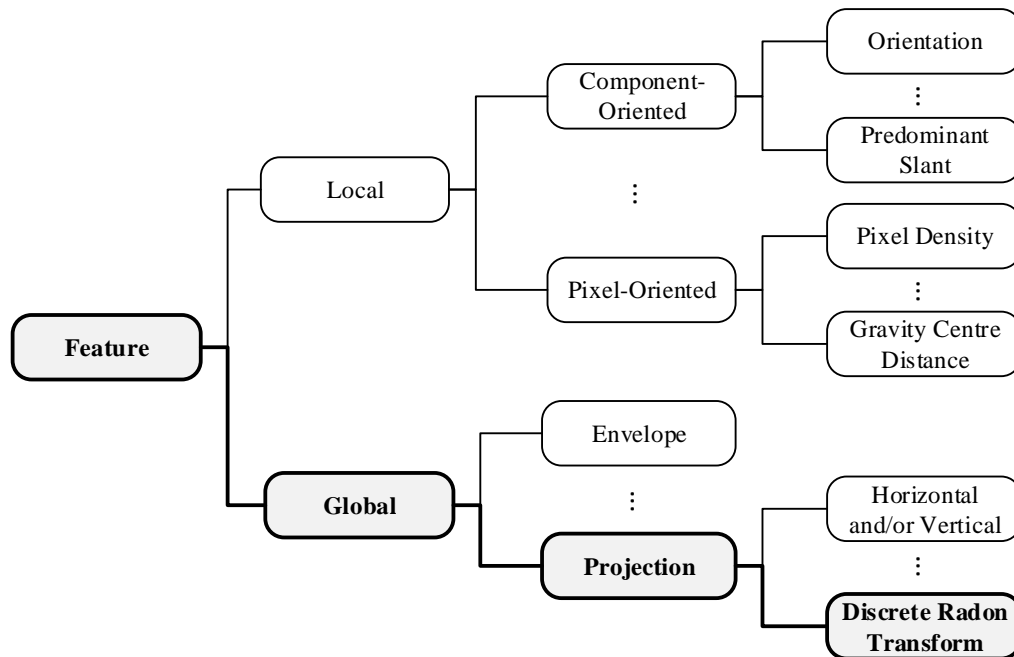


Figure 1.7: Categorisation of selected feature extraction techniques associated with *off-line* signature verification. Categories relevant to this study are emphasised in bold.

where each dimension corresponds to a specific feature, is known as feature extraction.

Although the descriptors utilised for pattern representation may be chosen arbitrarily, the selection of an effective feature extraction technique remains an issue that demands serious consideration, since the ultimate goal is to construct a feature-based pattern representation that maximally discriminates between the pattern classes in question. In an ideal scenario, each pattern class would occupy a compact, disjoint region in feature space. Several popular features utilised in the fields of off-line and on-line signature verification are categorised in Figures 1.7 and 1.8 respectively.

Depending on the application, it may be advisable to also include several data processing techniques into the feature extraction process, in order to maximise the efficacy of the resulting pattern representation. These techniques are generally concerned with improving the quality of the data presented for feature calculation, as well as compensating for the potential variability observed in different patterns belonging to the same pattern class. Within the context of signature representation, for instance, it is reasonable to expect that different signature samples, belonging to the same writer, may exhibit variations in terms of position, scale and/or rotation.

Modelling. Once a technique has been identified that is able to effectively construct a feature-based pattern representation, the efforts of a *classifier* are required, in order

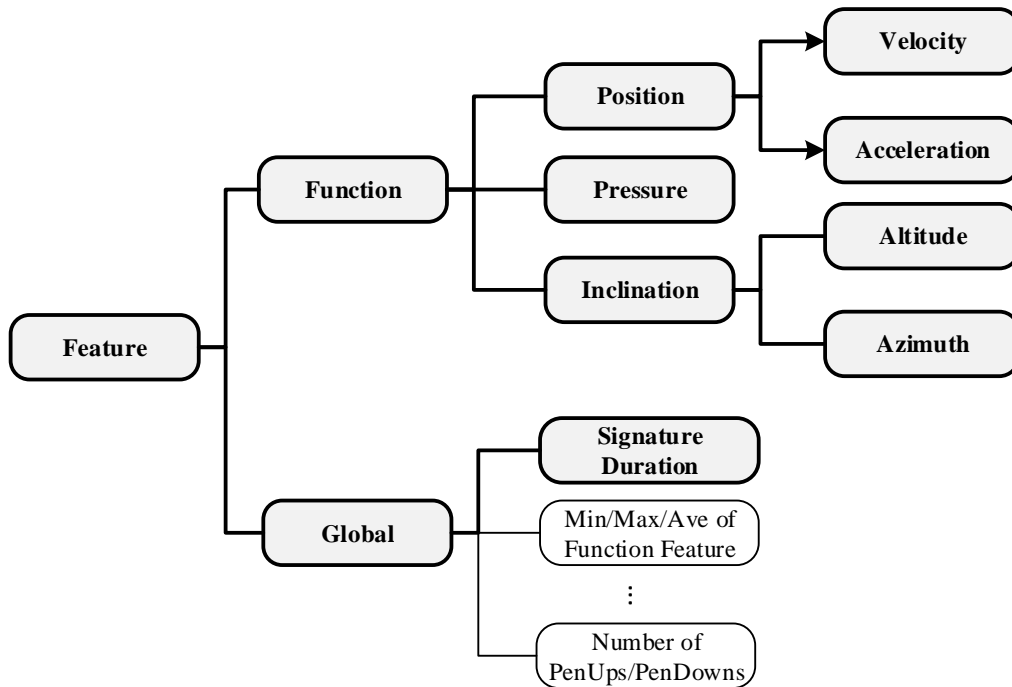


Figure 1.8: Categorisation of selected feature extraction techniques associated with *on-line* signature verification. Categories relevant to this study are emphasised in bold. Note that the arrow notation indicates that velocity and acceleration may be *derived from* position data.

to distinguish between the different pattern classes represented in feature space. The construction of such a classifier may be based on techniques as simple as the calculation of class-specific descriptive statistics, but more frequently involves the training of a more sophisticated statistical construct. Several popular classification techniques are categorised in Figure 1.9.

The successful training of any classifier requires a sufficiently large collection of historical patterns, referred to as a *training set*. The minimum amount of training data required is entirely dependent on the type of classification technique under consideration, although an increase in the size of the training set almost invariably results in a superior classifier. Furthermore, if the class label associated with each training sample is known beforehand, the process is referred to as *supervised learning*. In contrast, *unsupervised learning* occurs when the pattern classes need to be identified automatically, based on their relative measures of location and dispersion in feature space (i.e. clustering).

It should be noted that, since the systems developed in this study perform signature *verification* rather than *recognition*, the classification task is effectively reduced to a two-class problem. Specifically, a pattern is said to belong to either the *positive class* (i.e. the set of genuine signatures) or the *negative class* (i.e. the set of forgeries). As a result, certain

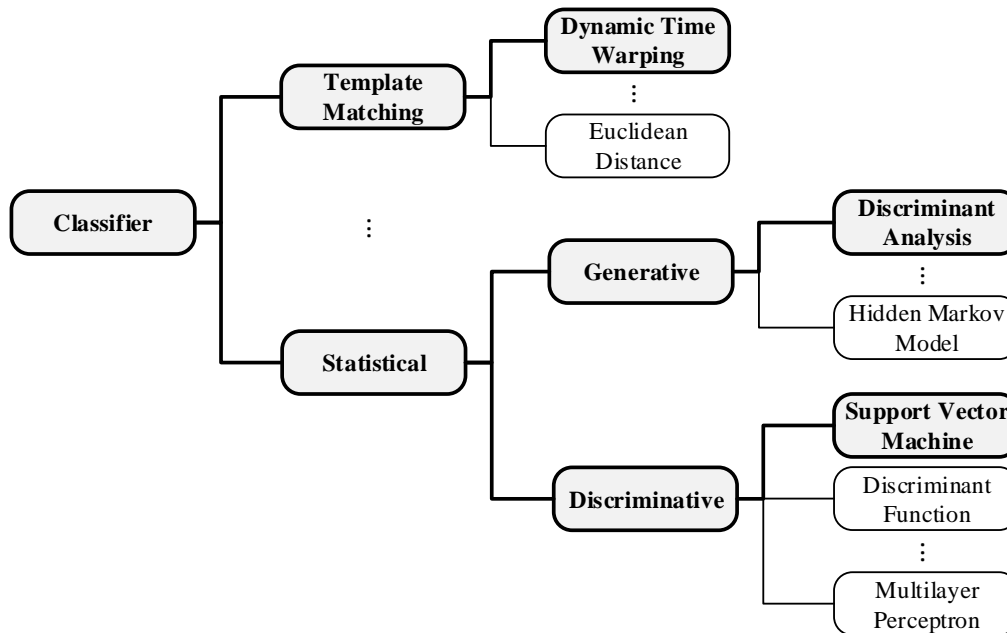


Figure 1.9: Categorisation of selected classification techniques associated with handwritten signature verification. Categories relevant to this study are emphasised in bold.

terminology is occasionally used interchangeably. For instance, when reference is made to a classifier (or classification) in this study, a verifier (or verification) is often implied.

A classifier typically receives as input the feature set extracted from a raw pattern and subsequently emits a numeric score. Depending on the specific type of classifier utilised, this score is typically representative of either:

- the similarity between the questioned pattern and a reference pattern,
- a likelihood based on the similarity between the questioned pattern and those used to train a probabilistic model, or
- an inverse distance measure relative to a decision boundary in feature space.

However, using a suitable transformation technique, any of the aforementioned scores may be expressed as a measure of *confidence* $s \in [0, 1]$, where a confidence of 1 denotes a perfect match, whilst a confidence of 0 indicates a complete mismatch. For the purposes of this discussion, we may assume, without loss of generality, that any classification event yields such a confidence measure.

In essence, the classifier training process therefore entails the analysis of feature sets extracted from training data, in order to uncover as much information as possible regarding

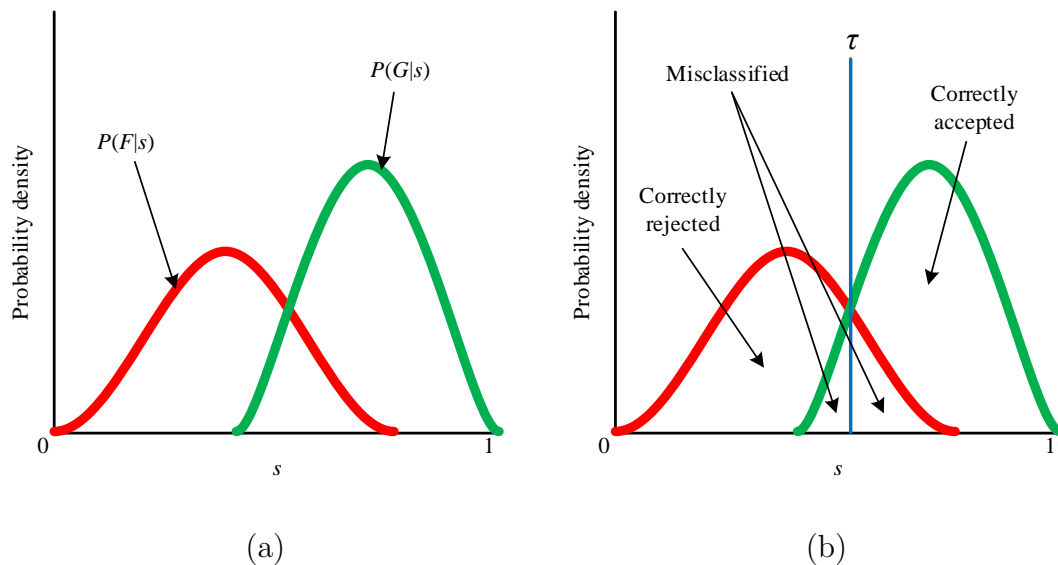


Figure 1.10: (a) Idealised confidence distributions for the positive class G and the negative class F . Note that the degree of overlap between $P(G|s)$ and $P(F|s)$ is proportional to the quality of the forgeries considered. (b) Class membership prediction by means of confidence thresholding, where a confidence score above (or below) the chosen threshold τ is considered indicative of a genuine (or forged) signature. It should be clear that, in a scenario where $P(G|s)$ and $P(F|s)$ are non-separable (as is often the case with *skilled* forgery detection), *any* value for τ inevitably results in a number of erroneous class membership predictions (i.e. *misclassifications*). Several suitable methods commonly utilised to quantify system performance are discussed in Section 1.2.7.

the confidence distributions associated with the pattern classes in question, as conceptualised in Figure 1.10 (a) within the context of signature verification. It is this empirical prior knowledge that enables a classifier to accurately predict the class membership of questioned patterns in the future.

The systems developed in this study utilise either a *quadratic discriminant* or a *support vector machine* (SVM) for classification purposes. Said classifiers are trained by means of supervised learning. It should be noted that each one of the proposed systems also utilises a *dynamic time warping* (DTW) algorithm. However, this algorithm is not employed in its traditional role as classifier, but rather as a *dichotomy transformation* for the purpose of obtaining a writer-independent signature representation, as explained in Section 1.2.6.

Testing

Assuming the successful completion of the preceding learning phase, the testing phase is concerned with the automatic assignment of a class label to any questioned pattern subsequently submitted for classification.

When a pattern of unknown origin is submitted for evaluation by a trained classifier, the evaluation process either produces a decision (i.e. a class label) or a confidence score,

depending on the specific *type* of classifier utilised. The former and latter classifiers are respectively referred to as *discrete* and *continuous* classifiers. In the case of a continuous classifier, an appropriate *confidence threshold* is imposed in order to ultimately obtain a decision, as conceptualised in Figure 1.10 (b).

Although the discriminative classifiers utilised in this study are typically associated with discrete outputs, we have opted for a modified approach that converts the initial classifier output, that is a *distance* measure relative to a predetermined decision boundary in feature space, into a *confidence score*. This conversion process is further discussed in Section 1.4.1. In other words, each system developed in this study employs a continuous classifier for the purposes of signature modelling and verification.

1.2.6 Writer-independent signature modelling

The majority of systems presented in the literature make use of a *writer-dependent* approach to signature modelling and verification, as conceptualised in Figure 1.11 (a). When this strategy is utilised, each writer ω enrolled into the system submits a set of K genuine training samples, in order to construct a model $\lambda^{(\omega)}$ in *feature space* that is specific to said writer. Any questioned sample $\mathbf{X}^{(q)}$ and claim of ownership ω subsequently presented to the system for verification, is then matched to $\lambda^{(\omega)}$ in order to obtain a class label $y \in \{1, -1\}$, such that

$$y = \lambda^{(\omega)}(\mathbf{X}^{(q)}), \quad (1.1)$$

where $y = 1$ (or $y = -1$) corresponds to the acceptance (or rejection) of the claim of ownership. This popular approach, however, has two notable disadvantages:

- A relatively large training set, which is not realistically obtainable in practical scenarios, is required to produce a sufficiently representative writer model.
- When a discriminative classifier (such as an SVM) is utilised for classification, *only* random forgeries may be used for model training, since it is not reasonable to assume that skilled forgeries will be available for *every* new writer enrolled into the system.

In contrast, a *writer-independent* approach (see Figure 1.11 (b)) aims to construct a *single* signature model λ that discriminates between two classes only, that is a positive and negative class that represents genuine and forged instances respectively. Most importantly, said instances may belong to *any* writer. A writer-independent system therefore attempts to model differences between positive and negative instances in a generic sense. It should be clear from Figure 1.11 (a) that a standard feature space representation is not suitable for such a modelling objective. Instead, a writer-independent model is constructed in *dissimilarity space* (Pekalska *et al.* (2002)).

The required dissimilarity representation is obtained by means of a *dichotomy transformation*, that is a process that compares a writer-specific positive or negative sample \mathbf{X} to a writer-specific *positive* reference sample $\mathbf{X}^{(k)}$ in order to produce a *dissimilarity vector* \mathbf{z} , such that

$$\mathbf{z} = D(\mathbf{X}^{(k)}, \mathbf{X}), \quad (1.2)$$

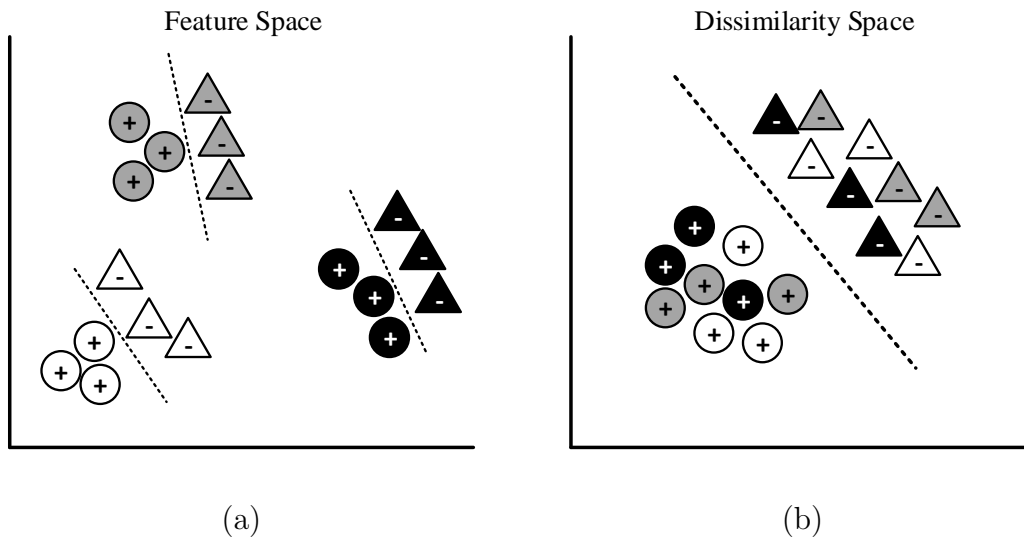


Figure 1.11: Conceptual comparison of the (a) *writer-dependent* and (b) *writer-independent* approaches to signature representation in feature space and dissimilarity space respectively, when three writers are considered. Positive and negative instances are indicated with “+” and “-” respectively.

where $D(\cdot)$ denotes any suitable distance measure. During system deployment, any newly enrolled writer ω submits a set of K genuine reference samples. Any questioned sample $\mathbf{X}^{(a)}$ and claim of ownership ω presented to the system for verification, is first subjected to the dichotomy transformation, in order to produce the dissimilarity vector $\mathbf{z}^{(a)}$. Subsequently, $\mathbf{z}^{(a)}$ is matched to λ , in order to accept or reject the claim of ownership, such that

$$y = \lambda(\mathbf{z}^{(a)}). \quad (1.3)$$

This approach therefore provides effective solutions to the problems of data scarcity and a lack of skilled forgeries for training purposes.

The systems developed in this study utilise a writer-independent approach to signature modelling and verification.

1.2.7 Performance metrics

A wide variety of quality performance measures (i.e. metrics that aim to quantify system proficiency) are considered throughout the literature. The most commonly used metrics include the *false rejection rate* (FRR), *false acceptance rate* (FAR), *average error rate* (AER), *equal error rate* (EER) and *accuracy* α .

In order to define the aforementioned performance measures, let us first consider the four possible outcomes of any verification experiment. If the instance submitted for verification is positive (i.e. a genuine signature), its correct classification constitutes a *true*

positive event, whilst its incorrect classification results in a *false negative* event. Conversely, if the instance is negative (i.e. a forged signature), its correct classification constitutes a *true negative* event, whilst its incorrect classification results in a *false positive* event. Given N instances, we denote the total number of true positive, false negative, true negative and false positive events with T^+ , F^- , T^- and F^+ , respectively. Furthermore, let n^+ and n^- denote the number of positive and negative instances considered for verification, respectively, such that $n^+ + n^- = N$.

The FRR quantifies the number of false negative events in relation to the number of positive instances submitted for verification, or

$$\text{FRR} = \frac{F^-}{n^+}. \quad (1.4)$$

Similarly, the FAR quantifies the number of false positive events in relation to the number of negative instances submitted for verification, or

$$\text{FAR} = \frac{F^+}{n^-}. \quad (1.5)$$

The average of the FRR and FAR is referred to as the AER. In contrast, the accuracy metric does not attempt to identify potential system weaknesses, but rather quantifies the system's overall ability to make a *correct* classification, such that

$$\alpha = \frac{T^+ + T^-}{N}. \quad (1.6)$$

It should be clear that each of the aforementioned performance metrics is inherently suited for the evaluation of *discrete* classifiers only. For *continuous* classifiers, the verification threshold τ is allowed to vary arbitrarily, where each unique threshold results in a unique discrete classifier. It should also be clear that an increased value of τ would invariably increase the FRR and decrease the FAR, as conceptualised in Figure 1.12 (a). It is therefore expected that, for some specific value of the verification threshold, the values of the FRR and FAR must coincide. This common value is known as the EER and its corresponding threshold value is denoted by τ_{EER} (see Figure 1.12 (a)).

Another popular platform for proficiency testing involves the depiction of system performance in *receiver operating characteristic* (ROC) space, that is the two-dimensional Cartesian space wherein the *true positive rate* (TPR) is plotted as a function of the *false positive rate* (FPR). Whilst the FPR is synonymous to the previously defined FAR, the TPR quantifies the number of true positive events in relation to the number of positive instances submitted for verification, or

$$\begin{aligned} \text{TPR} &= \frac{T^+}{n^+} \\ &= 1 - \text{FRR}. \end{aligned} \quad (1.7)$$

The corresponding pair of FPR and TPR values, resulting from the evaluation of a discrete classifier, is consequently depicted by a single point in ROC space. The performance

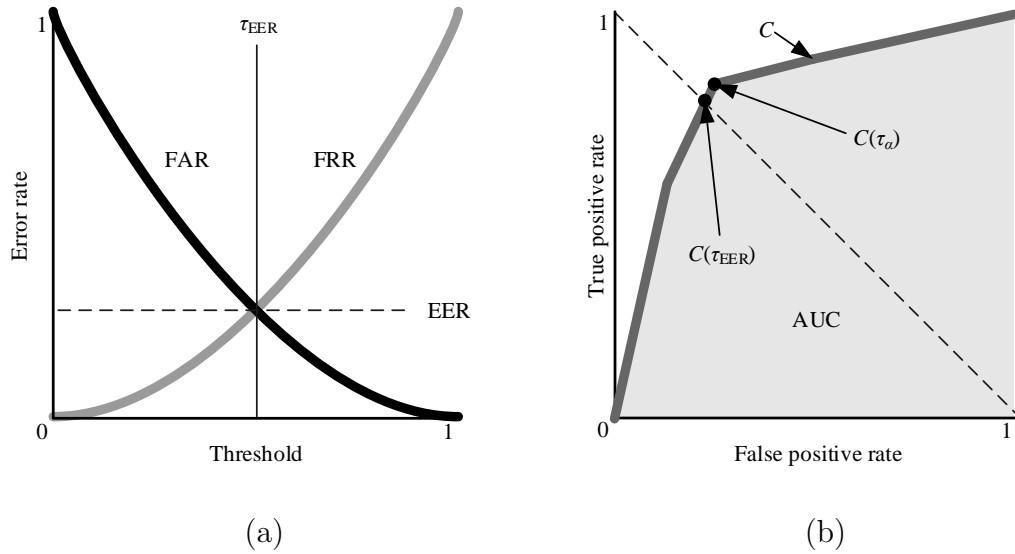


Figure 1.12: Conceptualisation of the system performance metrics considered in this study. (a) The FRR and FAR as functions of the verification threshold τ . Also indicated is the EER of the underlying continuous classifier. (b) Performance evaluation of a hypothetical classifier in ROC space. The continuous classifier C is depicted by a ROC curve, whilst each threshold-specific discrete classifier $C(\tau)$ is depicted by a single point in ROC space. Also indicated are the optimal α -based and EER-based discrete classifiers, denoted by $C(\tau_{\alpha})$ and $C(\tau_{EER})$ respectively.

of a continuous classifier is therefore represented by several threshold-specific ROC points that collectively form a so-called ROC curve. Such a ROC-based representation also allows one to determine the *area under curve* (AUC) metric for any given classifier, that is the area spanned by the convex hull of its resulting ROC point(s), as well as the points (0,0), (1,0) and (1,1), as illustrated in Figure 1.12 (b). Unlike the previously discussed performance measures, the AUC therefore facilitates system performance estimation for all possible verification thresholds. The AUC may also intuitively be interpreted as the probability that the system will rank a randomly chosen positive instance higher than a randomly chosen negative instance. For a comprehensive discussion on ROC-based performance analysis, the reader is referred to Fawcett (2006).

The systems developed in this study are evaluated both in terms of the AUC and EER performance measures. We are primarily concerned with the AUC, since this metric provides a stable and comprehensive indication of continuous classifier performance. We also consider the EER, however, since it remains the most commonly reported metric found in the literature. The EERs reported in this study therefore place the performance of each proposed system into a familiar context, whilst also providing a sensible platform for performance comparison with prior work.

1.3 Objectives

During the course of this study, we aim to achieve two primary objectives, namely the successful design and implementation of:

- a novel strategy for writer-independent *off-line* signature representation. We specifically aim to investigate the use of the *discrete Radon transform* (DRT) in conjunction with a DTW-algorithm for the respective purposes of feature extraction and the subsequent dichotomy transformation; and
- a novel strategy for the incorporation of writer-specific information into a writer-independent *off-line* or *on-line* signature modelling framework. We specifically aim to investigate the feasibility of a *writer-specific dissimilarity normalisation function*.

Furthermore, we aim to incorporate the above-mentioned concepts into the design of novel and proficient *off-line* and *on-line* signature verification systems.

1.4 Overview of this study

In this study we develop a set of four writer-independent handwritten signature verification systems, that is two off-line verification systems and two on-line verification systems. In both the off-line and on-line cases, each system employs either *quadratic discriminant analysis* (QDA) or an SVM for signature modelling and verification. As a result, each one of the aforementioned systems may be considered unique.

In this section we provide a condensed review regarding the design of these systems. We also briefly discuss the experimental protocol considered for system evaluation and highlight selected results achieved by the proposed systems.

An initial conceptualisation of the proposed QDA-based *off-line* signature verification system is presented in Swanepoel and Coetzer (2012), whilst the proposed SVM-based *off-line* signature verification system is introduced in Swanepoel and Coetzer (2013). An initial conceptualisation of the proposed SVM-based *on-line* signature verification system can be found in Swanepoel and Coetzer (2014).

1.4.1 System design

The design of a typical signature verification system developed in this study is conceptualised in Figure 1.13.

The proposed design framework adheres to the basic principles of the pattern recognition process illustrated in Figure 1.6 (see Section 1.2.5). It is clear from this conceptualisation that the development of each system may be divided into three key stages, namely *signature representation*, *signature modelling* and *signature verification*. We now provide an abridged overview of these key stages, whilst comprehensive discussions on said processes are reserved for Chapters 3–5.

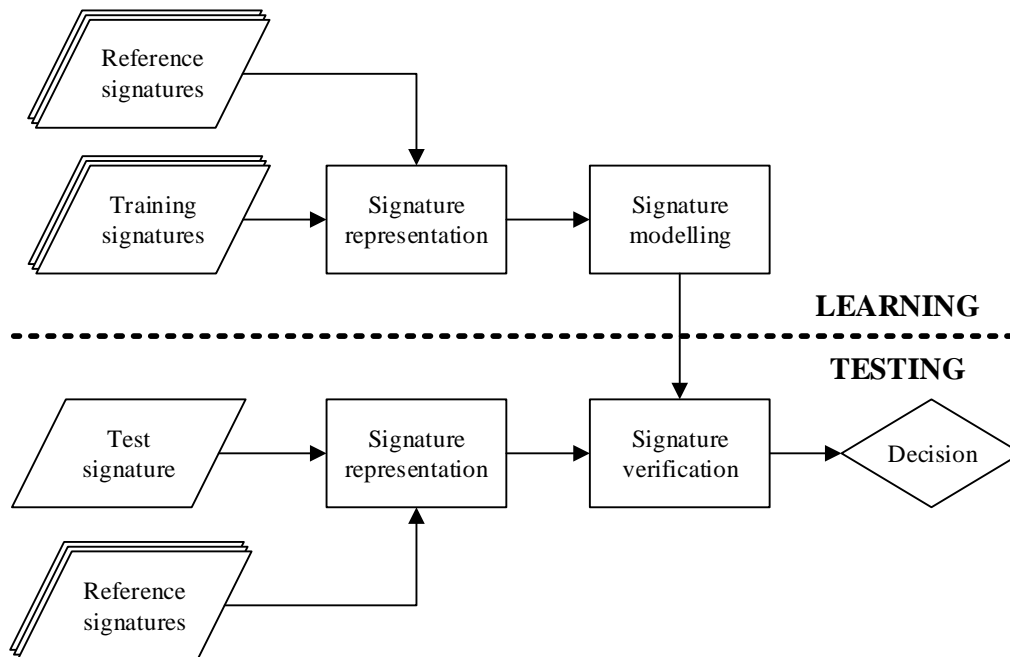


Figure 1.13: Schematic representation of a typical system developed in this study. Detailed schematics of the signature representation, signature modelling and signature verification processes are presented in Figures 1.14–1.17.

Signature representation

As mentioned earlier, the systems developed in this study are concerned with two fundamentally different types of signature data, that is off-line and on-line signatures. Therefore, in order to successfully convert raw signature data into a suitable representation in *feature space*, fundamentally different approaches to feature extraction are required. Furthermore, since we employ a *writer-independent* framework for signature modelling and verification, the initial feature space representation is subjected to a dichotomy transformation that yields the final signature representation in *dissimilarity space*.

Off-line signatures. The signature representation process utilised by a typical off-line system developed in this study is conceptualised in Figure 1.14.

Since the source data associated with an off-line sample is represented by a digitised image of the original pen-on-paper signature, several image processing techniques are first required in order to ensure efficient feature extraction. These techniques are concerned with *image binarisation*, *noise reduction* and *signature segmentation*. The issues of image binarisation and noise reduction are addressed using two tried and tested techniques, namely *Otsu's method* and the *median filter* respectively. In order to improve upon the traditional method of signature segmentation, namely signature extraction by means of

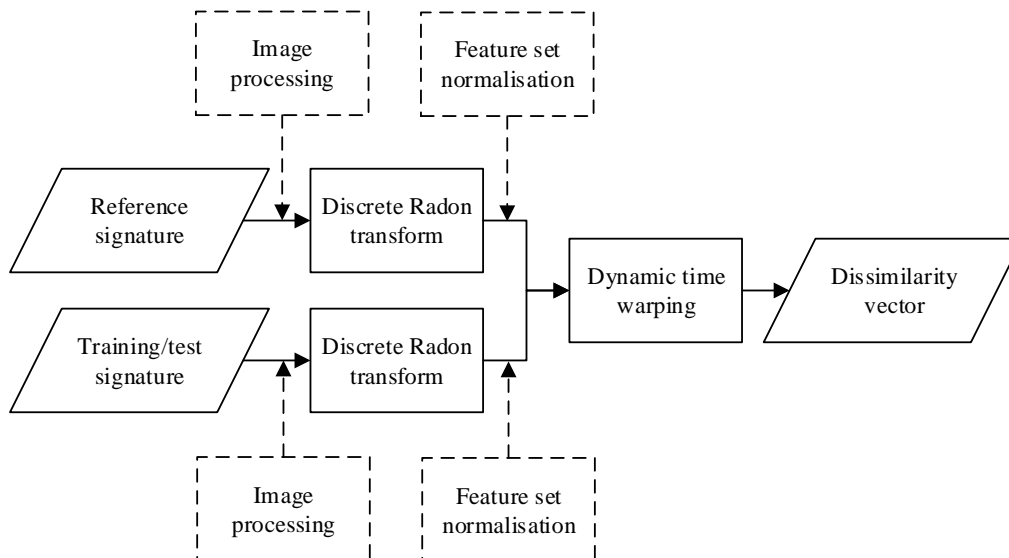


Figure 1.14: Schematic representation of the *off-line* signature representation process utilised in this study.

its bounding box, we propose that the signature images are extracted from the document background by removing all zero-valued image rows and/or columns. In Section 3.2.1 we show that this method is less sensitive to the presence of any residual noise, when compared to the bounding box method. The proposed method also ensures that any subsequently extracted features are invariant with respect to translation.

Once a suitable binary signature image has been obtained, the DRT is used to extract a set of projection profiles – each associated with one of T predetermined angles. These projection profiles are subsequently normalised in such a manner that each profile has the same predefined dimension d , whilst the feature set as a whole has a unit variance, thereby ensuring a feature set that is also invariant with respect to scale.

The off-line feature extraction technique utilised in this study therefore converts any raw signature image into a feature set consisting of T , d -dimensional feature vectors.

On-line signatures. The signature representation process utilised by a typical on-line verification system developed in this study is conceptualised in Figure 1.15.

Both of the on-line data sets considered in this study are captured using devices capable of recording the pen stroke coordinates, pen pressure, as well as pen angle relative to the writing surface. As is the case in the off-line scenario, the source data is first pre-processed in order to maximise the efficacy of the subsequent feature extraction process. Firstly, the pen stroke coordinates are normalised, in order to obtain spatial descriptors that are invariant with respect to scale and translation. Also, the set of pen pressure values are normalised with respect to the point of maximum pressure.

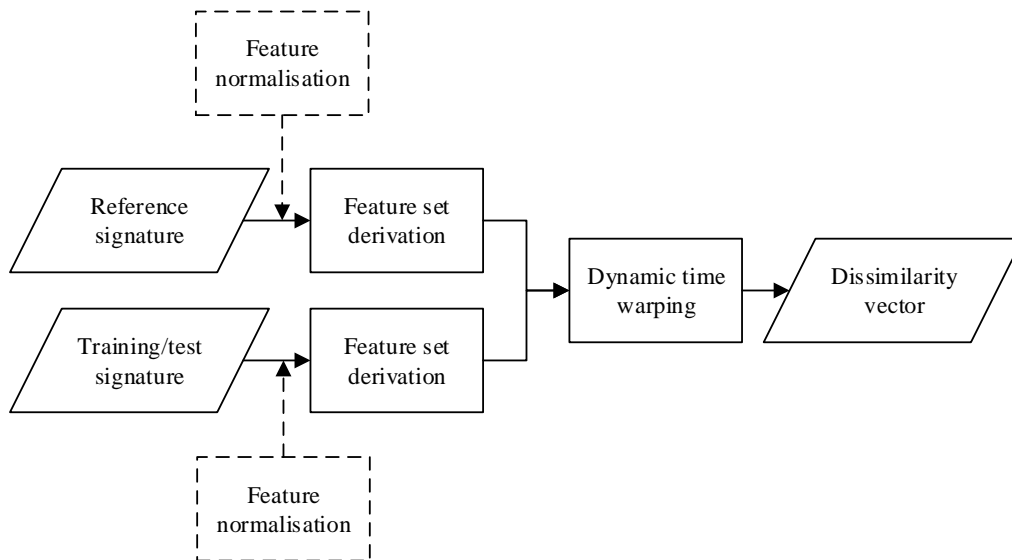


Figure 1.15: Schematic representation of the *on-line* signature representation process utilised in this study.

This initial set of normalised function features is used to derive several additional function features. Specifically, the first and second derivatives of each normalised function feature is calculated, in order to maximally exploit the temporal information associated with signature production. Finally, the signature duration, as described by the number of measurements captured for any given function feature, is also considered as a global descriptor.

The on-line feature extraction technique utilised in this study therefore converts any raw signature data into a feature set consisting of eighteen feature vectors with arbitrary dimension, as well as a single scalar feature.

Dichotomy transformation. The off-line and on-line feature extraction processes described above may be used to obtain an initial signature representation in *feature space*, which would be suitable for consideration within a *writer-dependent* signature modelling framework.

In order to convert this initial feature-based representation into a *writer-independent* signature representation in *dissimilarity space*, any feature set extracted from a training or questioned signature sample is compared to that extracted from a writer-specific genuine reference sample. In this manner, the systems developed in this study convert a specific feature set, composed of T feature vectors, into a single T -dimensional *dissimilarity vector*. Each element of this dissimilarity vector, in turn, represents the distance between a pair of

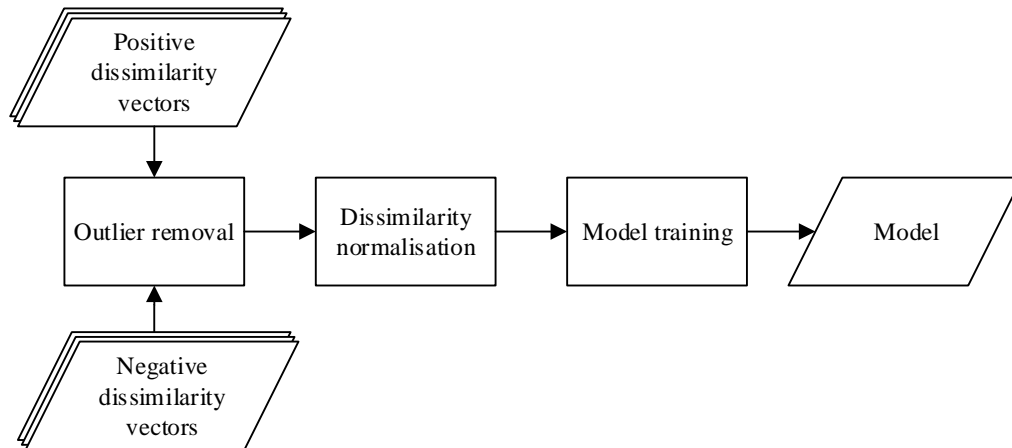


Figure 1.16: Schematic representation of the signature modelling process utilised in this study.

corresponding³ feature vectors. A DTW-algorithm is used to perform the aforementioned distance calculation.

Signature modelling

The signature modelling process utilised by a typical system developed in this study is conceptualised in Figure 1.16.

A collection of both positive and negative signature samples, that is required for the training of a writer-independent signature model, is obtained from a number of so-called *guinea-pig writers* in a *controlled* environment. These writers are considered representative of the general public and their signatures are used *only* for training purposes. Each writer submits a set of genuine reference samples, as well as a set of genuine training samples. A set of amateur skilled forgeries is also obtained for each writer.

For every guinea-pig writer considered, each positive and negative sample is compared to each reference sample, in order to produce a set of positive and negative dissimilarity vectors for training purposes. Every reference sample is also compared to every *other* reference sample, in order to obtain a set of positive dissimilarity vectors, which is subsequently used to determine dissimilarity statistics for said individual. These statistics, that include the feature-specific mean and standard deviation metrics, aim to quantify the acceptable level of variability expected for different genuine samples belonging to a specific writer and are retained for future use.

The entire collection of positive (or negative) dissimilarity vectors, obtained from *all*

³In the scenario where two *off-line* feature sets are presented for comparison, the corresponding feature vectors represent two projection profiles, that is one from the training or questioned sample and one from the reference sample, that were calculated from the same projection *angle*. In the *on-line* scenario, these feature vectors each contains the entire set of measurements recorded for the same *descriptor*.

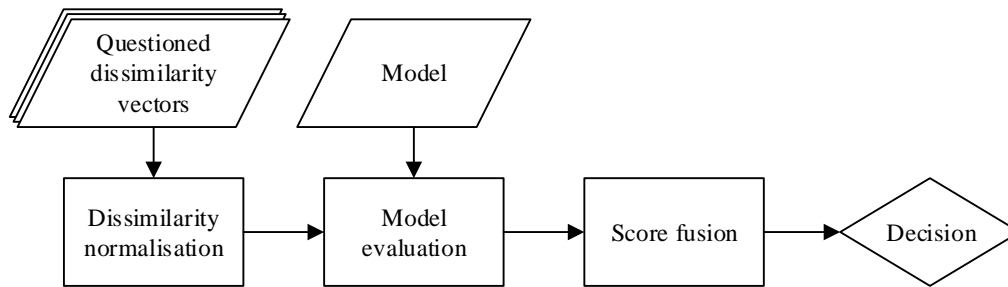


Figure 1.17: Schematic representation of the signature verification process utilised in this study.

the guinea-pig writers, is pooled into a positive (or negative) *dissimilarity set*. These dissimilarity sets are subsequently subjected to an *iterative outlier-pair removal* algorithm, in order to discard any training samples that are deemed misrepresentative of the positive and negative classes, whilst also ensuring that the training data remain balanced. Furthermore, in order to optimise the representation potential of the remaining samples in the training set, each dissimilarity vector undergoes appropriate normalisation. The systems developed in this study perform said normalisation on a *writer-specific* level, where the dissimilarity statistics computed earlier are used to rescale a logistic function, thereby yielding a separate normalisation function for each individual writer. The set of positive and negative dissimilarity vectors, belonging to a *specific* writer only, are subsequently normalised using his/her tailored normalisation function.

Finally, the normalised dissimilarity sets are used to train a universal classifier, that is either a quadratic discriminant or an SVM, which is retained for verification purposes.

Signature verification

The signature verification process utilised by a typical system developed in this study is conceptualised in Figure 1.17.

In order to qualify for consideration by the now functional verification system, any newly enrolled writer is required to submit a set of K genuine reference samples. When presented with an off-line or on-line signature sample (of unknown origin), as well as an associated claim of ownership, the system first compares the questioned sample to the entire set of reference samples belonging to the claimed owner. This process yields a set of K dissimilarity vectors. Each dissimilarity vector is subsequently normalised, using only the writer-specific dissimilarity statistics associated with the claimed owner.

Each of the normalised dissimilarity vectors is then individually presented to the trained model, in order to obtain a set of distance measures relative to the corresponding decision boundary in dissimilarity space. Each distance measure is converted, by means of a conventional logistic function, into a partial confidence score. A final confidence score, that represents the average of the set of K partial confidence scores, is then determined.

Finally, an appropriate threshold is applied to the final confidence score, in order to predict class membership.

1.4.2 Experimental results

The systems developed in this study are subjected to a rigorous experimental evaluation protocol, in order to ascertain their proficiency in the detection of *amateur skilled forgeries*. Several large signature databases are considered, whilst the experimental protocol ensures comprehensive and unbiased system performance estimates by employing both k -fold cross-validation and n -fold data randomisation.

The *off-line* systems are evaluated using *Dolfing's data set* (4530 samples from 51 writers) and the *MCYT-SignatureOff-75 subcorpus* (2250 samples from 75 writers). When the number of reference signatures available for each writer in Dolfing's data set is varied between five and fifteen, the optimal QDA-based system achieves an average AUC and EER of 99.39% and 3.43% respectively. When five and ten reference signatures are available for each writer in the MCYT-SignatureOff-75 subcorpus, these metrics deteriorate to 96.03% and 10.39% respectively. The optimal SVM-based system proves superior in terms of performance, yielding an average AUC (EER) of 99.43% (3.31%) on Dolfing's data set and 96.29% (9.93%) on the MCYT-SignatureOff-75 subcorpus. It should be noted that the discrepancy in terms of the system performance estimates obtained from these two data sets is largely explained by the fact that, unlike the MCYT-SignatureOff-75 subcorpus, Dolfing's data set contains *ideal* off-line signature samples i.e. samples that possess uniform stroke width and are completely free of background noise.

The *on-line* systems are evaluated using the *Philips database* (4530 samples from 51 writers) and the *MCYT-Signature-100 subcorpus* (5000 samples from 100 writers). When 5–15 reference signatures are available per writer, the QDA-based system achieves an average AUC (EER) of 99.66% (1.96%) on the Philips database and 99.38% (3.32%) on the MCYT-Signature-100 subcorpus. The SVM-based system again proves more proficient, yielding metrics of 99.87% (0.89%) on the Philips database and 99.49% (2.82%) on the MCYT-Signature-100 subcorpus.

In Chapter 6 we show that the results reported in this study compare favourably with previous results reported in the literature.

1.5 Contribution of this study

In this study we propose and develop several techniques that are novel within the context of automatic handwritten signature verification. These techniques include:

- **A DTW-based dichotomy transformation for writer-independent off-line signature representation.**

By matching the descriptors contained in DRT-based feature vectors according to similarity, rather than location, the proposed approach is able to minimise the adverse effects of intra-class variability and therefore aids in the construction of a su-

perior signature model. When compared to the traditional Euclidean distance-based dichotomy transformation, the DTW-based technique proposed in this study is shown to result in a statistically significant improvement in terms of system proficiency.

- **A writer-specific dissimilarity normalisation strategy for writer-independent handwritten signature modelling.**

By incorporating writer-specific information into the writer-independent signature modelling framework, the proposed approach is able to improve inter-class separability and therefore aids in the construction of a superior signature model. When compared to the traditional global normalisation strategy, the writer-specific strategy proposed in this study is shown to result in a statistically significant improvement in terms of system proficiency – when either the off-line or on-line scenarios are considered.

- **A writer-independent off-line signature verification system that is both novel and proficient.**

When compared to existing systems that were also evaluated on Dolfing’s data set and the MCYT-SignatureOff-75 subcorpus, the systems developed in this study are shown to outperform several of the most proficient systems documented in the literature.

- **A writer-independent on-line signature verification system that is both novel and proficient.**

When compared to existing systems that were also evaluated on the Philips database and the MCYT-Signature-100 subcorpus, the systems developed in this study are shown to outperform several of the most proficient systems documented in the literature.

Detailed discussions on the above-mentioned contributive components of the systems developed in this study are presented in Chapters 3–5, whilst a rigorous experimental validation of their impact on system proficiency is presented in Chapter 6.

In conclusion, since each of the novel techniques proposed in this study constitutes an improvement to the current state of the art, we assert that the work presented herein constitutes a significant contribution to the field of automatic handwritten signature verification.

1.6 Outline

The remainder of this dissertation is organised as follows:

Chapter 2: Literature Study presents concise discussions on selected previous works pertaining to handwritten signature verification, thereby providing the reader with a contextual perspective regarding the wide range of available techniques and associated levels of success achieved.

Chapter 3: Signature Representation discusses how the systems developed in this study convert raw signature data, associated with either off-line or on-line samples, into dissimilarity vectors that are suitable for consideration within a writer-independent signature modelling framework. This chapter also introduces the novel DTW-based dichotomy transformation for the purpose of off-line signature representation.

Chapter 4: Classifiers introduces and concisely describes the generative and discriminative classification techniques considered in this study.

Chapter 5: Signature Modelling and Verification presents additional data processing techniques performed on the initial dissimilarity sets and explains how the resulting sets are utilised, in conjunction with a suitable classification technique, to construct a robust writer-independent signature model. The verification protocol utilised by a trained model is also discussed in detail. This chapter also introduces the novel writer-specific dissimilarity normalisation strategy.

Chapter 6: Experiments describes the data sets and experimental protocol considered to evaluate the proficiency of the off-line and on-line systems developed in this study. A detailed system performance analysis is provided, as well as a performance comparison with previous systems proposed in the literature. We also empirically verify the contribution of the novel concepts proposed in this study.

Chapter 7: Conclusion and Future Work provides concluding remarks regarding the work presented in this study, whilst introducing several additional topics, deemed potentially beneficial to the systems developed herein, as part of a potential continuation of this study.

Chapter 2

Literature Study

“If I have seen further it is by standing on the shoulders of giants.”
- Isaac Newton (1642–1727)

2.1 Introduction

The field of automatic signature verification has enjoyed sustained interest over the past few decades. Over the years, numerous systems have been proposed in the literature. These systems utilise a wide variety of fundamentally different image processing and pattern recognition techniques for signature representation and modelling. In order to gain a historical perspective regarding advances in the field, the reader is referred to such surveys as Plamondon and Lorette (1989), Sabourin *et al.* (1992), Leclerc and Plamondon (1994), Hou *et al.* (2004), and Impedovo and Pirlo (2008). For an updated state of the art, as well as discussions on recent advances and open issues, the reader is referred to such surveys as Pal *et al.* (2011), Impedovo *et al.* (2012), and El-Henawy *et al.* (2013).

In this chapter we present concise discussions on selected works pertaining to off-line (Section 2.2) as well as on-line (Section 2.3) signature verification that are deemed relevant¹ to this study. These discussions are presented in chronological order and contain the following information: (i) the associated author(s) and year of publication; (ii) the feature extraction technique(s) and, where applicable, the dichotomy transformation utilised for signature representation; (iii) the classification technique(s) utilised for signature modelling and verification; (iv) the nature and composition of the data set(s) considered for experimental evaluation and (v) the experimental results reported by the author(s). The reader is reminded that, since the systems discussed in *this* chapter were evaluated on different data sets (using different experimental protocols), the results reported here may not necessarily be compared *directly*, but rather serve as a general indication of system proficiency.

¹Studies on handwritten signature verification documented in the literature are considered relevant to this study if they have similar objectives, utilise similar pattern recognition techniques or consider a similar data set for experimental evaluation.

2.2 Off-line signature verification

For our discussion of off-line signature verification systems, a distinction is made between those systems that utilise *writer-dependent* (Section 2.2.1) and *writer-independent* (Section 2.2.2) frameworks for signature modelling and verification. A *hybrid* system, that is a system that utilises a combination of writer-dependent *and* writer-independent approaches, is also discussed in Section 2.2.3.

2.2.1 Writer-dependent systems

During the early years of research into the development of automatic off-line signature verification systems, researchers focussed exclusively on the writer-dependent approach to signature model construction. We now present a concise survey of selected writer-dependent off-line signature verification systems.

Fierrez-Aguilar *et al.* (2004) illustrate the potential benefits of *information fusion* by developing three separate verification systems. The first system, referred to as the *global expert*, considers a set of 32 slant features (extracted by means of morphological operators) for the purpose of signature representation. These features are extracted from the *entire* signature image and are therefore considered to be global signature descriptors. The second system, referred to as the *local expert*, considers the same set of slant features. However, these local slant features are extracted from *sub-images* of the original signature image, that are obtained by means of a grid-based segmentation strategy. The global and local experts perform similarity computation using two fundamentally different techniques. The former utilises an inverse *Mahalanobis distance*, whilst the latter employs a left-to-right *hidden Markov model* (HMM). Both systems are evaluated on the MCYT-SignatureOff-75 subcorpus (henceforth referred to as MCYT-75). This subcorpus is discussed in more detail in Section 6.2.1. When five and ten genuine samples are available for model training, the global expert achieves equal error rates (EERs) of 21.84% and 18.93% respectively, whilst the more proficient local expert achieves EERs of 14.51% and 12.22% respectively. The third system, referred to as the *combined expert*, is obtained through fusion of the global and local experts. The employed fusion strategy involves a linear mapping of the similarity scores yielded by the two individual systems to the range $[0, 1]$, followed by the averaging of the *mapped* scores. When also evaluated on MCYT-75, the combined expert outperforms both the global *and* local experts by achieving EERs of 11.00% (for five training samples) and 9.28% (for ten training samples).

In Coetzer (2005), two fundamentally different verification systems are proposed. Both systems utilise the *discrete Radon transform* (DRT) for the purpose of global feature extraction. Signature modelling and verification are achieved using either a *dynamic time warping* (DTW) algorithm or a continuous HMM. A novel *ring-structured* HMM topology is also proposed. Unlike the traditional left-to-right HMM configuration (as utilised in e.g. Fierrez-Aguilar *et al.* (2004)), the proposed ring-structured HMM is able to facilitate *rotation invariant* signature modelling. Both the DTW-based system and the HMM-based system are evaluated on the Stellenbosch data set, which contains 660 genuine signatures,

132 amateur skilled forgeries and 132 casual forgeries obtained from 22 writers. In addition, the HMM-based system is also evaluated on Dolfig's data set (see Section 6.2.1). When only skilled forgeries and only casual forgeries from the Stellenbosch data set are considered, the DTW-based system achieves EERs of 18% and 4% respectively, whilst the HMM-based system achieves EERs of 17.7% and 4.5% respectively. When only skilled forgeries and only professional forgeries from Dolfig's data set are considered, the HMM-based system achieves EERs of 12.2% and 15% respectively. The results reported here are based on the assumption that fifteen genuine samples are available for model training.

The global and local experts proposed by Alonso-Fernandez *et al.* (2007) are similar in design to those previously proposed by Fierrez-Aguilar *et al.* (2004), in the sense that they also consider slant-based features for the construction of Mahalanobis distance-based and left-to-right HMM-based signature models. However, in addition to the 32 slant features detailed in Fierrez-Aguilar *et al.* (2004), the proposed systems also extract a further 30 slant features from the signature *envelope*. The resulting set of 62 features is utilised during signature modelling and verification. The proposed systems are also evaluated on MCYT-75 under similar experimental conditions to those considered in Fierrez-Aguilar *et al.* (2004). When five and ten genuine samples are available for model training, the global expert achieves EERs of 23.78% and 22.13% respectively. Again, the local expert proves superior and achieves EERs of 17.76% and 14.44% respectively. The fact that both of these systems are outperformed by those previously proposed in Fierrez-Aguilar *et al.* (2004) emphasises the importance of designing an effective feature extraction process. Specifically, the reported results confirm that the utilisation of an increased number of pattern descriptors does not necessarily improve the proficiency of the associated pattern recognition system. Furthermore, the fact that both the global and local experts employ a *generative* classification technique (see Section 4.2) for the purpose of signature modelling would suggest that their decreased performance, as a direct result of utilising an expanded feature set, may in all likelihood be due to the adverse effects of the *ugly duckling theorem* (see Section 3.5) and/or the *curse of dimensionality* (see Section 4.3.2).

The system proposed by Gilperez *et al.* (2008) examines binary signature images in terms of their apparent *stroke contours* – both internal and external. Signature representation is achieved by determining several histograms from local *contour directions*, *contour hinges* and directional *co-occurrence angles* (both horizontal and vertical) from a set of training samples, whereafter each histogram is normalised to a feature-specific *probability density function* (PDF). The authors note that this contour-based signature representation successfully circumvents any potential adverse effects related to variations in stroke width. During the subsequent verification stage, similarity calculation is performed by means of the *chi-squared distance* measure. Skilled forgeries from MCYT-75 are considered for system evaluation. When five genuine samples are available for model training (i.e. PDF estimation), the set of four feature-specific verification systems achieve EERs in the range 10.18%–12.71%. When the number of training samples is increased to ten, the system performance improves and EERs of 6.44%–10.00% are reported. Additional experiments reveal that no improvement in performance is witnessed when the feature-specific verification systems are combined into a single multi-feature verification system. The authors

suggest that this is most likely due to an existing correlation between the specific features utilised.

The systems proposed by Wen *et al.* (2009) are primarily aimed towards the construction of a rotation invariant signature representation in feature space. A binarised signature image is first segmented by means of a circular frame, so that the frame fully encloses the signature, whilst its center of gravity coincides with that of the signature. This is followed by the extraction of two ring-peripheral features, specifically the *ring-external-feature* (that is the background area between the segmentation frame and its closest pen strokes) and the *ring-internal-feature* (that is the background area between the center of gravity and its closest pen strokes). These ring-peripheral features are computed for several equally spaced segments within the circular frame, whereafter the entire set of computed feature values is combined to form a single feature vector. The authors note that, in order to ensure rotation invariance during signature modelling, a sufficiently large number of segments is required (from their experiments it appears that the use of 72 segments yields optimal system performance). Signature modelling and verification is achieved by using either an inverse Mahalanobis distance model or a ring-structured HMM. Both systems are evaluated on MCYT-75 using five genuine signature samples for writer-specific model construction. The optimal distance model and HMM-based systems achieve comparable EERs of 15.3% and 15.02% respectively.

Swanepoel and Coetzer (2010) propose an ensemble-based system that consists of eight so-called base classifiers. Each individual base classifier utilises one of four graphometric features, that is the *pixel density*, the *gravity center distance*, the *baseline orientation*, and the *predominant slant*. These features are extracted locally using a *flexible grid segmentation* scheme, that is a grid-based segmentation technique that allows adjacent grid cells to overlap by a predetermined factor. Each base classifier utilises either a DTW-algorithm or a left-to-right discrete HMM for signature model construction. When skilled forgeries from Dolfing's data set are considered for system evaluation, the individual base classifiers achieve EERs in the range 13.60%–18.55%. Furthermore, when the entire ensemble is combined using either score averaging or majority voting, the achieved EERs improve to 11.21% and 10.23% respectively. In each case, fifteen genuine samples are reserved for training purposes. The authors also show that the flexible grid-based signature segmentation strategy consistently and significantly outperforms the traditional rigid grid-based approach.

A novel extension of the HMM-based system previously proposed by Coetzer (2005) is investigated in Panton and Coetzer (2010). The proposed system also employs the DRT for feature extraction. However, where the systems proposed in Coetzer (2005) utilise the DRT to extract *global* shape descriptors, Panton and Coetzer (2010) utilise this algorithm to *locally* extract projection profiles from a set of overlapping circular retinæ. The aforementioned signature representation is subsequently used to construct an *ensemble* of ring-structured HMMs – one associated with each retina. Finally, the verification of a questioned signature sample is performed through a majority vote of the region-specific HMMs in the trained ensemble. When skilled forgeries from Dolfing's data set are considered for system evaluation, the proposed system achieves an EER of 8.6% when fifteen

genuine samples are available for model training. The reported results indicate a significant improvement in proficiency when compared to those reported for the previously proposed global DRT-based system. Furthermore, this study makes an important observation regarding the impact of the experimental protocol on the reliability of system performance evaluation. In their initial experimental setup, which incorporates *k*-fold *cross-validation* – a technique commonly utilised in the literature for the purpose of avoiding an over-fitted system performance estimation – it is found that the achieved results remain sensitive to the *ordering* of the writers considered for training and evaluation. In order to address this issue, the ordering of the writers is *randomised* prior to cross-validation, thereby nullifying the influence of writers associated with atypically proficient or poor system performance. The authors suggest that this process of so-called *data shuffling* should be repeated several times in order to obtain reliable system performance estimates.

In contrast to the predominantly employed strategy for off-line signature representation, that is the extraction of global and/or local descriptors from a *binarised* signature image, Vargas *et al.* (2011) propose a feature extraction technique that is rooted in statistical texture analysis of *grey-level* images. The proposed method involves the computation of the *grey-level co-occurrence matrix* together with histograms of several *local binary patterns* (LBPs) as *pseudo-dynamic* global signature descriptors. In order to maximise the efficacy of the aforementioned feature extraction technique, any signature image presented for analysis first undergoes several pre-processing stages, including *histogram displacement* and *background removal*². Those feature sets extracted from positive and negative³ samples associated with a specific writer are used to construct a *support vector machine* (SVM) classifier with a *radial basis function* (RBF) kernel (see Section 4.4). The proposed system is evaluated using skilled forgeries from both the MCYT-75 and GPDS-100 subcorpora. When five and ten genuine samples are available for model training, the system achieves EERs of 12.02% and 8.80% respectively on the former data set. When evaluated on the latter data set, EERs of 12.06% (for five training samples) and 9.02% (for ten training samples) are reported.

A further study into the potential benefits associated with the consideration of grey-level information during signature image analysis is presented in Ferrer *et al.* (2012), with special focus on the investigation of the robustness of grey-level features when confronted with a complex document background. To this end, MCYT-75 and the well-known

²It is worth mentioning that a clear distinction should be made between the processes of image binarisation and image background removal. Binarisation algorithms are aimed towards the categorisation of image pixels as belonging either to the foreground or background. In the resulting binarised image, foreground pixels are typically represented by ones, whilst background pixels are typically represented by zeros. A background removal algorithm also aims to detect the image background and subsequently represents the identified background pixels with zero-values. However, following successful background removal, the remaining pixels with non-zero values not only indicate the location of foreground pixels (that are indicative of the apparent pen strokes), but also their associated grey-level intensities (that are indicative of the associated pen pressure).

³The reader is reminded that, within the context of *writer-dependent* signature modelling by means of a *discriminative* classification technique, the negative samples associated with a specific writer are represented by *positive* samples belonging to *other* writers (i.e. random forgeries).

GPDS960GraySignature corpus are considered for the purpose of system evaluation, as well as several *artificially generated* signature databases. The samples contained in the artificial databases are obtained by *blending* the original signature images from the aforementioned signature corpora with real-world document images that possess varying levels of distortion. Following a background removal process, histograms of the LBPs, *local derivative patterns* and *local directional patterns* are extracted and subsequently used to construct either a *nearest neighbour* classifier or an SVM (several SVM kernels are investigated). Numerous experimental results are reported in order to investigate not only the discriminative potential of each feature-classifier pair, but also the influence of different document distortion levels on system proficiency. Ultimately, the SVM with a *positive definite chi-square kernel* is found to be the most robust classifier, whilst the local derivative pattern is found to be the most discriminating feature. Under optimal operating conditions (i.e. when no distortion of the document background occurs), the aforementioned feature-classifier pair achieves EERs of 10.97% and 16.85% on MCYT-75 and the GPDS960GraySignature corpus, respectively, when five training samples are available for model construction. Furthermore, the authors show that the expected level of success achievable by each system is inversely proportional to the level of distortion present in the document image background.

2.2.2 Writer-independent systems

Although the development of writer-dependent signature verification systems remains a very active field of research, the use of writer-independent frameworks for signature modelling has gained increased popularity in recent years. We now present a concise survey of selected writer-independent off-line signature verification systems.

The system proposed by Santos *et al.* (2004) employs the so-called *questioned document expert's approach* to signature verification. In the expert's approach, any questioned signature is first individually compared to each one of the genuine samples contained within the appropriate writer-specific reference set, where each such comparison yields a *partial decision*. The *final decision* regarding the authenticity of the questioned sample is subsequently obtained by means of *decision fusion*, usually a majority vote of the set of partial decisions. Following the segmentation of the signature image, by means of a rigid grid composed of 8×20 square grid cells, the proposed system extracts four different graphometric features from each grid cell. Similar descriptors extracted from different grid cells are subsequently concatenated, which produces a set of four 160-dimensional feature vectors. An Euclidean distance-based dichotomy transformation is subsequently performed, in order to obtain a 4-dimensional dissimilarity vector for the purpose of writer-independent signature representation in dissimilarity space. Signature modelling is achieved by means of a *multilayer perceptron* (MLP). The system is evaluated on an unspecified signature corpus that contains 9600 samples obtained from 240 writers. When only skilled forgeries are considered, this evaluation results in a false rejection rate (FRR) of 10.33% and a false acceptance rate (FAR) of 15.67%. Furthermore, the authors conclude that the utilisation of a writer-independent approach to signature model construction is able to successfully reduce the number of genuine signatures required per writer for the purposes of train-

ing and validation, whilst their proposed writer-independent verification system is able to efficiently absorb new writers without the need to generate additional personal models.

In Batista *et al.* (2010), a two-stage verification system is proposed that employs a novel approach to dissimilarity calculation. In the first stage, a grid-based segmentation scheme is used to extract local pixel density information from a grey-level signature image. However, where a typical writer-independent system would compare the feature sets obtained from writer-specific training and reference samples by means of a suitable distance measure, the proposed system uses the feature sets extracted from positive and negative training samples to construct several writer-dependent discrete left-to-right HMMs – each with a different number of states and/or codebook sizes, and each associated with either the genuine class or the imposter class. The combination of the likelihood scores emitted by these HMMs serves as a writer-independent signature representation in dissimilarity space, as opposed to (for example) the dissimilarity measures yielded by an Euclidean distance method. In the second stage, the aforementioned HMM-based dissimilarity representation is used to train two fundamentally different writer-independent classifiers, that is either a *Gentle AdaBoost* classifier or an SVM ensemble of which the decisions are combined through majority voting. Random, casual and skilled forgeries from the Brazilian database, that contains 7920 samples obtained from 168 writers (108 writers provided 40 samples each, whilst 60 writers provided 60 samples each), are considered for system evaluation. When twenty genuine signatures and twenty random forgeries are available per writer for training purposes, the AdaBoost-based system and the SVM-based system achieve average error rates (AERs) of 6.54% and 6.21% respectively.

In Kumar *et al.* (2011), a novel feature extraction technique is proposed that is based on the so-called *surroundedness* property of each pixel in a binary signature image and is said to capture both shape and texture information. The absolute difference between corresponding feature vector elements is used as a dichotomy transformation, whereafter an appropriate feature selection technique (several candidate techniques are investigated) is applied. Two discriminative classifiers, an MLP and an SVM, are subsequently constructed for the purposes of signature modelling and verification. When evaluated on the well-known CEDAR and GPDS300 signature corpora, the authors report that their MLP-based verification system proves superior in terms of classification accuracy – achieving 91.67% and 86.24% success on CEDAR and GPDS300 respectively.

2.2.3 Hybrid systems

When designing a signature verification system, one generally decides beforehand to utilise *either* a writer-dependent *or* a writer-independent approach to signature model construction. However, in a recent paper by Eskander *et al.* (2012) it is suggested that the simultaneous utilisation of *both* of these fundamentally different signature representation platforms may in fact prove complimentary, resulting in a so-called *hybrid approach*.

The hybrid system proposed in Eskander *et al.* (2012) is constructed in two separate stages. In the first stage, an initial universal signature representation is obtained in *dissimilarity space*. This is achieved through the subtraction of writer-specific feature vectors,

each composed of several features based on *extended-shadow-codes* and directional PDFs. This initial representation is subsequently used for writer-independent feature selection by means of a *boosting feature selection* algorithm. In the second stage, the same boosting process is applied to construct a unique model for each writer in *feature space*, but using only the features that proved maximally discriminant in the preceding feature selection stage. The authors report that this hybrid system achieves verification performance which is comparable to that of existing writer-independent systems, whilst computational complexity is significantly improved. When five genuine training samples are available per writer, an AER of 5.38% is achieved on the Brazilian database.

To the best of our knowledge, the signature verification system detailed above represents the only system in the current literature that utilises such a hybrid architecture.

2.3 On-line signature verification

For our discussion of on-line signature verification systems, once again a distinction is made between those systems that utilise writer-dependent (Section 2.3.1) and writer-independent (Section 2.3.2) frameworks for signature modelling and verification.

2.3.1 Writer-dependent systems

An exhaustive search of the literature reveals that practically *all* previously proposed on-line signature verification systems utilise a writer-dependent approach to signature model construction. We now present a concise survey of selected writer-dependent on-line signature verification systems.

The system proposed by Dolfig *et al.* (1998) constructs a unique left-to-right continuous HMM for each enrolled writer. Each HMM is constructed using fifteen genuine training signatures, where each training sample is represented by a combination of 32 different spatial, temporal and contextual features. These features are extracted at the stroke level, which requires a pre-processing stage that segments a signature sample according to its velocity profile. The collection of the Philips signature database (see Section 6.2.1) is also facilitated in this study. When the proposed system is evaluated on the *subset* of 24 writers for whom *professional* forgeries are available, the authors report EERs of 2.33%, 2.88% and 2.33% when only home-improved, only over-the-shoulder and only professional forgeries are respectively considered. When skilled forgeries (that includes both home-improved and over-the-shoulder forgeries) from the *entire* set of 51 writers are considered for evaluation, an improved EER of 1.90% is achieved. Furthermore, this study also compares the discriminative potential of spatial features with that of temporal features. This comparison, performed by means of *linear discriminant analysis* (LDA), clearly shows that the discriminative potential of temporal features is substantially superior to that of spatial features. This disparity between the discriminative potential of spatial and temporal features also explains why said system achieves comparable EERs for home-improved and professional forgeries (i.e. instances where the forger only has access to an off-line sam-

ple and can therefore not infer reliable temporal information), whilst over-the-shoulder forgeries (i.e. instances where the forger witnesses an authentic signing event and is therefore able to mimic several temporal properties during signature reproduction) prove more difficult to detect.

Le Riche (2000) also investigates an HMM-based approach to on-line signature modelling. The proposed system first normalises the input signals in order to obtain an initial feature set that is invariant with respect to translation, scale and rotation. Several additional features relating to pen velocity and acceleration are also derived from the initial feature set, which completes the feature extraction stage. Signature modelling is achieved by means of first order left-to-right HMMs with duration modelling. When evaluated on skilled forgeries from the Philips signature database, the proposed system achieves an EER of 1.02% when fifteen training samples are available per writer. The fact that this HMM-based system outperforms the HMM-based system proposed by Dolfing *et al.* (1998), when evaluated on the same data set, supports the author's suspicions that the segmentation of handwriting samples into individual strokes may be redundant (or even detrimental) within the context of on-line signature verification. According to the author, this pre-processing stage was most likely inherited from existing systems that were originally aimed towards *handwriting recognition* (as is the case for the system proposed in Dolfing *et al.* (1998)).

Many signature verification systems documented in the literature utilise HMMs to construct writer-specific signature models and subsequently employ the efforts of the *Viterbi algorithm* to compute a likelihood score from a questioned signature for the purpose of verification. However, the study documented in Van *et al.* (2004) and Van *et al.* (2007) suggests that the signature modelling capabilities of an HMM may be exploited even further by considering not only the likelihood yielded by the Viterbi algorithm, but also the corresponding *Viterbi path* produced during likelihood computation. Their proposed system extracts sixteen gesture related features and nine local shape related features from the original input signal. Following appropriate normalisation, these 25 features are used to construct a continuous left-to-right HMM from five genuine training samples. As mentioned earlier, the verification of a subsequently presented questioned sample involves the computation (by means of the Viterbi algorithm) of the likelihood that the questioned sample was generated by the model belonging to the claimed writer. In addition to the resulting likelihood score, the system also converts the Viterbi path into a so-called *segmentation score*. The final similarity score is calculated as the average of the likelihood score and the segmentation score. System performance evaluation is conducted using skilled forgeries from the BIOMET, Philips, SVC2004 and MCYT-Signature-100 (henceforth referred to as MCYT-100) signature corpora, which yields optimal EERs of 2.33%, 3.25%, 4.83%, and 3.37% respectively.

Lumini and Nanni (2009) propose a novel ensemble-based approach to on-line signature modelling and verification. Several candidate ensembles are presented – each constructed by training a collection of classifiers on different subsets of the larger training set. The set of classifiers within a specific ensemble employ the same classification technique (seven techniques are considered in total). The pool of training data is constructed by merging the results obtained from employing two fundamentally different techniques, that is the

random subspace method (as detailed in e.g. Ho (1998)) applied to a set of 100 so-called *original* features (as described in Fierrez-Aguilar *et al.* (2005) and Nanni and Lumini (2005)), as well as *sequential forward floating selection* (as detailed in e.g. Pudil *et al.* (1994)) applied to a set of 1000 *artificial* features. The artificial features are generated by means of *over-complete feature combination*, that is the iterative application of a random number of randomly selected mathematical operators (seventeen operators are considered in total) on the original feature set. When a questioned signature sample is subsequently presented for verification, said sample is evaluated by each individual classifier belonging to a specific ensemble, whereafter the final similarity score yielded by the entire ensemble is obtained by means of the *max rule*. The entire set of seven classifier-specific ensembles is evaluated separately on MCYT-100. When five genuine training samples per writer are available for model construction, the proposed *single-ensemble* systems achieve EERs in the range 5.4%–9.4%. Furthermore, it is shown that an improvement in verification proficiency is possible when the different ensembles are fused by means of the *sum rule*. Under similar experimental conditions, the optimal *multi-ensemble* system outperforms all of the single-ensemble systems and achieves an EER of 4.4%.

Montalvão *et al.* (2010) investigate the impact of *data scarcity* on the success of a *generative* signature modelling strategy, that is a technique that requires the estimation of (typically multivariate) PDFs for model construction. The proposed system uses a relatively small set of five training samples to obtain a *Gaussian mixture model* (GMM) for each enrolled writer. Signature samples are represented by a set of four features that are indicative of pen position (both horizontal and vertical), pen pressure and time. In order to ascertain the influence of the training set size on the reliability of the resulting covariance estimate, five different levels of model regularisation are also considered, such that data variability is modelled using either full, diagonal, multi-scalar, single-scalar or Parzen covariance estimates. System evaluation is performed on MCYT-100. The authors report that increased regularisation of their GMM-based model (i.e. a reduction in the complexity of the feature correlations to be estimated) consistently results in improved system performance. In fact, the optimal average EER of 4.5% is achieved when employing *single-scalar covariance estimation*, that is when the same scalar covariance is considered for each Gaussian in the mixture. This paper therefore serves as a caution against overly complex covariance estimation when working with limited training data, since this is likely to invoke the *curse of dimensionality* (see Section 4.3.2). As explained in Section 1.2.6, the systems developed in this study attempt to address the issue of data scarcity (and thereby avoid the curse of dimensionality) by adopting a writer-independent approach to signature modelling, and in so doing pool the training samples associated with many different writers into a single (and substantially larger) universal training set.

The system design proposed in Sae-Bae and Memon (2013) is focussed on computational efficiency and information security. Feature extraction is achieved by first decomposing the original time-series signature data into Cartesian vectors, from which several higher-order derivatives are subsequently computed. The vector sequences obtained in the first stage are then converted into a polar coordinate representation, whereafter several histogram-based features are derived from the aforementioned polar representation. The authors note that

this histogram-based feature set is irreversible and therefore ensures the protection of a writer's privacy in the event that his/her stored template is somehow uncovered. The final signature representation is obtained by means of a writer-specific uniform quantisation of each feature component. Any questioned signature that is presented for verification is first appropriately quantised, after which the *Manhattan distance* between the questioned and template quantised vectors is used to determine its authenticity. The proposed system is evaluated using skilled forgeries from MCYT-100 and consequently achieves EERs in the range 5.74%–2.72% when 3–20 genuine samples are available per writer for model training. When only random forgeries from the aforementioned corpus are considered, the system performance improves and EERs of 1.43%–0.35% are achieved.

2.3.2 Writer-independent systems

Unlike the paradigm shift witnessed in the field of off-line signature verification, the development of writer-independent *on-line* signature verification systems has received little to no attention. In fact, to the best of our knowledge, only Ibrahim *et al.* (2007) and Muramatsu and Matsumoto (2009) have previously investigated this topic.

The system proposed in Ibrahim *et al.* (2007) first obtains an initial feature set that describes the pen trajectory (which is decomposed into horizontal and vertical components), pen speed (i.e. the magnitude of pen velocity), as well as pen pressure. Several pre-processing techniques are performed in order to achieve a feature-based signature representation that is invariant with respect to translation, scale and rotation. The pen trajectory information is then used to compute the *absolute angle* associated with each individual trajectory point, that is the angle between the position of said point and the center of gravity. The set of absolute angles is subsequently used to partition the original features into so-called low-angle and high-angle subsets, which yields a total of eight feature vectors. Only the *most stable* pair of features, that is the two features with the smallest standard deviation, is retained for signature representation. The *retained* feature set is transformed into a dissimilarity-based representation by means of the Euclidean distance, whereafter a linear decision boundary is determined in dissimilarity space. An unspecified signature database, that contains 21250 samples (15000 genuine signatures and 6250 skilled forgeries) from 25 writers, is considered for system evaluation, which consequently yields an EER of 1.3%. The number of training signatures per writer is also not specified.

In Muramatsu and Matsumoto (2009), a writer-independent system is proposed that also takes *user individuality* into account, by employing a so-called *user-specific global-parameter fusion* model. Following the translation and scale normalisation of the captured pen coordinates, the magnitude and directional components of the pen velocity are also calculated. A DTW-algorithm is used to convert each feature set into a writer-independent dissimilarity vector. In addition, a *user-dependent mean vector* is calculated for each writer, that is a vector which depicts the average dissimilarity between the reference signatures belonging to a *specific* writer. The final signature representation is obtained through *concatenation* of the dissimilarity vector and the appropriate mean vector. The authors assert that this concatenation process is able to *personalise* each dissimilarity vector with

respect to its owner (or claimed owner) prior to signature modelling (or verification). Model construction is achieved through the application of an AdaBoost algorithm to a set of initial simple perceptrons. Signature samples obtained from the BIOMET multimodal database are used for training purposes, whilst several different signature corpora are considered for system performance estimation. When evaluated separately on the MCYT-100, SVC2004 and MyIdea signature corpora, the proposed system achieves EERs of 4.0%, 8.6%, and 6.1% respectively.

2.4 Concluding remarks

In this chapter we presented condensed discussions on a wide variety of off-line and on-line signature verification systems considered relevant to this study.

Several of the works discussed here are of particular interest to this study, since these works have directly led to, or indirectly served as inspiration for, the development of the novel techniques proposed in this study. For instance, the combined use of the DRT and a DTW-algorithm as a feature-classifier pair (as proposed in Coetzer (2005)) inspired the development of the novel writer-independent off-line signature representation strategy proposed in this study (see Sections 3.2.2 and 3.4). Also, the dissimilarity normalisation function proposed in this study (see Section 5.2.2), that incorporates dissimilarity statistics inferred from the reference set belonging to a specific writer, is an adaptation of the normalisation function proposed in Swanepoel and Coetzer (2010). Furthermore, the verification protocol utilised by the systems developed in this study (see Section 5.3) is based on the questioned document expert's approach outlined in Santos *et al.* (2004). Finally, the scientifically rigorous experimental protocol considered for the evaluation of the systems developed in this study (see Section 6.2.2) is similar to the protocol proposed in Panton and Coetzer (2010).

At present, there is no single standard, internationally accepted handwritten signature corpus or evaluation protocol for the purpose of performance benchmarking. For the most part we have consequently limited our literature survey to the discussion of systems that were evaluated on a data set that is also considered in this study. Although many other groundbreaking signature verification systems have been proposed in the literature (see e.g. Impedovo and Pirlo (2008)), only those systems evaluated on similar data sets provide a realistic platform for comparison. This comparison is presented in Chapter 6. However, the reader is reminded that even subtle differences in the experimental protocol utilised for system evaluation prevents one from being able to *directly* compare reported results – even when the same data set is considered. Nevertheless, such a comparison aids in placing the system performance estimates reported in this study into perspective.

In the next chapter we initiate our discussions on the design and implementation of the handwritten signature verification systems developed in this study.

Chapter 3

Signature Representation

“Measure what is measurable, and make measurable what is not so.”
- Galileo Galilei (1564–1642)

3.1 Introduction

The process of signature representation constitutes one of the three fundamental stages in the successful development of a signature verification system, as discussed in Section 1.2.5. In fact, the ability (or lack thereof) of *any* pattern recognition system to successfully quantify the difference between multiple pattern classes is entirely dependent on the level of separability of these classes in feature space. It is therefore of critical importance to obtain a feature-based pattern representation that maximally discriminates between the classes in question. Furthermore, in order to utilise a *writer-independent* signature modelling strategy, it is required that the initial feature-based signature representation be converted into a dissimilarity-based representation by means of an appropriate dichotomy transformation.

In this chapter we discuss how the systems developed in this study obtain a writer-independent dissimilarity-based representation from raw signature data. Two fundamentally different signature acquisition methods are considered, namely those associated with *off-line* signatures (Section 3.2) and *on-line* signatures (Section 3.3). In each case, we provide a brief overview of the acquisition process and the nature of the captured raw data. We also discuss the need to incorporate suitable data pre-processing techniques that aim to maximise the efficacy of the subsequent feature extraction processes. Several algorithms are presented in order to fulfil these pre-processing requirements.

Most notably, in Section 3.4 we propose the utilisation of a DTW-based dichotomy transformation for the purposes of *both* off-line and on-line signature representation. The use of such a dichotomy transformation is novel within the context of *off-line* signature representation. We discuss how this approach is able to produce dissimilarity vectors in a more robust manner than the traditional approaches documented in the literature, thereby completing the writer-independent signature representation process.

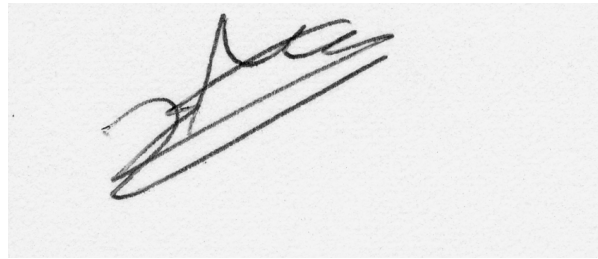


Figure 3.1: Example of a grey-level intensity image that contains an off-line signature sample.

3.2 Off-line signatures

In this section we discuss how the systems developed in this study convert the raw data extracted from an *off-line* signature into a suitable representation in feature space.

3.2.1 Signature acquisition and pre-processing

The acquisition of an off-line signature sample typically involves its extraction from a digitised document image. The task of signature detection and extraction is, however, by no means a trivial endeavour. Several suitable techniques have been documented in the literature, such as the methods proposed by Zhu *et al.* (2007), Mandal *et al.* (2011), and Ahmed *et al.* (2012). In this study we assume the successful completion of such an extraction process. The raw off-line signature data therefore consists of the $m \times n$ grey-level intensity image I , that is an image with mn pixel values $I_i \in [0, 255]$, that contains no machine-printed nor handwritten data other than the signature itself. An example of such an image is provided in Figure 3.1.

Unlike the raw data associated with on-line signatures, as discussed later in this chapter, an off-line signature sample is completely devoid of temporal information. The task of feature extraction therefore entails the derivation of maximally discriminant descriptors, using only the spatial information supplied by the apparent¹ pen stroke coordinates. As a result, the problem of efficient off-line signature representation constitutes a considerably more difficult task than that of its on-line counterpart.

In order to optimally prepare the raw data for feature extraction, it is advisable to include several pre-processing stages. The image processing techniques utilised during these stages are typically concerned with *image binarisation*, *noise reduction* and *signature segmentation*.

Image binarisation

The first stage in pre-processing involves the binarisation of the original grey-level image. This process aims to yield an image representation that distinguishes between only two

¹Since the only data available in a digitised document image consists of unlabelled pixel intensity values, an automated system can never be truly certain about which pixels undoubtedly represent pen strokes.

pixel types, namely those pixels that indicate the foreground (i.e. pen strokes) and those that indicate the background (i.e. the writing surface). The successful completion of such an image binarisation process therefore nullifies the influence of the writing instruments used to produce the signature and consequently allows for efficient signature shape analysis. Furthermore, the binarisation process also aids in the successful completion of the signature segmentation stage, as we shall discuss later in this section.

As mentioned earlier, the original input image I consists of pixels that indicate grey-level intensities in the range $[0, 255]$, where a value of 0 corresponds to a black pixel, whilst a value of 255 denotes a white pixel. The simplest method for obtaining a binary image $I^{(B)}$ from the grey-level image I is the application of a *global threshold* $\rho \in [0, 255]$, such that

$$I_i^{(B)} = \begin{cases} 1 & \text{if } I_i \leq \rho \\ 0 & \text{if } I_i > \rho. \end{cases} \quad (3.1)$$

In the resulting binary image, a value of 1 is associated with a foreground pixel, whilst a value of 0 indicates a background pixel. The success of such a global thresholding strategy is therefore dependent on a sensible value for ρ . The systems developed in this study determine ρ through the well-known *Otsu method*. According to this strategy, the optimal value of the global threshold maximises the inter-class variance between foreground and background pixels, as outlined in Otsu (1979). A typical result yielded by this method is illustrated in Figure 3.2 (a).

Noise reduction

As discussed earlier, the primary objective of the binarisation stage is to produce an image that clearly distinguishes between pixels indicative of pen strokes as opposed to the document background.

It is, however, entirely possible (and in fact an all too regular occurrence) that several of the pixel values obtained during the binarisation stage are misrepresentative. For instance, foreign objects such as dust or ink residue may be erroneously binarised as foreground pixels. In contrast, pen strokes produced with relatively little pressure (and therefore represented by pixels with a relatively low intensity) may be erroneously binarised as background pixels.

These erroneously binarised pixels, or *noise*, may potentially have a negative influence on both the signature segmentation and feature extraction processes, since they constitute inaccurate pen stroke information and are therefore not optimally descriptive of the signature in question. It is therefore of paramount importance that these pixel anomalies be corrected prior to further analysis.

The systems developed in this study utilise the well-known *median filter* for noise reduction. This method uses a 3×3 pixel window to iteratively pass through a binary image, subsequently replacing the value of the center pixel with the median value of its neighbouring pixels. The median filter has been shown to be especially well suited for the reduction of impulse noise, although it may be less effective in dealing with relatively larger

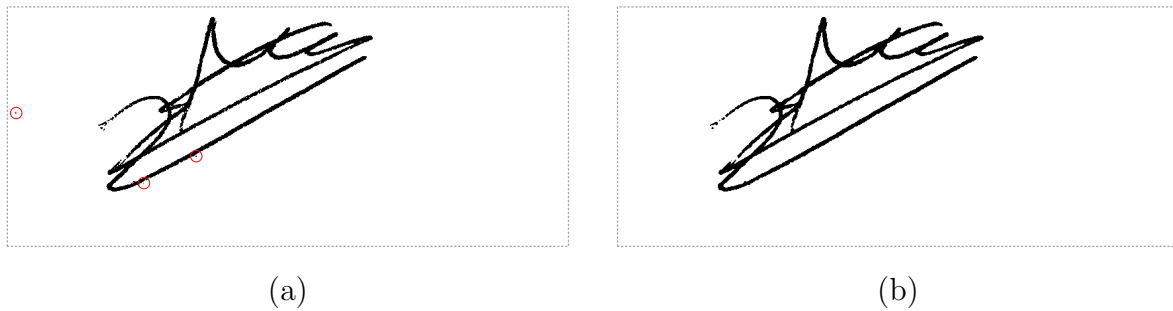


Figure 3.2: Image binarisation and noise reduction. Images obtained after (a) Otsu's binarisation method and (b) the median filter are successively applied to the image depicted in Figure 3.1. The dashed borders indicate the boundary of the original grey-level image. Note that the median filter successfully removes several traces of impulse noise (encircled in red) and also *partially* repairs selected pen stroke segments that were decimated during the preceding binarisation stage.

noise regions. A comprehensive discussion of the median filter is found in Gonzales and Woods (2002), whilst a typical result yielded by this method is illustrated in Figure 3.2 (b).

Signature segmentation

The final stage in pre-processing is concerned with the segmentation of the signature sub-image from the document background. This process discards any remaining background pixels surrounding the signature region, thereby ensuring translation invariance prior to feature extraction².

The simplest method for determining the sub-image that isolates the signature region involves obtaining the *bounding box* of all non-zero pixels. The resulting segmented sub-image consists of all pixels within this bounding box. The implementation of this method is quite straightforward and the obtained results may prove accurate in many instances. However, the efficacy of this method may be severely impeded by the presence of any remaining noise not corrected during the noise reduction stage, especially if this noise is located near the boundary regions of the image, as illustrated in Figure 3.3 (c).

In order to minimise the adverse effects of any remaining noise, the systems developed in this study segment the signature region by discarding all the rows and/or columns that contain *only* zero-values. This method therefore effectively *attaches* any remaining noise to the actual signature sub-image, thereby yielding a significantly more accurate representation of the signature in question, as illustrated in Figure 3.3 (d). It should be clear that this method also inherently ensures translation invariance.

²The motivation for including a signature segmentation stage is entirely dependent on the feature extraction technique utilised. For example, if one were to consider local binary patterns (LBPs) for signature representation (see e.g. Nicolaou *et al.* (2013)), a segmentation stage would prove redundant, since LBP-based features are inherently translation invariant. The projection-based feature extraction technique considered in this study (see Section 3.2.2) does, however, require the efforts of a segmentation stage in order to achieve a translation invariant feature set.

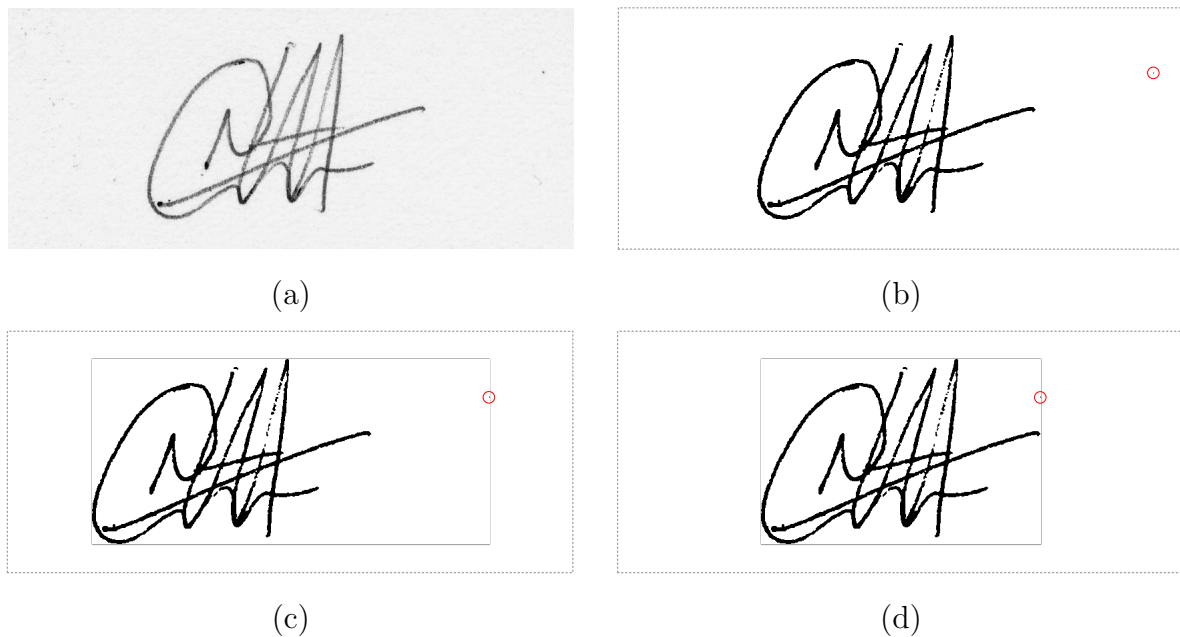


Figure 3.3: Signature segmentation. (a) Grey-level image of a signature sample surrounded by traces of ink residue and (b) its resulting binary representation obtained after noise reduction. Although the pre-processing stages effectively dealt with most of the document degradation, a small portion of noise is still present near the image boundary (encircled in red). (c)-(d) Signature segmentation results, where the solid border indicates the sub-image extracted using (c) the traditional bounding box method and (d) the proposed method of discarding all zero-valued rows and/or columns. This example clearly illustrates the shortcomings of the bounding box method, as well as the ability of the proposed segmentation technique to minimise the adverse affects associated with sub-optimal noise reduction.

The pre-processing stages discussed in this section therefore convert any grey-level document image, that contains a signature sample, into a binary image that represents the signature region only, and contains minimal noise.

3.2.2 Feature extraction and normalisation

Once a suitable binary image representation has been obtained, a practically countless number of suitable local and/or global feature extraction techniques are available for consideration (see e.g. Impedovo and Pirlo (2008)).

In particular, the use of *projection profiles* for off-line signature feature extraction has proven a popular technique and is commonly utilised in the literature, since this method successfully captures global shape information. However, many systems proposed in the literature (e.g. Fang *et al.* (2003); Piyush Shanker and Rajagopalan (2007); Jayadevan *et al.* (2009)) rely solely on the horizontal and vertical projection profiles for signature description.

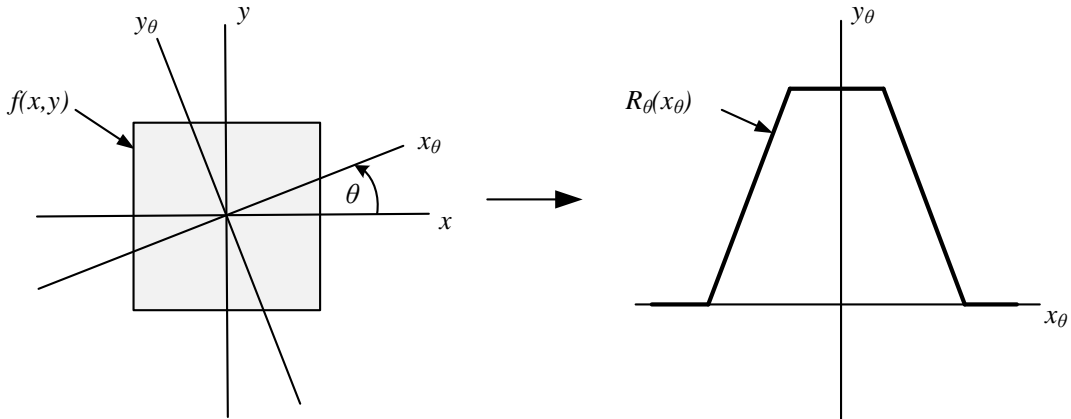


Figure 3.4: Geometric interpretation of the Radon transform. Each projection $R_\theta(f(x, y))$ constitutes the line integral of $f(x, y)$ parallel to the y_θ -axis.

The off-line systems developed in this study extract pen stroke projection profiles by means of the *discrete Radon transform* (DRT), that is a discrete approximation of the transform originally proposed in Radon (1917)³, since it enables the use of an expanded projection angle set and therefore constitutes a natural progression of the projection-based method. The DRT has been shown to be well suited for signature representation in e.g. Coetzer *et al.* (2004).

The Radon transform of a function $f(x, y)$ is defined as the line integral of f parallel to the y_θ -axis, that is the y -axis rotated by an angle θ , such that

$$\mathcal{R}_\theta(x_\theta) = \int_{-\infty}^{\infty} f(x_\theta \cos \theta - y_\theta \sin \theta, x_\theta \sin \theta + y_\theta \cos \theta) dy_\theta, \quad (3.2)$$

where

$$\begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (3.3)$$

as conceptualised in Figure 3.4.

A DRT-based projection profile is calculated using a set of non-overlapping, equidistant projection beams, as conceptualised in Figure 3.5. The DRT of an $m \times n$ image I is calculated for T angles using N_φ projection beams per angle. The cumulative intensity of the pixels that lie within the j^{th} beam, referred to as the j^{th} beam-sum R_j , is calculated as follows,

$$R_j = \sum_{i=1}^{mn} \alpha_{ij} I_i, \text{ for } j = 1, 2, \dots, N_\varphi T, \quad (3.4)$$

³An English translation of the original paper by Radon (1917), published in German, may be found in Radon (1986).

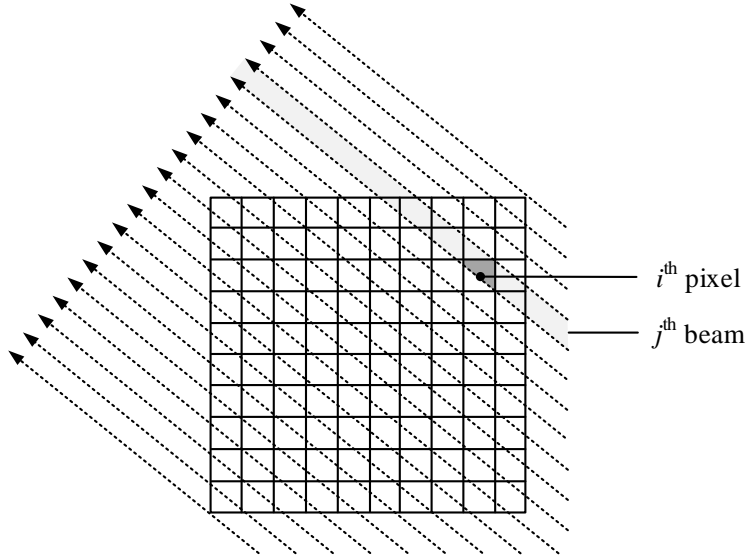


Figure 3.5: Conceptualisation of the method used to calculate the discrete Radon transform for a specific angle θ with $\alpha_{ij} \approx 0.6$. This implies that the j^{th} beam overlaps approximately 60% of the i^{th} pixel, as indicated by the dark grey shaded region.

where I_i denotes the intensity of the i^{th} pixel, whilst α_{ij} denotes the contribution of the i^{th} pixel to the j^{th} beam-sum. Each element of the resulting projection profile associated with a specific angle θ therefore corresponds to a specific beam-sum calculated at an angle θ . The value of α_{ij} is obtained using two-dimensional interpolation. The overall accuracy of the resulting DRT is therefore determined by the number of projection angles considered, the number of projection beams utilised per angle, as well as the accuracy of the interpolation method. More detailed discussions of the DRT-based method can be found in e.g. Beylkin (1987) and Toft (1996).

When presented with a binary signature image, the systems developed in this study utilise the DRT in order to obtain a projection profile \mathbf{R}_θ for each angle in the projection angle set $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_T\}$, that contains T equally distributed angles in the range $[0, \pi]^4$. However, since the number of beams required to obtain a complete projection, as well as the intensity of each resulting beam-sum, is proportional to the dimensions of the input image, each profile is subsequently converted into a *normalised* projection profile $\bar{\mathbf{R}}_\theta$ by employing two separate techniques. Firstly, the projection profile is linearly interpolated to possess a fixed, predetermined dimension d . This process ensures scale invariance

⁴Initially, $T + 1$ equally distributed angles in the range $[0, \pi]$ are considered. The projection associated with an angle of π is then discarded, since it is essentially equivalent to the projection associated with an angle of 0.

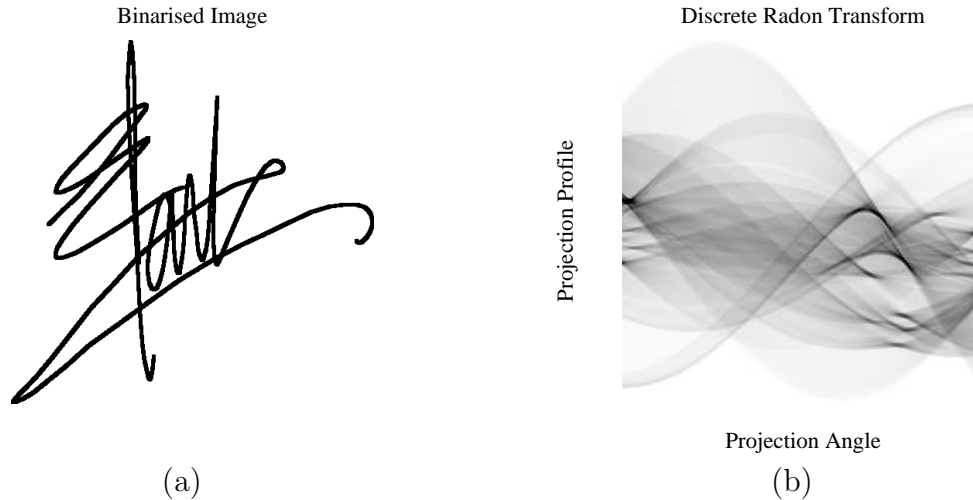


Figure 3.6: Example of (a) a binarised signature image and (b) its resulting DRT-based representation for $T = 180$. Each column of the DRT constitutes a feature vector $\bar{\mathbf{R}}_{\theta}$, that is the *normalised* projection profile associated with a specific projection angle θ .

in the direction *perpendicular* to the projection beams. The entire set of interpolated projections is subsequently normalised to possess unit variance, in order to ensure scale invariance in the direction *parallel* to the projection beams. The aforementioned normalisation techniques therefore produce a scale invariant feature set $\mathbf{X} = \{\bar{\mathbf{R}}_{\theta_1}, \bar{\mathbf{R}}_{\theta_2}, \dots, \bar{\mathbf{R}}_{\theta_T}\}$, as illustrated in Figure 3.6.

The systems developed in this study therefore consider a total of T projection-based features for off-line signature representation in *feature space*, where T may be chosen arbitrarily. In Section 3.4 we discuss how this *feature*-based representation may be converted into a *dissimilarity*-based representation, which is suitable for incorporation into a writer-independent signature modelling framework.

3.3 On-line signatures

In this section we discuss how the systems developed in this study convert the raw data extracted from an on-line signature into a suitable representation in feature space.

3.3.1 Signature acquisition and pre-processing

On-line signature samples are produced using specialised hardware, typically a tablet and electronic pen. As a result, several key descriptors are already recorded during the signature acquisition process. However, since the wide variety of electronic devices currently available on the market vary greatly in terms of price (and therefore also in their level of sophistication), the type of descriptors extracted are entirely dependent on the hardware considered. For instance, the pen stroke coordinates are captured by *all* devices. In addi-

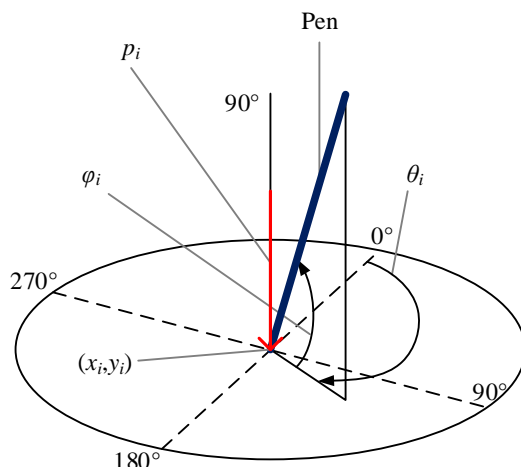


Figure 3.7: Conceptualisation of the measurements recorded during on-line signature acquisition at time i . These signature descriptors may include the pen stroke coordinates (x_i, y_i) , the pen pressure p_i , as well as the pen orientation, as described in terms of the azimuth angle θ_i and the altitude angle φ_i .

tion, many devices are also able to capture the pen pressure, whilst selected devices also measure the pen orientation relative to the writing surface.

The on-line signature databases considered in this study were captured using devices able to capture *all* of the above-mentioned descriptors. These data sets are discussed in detail in Section 6.2.1. Consequently, following the successful recording of a signing event, each point sampled at time i is represented by the 5-tuple $(x_i, y_i, p_i, \theta_i, \varphi_i)$, whilst the entire signature is represented by the feature vectors \mathbf{x} and \mathbf{y} (horizontal and vertical pen positions respectively), \mathbf{p} (axial pen pressure), as well as $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ (azimuth and altitude angles, respectively, relative to the writing surface), as illustrated in Figure 3.7. The feature vector dimension d corresponds to the number of points sampled during acquisition. We henceforth use \mathbf{f} to denote an arbitrary d -dimensional feature vector.

In order to ensure a sensible analysis, several of the aforementioned feature vectors require *normalisation*. For example, although the relative pen trajectories associated with different signature samples belonging to the same writer should remain fairly consistent, it is reasonable to expect that the raw pen strokes captured during acquisition may vary in terms of both position and scale, since the physical dimensions of the utilised device may vary arbitrarily. Furthermore, different applications may require the user to sign within differently demarcated areas. A person generally adapts his/her signature to fit into such a specified area. The consequent variation in terms of translation and scale is illustrated in Figure 3.8 (a).

In order to address this issue, the pen stroke coordinates \mathbf{x} and \mathbf{y} are first shifted in

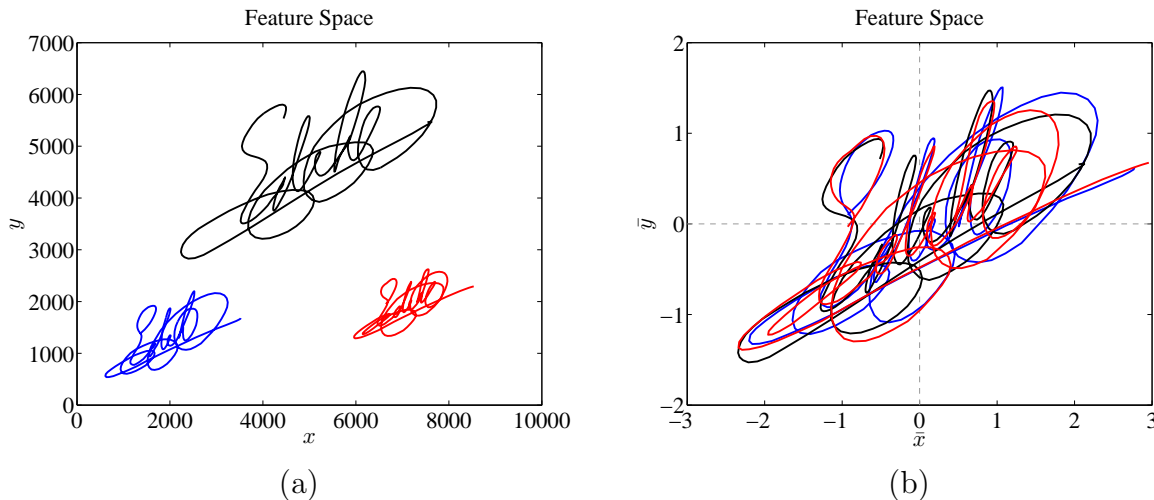


Figure 3.8: Translation and scale normalisation. (a) Superposition of the pen position features \mathbf{x} and \mathbf{y} associated with three signature samples belonging to the same writer. Since their positions and sizes differ significantly, these samples are not yet fit for direct comparison. (b) Superposition of the normalised pen position features $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, that is the features depicted in (a) following successful translation and scale normalisation.

such a way that the center of gravity of the entire signature is relocated to the origin. This process ensures *translation* invariance. Furthermore, each coordinate value is divided by the average magnitude of all the *translated* pen stroke coordinates, as prescribed in Le Riche (2000). This process, which is equivalent to signature rescaling by a factor equal to the standard deviation in the xy -plane, ensures *scale* invariance. The *normalised* pen stroke coordinates, denoted by $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, are therefore determined such that

$$\bar{x}_i = \frac{d(x_i - \mu(\mathbf{x}))}{\sqrt{(x_i - \mu(\mathbf{x}))^2 + (y_i - \mu(\mathbf{y}))^2}}, \quad (3.5)$$

$$\bar{y}_i = \frac{d(y_i - \mu(\mathbf{y}))}{\sqrt{(x_i - \mu(\mathbf{x}))^2 + (y_i - \mu(\mathbf{y}))^2}}, \quad (3.6)$$

where $\mu(\mathbf{f})$ denotes the mean of \mathbf{f} . The effect of this pen stroke normalisation strategy is illustrated in Figure 3.8 (b).

It is also advisable to normalise the pressure signal \mathbf{p} , since it may be dependent on the sensitivity of the device. We therefore consider the *relative* pen pressure $\bar{\mathbf{p}} \in [0, 1]$ such that

$$\bar{p}_i = \frac{p_i}{\max(\mathbf{p})}. \quad (3.7)$$

The azimuth and altitude angles remain unchanged, since they are measured from the point of contact between the pen and writing surface, and are therefore already scale and translation invariant. Following the successful completion of the feature normalisation processes discussed above, an *initial feature set* is obtained, as presented in Table 3.1.

Table 3.1: The initial set of *five* function features in \mathbb{R}^d , that is the signature data captured during the acquisition process after appropriate normalisation.

Feature	Description
\bar{x}	Normalised horizontal pen position
\bar{y}	Normalised vertical pen position
\bar{p}	Relative pen pressure
θ	Azimuth angle of the pen relative to the writing surface
φ	Altitude angle of the pen relative to the writing surface

3.3.2 Feature extraction

Although the initial feature set may suffice for basic model construction, various additional descriptors may be calculated in order to further exploit the discriminative potential of the initial features recorded.

Consider, for example, the spatial features \bar{x} and \bar{y} . Whilst each of these features indicates pen *position* in a specific direction, one may obtain additional information relating to the pen *trajectory* by calculating the *path tangent angle* ϑ as follows,

$$\vartheta_i = \arctan \left(\frac{\bar{y}_i - \bar{y}_{i-1}}{\bar{x}_i - \bar{x}_{i-1}} \right). \quad (3.8)$$

Even more valuable than this additional spatial feature, however, is the temporal information inherently contained in each of the initial feature vectors, since each such feature vector represents a descriptor *sequence*.

In order to quantify the aforementioned temporal information, let $\dot{\mathbf{f}}$ and $\ddot{\mathbf{f}}$ respectively denote the first and second derivatives of \mathbf{f} . Note that $\dot{\bar{x}}$ and $\ddot{\bar{x}}$ for instance represent the velocity and acceleration, respectively, in the horizontal direction. A similar argument follows for the remainder of the initial feature vectors. These temporal features are considered invaluable throughout the literature, since they have proven to be considerably more difficult to mimic than spatial signature characteristics, even when the forger possesses forensic expertise (Dolfing *et al.* (1998); Houmani *et al.* (2012)). It is of course reasonable to expect that the typical forger would be unable to mimic the temporal properties of a signature, given the fact that they are, unlike its spatial properties, near impossible to derive from an off-line⁵ sample. The final set of function features is presented in Table 3.2. Finally, it is important to realise that the length of the function features listed in Table 3.2 may in itself be considered a *global* signature descriptor, since this value is proportional to the length of the signature. Let γ denote the sampling frequency of the device utilised for signature acquisition. The feature vector dimension d may consequently be used to obtain the signature duration \bar{d} as follows,

$$\bar{d} = \frac{d}{\gamma}. \quad (3.9)$$

⁵Although this discussion focuses on the problem of *on-line* signature analysis, a forger typically only has access to *off-line* samples for the purpose of practising signature reproduction.

Table 3.2: The final set of *eighteen* function features. Each vector $\mathbf{f} \in \mathbb{R}^d$ is obtained from the initial feature set listed in Table 3.1.

Feature	Description
\bar{p}	Relative pen pressure
$\dot{\bar{p}}$	First derivative of relative pen pressure
$\ddot{\bar{p}}$	Second derivative of relative pen pressure
x	Horizontal pen position
\dot{x}	Horizontal pen velocity
\ddot{x}	Horizontal pen acceleration
\bar{y}	Vertical pen position
$\dot{\bar{y}}$	Vertical pen velocity
$\ddot{\bar{y}}$	Vertical pen acceleration
ϑ	Path tangent angle
$\dot{\vartheta}$	Path tangent angular velocity
$\ddot{\vartheta}$	Path tangent angular acceleration
θ	Azimuth angle
$\dot{\theta}$	Azimuth angular velocity
$\ddot{\theta}$	Azimuth angular acceleration
φ	Altitude angle
$\dot{\varphi}$	Altitude angular velocity
$\ddot{\varphi}$	Altitude angular acceleration

The systems developed in this study therefore consider a total of *eighteen* function features and *one* global feature for on-line signature representation in *feature space*. In the next section we discuss how this feature-based representation may be converted into a dissimilarity-based representation, which is suitable for incorporation into a writer-independent signature modelling framework.

3.4 Dichotomy transformation

Although the raw data associated with off-line and on-line signature samples are fundamentally different, both of the feature extraction techniques discussed in Sections 3.2.2 and 3.3.2 yield a feature set that contains T , d -dimensional feature vectors⁶ (see Table 3.3) for every signature sample enrolled into the system. In a *writer-dependent* verification scenario, one would use the collection of feature sets associated with a specific writer to construct a signature model unique to this individual. However, in order to facilitate the construction of a sensible *writer-independent* signature model, the efforts of a dichotomy transformation are required. This process converts the entire collection of feature sets, belonging to all the

⁶For the purposes of this discussion, we also treat the scalar feature associated with on-line signatures, that is the *signature duration*, as a trivial, one-dimensional feature *vector*.

Table 3.3: Interpretation of the feature set dimensions associated with off-line and on-line signatures.

Feature set dimension	Off-line signature	On-line signature
Feature set length (T)	Number of projection angles	Number of features
Feature vector dimension (d)	Projection profile length	Number of sample points

writers considered, from *feature space* into a universal dissimilarity-based representation in *dissimilarity space*.

Given a feature set $\mathbf{X}^{(k)}$, extracted from a positive reference signature belonging to a specific writer, any other feature set $\mathbf{X}^{(q)}$ that is claimed to belong to this writer can be converted into a dissimilarity vector $\mathbf{z}^{(q,k)}$ by calculating the dissimilarity between each pair of corresponding feature vectors contained in $\mathbf{X}^{(q)}$ and $\mathbf{X}^{(k)}$. It is proposed in Santos *et al.* (2004) that, for the development of an off-line verification system that considers graphometric features, this conversion can be achieved using the Euclidean distance, such that

$$\begin{aligned} \mathbf{z}^{(q,k)} &= D_{\text{Eucl}}(\mathbf{X}^{(k)}, \mathbf{X}^{(q)}) \\ &= \bigcup_{t=1}^T \sqrt{\left(\mathbf{x}_t^{(k)} - \mathbf{x}_t^{(q)}\right)' \left(\mathbf{x}_t^{(k)} - \mathbf{x}_t^{(q)}\right)}, \end{aligned} \quad (3.10)$$

where \bigcup and $'$ denote the vector concatenation and transpose operators, respectively, whilst $\mathbf{x}_t^{(q)} \in \mathbf{X}^{(q)}$ and $\mathbf{x}_t^{(k)} \in \mathbf{X}^{(k)}$ denote the t^{th} pair of corresponding feature vectors.

Although this Euclidean distance-based dichotomy transformation successfully quantifies the dissimilarity between two feature vectors, its element-based vector matching approach may be sensitive to slight variations in genuine signatures belonging to a specific writer (see e.g. Figure 3.8 (b)). This phenomenon, known as *intra-class variability*, is a common occurrence in signature modelling and may impede system proficiency if not addressed properly.

Consider, for example, the dichotomisation of the DRT-based feature sets yielded by the feature extraction process discussed in Section 3.2.2. When utilising a projection-based feature extraction technique, intra-class variability is generally manifested in the form of lateral shifting of the local minima and/or maxima within a specific projection profile, as conceptualised in Figure 3.9. It should be clear from Figure 3.9 (a) that, even when only minor shifting occurs, the use of an Euclidean-based approach may produce a significant misrepresentation of the dissimilarity between two projection profiles.

In order to address this issue, we propose that the dissimilarity between two DRT-based feature vectors be obtained by means of a *dynamic time warping* (DTW) algorithm. A similar dichotomy transformation is proposed in Muramatsu and Matsumoto (2009), where it is successfully utilised for the construction of dissimilarity vectors from *on-line* feature sets (similar to those yielded by the feature extraction process discussed in Section 3.3.2). When compared to the Euclidean distance, a DTW-based approach offers two notable advantages:

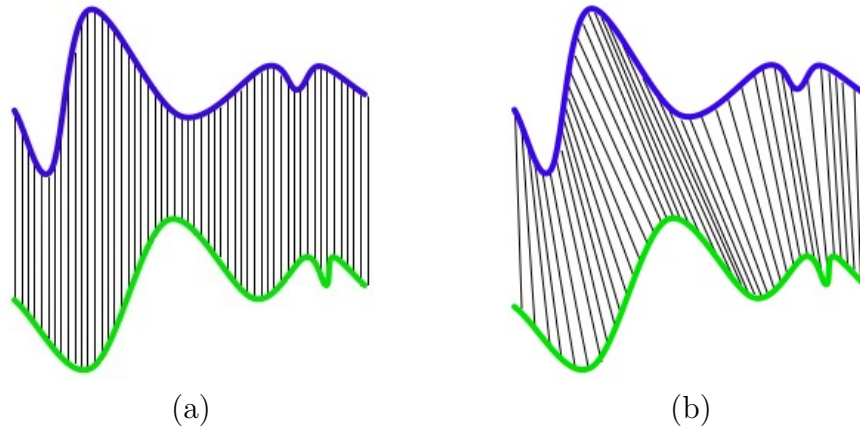


Figure 3.9: Conceptual comparison of the feature correspondences considered during dissimilarity vector construction when either (a) the Euclidean distance or (b) a DTW-algorithm is utilised. Note that, unlike the Euclidean distance-based approach, a DTW-algorithm is able to detect (and subsequently compensate for) non-linearly misaligned features, thereby producing a considerably more reliable measure of dissimilarity between the feature vectors submitted for comparison.

1. It enables a dissimilarity calculation between two vectors with different dimensions.

This algorithm is therefore able to obtain dissimilarity vectors from the on-line feature sets, which contain feature vectors that are all but *guaranteed* to have different dimensions, without any additional data processing in terms of length normalisation.

Since the off-line feature sets contain DRT-based feature vectors with a *standardised* dimension, this property is not exploited by the off-line systems developed in this study.

2. Prior to matching, it non-linearly aligns vector elements based on similarity, as illustrated in Figure 3.9 (b). The resulting dissimilarity measure corresponds to the Euclidean distance between the *aligned* vectors.

This property is of critical importance to both the on-line and off-line systems developed in this study. By matching features according to similarity rather than location, a DTW-based approach to dissimilarity calculation is able to compensate for reasonable intra-class variability i.e. signatures may be compared according to relative pen stroke shape as opposed to absolute pen stroke position.

Using the proposed DTW-based approach, the dissimilarity between $\mathbf{X}^{(q)}$ and $\mathbf{X}^{(k)}$ is calculated as follows,

$$\begin{aligned} z^{(q,k)} &= D_{\text{DTW}}(\mathbf{X}^{(k)}, \mathbf{X}^{(q)}) \\ &= \bigcup_{t=1}^T \mathcal{D}(\mathbf{x}_t^{(k)}, \mathbf{x}_t^{(q)}), \end{aligned} \quad (3.11)$$

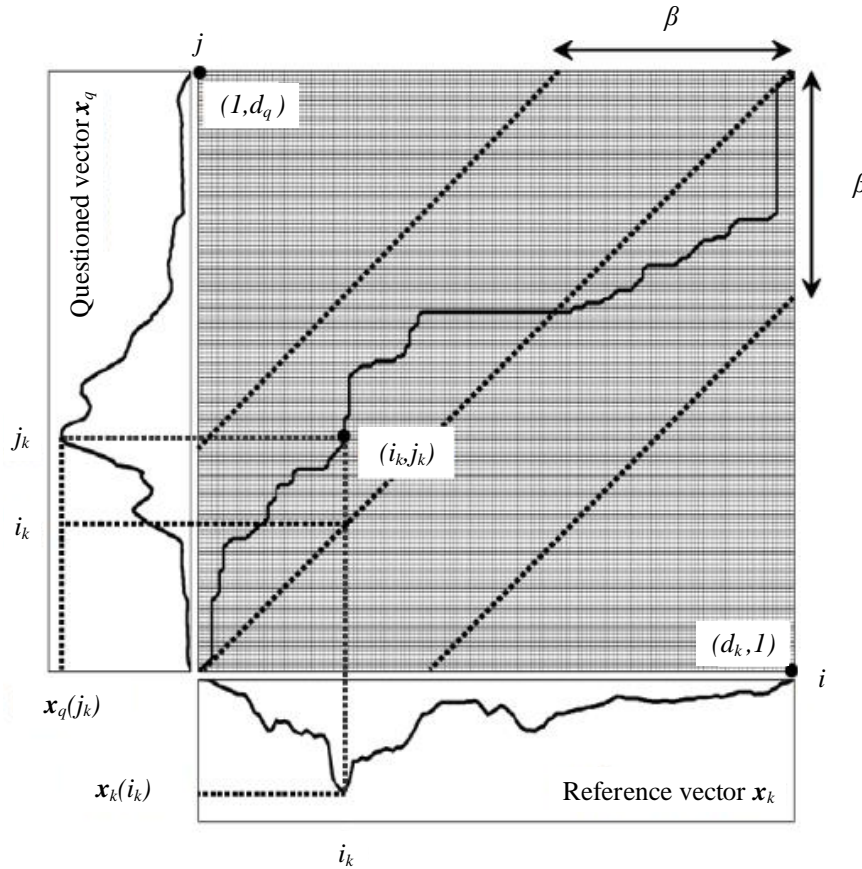


Figure 3.10: Conceptualisation of the dynamic time warping algorithm considered in this study. The algorithm identifies similar elements contained in the reference vector \mathbf{x}_k and questioned vector \mathbf{x}_q and subsequently constructs an optimal path between said vectors, based on these feature similarities. The resulting distance measure is calculated between elements matched according to the optimal path, as opposed to simply using corresponding elements. The *bandwidth* β restricts the search space and is used to regulate both the flexibility and computational requirements of the alignment process. Note that when $\beta = 0$, this algorithm is equivalent to utilising the Euclidean distance.

where $\mathcal{D}(\mathbf{f}, \mathbf{g})$ denotes the DTW-based distance between two arbitrary vectors \mathbf{f} and \mathbf{g} . A wide variety of DTW-algorithms are documented in the literature (Rabiner and Schmidt (1980); Keogh and Pazzani (2001); Henniger and Muller (2007); Jayadevan *et al.* (2009)). A detailed discussion of the *specific* DTW-algorithm considered in this study is presented in Appendix A, whilst a graphical conceptualisation of the algorithm is provided in Figure 3.10.

The DTW-based dichotomy transformation discussed in this section is therefore able to convert any two *sets* of T feature vectors into a *single* T -dimensional dissimilarity

vector, which is suitable for consideration within a writer-independent signature modelling framework.

3.5 Concluding remarks

In this chapter we discussed how the systems developed in this study convert raw signature data associated with two fundamentally different signature acquisition methods, namely off-line and on-line methods, into a dissimilarity-based representation suitable for writer-independent signature modelling.

In the off-line case we discussed how the expected quality of features extracted from a typical signature sample is in fact largely dependent on the successful completion of several image pre-processing techniques. Such techniques include those concerned with image binarisation, noise reduction and signature segmentation. Suitable methods were presented in order to address these issues. We also explained how the DRT may subsequently be used to extract a set of projection profiles from any given binary signature image, thereby yielding a writer-dependent feature set. It is worth mentioning that the image processing techniques discussed in this chapter are considered to be tried and tested, and therefore commonly utilised throughout the literature. However, none of these techniques are necessarily considered *optimal* for their respective tasks. Advancements are continually being made in the field of document processing and state-of-the-art techniques are proposed on a regular basis. An in-depth investigation into such advanced image processing techniques is, however, deemed outside the scope of this study.

In the on-line case we discussed how relatively little pre-processing of the raw data is required, whilst a wide variety of additional function and/or global features may easily be calculated from the initial feature set yielded by the acquisition process. It is worth noting that a great many additional features may also be calculated using the originally captured signature data, as detailed in e.g. Dolfing *et al.* (1998), Ketabdar *et al.* (2005), and Van *et al.* (2007). One should, however, be cautious not to include *too many* features in the final signature representation, in order to avoid the adverse effects of the so-called *ugly duckling theorem* postulated in Watanabe (1985), which states that it is possible to make two arbitrary patterns *similar* by encoding them with a sufficiently large number of redundant features.

Following the completion of appropriate feature normalisation techniques, we showed that both the off-line and on-line signature feature extraction strategies are able to produce feature sets that are invariant with respect to translation and scale. We do not address the issue of rotation invariance in this study. The assumption of a relatively consistent signature orientation is considered reasonable within the context of this study, since the signing area of e.g. a cheque is generally bounded and has a fixed orientation. Consequently, the writer has very limited freedom regarding the orientation of his/her signature. Nevertheless, several methods for achieving rotation invariance are documented in the literature. One such method, proposed in Panton and Coetzer (2010), employs the efforts of a ring-structured HMM and *Viterbi alignment* in order to estimate the most probable

angle of rotation of a questioned signature, relative to a known reference sample. The rotation of the questioned signature is subsequently corrected for prior to feature extraction. The development of such an advanced classifier for the sole purpose of achieving rotation invariance is, however, also deemed outside the scope of this study.

After successful completion of the feature extraction process, the efforts of a dichotomy transformation are required in order to obtain a writer-independent signature representation in dissimilarity space. In order to improve upon the traditional Euclidean distance based method, we proposed the use of a DTW-based dichotomy transformation for obtaining both an off-line *and* on-line dissimilarity representation. By employing a DTW-algorithm, the proposed technique is able to non-linearly align feature vectors prior to matching and therefore improves the robustness of the dissimilarity vectors generated, and in so doing compensate for reasonable intra-class variability. The contribution of the proposed off-line dissimilarity representation technique to the field of handwritten signature analysis is experimentally verified in Chapter 6.

In the next chapter, we introduce the classification techniques employed in this study, namely discriminant analysis and support vector machines.

Chapter 4

Classifiers

“It is far better to foresee, even without certainty, than not to foresee at all.”
- Henri Poincaré (1854–1912)

4.1 Introduction

The signature representation techniques discussed in the previous chapter provide an effective platform for distinguishing between the positive and negative classes (i.e. authentic signatures and forgeries) in dissimilarity space, thereby completing the feature extraction stage within the larger *learning* process. During the next stage of the learning (or testing) process, namely signature modelling (or verification), the efforts of a suitable mathematical mechanism, or *classifier*, are required in order to interpret the aforementioned signature representation.

In this chapter we discuss two fundamentally different classification techniques, namely *discriminant analysis* (DA) and *support vector machines* (SVMs). In Section 4.2 we first discuss the fundamental differences between *generative* and *discriminative* models for pattern recognition. We then present abridged discussions on the specific classification techniques considered in this study, that is a DA-based method (Section 4.3) and an SVM-based method (Section 4.4), and illustrate how these techniques may be utilised for distinguishing between positive and negative patterns.

4.2 Overview

As explained in Bishop (2006), any classification technique that is rooted in probability theory (and is primarily aimed towards minimising its misclassification rate) typically assigns a questioned pattern \mathbf{x} to the pattern class y_i that maximises the class-specific *posterior probability* $P(y_i|\mathbf{x})$. Furthermore, any statistical classification technique may be broadly categorised as either a *generative* or a *discriminative* approach to pattern modelling and recognition.

The generative approach is aimed towards estimating each *class prior probability* and *class-conditional density* from a set of labelled training samples, in order to ultimately infer the *posterior probability* associated with each class by employing *Bayes' theorem*. A generative classifier therefore models the joint distribution of pattern descriptor variables (i.e. classifier input) and target variables (i.e. classifier output). Popular generative classifiers include *hidden Markov models* (HMMs), *Gaussian mixture models* (GMMs) and *naive Bayes classifiers*.

In contrast, a discriminative classifier constructs a parametric model for the *posterior probabilities* and subsequently infers optimal values for said parameters from a set of labelled training samples, in order to obtain a maximally discriminant *decision boundary* in feature space. Popular discriminative classifiers include *discriminant functions*, *multilayer perceptrons* (MLPs) and SVMs.

The most appropriate modelling approach for any given application is generally dependent on its deployment environment, since both generative and discriminative classifiers have specific advantages and limitations. A detailed comparison between the generative and discriminative approaches to pattern modelling, within the context of object recognition, can be found in Ulusoy and Bishop (2005).

Within the context of *signature verification*, the majority of systems proposed in the literature employ either a template matching technique or a generative classifier for the purposes of signature modelling and verification, whilst a relatively small percentage of existing systems employ the efforts of a discriminative classification technique. This is mainly due to the fact that, prior to the advent of *writer-independent* signature modelling strategies, the utilisation of discriminative classifiers was generally considered inappropriate for the purpose of *skilled*¹ forgery detection. It is therefore reasonable to expect that the utilisation of discriminative classifiers for the purpose of signature model construction will become more popular in future, when the writer-independent approach is more commonly adopted. In this study we investigate both generative *and* discriminative approaches to signature modelling and verification.

Traditionally, the respective outputs of generative and discriminative classifiers are expected to be continuous (i.e. a similarity score) and discrete (i.e. a class label). However, one can easily augment a discriminative method of classification to produce a continuous *confidence score*. This may be achieved by not only considering the *location* of a questioned pattern relative to the decision boundary, but also its *distance* from said boundary. It should be clear that the confidence associated with a discriminative classification event is

¹In order to facilitate the successful training of a discriminative classifier, the availability of *both* positive and negative training samples is strictly required. Within the context of the traditional *writer-dependent* approach, it is not reasonable to expect that samples of skilled forgeries would be available for *every* new writer enrolled into the system and one is therefore *forced* to consider *random forgeries* for training purposes. However, since the *writer-independent* approach involves a once-off training process that yields one universal signature model, the availability of samples of skilled forgeries is *only* required during training of the initial model, whilst any newly enrolled writers only need to submit positive reference samples. A discriminative classification technique is therefore appropriate within the context of a writer-independent signature modelling framework.

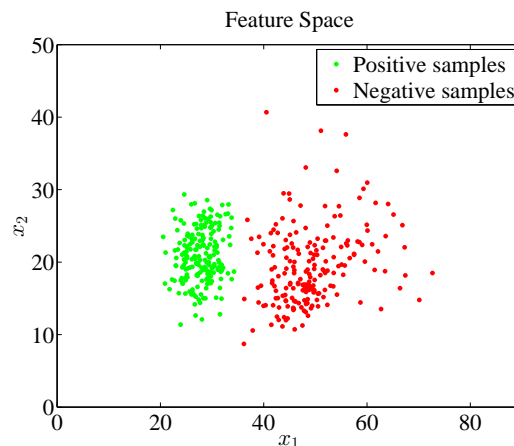


Figure 4.1: Distribution of hypothetical positive and negative training samples in a two-dimensional feature space.

directly proportional to the aforementioned distance measure. The systems developed in this study employ such an augmented approach to discriminative classification.

In order to aid in our discussions of the DA-based and SVM-based methods considered in this study, we also illustrate the implementation of these techniques on an artificially generated set of positive and negative training samples, as shown in Figure 4.1.

Finally, it should be noted that, although the systems developed in this study convert raw signature data into a pattern representation in *dissimilarity space*, we may, without loss of generality, present our discussions and/or demonstrations in this chapter within the traditional context of *feature space*.

4.3 Discriminant analysis

Discriminant classifiers employ a *generative* approach to pattern modelling. Specifically, a discriminant classifier uses a collection of training patterns to model each relevant pattern class as a Gaussian probability density function (PDF) and subsequently relies on the measure of similarity computed between a questioned pattern and each estimated PDF in order to predict class membership.

4.3.1 Overview

For comprehensive discussions on the DA-based classification method, the reader is referred to such works as Klecka (1980), Friedman (1989), and McLachlan (2004).

As mentioned in Section 4.2, a generative classifier is aimed towards the inference of class posterior probabilities $P(y_i|\mathbf{x})$ by means of Bayes' theorem, which is defined for

multi-class classification scenarios as

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)P(y_i)}{\sum_j P(\mathbf{x}|y_j)P(y_j)}, \quad (4.1)$$

where $P(y_i)$ and $P(\mathbf{x}|y_i)$ denote the prior probability and class-conditional density associated with class y_i respectively. The systems developed in this study are only concerned with two classes, that is a positive class G and a negative class F . We also assume *equal prior probabilities*, that is $P(G) = P(F) = 0.5$, in order to ensure *unbiased*² classifiers. These application-specific assumptions lead to the following simplification of the posterior probability estimates,

$$P(G|\mathbf{x}) = \frac{P(\mathbf{x}|G)}{P(\mathbf{x}|G) + P(\mathbf{x}|F)}, \quad (4.2)$$

$$P(F|\mathbf{x}) = \frac{P(\mathbf{x}|F)}{P(\mathbf{x}|G) + P(\mathbf{x}|F)}. \quad (4.3)$$

It is clear from (4.2) and (4.3) that the posterior probabilities of the positive and negative classes are entirely dependent on the class-conditional densities of said classes. As mentioned earlier, the class-conditional density of class y_i is represented by a multivariate Gaussian PDF, defined for $\mathbf{x} \in \mathfrak{R}^d$ as

$$\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}, \quad (4.4)$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ denote the mean vector and covariance matrix respectively. The class statistics $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ therefore uniquely define y_i . Given a set of N_i training samples, known to belong to y_i and denoted by $\mathbf{X}_i = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_i}\}$, these class statistics may be determined as follows,

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_j, \quad (4.5)$$

$$\boldsymbol{\Sigma}_i = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)'. \quad (4.6)$$

Since the systems developed in this study are only concerned with the positive class (defined by $\boldsymbol{\mu}^+$ and $\boldsymbol{\Sigma}^+$ estimated from \mathbf{X}^+) and the negative class (defined by $\boldsymbol{\mu}^-$ and $\boldsymbol{\Sigma}^-$ estimated

²In most deployment scenarios, the likelihood of encountering an authentic or a fraudulent signature is not equal. In a banking environment, for example, one would expect $P(G) \approx 1$ and $P(F) \approx 0$. In such a scenario, one may achieve near-optimal performance by simply accepting every questioned signature as authentic. However, such a strategy of *blind acceptance* would obviate the need for any human and/or machine-based signature verification and consequently render the task of forgery detection obsolete. Furthermore, in order to arrive at a truly *objective* conclusion regarding the authenticity of a questioned signature, one has to assume that any sample in question is equally likely to be authentic or fraudulent.

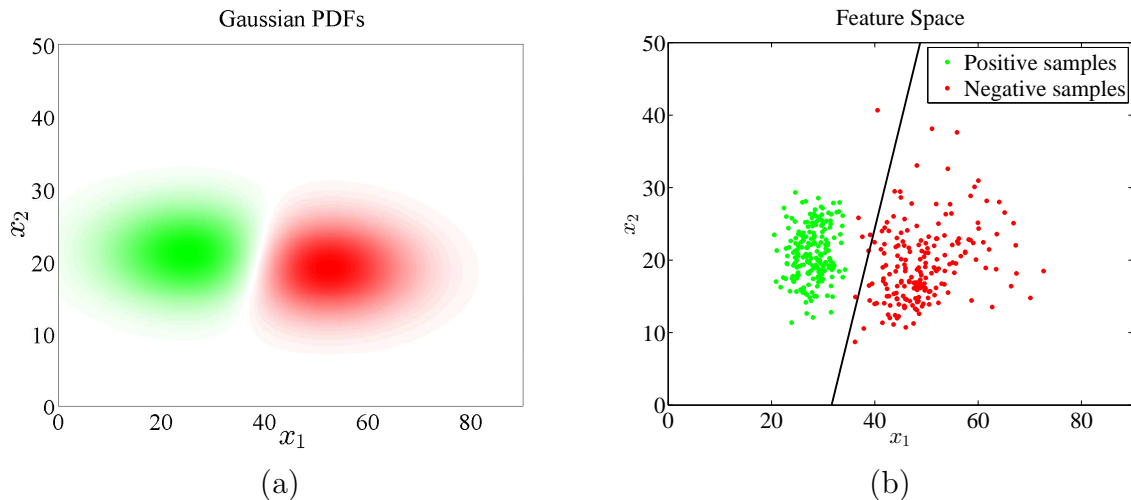


Figure 4.2: Implementation of a linear discriminant. (a) Multivariate Gaussian PDFs estimated from the samples depicted in Figure 4.1, using class-specific means and a *pooled* covariance estimate. (b) The resulting linear decision boundary $f_{\text{LDA}}(\mathbf{x}) = 0$. Note that, although the positive and negative classes are clearly linearly separable, several negative samples are misclassified.

from \mathbf{X}^-), a convenient formulation for the DA-based class membership function $f_{\text{DA}}(\mathbf{x})$ can be obtained as follows,

$$f_{\text{DA}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \boldsymbol{\mu}^+, \boldsymbol{\Sigma}^+) - \mathcal{N}(\mathbf{x}, \boldsymbol{\mu}^-, \boldsymbol{\Sigma}^-). \quad (4.7)$$

Using the formulation in (4.7), \mathbf{x} is classified as belonging to G if $f_{\text{DA}}(\mathbf{x}) \geq 0$. Similarly, \mathbf{x} is classified as belonging to F if $f_{\text{DA}}(\mathbf{x}) < 0$.

Depending on how the *variability* of each class is modelled, the above method may be used to perform either *linear discriminant analysis*³ (LDA) or *quadratic discriminant analysis* (QDA).

An LDA-based classifier models the positive and negative classes with a *pooled covariance* estimate, that is $\boldsymbol{\Sigma}^+ = \boldsymbol{\Sigma}^- = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is estimated from the *entire* set of training samples. This *assumption of homoscedasticity* does therefore not address the variability of the data in any way and consequently yields a linear decision boundary, as illustrated in Figure 4.2. As a result, when utilising a *linear discriminant* for pattern class modelling, the similarity between a questioned sample and each class-specific PDF is based on the *Euclidean* distance measure.

In contrast, a QDA-based classifier models the two classes G and F using *class-specific covariance* estimates. Since this method is able to quantify both the location and variability of the pattern classes in feature space, a non-linear decision boundary is produced, as illustrated in Figure 4.3. As a result, when utilising a *quadratic discriminant* for pattern class modelling, the similarity between a questioned sample and each class-specific PDF is based on the *Mahalanobis* distance measure.

³LDA is also commonly referred to as *Fisher's discriminant*, since this approach to pattern classification was originally proposed in Fisher (1936).

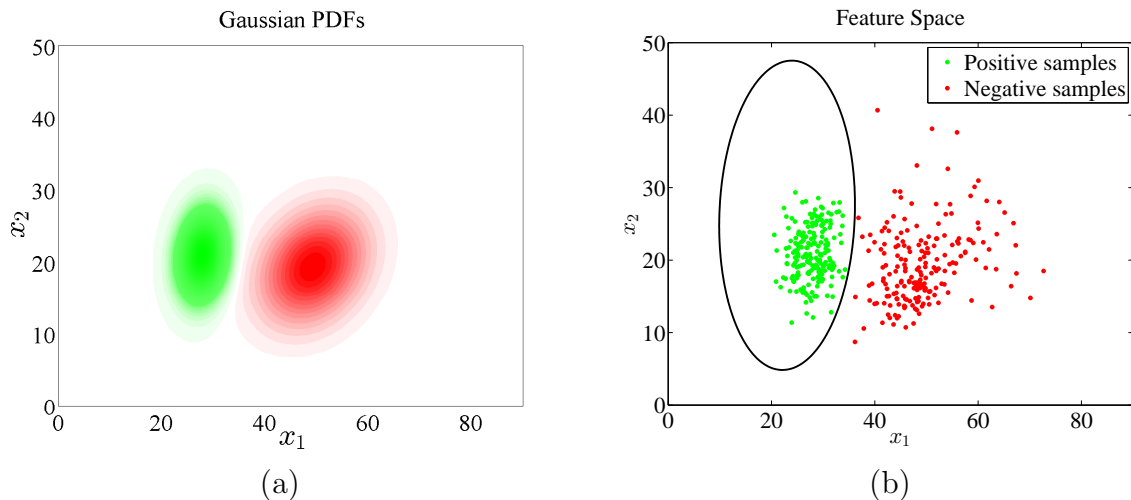


Figure 4.3: Implementation of a quadratic discriminant. (a) Multivariate Gaussian PDFs estimated from the samples depicted in Figure 4.1, using class-specific means and *class-specific* covariance estimates. (b) The resulting quadratic decision boundary $f_{\text{QDA}}(\mathbf{x}) = 0$. Note that, unlike the linear discriminant depicted in Figure 4.2 (b), the quadratic discriminant is able to successfully separate the positive and negative classes.

4.3.2 The curse of dimensionality

It is evident from the examples presented in Figures 4.2–4.3 that the ability to address class-specific variability lends much needed flexibility to the resulting model. One may therefore expect that the utilisation of a QDA-based approach will *always* result in superior classification performance, when compared to its linear counterpart. The flexibility provided by the quadratic discriminant is, however, offered at a premium.

It should be clear from (4.4) that, in order to model each pattern class with a Gaussian PDF, each class-specific covariance matrix needs to be invertible. However, when limited data are available for covariance estimation, the resulting covariance matrices may become unreliable (in which case the predictive ability of the resulting classifier may decrease substantially), or even singular (in which case it is not possible to construct a classifier at all). Furthermore, as the feature space dimension (and therefore also the number of feature correlations that need to be estimated) increases, the number of training samples required for reliable covariance estimation also increases *dramatically*. This phenomenon is commonly referred to as the *curse of dimensionality*. In order to avoid this potentially serious impediment to successful pattern class modelling, one is therefore required to either maximise the number of training samples or, alternatively, minimise the number of features considered.

The systems developed in this study do not employ any techniques aimed towards dimension reduction. We consequently examine (in Section 6.3) whether or not the DA-based systems developed in this study are susceptible to the curse of dimensionality.

4.4 Support vector machines

SVMs employ a *discriminative* approach to pattern modelling. Specifically, the decision boundary yielded by an SVM classifier represents the separating hyperplane in feature space that maximises the classification *margin*, that is the Euclidean distance between said hyperplane and its closest (positive or negative) training sample. The *structural risk minimisation* principle (Guyon *et al.* (1992)) guarantees that only one such *maximally separating hyperplane* exists.

4.4.1 Overview

For a comprehensive discussion on the SVM-based classification method, the reader is referred to such works as Vapnik (1995) and Cortes and Vapnik (1995), whilst a practical guide to SVM design and implementation can be found in e.g. Burges (1998), Hsu *et al.* (2003), and Ben-Hur and Weston (2010).

Given a set of N training samples $\mathbf{x}_i \in \mathfrak{R}^d$, each associated with a known class label $y_i \in \{1, -1\}$ (i.e. each sample \mathbf{x}_i belongs either to the positive class G or the negative class F), the objective of an SVM classifier is to find the optimal solution to the following minimisation problem,

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}' \mathbf{w} \\ \text{subject to} \quad & y_i (\mathbf{w}' \mathbf{x}_i + b) \leq 1 \\ & i = 1, \dots, N. \end{aligned} \tag{4.8}$$

Solving (4.8) yields the *weight vector* \mathbf{w} and *bias* b , which in turn determines the orientation and location, respectively, of the hyperplane $f(\mathbf{x}) \in \mathfrak{R}^{d-1}$ that maximally separates G and F in feature space. Furthermore, the *dual* formulation of (4.8) reveals that the optimal weight vector is determined from a specific subset of training samples, that is those samples located on the *margin* of the separating hyperplane. These training samples are referred to as the *support vectors* (see Figure 4.4 (a)). The successful training of an SVM classifier ultimately yields a linear membership function in feature space, such that

$$f(\mathbf{x}) = \mathbf{w}' \mathbf{x} + b. \tag{4.9}$$

Using the so-called *hard-margin* SVM formulation in (4.9), \mathbf{x} is classified as belonging to G if $f(\mathbf{x}) \geq 0$. Similarly, \mathbf{x} is classified as belonging to F if $f(\mathbf{x}) < 0$.

The hard-margin formulation described above inherently makes the assumption that the training data is linearly separable in feature space, as conceptualised in Figure 4.4 (a). In most practical applications, this is a naive assumption. It is therefore often desirable to utilise a *soft-margin* SVM formulation. This is achieved by modifying (4.8) to include a set of N *slack variables* $\xi_i \geq 0$, thereby allowing for *margin errors* (where $0 \leq \xi_i \leq 1$) and *misclassifications* (where $\xi_i > 1$), as conceptualised in Figure 4.4 (b). The introduction of

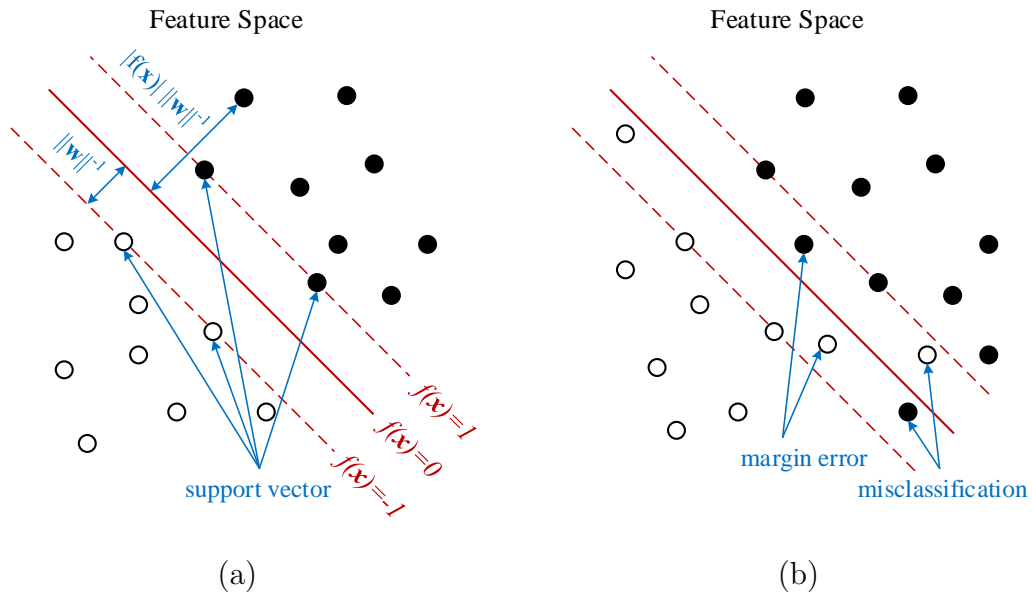


Figure 4.4: Conceptualisation of the maximally separating hyperplane and its associated margin in a hypothetical two-dimensional feature space for (a) linearly separable and (b) linearly non-separable training data.

these additional variables also results in a modified minimisation objective as follows,

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \mathbf{w}' \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & y_i (\mathbf{w}' \mathbf{x}_i + b) \leq 1 - \xi_i \\ & \xi_i \geq 0 \\ & i = 1, \dots, N, \end{aligned} \quad (4.10)$$

which in turn also results in a modified membership function,

$$f(\mathbf{x}) = \mathbf{w}' \mathbf{x} + b + C \sum_{i=1}^N \xi_i. \quad (4.11)$$

The soft-margin approach therefore introduces an additional internal parameter, namely the *regularisation parameter* $C > 0$, which is used to quantify the penalty associated with potential margin errors and misclassifications. The significance of C is further discussed in the next section.

Although it should be clear that an SVM classifier invariably aims to construct a maximally separating *linear* decision boundary, one may easily adapt (4.11) in order to perform *non-linear* classification. This is achieved by employing the so-called *kernel trick*, where the original data in *feature space* is mapped into a (potentially higher dimensional) *kernel space*, wherein the *mapped* data may be linearly separated. This data transformation,

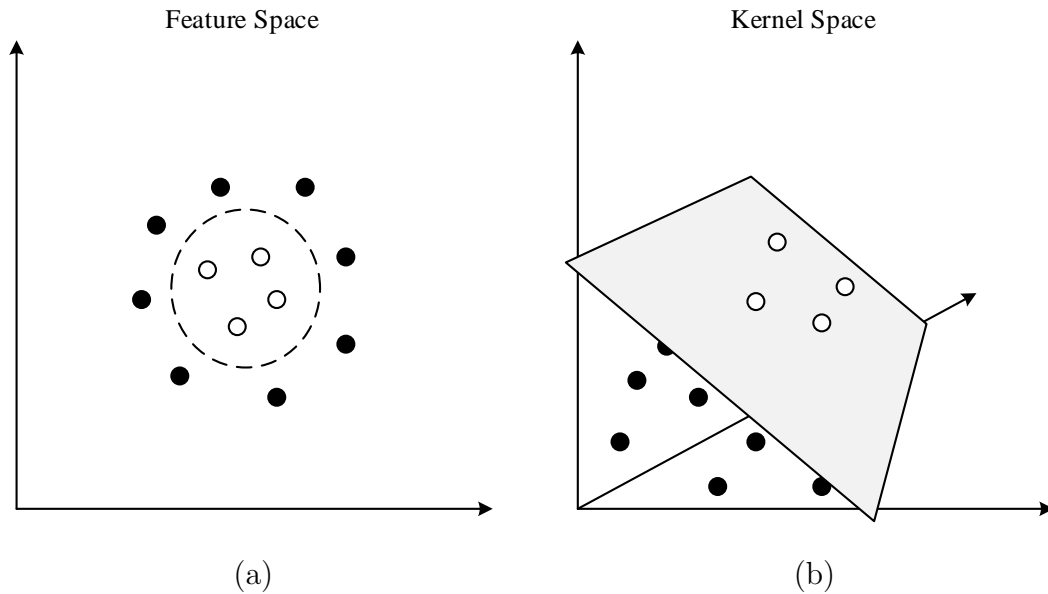


Figure 4.5: Conceptualisation of the kernel trick. (a) Positive and negative training samples in a hypothetical two-dimensional *feature space*. There clearly exists no linear decision boundary (i.e. a straight line) capable of separating the two classes. (b) Non-linear mapping of the training samples depicted in (a) into a hypothetical three-dimensional *kernel space*, wherein it becomes possible to obtain a separating hyperplane. The inverse mapping of the hyperplane indicated in (b) corresponds to the dashed line indicated in (a), which successfully separates the two classes in feature space.

conceptualised in Figure 4.5, is performed using a *kernel function*, defined for $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d$ as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)' \phi(\mathbf{x}_j), \quad (4.12)$$

where $\phi(\mathbf{x})$ denotes any suitable mapping function. The maximally separating hyperplane in kernel space effectively translates into a non-linear decision boundary in the original feature space, where the resulting SVM-based membership function may be expressed as follows,

$$f_{\text{SVM}}(\mathbf{x}) = \mathbf{w}' \phi(\mathbf{x}) + b + C \sum_{i=1}^N \xi_i. \quad (4.13)$$

In a scenario where a suitable hyperplane can be found in the original feature space, a *linear kernel* is used, such that

$$K_{\text{LIN}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i' \mathbf{x}_j. \quad (4.14)$$

It is evident from (4.12) and (4.14) that the use of a linear kernel corresponds to a *trivial mapping* wherein the kernel space is equivalent to the original feature space.

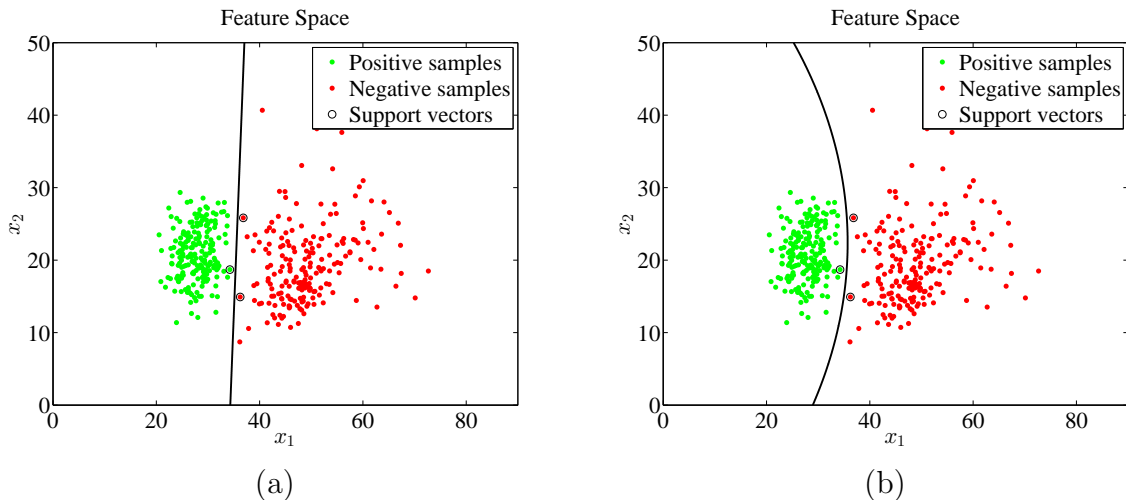


Figure 4.6: Implementation of a soft-margin SVM. Decision boundaries obtained from the training samples depicted in Figure 4.1 when (a) a linear kernel and (b) an RBF kernel are utilised. Since the training samples are linearly separable in feature space, both the linear and RBF kernels yield boundaries that are able to successfully separate the two classes. Note that the orientation (and also the curvature in the case of the RBF kernel) of each boundary is determined by the value specified for the internal parameter C (and also γ), which is further discussed in Section 4.4.2.

A variety of popular *non-linear* kernel functions are documented in the literature, including the *polynomial*, *radial basis function* (RBF) and *sigmoid* kernels. Although the selection of the most suitable kernel function is often a process of trial and error, Hsu *et al.* (2003) suggest that the RBF kernel generally provides an effective non-linear decision boundary and should therefore prove a sound first choice in most instances. The RBF kernel is defined as

$$K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\gamma^2}\right), \quad (4.15)$$

where the internal parameter $\gamma > 0$ determines the *kernel width* and controls the non-linear flexibility of the resulting classifier. The significance of γ is further discussed in the next section.

Popular algorithms for SVM training include *least squares approximation* (Suykens and Vandewalle (1999)) and the *sequential minimal optimisation* (SMO) algorithm (Platt (1998)). Graphical examples of the decision boundary obtained by implementing a linear kernel and an RBF kernel soft-margin SVM are presented in Figure 4.6.

4.4.2 SVM parameter selection

The level of success achievable by an SVM classifier is entirely dependent on the optimality of the parameters that describe its separating hyperplane. As explained in the previous

section, appropriate values for the weight vector \mathbf{w} and bias b are determined automatically by means of a suitable SVM training algorithm. Prior to training, however, it is required that predetermined values be specified for the *internal* SVM parameters, that is the regularisation parameter C and, if applicable, the RBF kernel width γ . Although the values of C and γ may be chosen arbitrarily, these parameters may potentially have a significant impact on the optimality of the resulting decision boundary, as we now discuss in more detail.

The regularisation parameter C specifies the penalty associated with margin errors and/or misclassifications. For instance, when $C \rightarrow \infty$, the optimal solution to (4.10) yields values for \mathbf{w} and b that shift the separating hyperplane in such a way that these errors are (if possible) totally avoided. Conversely, when $C \rightarrow 0$, those samples closest to the hyperplane are effectively ignored, in which case the optimal solution maximises the margin in terms of the *remaining* training samples. The parameter C is therefore said to control both the *orientation* and *margin width* associated with the resulting SVM decision boundary. One may therefore be tempted to always specify a relatively large value for C , thereby ensuring optimal classifier performance during the training phase. However, Ben-Hur and Weston (2010) explain that it is often advisable to diminish the importance associated with misclassifications (and especially margin errors) on training data, in order to improve the generalisation potential of the resulting classifier, as conceptualised in Figure 4.7.

The kernel width parameter γ essentially determines the non-linear flexibility of the decision boundary. It should be clear from (4.15) that, for any two data points \mathbf{x}_i and \mathbf{x}_j , the value of the RBF kernel function $K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j)$ is proportional to γ , whilst it is inversely proportional to the distance between \mathbf{x}_i and \mathbf{x}_j . The parameter γ therefore determines the *region of influence* of each data sample, that is the region in feature space for which \mathbf{x}_i yields a non-zero kernel value. The regions of influence associated with the *support vectors*, in turn, determine the *curvature* of the resulting SVM decision boundary, as explained in Ben-Hur and Weston (2010). When $\gamma \rightarrow \infty$, the decision boundary is smoothed to such an extent that it approaches linearity, whilst a value of $\gamma \approx 0$ results in an excessively curved decision boundary that over-fits the training data, as illustrated in Figure 4.8.

Since both C and γ have such a significant influence on the decision boundary yielded by an SVM, it is of the utmost importance that every effort is made to assign sensible values to these internal parameters. Unfortunately, no analytical method capable of identifying the optimal values for C and γ currently exists. The majority of works found in the literature (e.g. Min and Lee (2005); Kumar *et al.* (2011)) therefore perform SVM parameter selection by means of a *grid search*, that is a brute force search for the best *combination* of values for C and γ , as indicated by an appropriate performance metric which is estimated for each resulting classifier candidate. Provided that a sufficiently fine grid, i.e. an adequately comprehensive set of possible parameter value combinations, is considered during the search, this method guarantees the optimality of the parameter values identified. However, as is generally the case with brute force methods, such a grid search is computationally exhaustive. Furthermore, when used within a highly repetitive experimental protocol, as is the case in this study (see Section 6.2.2), the utilisation of a fine grid search for the purpose of parameter selection is not computationally feasible.

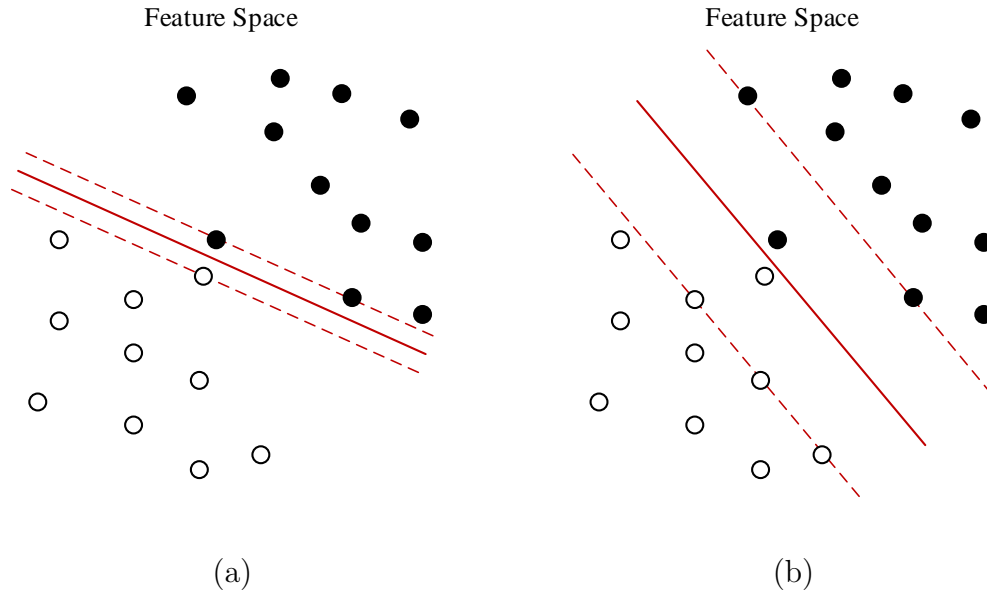


Figure 4.7: Significance of the regularisation parameter C . Optimal decision boundaries for the separation of two classes in a hypothetical two-dimensional feature space when (a) relatively large and (b) relatively small values for C are specified. Although the boundary in (a) reduces the number of margin errors on the training set, the boundary in (b) provides a more sensible overall separation of the two classes and is therefore expected to prove superior in the classification of future questioned samples.

Various alternative strategies, aimed towards alleviating the computational overhead typically associated with SVM parameter selection, are documented in the literature. Notable variations of the traditional grid search include the *pattern search* (Momma and Bennett (2002)) and the *self-tuning grid search* (Staelin (2003)). Several strategies have also been proposed that utilise a fundamentally different approach to that of a grid-based search, including *genetic algorithms* (Huang and Wang (2006)), *particle swarm optimisation* (Guo *et al.* (2008)), as well as *simulated annealing* (Lin *et al.* (2008)). The fact that such a wide variety of methods have been proposed in the literature confirms that SVM parameter selection is by no means a trivial endeavour.

The SVM-based systems developed in this study estimate suitable values for C and γ by means of the heuristic approach outlined in Mattera and Haykin (1999). Firstly, *every* training sample (which may be a positive or negative sample) is compared to every *other* training sample, in order to obtain the complete set \mathbf{X}_D of pairwise Euclidean distances between all the training samples, such that

$$\mathbf{X}_D = \bigcup_{\substack{\mathbf{x}_i, \mathbf{x}_j \\ i > j}} \|\mathbf{x}_i - \mathbf{x}_j\|. \quad (4.16)$$

The parameters C and γ are subsequently assigned the *maximum* and *median* values of the

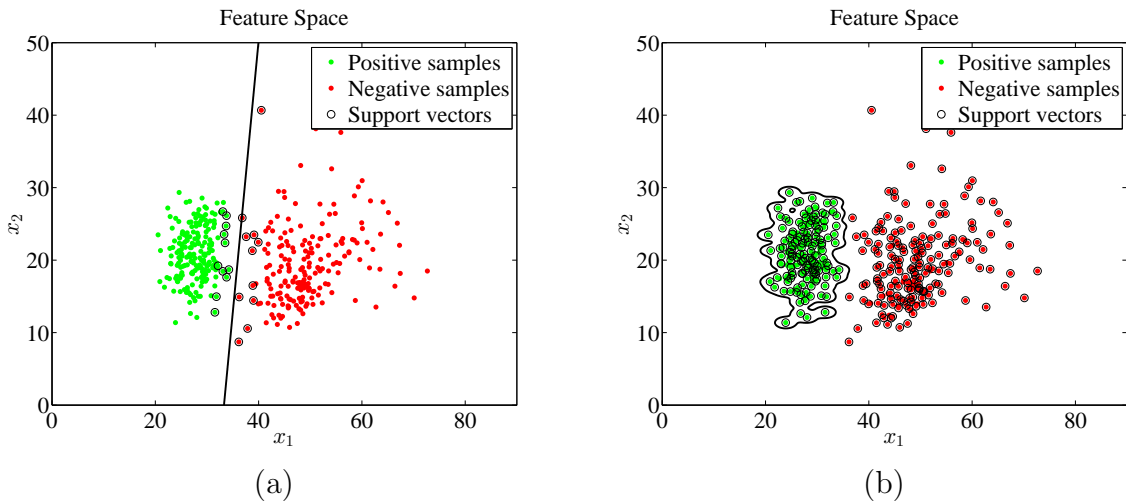


Figure 4.8: Significance of the RBF kernel width parameter γ . Decision boundaries obtained from the training samples depicted in Figure 4.1 when (a) $\gamma = 10000$ and (b) $\gamma = 0.1$. The boundary in (a) is over-smoothed to such an extent that it resembles the boundary obtained when employing a linear kernel (see Figure 4.6 (a)), whilst the excessive non-linear flexibility of the boundary in (b) clearly over-fits the training set – this is easily confirmed by observing that each of the training samples is also identified as a support vector i.e. *every* training sample contributes towards the optimal solution.

set \mathbf{X}_D respectively. Although this method does not guarantee that the obtained parameter values are optimal, it does reduce the computational overhead *substantially*. This heuristic approach to SVM parameter selection is also utilised successfully in e.g. Kan and Shelton (2008).

Since both parameter values are estimated directly from the training data, Cherkassky and Ma (2004) caution that the success of this parameter selection strategy may be adversely affected by outliers in the training data. The issue of outlier detection and removal is addressed in the next chapter.

4.5 Concluding remarks

In this chapter we introduced two fundamentally different classification techniques as candidates for incorporation into the systems developed in this study. For each of these classifiers, both linear and non-linear configurations were discussed. We also illustrated the nature of the decision boundaries yielded by these different classifiers on an artificially generated data set that contains samples representative of two pattern classes.

It should be noted that, in order to simplify the illustrations presented in this chapter, a small data set that contains linearly separable, two-dimensional samples was used for *demonstration* purposes. However, in the next chapter we will show that, within the context of the systems developed in this study, a typical training set is expected to contain a

large number of non-separable positive and negative samples. Furthermore, typical training samples are not limited to a two-dimensional pattern representation. As explained in the previous chapter, the dimension of the dissimilarity vectors utilised for signature modelling may vary in accordance with the value specified for the system parameter T (see Section 3.4).

It is also important to recall from the previous chapter that, when a *dissimilarity*-based approach to pattern representation is utilised, a dissimilarity vector obtained from a typical positive sample is expected to comprise elements located relatively close to the origin in dissimilarity space, whilst the dissimilarity values representative of a typical negative sample are expected to be located further from the origin – in one or more dimensions (depending on the number of features). For this reason, the *non-linear* classifier configurations discussed in this chapter, that is the quadratic discriminant and the RBF-kernel SVM (or R-SVM), are deemed most appropriate for incorporation into the systems developed in this study.

In the next chapter we explain how the efforts of the non-linear classifiers (discussed in this chapter) may be used in conjunction with a dissimilarity-based signature representation strategy (discussed in the previous chapter) to construct a framework for successful writer-independent signature modelling and verification.

Chapter 5

Signature Modelling and Verification

“The oldest, shortest words - ‘yes’ and ‘no’ - are those which require the most thought.”
- Pythagoras (582BC–497BC)

5.1 Introduction

In Chapter 3 we explained how raw signature data may be converted into a dissimilarity vector that is suitable for consideration within a writer-independent verification framework. In the previous chapter we discussed *quadratic discriminant analysis* (QDA) and *radial basis function kernel support vector machines* (R-SVMs) and demonstrated how these classification techniques may be used for the purpose of discriminating between two pattern classes.

In this chapter we discuss how sets of dissimilarity vectors may be used in conjunction with the previously discussed classification techniques, in order to construct an efficient framework for writer-independent signature modelling and verification. We also introduce two post-processing techniques, namely an *iterative outlier removal* (IOR) algorithm (Section 5.2.1) and a novel *writer-specific dissimilarity normalisation* strategy (Section 5.2.2). These algorithms are implemented prior to model training (Section 5.2.3), in order to improve the robustness of the resulting signature model.

Finally, in Section 5.3 we discuss the verification protocol associated with a trained model, thereby completing the design of the writer-independent handwritten signature verification system.

5.2 Signature modelling

The systems developed in this study use a collection of dissimilarity vectors, obtained from samples of genuine signatures *and* skilled forgeries belonging to *several different writers*, for signature modelling. In order to produce these dissimilarity vectors, each writer is required to submit K genuine signatures during enrolment, that serve as a writer-specific reference set.

Prior to model construction, signature samples are obtained from a set of so-called *guinea pig writers* in a *controlled* environment. These writers are considered representative of the general public, and their signatures are used for training purposes only. Given K reference signatures and N labelled training signatures (that include an equal number of positive and negative samples) for each of the Ω guinea pig writers, a set of dissimilarity vectors is generated for each writer by computing $\mathbf{z}^{(n,k)}$ for $n = \{1, 2, \dots, N\}$ and $k = \{1, 2, \dots, K\}$.

We henceforth use the simplified notation $\mathbf{z}^{(i)}$ to denote an arbitrary dissimilarity vector and refer to dissimilarity vectors that represent genuine signatures and forgeries as being *positive* and *negative* respectively. Furthermore, let $\mathbf{Z}^+ = \{\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(N^+)}\}$ denote the set that contains the $N^+ = \frac{NK\Omega}{2}$ positive dissimilarity vectors obtained from *all* the guinea pig writers. Similarly, \mathbf{Z}^- contains the negative dissimilarity vectors obtained from said writers.

5.2.1 Outlier removal

The sets \mathbf{Z}^+ and \mathbf{Z}^- provide a sufficient platform from which a decision boundary may be obtained for verification purposes. However, the potential inclusion of *outliers*, that is the inclusion of positive and negative samples that differ substantially from their respective class means, may yield a sub-optimal decision boundary as a result of model over-fitting. Furthermore, as mentioned in Section 4.4.2, the efficacy of the method considered for SVM parameter estimation may be severely compromised by the presence of outliers in the training data.

In order to address this issue, we propose the incorporation of an IOR algorithm prior to model training. This process ensures that only those samples that are sufficiently representative of positive and negative instances across all prospective writers are used for model construction. The IOR algorithm for samples in \mathbf{Z}^+ proceeds as follows:

1. Determine the *composite* mean $\boldsymbol{\mu}^+$ and standard deviation $\boldsymbol{\sigma}^+$ from the *feature-specific* means and standard deviations as follows,

$$\boldsymbol{\mu}^+ = \{\mu(\mathbf{Z}_1^+), \mu(\mathbf{Z}_2^+), \dots, \mu(\mathbf{Z}_T^+)\}, \quad (5.1)$$

$$\boldsymbol{\sigma}^+ = \{\sigma(\mathbf{Z}_1^+), \sigma(\mathbf{Z}_2^+), \dots, \sigma(\mathbf{Z}_T^+)\}, \quad (5.2)$$

where, $\mu(\mathbf{f})$ and $\sigma(\mathbf{f})$ respectively denote the mean and standard deviation of an arbitrary vector \mathbf{f} , whilst

$$\mathbf{Z}_t^+ = \{z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(N^+)}\} \quad (5.3)$$

denotes the vector that contains all the t^{th} feature values of all the positive dissimilarity vectors.

2. Locate $\mathbf{z}^{(\oplus)} \in \mathbf{Z}^+$, that is the positive dissimilarity vector which is furthest from $\boldsymbol{\mu}^+$ based on the Euclidean distance $\|\mathbf{z}^{(i)} - \boldsymbol{\mu}^+\|$ for $i = 1, 2, \dots, N^+$.

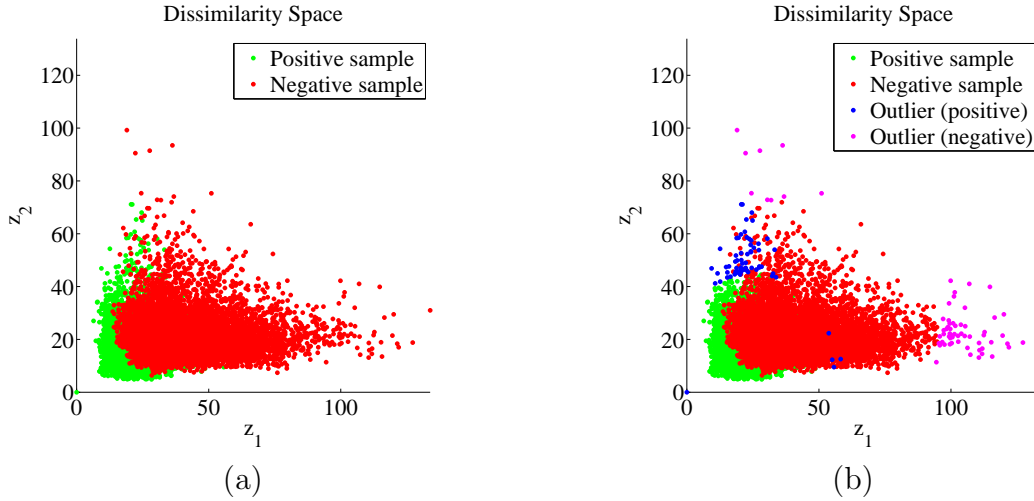


Figure 5.1: (a) Typical representation of negative samples *superimposed* onto positive samples in dissimilarity space for $T = 2$, $N = 40$ and $K = 10$. (b) Outliers detected (and subsequently removed) by the IOR algorithm.

If $\|z^{(\oplus)} - \mu^+\| \leq 3\|\sigma^+\|$, terminate the algorithm. Otherwise, remove $z^{(\oplus)}$ from Z^+ . Similarly, also locate and remove the least representative *negative* dissimilarity vector $z^{(\ominus)}$ from Z^- , regardless of its distance to μ^- , and return to Step 1.

A similar algorithm is implemented for samples in Z^- . Note that upon the detection and subsequent removal of an outlier, the least representative sample from the opposite class is *also* removed without question. This ensures that the sets Z^+ and Z^- remain balanced. Also note that only one such outlier-pair is removed during a single iteration, after which the class statistics μ and σ are recalculated. This is to prevent excessive outliers from yielding misrepresentative class statistics and therefore ensures that no samples are removed unnecessarily. A graphical illustration of the IOR algorithm, after being applied to typical examples of Z^+ and Z^- , is presented in Figure 5.1.

Apart from improving the reliability of the SVM parameter estimation process by removing potentially misrepresentative training samples, the rationale behind this outlier removal strategy is as follows:

- The inclusion of positive samples that differ very little from their associated reference signatures would cause the trained system to be biased toward rejection.
- Similarly, the inclusion of negative samples that differ substantially from their associated reference signatures would yield a model biased toward acceptance.

By discarding these extreme samples prior to training, a decision boundary may be found that optimally discriminates between positive and negative instances located in the region of uncertainty i.e. the region in dissimilarity space where Z^+ and Z^- become non-separable.

5.2.2 Dissimilarity normalisation

Despite the IOR algorithm's efforts to improve class separability, a significant degree of overlap between the positive and negative classes is still expected in dissimilarity space (see Figure 5.1 (b)), since the dissimilarity vectors in \mathbf{Z}^- represent *skilled* forgeries that are not easily distinguishable from the genuine signature samples represented in \mathbf{Z}^+ . In order to maximally separate these sets, each vector in \mathbf{Z}^+ and \mathbf{Z}^- requires appropriate normalisation.

A wide variety of suitable normalisation techniques are documented in e.g. Snelick *et al.* (2005) and Jain *et al.* (2005). Popular strategies include the *min-max*, *z-score* and *sigmoid* normalisation functions. One such sigmoid function is the well-known *logistic function*, defined for $x \in \mathfrak{R}$ as

$$L(x) = [1 + \exp(-x)]^{-1}. \quad (5.4)$$

This monotonically increasing scalar function maps any value in the domain $x \in (-\infty, 0)$ to a value in the range $L(x) \in (0, 0.5)$, whilst $x \in (0, \infty)$ is mapped to $L(x) \in (0.5, 1)$ and $L(0) = 0.5$. Note that, although $L(x)$ is defined for all $x \in \mathfrak{R}$, it is generally deemed sufficient to consider the diminished domain $x \in [-6, 6]$ for numerical applications, as illustrated in Figure 5.2 (a). We henceforth refer to the value of x for which $L(x) = 0.5$ as the *critical value*.

Since the dissimilarity values yielded by the DTW-based dichotomy transformation (see Section 3.4) are restricted to the domain $x \in [0, \infty)$, the systems developed in this study perform dissimilarity normalisation using a *modified* logistic function, defined as

$$\eta(x, c) = \left[1 + \exp\left(c - \frac{6x}{c}\right) \right]^{-1}, \quad (5.5)$$

where $c \in \mathfrak{R}$ denotes a predefined *scaling factor*. This modification shifts and rescales the numerically significant domain of the conventional logistic function from $x \in [-6, 6]$ to $x \in [0, 2c]$. Consequently, the critical value is also relocated from $x = 0$ to $x = c$, as illustrated in Figure 5.2 (b).

The normalisation function defined in (5.5) may subsequently be used to convert any T -dimensional dissimilarity vector $\mathbf{z} = \{z_1, z_2, \dots, z_T\}$ into a normalised dissimilarity vector $\bar{\mathbf{z}}$ as follows,

$$\bar{\mathbf{z}} = \bigcup_{t=1}^T \eta(z_t, \mu_t + \sigma_t), \quad (5.6)$$

$$\mu_t = \frac{1}{N^+} \sum_{i=1}^{N^+} z_t^{(i)}, \quad (5.7)$$

$$\sigma_t = \sqrt{\frac{1}{N^+ - 1} \sum_{i=1}^{N^+} (z_t^{(i)} - \mu_t)^2}. \quad (5.8)$$

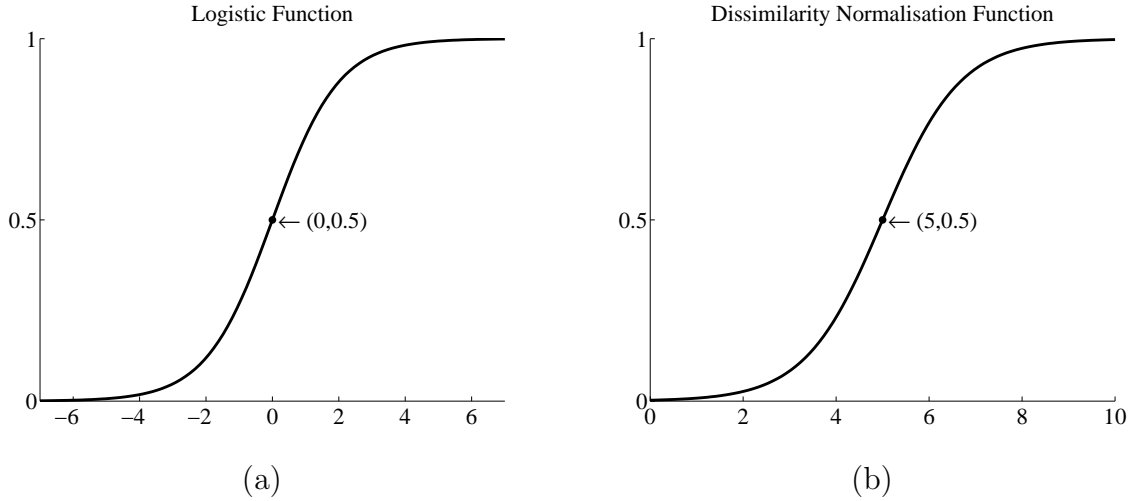


Figure 5.2: (a) The conventional logistic function. (b) The dissimilarity normalisation function utilised by the systems developed in this study for $c = 5$.

This *global dissimilarity normalisation* strategy is illustrated in Figure 5.3, whilst a typical result yielded by this method (for a scenario where $T = 2$) is presented in Figure 5.5 (a). It is clear that this strategy successfully rescales any arbitrary dissimilarity value to the interval $[0, 1]$. However, since the normalisation function is monotonic, it is unable to successfully address the overlap of \mathbf{Z}^+ and \mathbf{Z}^- .

In order to address this shortcoming of the global normalisation strategy, we propose the incorporation of dissimilarity statistics into the normalisation function on a *writer-specific* level. For every writer ω , the statistics $\boldsymbol{\mu}^{(\omega)}$ and $\boldsymbol{\sigma}^{(\omega)}$ are estimated by considering *only* the $N^{(\omega)} = \frac{K^2 - K}{2}$ unique dissimilarity vectors generated when every reference signature belonging to writer ω is compared to every other reference signature belonging to the *same* writer. Any dissimilarity vector, that is obtained by using a reference sample belonging to writer ω , is subsequently normalised using a writer-specific normalisation function as follows,

$$\bar{\mathbf{z}} = \bigcup_{t=1}^T \eta(z_t, \mu_t^{(\omega)} + \sigma_t^{(\omega)}), \quad (5.9)$$

$$\mu_t^{(\omega)} = \frac{1}{N^{(\omega)}} \sum_{\substack{i=1 \\ j>i}}^K z_t^{(i,j)}, \quad (5.10)$$

$$\sigma_t^{(\omega)} = \sqrt{\frac{1}{N^{(\omega)} - 1} \sum_{\substack{i=1 \\ j>i}}^K \left(z_t^{(i,j)} - \mu_t^{(\omega)} \right)^2}. \quad (5.11)$$

This novel *writer-specific dissimilarity normalisation* strategy is illustrated in Figure 5.4, whilst a typical result yielded by this method (for a scenario where $T = 2$) is presented in Figure 5.5 (b).

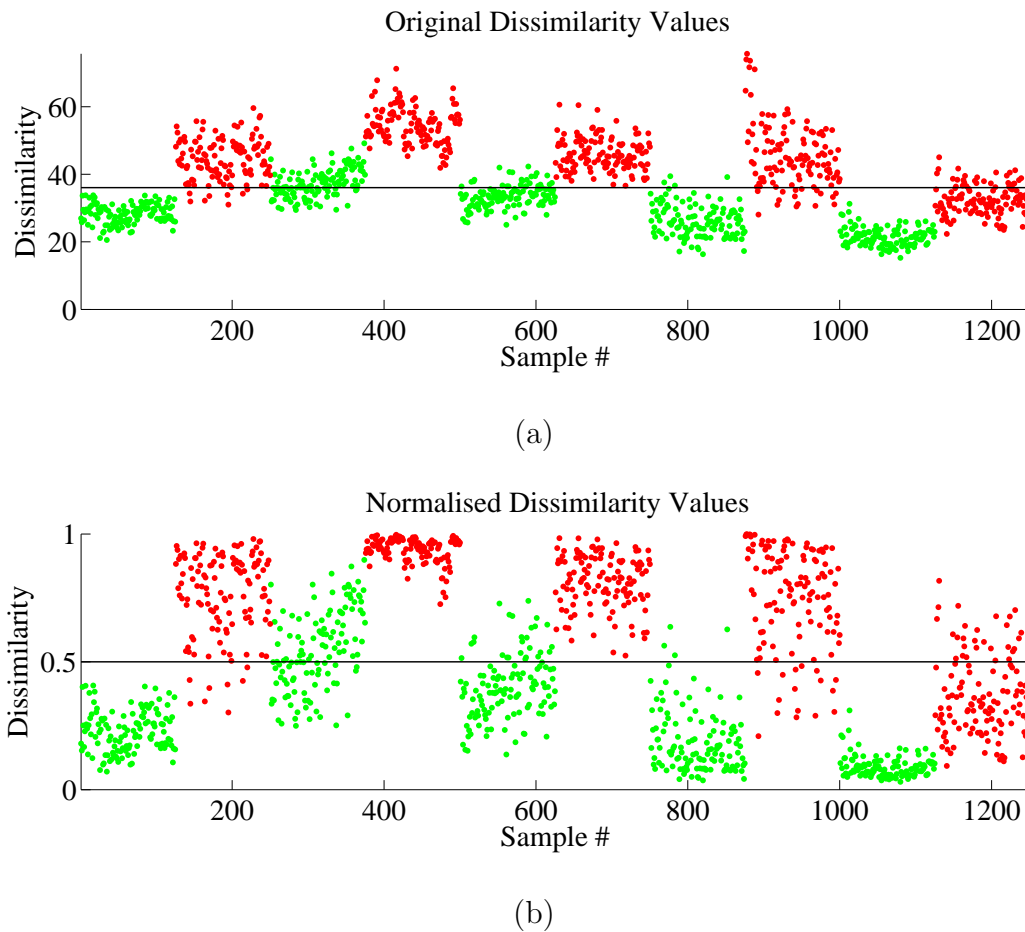


Figure 5.3: The *global* dissimilarity normalisation strategy. (a) Dissimilarity values representative of positive and negative samples obtained from five different writers. Also included is the critical value $z = \mu + \sigma$, as obtained from *all* the samples. (b) Normalised dissimilarity values obtained using the *global* normalisation function $\eta(z, \mu + \sigma)$, as well as the mapped critical value $\eta(\mu + \sigma, \mu + \sigma) = 0.5$. Note that, although the original dissimilarity values have been successfully rescaled, no improvement is observed in terms of class separability.

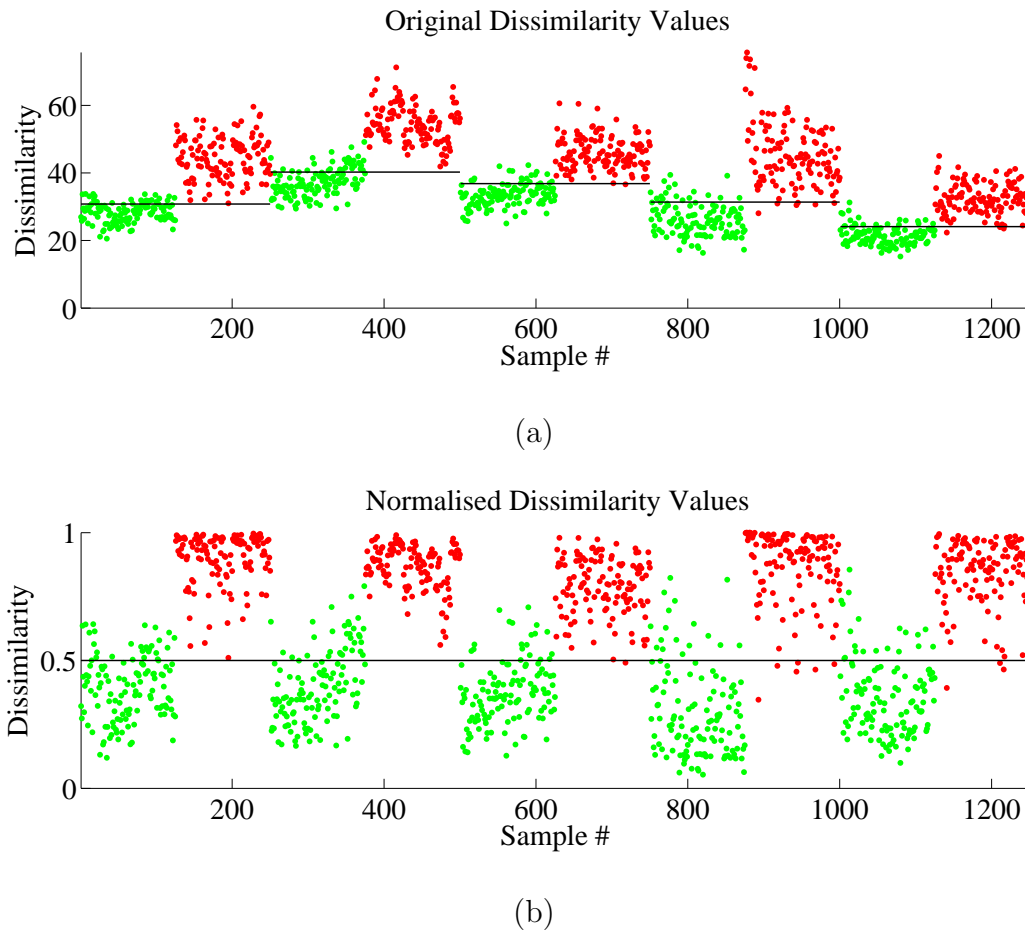


Figure 5.4: The *writer-specific* dissimilarity normalisation strategy. (a) Dissimilarity values as also depicted in Figure 5.3 (a). Also included are the writer-specific critical values $z = \mu^{(\omega)} + \sigma^{(\omega)}$ for $\omega = 1, 2, \dots, 5$, as obtained from the samples belonging to writer ω only. (b) Normalised dissimilarity values obtained using the *writer-specific* normalisation function $\eta(z, \mu^{(\omega)} + \sigma^{(\omega)})$ for each writer separately, as well as their associated mapped critical values $\eta(\mu^{(\omega)} + \sigma^{(\omega)}, \mu^{(\omega)} + \sigma^{(\omega)}) = 0.5$. Note that this strategy not only successfully rescales the original dissimilarity values, but also significantly improves overall class separability when compared to the results illustrated in Figure 5.3 (b).

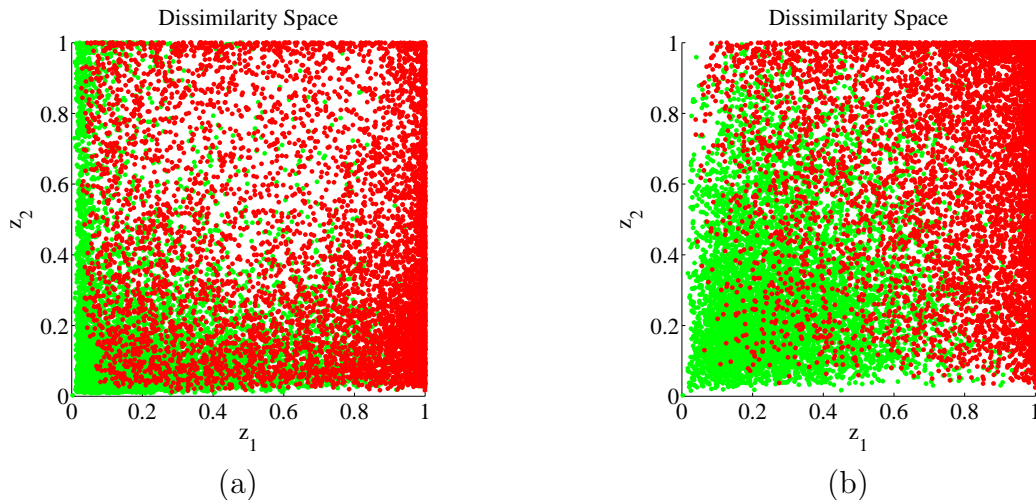


Figure 5.5: Comparison of (a) the *global* dissimilarity normalisation strategy and (b) the *writer-specific* dissimilarity normalisation strategy, when applied to the *retained samples* depicted in Figure 5.1 (b).

It is clear from Figures 5.3–5.5 that when the proposed writer-specific dissimilarity normalisation strategy is employed, the resulting separation of positive and negative samples for each *individual* writer leads to improved global separation of positive and negative samples across the *entire* set of writers, since only strictly relevant information is used during the normalisation process.

5.2.3 Model training

Following the successful completion of the IOR and dissimilarity normalisation algorithms, the sets $\bar{\mathbf{Z}}^+$ and $\bar{\mathbf{Z}}^-$ are deemed fit for consideration as training data for the construction of a writer-independent signature model.

In this section we explain how the efforts of the QDA-based and R-SVM-based classification techniques (discussed in the previous chapter) may be harnessed in order to determine an optimal decision boundary between $\bar{\mathbf{Z}}^+$ and $\bar{\mathbf{Z}}^-$. This decision boundary, which in essence constitutes the signature model λ , is retained for the subsequent verification of questioned signatures.

QDA-based systems

The QDA-based class membership function is completely described by the mean vectors and covariance matrices that are representative of the positive and negative classes. The estimation of these class statistics from the training data, using (4.5) and (4.6), therefore constitutes the entire model construction process.

Consequently, the statistics $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}^-$ are estimated from $\bar{\mathbf{Z}}^+$ and $\bar{\mathbf{Z}}^-$ respectively. Furthermore, since the QDA-based approach incorporates class specific covariances, the

statistics Σ^+ and Σ^- are also separately estimated from $\bar{\mathbf{Z}}^+$ and $\bar{\mathbf{Z}}^-$ respectively. This process yields the QDA-based model parameter set

$$\lambda_{\text{QDA}} = \{\boldsymbol{\mu}^+, \boldsymbol{\mu}^-, \Sigma^+, \Sigma^-\}, \quad (5.12)$$

which is retained for the future verification of a questioned signature sample, as discussed in Section 5.3.

R-SVM-based systems

The R-SVM-based class membership function is described by several parameters. Whilst the weight vector and bias are estimated from the training data by means of an appropriate training algorithm, the specification of a predetermined regularisation parameter and kernel width is required prior to training.

In the previous chapter we discussed a heuristic method for the estimation of suitable values for the internal parameters C and γ . After calculating the pairwise distances between all the samples in $\bar{\mathbf{Z}}^+ \cup \bar{\mathbf{Z}}^-$, we set C and γ equal to the maximum and median, respectively, of these calculated distances. The *sequential minimal optimisation* (SMO) algorithm is subsequently used to find optimal values for the weight vector \mathbf{w} and bias b . This process yields the R-SVM-based model parameter set

$$\lambda_{\text{R-SVM}} = \{\mathbf{w}, b, C, \gamma\}, \quad (5.13)$$

which is retained for the future verification of a questioned signature sample, as discussed in the next section.

5.3 Verification

Following the successful completion of the model training process, the resulting writer-independent signature model may be used to verify the authenticity of any subsequent questioned signature sample.

When presented with such a questioned signature sample $S_q^{(\omega)}$, that is a sample of unknown origin and claimed to belong to writer ω , the systems developed in this study perform the task of signature verification as follows:

1. The raw data obtained from $S_q^{(\omega)}$ is first converted into the appropriate *feature set*.
2. This feature set is subsequently compared to those extracted from each of the K reference signatures belonging to writer ω , in order to produce a set of *dissimilarity vectors* $\mathbf{z}_q^{(i)}$ for $i = 1, 2, \dots, K$. This reference set is also used to obtain the dissimilarity statistics $\boldsymbol{\mu}^{(\omega)}$ and $\boldsymbol{\sigma}^{(\omega)}$. These statistics are then used to convert each dissimilarity vector into a *normalised* dissimilarity vector $\bar{\mathbf{z}}_q^{(i)}$, which is suitable for consideration by the trained model.

3. Each normalised dissimilarity vector is presented to the model individually, in order to obtain a signed distance measure $D^{(i)} \in \mathfrak{R}$ relative to the corresponding decision boundary in dissimilarity space, such that

$$D^{(i)} = f(\bar{\mathbf{z}}^{(i)}, \lambda), \quad (5.14)$$

where $f(\cdot, \lambda)$ denotes the QDA-based or R-SVM-based class membership function, as described by the parameter set associated with the trained model λ .

4. Each distance measure is then converted into a *partial confidence score* $s^{(i)} \in [0, 1]$, using the conventional logistic function, such that

$$s^{(i)} = [1 + \exp(-D^{(i)})]^{-1}. \quad (5.15)$$

The entire set of partial confidence scores is then averaged, which produces the *final confidence score* $s^* \in [0, 1]$, such that

$$s^* = \frac{1}{K} \sum_{i=1}^K s^{(i)}. \quad (5.16)$$

5. Finally, a global¹ threshold $\tau \in [0, 1]$ is imposed on s^* to predict class membership, such that $S_q^{(\omega)}$ is accepted as genuine if and only if $s^* \geq \tau$.

The verification protocol described above is an adaptation of the *questioned document expert's approach*, as discussed in Santos *et al.* (2004). This approach requires any questioned signature sample to be compared to multiple known genuine reference samples, whereafter the resulting dissimilarity vectors are presented to a trained signature model. This is of course identical to the approach considered in this study. Notably, however, the traditional expert's approach subsequently uses the model output (in this case a distance measure relative to the decision boundary) to obtain partial *decisions*. The final decision regarding the authenticity of the questioned sample is obtained through fusion of the partial decisions, usually by means of a majority vote.

In this study we elect to use the distance measures yielded by the model to obtain *confidence scores* instead of decisions, which in turn facilitates the use of score-level fusion instead of decision-level fusion. The rationale behind this approach is as follows:

- Firstly, this strategy produces continuous classifier output, which adds to the robustness of the resulting verification system in that the system may easily adapt to different operating conditions (Coetzer *et al.* (2012)).

¹Although the same threshold value is considered for all questioned samples, claimed to belong to *any* writer, this value is imposed on distance measures calculated from *normalised* dissimilarity vectors. When utilising the writer-specific dissimilarity normalisation strategy, the threshold value effectively also becomes writer-specific.

- More importantly, however, the use of partial confidence scores also enables the final decision-making process to take into consideration the level of certainty associated with each individual classification.

We illustrate this maximal exploitation of the model output in the following example.

Example

Consider a hypothetical scenario where a questioned signature sample, claimed to belong to a specific client who provided three genuine reference samples during an initial enrolment phase, is presented for authentication. Suppose that, following the presentation of the corresponding dissimilarity vectors to the trained model, the following set of partial confidence scores $\{0.48, 0.49, 0.86\}$ is obtained. By imposing a threshold of 0.50, the traditional expert's approach will yield the following partial decisions $\{0, 0, 1\}$, which will result in a final (majority vote) decision of 0, which constitutes a rejection. In contrast, the approach proposed in this study will first determine the final confidence score $\frac{0.48+0.49+0.86}{3} = 0.61$. By again imposing a threshold of 0.50, this will result in a final decision of 1, which constitutes an acceptance.

In the scenario described above, the *acceptance* of the hypothetical questioned sample is considered the more sensible result, since *only* the partial acceptance is made with any degree of certainty, whilst the confidence scores associated with the two partial rejections are approximately equal to 0.50, which indicates total ambiguity. It is furthermore very unlikely that an amateur forger would be able to produce a sample associated with a (very high) partial confidence score of 0.86.

5.4 Concluding remarks

In this chapter we discussed how the efforts of a dissimilarity-based signature representation protocol and an appropriate classifier may be combined, in order to construct a writer-independent signature model, which may be used to verify the authenticity of subsequently presented questioned signature samples. We also introduced an algorithm for efficient outlier detection and removal, which aids in the construction of a robust model.

Most notably, we proposed a novel dissimilarity normalisation strategy, that incorporates writer-specific information into the writer-independent modelling framework, and showed that this technique is capable of significantly improving class separability prior to model construction. It is worth noting that, despite the fact that the proposed strategy utilises writer-dependent information to improve the separability of the positive and negative classes, the resulting model constructed from these classes remains completely writer-independent. Only the set of K positive reference samples is considered when the writer-specific normalisation statistics are determined. It should also be noted that the proposed approach to dissimilarity normalisation is only feasible if $K > 1$. For $K = 1$, sensible values for $\mu^{(\omega)}$ and $\sigma^{(\omega)}$ can not be obtained and one is forced to employ the global

normalisation strategy. The contribution of this novel technique to the field of handwritten signature analysis is verified experimentally in the next chapter.

Finally, in this chapter we also discussed a suitable verification protocol for questioned signature authentication, thereby completing the design of the systems developed in this study. From this discussion it should be clear that, although the QDA-based and R-SVM-based methods for obtaining the optimal decision boundary in dissimilarity space are fundamentally different, the underlying verification protocols utilised by the QDA-based and R-SVM-based systems are identical.

In the next chapter we present a rigorous experimental evaluation of the verification proficiency expected from each of the systems developed in this study.

Chapter 6

Experiments

“Reasoning draws a conclusion, but does not make the conclusion certain, unless the mind discovers it by the path of experience.”
- Roger Bacon (1214–1294)

6.1 Introduction

In Chapters 3–5 we presented several pattern recognition techniques, that included a novel *dynamic time warping* (DTW) based dichotomy transformation for writer-independent off-line signature representation and a novel writer-specific dissimilarity normalisation strategy, which are geared towards the proficient writer-independent verification of handwritten signatures. In this chapter, we investigate the merits of these techniques by means of a rigorous experimental evaluation protocol.

This protocol is implemented by considering four different signature corpora that contain two fundamentally different types of signature data, that is those containing off-line and on-line signature samples. In Section 6.2 we discuss the experimental setup, which includes the data sets and experimental protocol considered for system evaluation. In Section 6.3 we report and discuss the experimental results. In Section 6.4 we compare the results reported in this chapter with those reported in the literature, in order to place the performance of the systems developed in this study into perspective.

Finally, in Section 6.5 we ascertain the significance of the novel techniques proposed in this study, in order to determine their contribution to the current state of the art.

6.2 Experimental setup

Unfortunately, there exists at present no single, universally accepted signature data corpus nor an associated experimental protocol for the purpose of global system performance benchmarking. As a result, system performance metrics reported in the literature only serve as a general indication of expected proficiency and should therefore not be considered absolute. Even when the *same* data set is considered for system evaluation, a *direct*

comparison of the performance reported for different systems proposed in the literature may not be sensible, unless the experimental protocols considered to obtain these performance estimations are *also* identical.

Furthermore, since experimental conditions associated with system performance estimation may vary arbitrarily, it becomes of the utmost importance that researchers devise their experimental setups in such a way that *credible* performance estimations are obtained. In this section we discuss the experimental setup considered for the evaluation of the systems developed in this study.

6.2.1 Data

In order to ensure the credibility of the results reported in this chapter, system evaluations are performed using *four* separate data sets. Two of these data sets are used for proficiency testing of the *off-line* verification systems developed in this study, whilst the other two data sets are considered for the evaluation of the *on-line* systems.

Experiments relating to off-line signature verification are conducted using *Dolfing's data set* (Coetzer (2005)) and the *MCYT-SignatureOff-75 subcorpus*, whilst experiments relating to on-line signature verification are conducted using the *Philips database* (Dolfing *et al.* (1998)) and the *MCYT-Signature-100 subcorpus*. Both MCYT subcorpora were obtained from the *MCYT bimodal baseline corpus* (Ortega-Garcia *et al.* (2003)).

It is worth noting that the data sets considered in this study represent only a small subset of the wide variety of well-known signature databases documented in the literature. Notable examples include the *CEDAR signature database* (Kalera *et al.* (2004)), the *SVC2004 Task 2 data set* (Yeung *et al.* (2004)), the *BIOMET signature subcorpus* (Garcia-Salicetti *et al.* (2007)), as well as the *MyIdea signature subcorpus* (Dumas *et al.* (2005)). Countless other data sets are also documented in the literature, although these (generally smaller) signature databases are not necessarily made available to the research community. Perhaps the largest, most well-known signature database, however, is the *GPDS960signature database* (Vargas *et al.* (2007)), which contains 51840 off-line samples obtained from 960 writers. This database has been considered extensively by countless researchers the world over and is therefore well-suited for performance benchmarking. However, due to privacy concerns, the GPDS960 corpus is no longer available to the research community and has recently been replaced by the *GPDSsyntheticSignature database* (Ferrer *et al.* (2012)).

We now present detailed discussions on the nature and composition of the data sets considered *in this study*.

Philips database / Dolfing's data set

The *Philips database* contains 4800 *on-line* signature samples from 51 writers and comprises 1530 genuine signatures, 3000 amateur skilled forgeries and 270 professional skilled forgeries. The amateur skilled forgeries may further be sub-categorised into so-called *home-improved* (1530 samples) and *over-the-shoulder* (1470 samples) forgeries. The home-improved forgeries were produced by forgers who had in their possession an *off-line* sample



Figure 6.1: Selected samples from Dolfig's data set. Note that all the signatures depicted here have the same uniform stroke width. Since the original images have been rescaled for improved representation, the aforementioned property is not always clearly visible.

of an authentic signature, as well as ample time to practice its reproduction. The over-the-shoulder forgeries were produced by forgers who witnessed a legitimate signing event and then attempted to reproduce the signature immediately afterwards. Since we do not consider professional forgeries in this study, 30 genuine signatures and 60 forgeries are therefore available per writer, with the exception of two writers, for whom only 30 forgeries are available for experimental purposes.

The signature samples were acquired by means of a Philips Advanced Interactive Display with a sampling frequency of 160Hz. At each sample point, this device measures the horizontal and vertical pen tip position, the axial pen pressure, as well as the pen orientation (in terms of the azimuth and altitude angles) relative to the writing surface.

The composition of *Dolfig's data set* is identical to that of the Philips database, since the *off-line* samples represented in Dolfig's data set have in fact been *artificially constructed* from the *on-line* signature data contained in the Philips database. Although these signatures were originally captured on-line, it is shown in Coetzer (2005) that each sample may easily be converted into a suitable off-line¹ representation. A detailed discussion of the conversion algorithm can be found in Coetzer (2005), whilst Figure 6.1 illustrates typical images yielded by the conversion process.

Despite the fact that this data set has not been obtained through the traditional off-line acquisition method, namely the digitisation of *pen-on-paper* signatures by means of a scan-

¹An off-line signature sample yielded by this conversion process is said to be *ideal*, since the image is completely free of background noise, whilst a uniform pen stroke width is also guaranteed.

ning device, it remains suitable for the evaluation of the systems developed in this study. Furthermore, there is a sound argument that the use of ideal data for such an evaluation is in fact *optimal*, since document image pre-processing is not considered a primary objective of this study. The use of ideal off-line signature data circumvents the issue of inadequate document image pre-processing and therefore allows one to *independently* investigate the significance of the signature modelling techniques proposed in this study under optimal operating conditions.

Nevertheless, in order to ascertain the system performance expected in a realistic deployment scenario, we also consider an off-line data set that was obtained through *traditional* acquisition methods, as is discussed in the next section.

MCYT subcorpora

As previously mentioned, both the MCYT-Signature-100 and MCYT-SignatureOff-75 subcorpora constitute randomly selected subsets of the much larger MCYT bimodal baseline corpus – a biometric database that contains fingerprint and signature data from 330 contributors across 4 different Spanish sites. Note that, in the interest of brevity, we henceforth refer to the MCYT-Signature-100 subcorpus and the MCYT-SignatureOff-75 subcorpus simply as MCYT-100 and MCYT-75 respectively.

In order to construct these signature databases, a set of genuine and forged samples was collected for each contributing writer over several sessions. During each session, each writer was instructed to produce five samples of his/her own signature, as well as five imitations of another writer's signature. In order to produce these imitations, the writer was supplied with several off-line samples of a genuine signature, as well as ample time to practise the forgery thereof. As a result, $5N_s$ genuine signatures and $5N_s$ forgeries (produced by N_s different forgers) were collected for each writer, where N_s denotes the number of sessions considered. MCYT-100 and MCYT-75 contain samples that were captured over five and three sessions respectively. As a result, MCYT-100 contains 5000 samples (2500 genuine signatures and 2500 amateur skilled forgeries) from 100 writers, whilst MCYT-75 contains 2250 samples (1125 genuine signatures and 1125 amateur skilled forgeries) from 75 writers.

The data collection protocol considered in the MCYT project facilitated the simultaneous acquisition of on-line *and* off-line signature samples. This was achieved by placing a paper template over a pen tablet and subsequently capturing the signature by means of an inking pen. The device, a WACOM Intous A6 USB pen tablet with a sampling frequency of 100Hz, measured the horizontal and vertical pen tip positions, the axial pen pressure, as well as the pen orientation relative to the writing surface. The on-line signature acquisition process was completed automatically by the capturing device. In order to complete the off-line signature acquisition process, the paper template was digitised using a flatbed scanner with a resolution of 600dpi. Several examples of signature images yielded by this process are presented in Figures 6.2–6.4.

Unlike the ideal off-line samples contained in Dolfing's data set (see Figure 6.1), these images have therefore been obtained using the traditional off-line signature acquisition method. As a result, in order to maximise the efficacy of the signature modelling and ver-



Figure 6.2: Examples of typical signature images contained in MCYT-75.

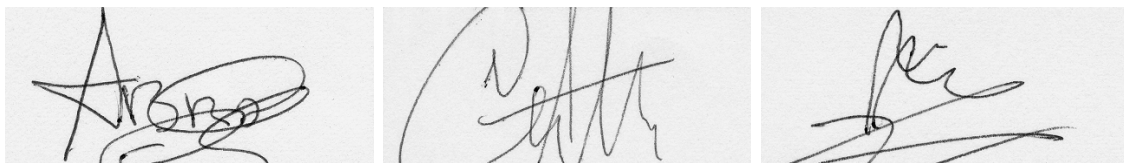


Figure 6.3: Examples of partial signature images contained in MCYT-75. These incomplete samples, a result of image *cropping*, were most likely obtained as a result of writers who signed outside a designated signing area.

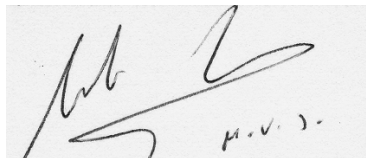


Figure 6.4: Example of an incorrectly extracted signature image contained in MCYT-75. The dark region visible on the right is in all likelihood a result of incorrect page positioning during the digitisation process.

ification processes, the samples contained in MCYT-75 require the efforts of several image pre-processing techniques prior to model construction and authentication. As mentioned in Section 3.2.1, these pre-processing stages are primarily concerned with image binarisation, noise reduction and signature segmentation. The successful completion of these image processing techniques results in enhancements that compensate for variations in ink intensity and the presence of ink residue, as is evident in Figures 6.2–6.4.

However, the digitisation process described earlier also yields several *partial* signature images (see Figure 6.3), as well as *incorrectly extracted* signature images (see Figure 6.4). Since the systems developed in this study do not currently employ any pre-processing techniques capable of addressing these issues, it is reasonable to expect that these problematic samples will hinder the successful completion of the signature image analysis stage and consequently impede system performance. Although the incorporation of such additional image processing techniques is considered outside the scope of this study, *all* the samples

contained in MCYT-75 are considered during system performance evaluation.

6.2.2 Protocol

In order to ensure that the results reported in this study represent a *comprehensive* and *unbiased* estimation of system performance, the experimental protocol incorporates both *k-fold cross-validation* and *n-fold data randomisation*, and proceeds as follows:

- *k*-fold cross-validation:

Given a data set that contains samples from Ω writers, the data set is first partitioned into k equal subsets. Each subset, in turn, is used as an *evaluation set* representative of $\frac{\Omega}{k}$ writers, whilst the samples from the remaining $\frac{\Omega(k-1)}{k}$ writers constitute the *training set*. Each of these individual evaluations is referred to as a single *run*, whilst the entire set of k runs constitutes a single *trial*.

This process ensures that signature samples from *different* writers are used for the purposes of model training and evaluation, thereby avoiding a potentially over-fitted² performance estimate, whilst still ensuring that *all* the writers represented in the data set are considered for evaluation.

Each system evaluation reported in this chapter involves $k = 3$ runs per trial.

- *n*-fold data randomisation:

The cross-validation process described above is repeated n times. For each trial, the *order* of the writers is *randomised* prior to data set partitioning.

It should be clear that each run yields a performance estimate that is obtained by considering a relatively small subset of the signature corpus. It is therefore entirely possible that any discrepancy in terms of the quality of the forgeries contained in the evaluation set, as compared to those contained in the training set, may potentially result in misrepresentative performance estimates. Consider, for example, the scenario where a model is trained using predominantly high-quality forgeries and is subsequently evaluated using predominantly low-quality forgeries. Such an evaluation would yield an optimistically biased performance estimate. In contrast, a model trained using low-quality forgeries and evaluated using high-quality forgeries would indicate unrealistically poor performance.

When repeated a sufficient number of times, this randomised allocation of writers to the training and evaluation sets therefore nullifies the influence of *outliers* – in this case writers whose inclusion into the evaluation set results in atypically high or low performance estimates.

Each system evaluation reported in this chapter involves $n = 100$ trials.

²An over-fitted model may indicate promising performance during system development, but lack generalisation potential during system deployment.

Table 6.1: Summary of the data set partitioning considered in this study, including the number of writers in the data set (Ω), the number of writers in the training set (Ω_T), the number of writers in the evaluation set (Ω_E), the number of genuine samples per writer (N_ω^+), as well as the number of forged samples per writer (N_ω^-).

Data set	Ω	Ω_T	Ω_E	N_ω^+	N_ω^-
Dolfing's data set / Philips database	51	34	17	30	60
MCYT-SignatureOff-75 subcorpus	75	50	25	15	15
MCYT-Signature-100 subcorpus	100	67	33	25	25

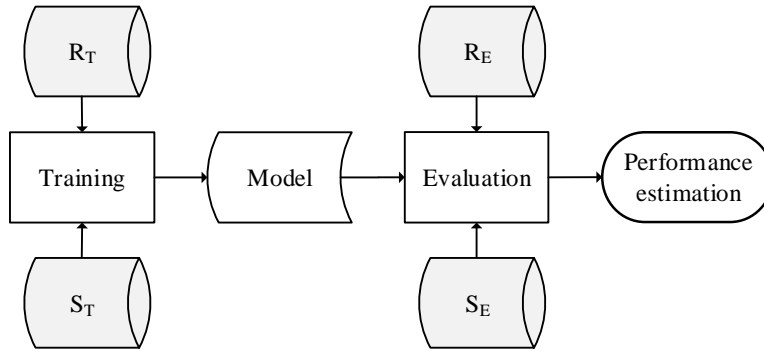


Figure 6.5: Schematic representation of the experimental protocol utilised to estimate system performance for a *single* run within a *single* trial.

We now present a detailed discussion of the experimental protocol associated with a *single* run within a *single* trial. Note that, prior to experimentation, the data set is first partitioned into two disjoint subsets, namely a *training set* and an *evaluation set*, as summarised in Table 6.1.

During model training, *only* the training set is used. For every writer, K genuine signatures are reserved for inclusion in the reference set R_T , whilst $N_\omega^+ - K$ genuine signatures and $N_\omega^- - K$ forgeries constitute the set S_T considered for training. As a result, $N_\omega^- - N_\omega^+ + K$ forgeries per writer remain unused, in order to ensure balanced training data. The set of K reference signatures is used to obtain $K(N_\omega^+ - K)$ positive and $K(N_\omega^- - K)$ negative dissimilarity vectors. The entire set of *normalised* positive and negative dissimilarity vectors, obtained from all Ω_T writers in the training set, is used to determine the optimal decision boundary, which is retained for subsequent verification.

During model evaluation, *only* the evaluation set is used. For every writer, K genuine signatures are once again reserved for the reference set R_E , whilst $N_\omega^+ - K$ genuine signatures and N_ω^- forgeries constitute the set S_E considered for verification. The entire set of genuine signatures and forgeries, obtained from all Ω_E writers in the evaluation set, is used to gauge system performance. This protocol is conceptualised in Figure 6.5.

Table 6.2: The set of experimental parameter values considered for off-line system evaluation.

Parameter	Value(s)
d	128
T	{2, 4, 8, 16, 32, 64, 128, 256}
K	{5, 10, 15}

6.3 Results

Since the experimental protocol discussed in the previous section involves $k = 3$ runs per trial and $n = 100$ trials, the resulting performance estimates reported in this chapter are obtained from $kn = 300$ separate system evaluations. During each one of these evaluations, system performance is gauged using the *area under curve* (AUC) and *equal error rate* (EER) performance metrics, as discussed in Section 1.2.7.

In the interest of brevity, we henceforth use the acronyms QDS and SVMS to denote a *quadratic discriminant analysis* (QDA) based system and a *radial basis function* (RBF) kernel *support vector machine* (SVM) based system, respectively.

6.3.1 Off-line verification systems

In Section 3.2.2 we explained that, in order for the off-line systems developed in this study to convert a signature image into a *discrete Radon transform* (DRT) based feature set, it is required that predetermined values be specified for two system *hyperparameters*, namely the projection profile length d and the projection angle set size T . In addition to this, we also wish to investigate the relationship between system proficiency and the reference set size K , although it should be clear that this value does not technically represent a system hyperparameter, but rather a deployment parameter³. System performance estimates are therefore obtained for each possible combination of the experimental parameter values listed in Table 6.2.

In order to avoid an excessively verbose report on system performance, the optimal value for d was predetermined experimentally and consequently remains fixed at $d = 128$. For $d > 128$, we found that no significant increase in system proficiency is witnessed, whilst computational cost increases dramatically. Also note the manner in which the values considered for T are increased, namely $T = 2^n$ for $n = 1, 2, \dots, 8$. This systematic increase is an important property of the experimental setup, since it ensures that each higher resolution projection angle set includes at least *all* of the angles considered in *any* of the lower resolution sets. If this is not the case, e.g. when the values of T are increased incrementally, it is entirely possible that several significant projection angles, specifically those associated with highly discriminative projection profiles, may be included into (or

³The number of reference signatures required for successful writer enrolment may be specified arbitrarily by the deployment entity (e.g. a banking institution). A set of $K = 5$ reference signatures is deemed reasonable for a realistic deployment scenario, whilst larger reference set sizes are also considered for the purpose of performance comparisons with previously developed systems.

excluded from) certain projection angle sets purely by coincidence.

Let $\mu_{\text{AUC}}^{(K,T)}$ denote the mean AUC, achieved for the entire set of 300 system evaluations, for specific values of K and T . Also, let $\mu_{\text{AUC}}^{(K)}$ (or $\mu_{\text{AUC}}^{(T)}$) denote the mean AUC for a specific value of K (or T) that is averaged over all possible values of T (or K), as listed in Table 6.2. Finally, let μ_{AUC} denote the *overall* mean AUC, which represents the average system performance for all possible values of K and T . The performance metrics $\mu_{\text{EER}}^{(K,T)}$, $\mu_{\text{EER}}^{(K)}$, $\mu_{\text{EER}}^{(T)}$ and μ_{EER} are similarly defined in terms of the mean EER.

The mean AUC-based and mean EER-based performance estimates obtained for the QDS and SVMS, when evaluated on Dolfing’s data set and MCYT-75, are presented in Tables 6.3–6.6. From these results it is clear that, under optimal operating conditions, the SVMS outperforms the QDS. This is the case for both of the data sets considered. It is also clear that the system evaluations performed on Dolfing’s data set indicate a much higher level of system proficiency than those performed on MCYT-75. This is an expected result, since the former contains *ideal* signature images, as explained in Section 6.2.1. Since these two corpora contain very different types of off-line signature data, the system performance estimates obtained for Dolfing’s data set and MCYT-75 should not be analysed from the same perspective, but rather be seen as indicative of an *ideally attainable* and a *realistically expected* performance, respectively.

Furthermore, the performance metrics reported in Tables 6.3–6.6 also reveal several insights into the influence of the system parameters K and T . When either the QDS or the SVMS is considered, a strong correlation is evident between the reference set size K and the verification proficiency of the resulting system. This is a sensible result, since K determines both the representation potential of the writer-specific normalisation statistics (see Section 5.2.2), as well as the size of the partial scoring pool (see Section 5.3) – two key aspects of the signature modelling and authentication processes. However, the impact of the projection angle set size T on verification proficiency is not so straightforward, as is illustrated in Figure 6.6. For relatively small projection angle set sizes, that is for $T \in [2, 16]$, an increase in the number of projection angles consistently results in an improved system performance. This is the case for both the QDS and SVMS. As the projection angle set size is increased further, that is when $T > 16$, this upward trend in system performance continues for the SVMS, although the magnitude of the observed improvement gradually diminishes. In the case of the QDS, however, a notable increase in T results in a substantial *decrease* in verification proficiency.

It is suggested in Swanepoel and Coetzer (2012) that this decrease in the performance of the QDS is most likely due to data redundancy, since the influence of maximally discriminant projection profiles is reduced when a relatively expansive projection angle set is utilised. This paper was, however, published during the early stages of this study and, more importantly, prior to the development of the SVMS. Since we now find that *only* the QDS is severely impeded by notably increasing T , coupled with the fact that the utilisation of a larger projection angle set consistently *improves* the performance of the SVMS, this suggests that it is not the inclusion of non-essential projection profiles that impedes system performance, but rather an inherent high-dimensional modelling deficiency of the

Table 6.3: Average AUCs and EERs achieved by the QDS when evaluated on Dolfig’s data set. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.

$\mu_{\text{AUC}}^{(K,T)}$ (%)		K			$\mu_{\text{AUC}}^{(T)}$	$\mu_{\text{EER}}^{(K,T)}$ (%)		K			$\mu_{\text{EER}}^{(T)}$
		5	10	15				5	10	15	
T	2	97.67	98.64	98.80	98.37	T	2	8.05	5.81	5.50	6.45
	4	98.51	99.29	99.41	99.07		4	6.53	4.01	3.48	4.67
	8	98.81	99.43	99.50	99.25		8	5.55	3.45	3.19	4.06
	16	98.98	99.57	99.61	99.39		16	4.71	2.86	2.72	3.43
	32	98.91	99.58	99.62	99.37		32	5.12	2.88	2.76	3.58
	64	98.68	99.49	99.60	99.26		64	5.58	3.47	2.94	3.99
	128	97.44	98.94	99.20	98.53		128	7.95	5.03	4.18	5.72
	256	94.28	96.92	97.55	96.25		256	12.50	9.08	7.94	9.84
	$\mu_{\text{AUC}}^{(K)}$	97.91	98.98	99.16	98.69		$\mu_{\text{EER}}^{(K)}$	7.00	4.57	4.09	5.22

Table 6.4: Average AUCs and EERs achieved by the SVMs when evaluated on Dolfig’s data set. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.

$\mu_{\text{AUC}}^{(K,T)}$ (%)		K			$\mu_{\text{AUC}}^{(T)}$	$\mu_{\text{EER}}^{(K,T)}$ (%)		K			$\mu_{\text{EER}}^{(T)}$
		5	10	15				5	10	15	
T	2	97.25	98.50	98.65	98.14	T	2	8.47	6.11	5.78	6.79
	4	98.06	99.15	99.30	98.84		4	6.87	4.31	3.73	4.97
	8	98.23	99.28	99.42	98.98		8	6.06	3.65	3.39	4.36
	16	98.61	99.50	99.59	99.23		16	5.41	3.13	2.83	3.79
	32	98.72	99.52	99.61	99.28		32	5.08	3.00	2.65	3.58
	64	98.89	99.58	99.66	99.38		64	4.86	2.95	2.61	3.47
	128	98.99	99.61	99.66	99.42		128	4.52	2.89	2.52	3.31
	256	99.02	99.59	99.66	99.42		256	4.41	2.95	2.56	3.31
	$\mu_{\text{AUC}}^{(K)}$	98.47	99.34	99.44	99.09		$\mu_{\text{EER}}^{(K)}$	5.71	3.62	3.26	4.20

Table 6.5: Average AUCs and EERs achieved by the QDS when evaluated on MCYT-75. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.

$\mu_{\text{AUC}}^{(K,T)}$ (%)		K		$\mu_{\text{AUC}}^{(T)}$	$\mu_{\text{EER}}^{(K,T)}$ (%)		K		$\mu_{\text{EER}}^{(T)}$
		5	10				5	10	
T	2	90.63	92.74	<i>91.68</i>	T	2	16.76	15.18	<i>15.97</i>
	4	93.57	95.29	<i>94.43</i>		4	13.96	11.58	<i>12.77</i>
	8	94.85	96.46	<i>95.65</i>		8	11.81	9.50	<i>10.65</i>
	16	95.45	96.60	96.03		16	11.09	9.68	10.39
	32	95.39	96.24	<i>95.81</i>		32	11.36	10.41	<i>10.88</i>
	64	94.22	95.70	<i>94.96</i>		64	12.97	11.34	<i>12.15</i>
	128	92.08	94.68	<i>93.38</i>		128	14.36	12.44	<i>13.40</i>
	256	88.27	92.76	<i>90.51</i>		256	18.25	13.90	<i>16.08</i>
	$\mu_{\text{AUC}}^{(K)}$	<i>93.06</i>	<i>95.06</i>	<i>94.06</i>		$\mu_{\text{EER}}^{(K)}$	<i>13.82</i>	<i>11.75</i>	<i>12.79</i>

Table 6.6: Average AUCs and EERs achieved by the SVMs when evaluated on MCYT-75. The system performance estimates associated with the optimal number of projection angles, as indicated by $\mu_{\text{AUC}}^{(T)}$, are emphasised in bold.

$\mu_{\text{AUC}}^{(K,T)}$ (%)		K		$\mu_{\text{AUC}}^{(T)}$	$\mu_{\text{EER}}^{(K,T)}$ (%)		K		$\mu_{\text{EER}}^{(T)}$
		5	10				5	10	
T	2	89.70	92.15	<i>90.93</i>	T	2	17.54	15.66	<i>16.60</i>
	4	92.27	94.25	<i>93.26</i>		4	14.07	12.41	<i>13.24</i>
	8	94.00	95.76	<i>94.88</i>		8	12.45	10.26	<i>11.35</i>
	16	95.01	96.52	<i>95.77</i>		16	11.35	9.62	<i>10.49</i>
	32	95.43	96.68	<i>96.05</i>		32	10.97	9.54	<i>10.26</i>
	64	95.59	96.77	<i>96.18</i>		64	10.80	9.56	<i>10.18</i>
	128	95.73	96.69	<i>96.21</i>		128	10.54	9.42	<i>9.98</i>
	256	95.85	96.73	96.29		256	10.37	9.48	9.93
	$\mu_{\text{AUC}}^{(K)}$	<i>94.20</i>	<i>95.69</i>	<i>94.95</i>		$\mu_{\text{EER}}^{(K)}$	<i>12.26</i>	<i>10.75</i>	<i>11.50</i>

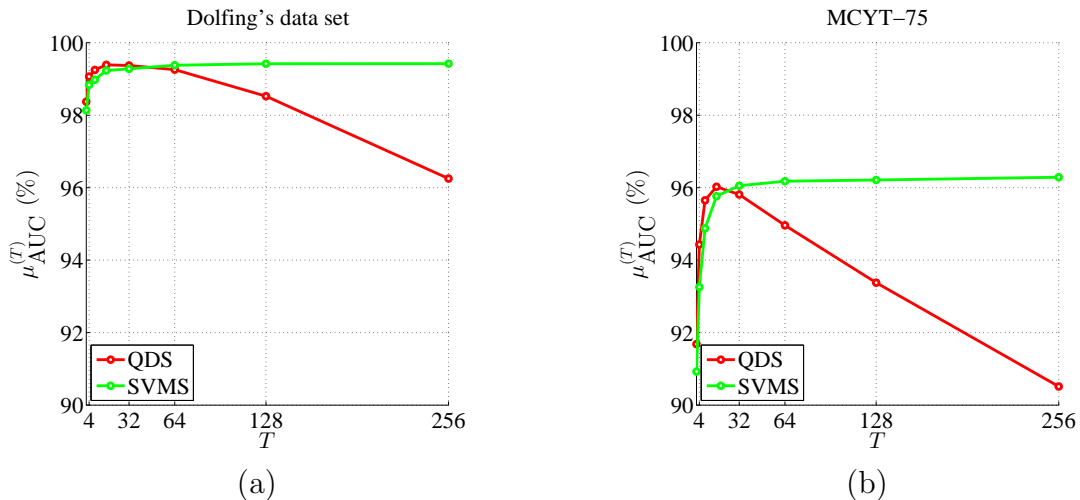


Figure 6.6: Average AUCs achieved by the QDS and SVMS, when evaluated on (a) Dolfig's data set and (b) MCYT-75, as a function of the projection angle set size T .

QDS. This assertion is supported by e.g. Bouveyron *et al.* (2007), where it is explained that it becomes increasingly difficult to obtain reliable class-specific covariance estimates when the number of training samples becomes small relative to the modelling dimension. In other words, the QDS falls prey to the curse of dimensionality (see Section 4.3.2).

It should of course be clear that, since the deterioration in performance of the QDS is in all likelihood due to the curse of dimensionality, the optimal size of the projection angle set reported in this section, that is $T \approx 16$, should not be considered absolute. If a significantly greater number of training samples were available for model construction, it is entirely reasonable to expect that the QDS would become more adept at constructing a successful model from higher-dimensional dissimilarity vectors. In contrast, if the number of available training samples were to decrease, one may reasonably expect that the curse of dimensionality will impede the performance of the QDS for $T < 16$.

Since the SVMS is not affected by this phenomenon, the optimal value of $T = 256$ may also be considered reliable for other deployment environments. However, if limited computational resources were to become relevant during deployment, the results reported in this section indicate that $T \approx 64$ should provide a satisfactory trade-off between system accuracy and system overhead. As T increases from 64 to 256, the improvement in system performance becomes less pronounced, whilst the computational requirements increase significantly.

6.3.2 On-line verification systems

Unlike the off-line systems evaluated in the previous section, the on-line verification systems developed in this study do not require the specification of any predetermined values prior to model construction. As explained in Section 3.3.2, the on-line signature representation process utilised in this study converts any feature set, comprising nineteen feature vectors

Table 6.7: Average AUCs and EERs achieved by the QDS and SVMS when evaluated on the Philips database.

$\mu_{\text{AUC}}^{(K,19)}$ (%)	K			μ_{AUC}	$\mu_{\text{EER}}^{(K,19)}$ (%)	K			μ_{EER}
	5	10	15			5	10	15	
QDS	99.52	99.70	99.84	<i>99.66</i>	QDS	2.18	2.04	1.65	<i>1.96</i>
SVMS	99.83	99.89	99.91	<i>99.87</i>	SVMS	1.19	0.85	0.65	<i>0.89</i>

Table 6.8: Average AUCs and EERs achieved by the QDS and SVMS when evaluated on MCYT-100.

$\mu_{\text{AUC}}^{(K,19)}$ (%)	K			μ_{AUC}	$\mu_{\text{EER}}^{(K,19)}$ (%)	K			μ_{EER}
	5	10	15			5	10	15	
QDS	99.14	99.47	99.53	<i>99.38</i>	QDS	4.12	3.33	2.51	<i>3.32</i>
SVMS	99.34	99.55	99.58	<i>99.49</i>	SVMS	3.24	2.72	2.49	<i>2.82</i>

of the same arbitrary dimension d , into a dissimilarity vector with fixed dimension $T = 19$. The resulting verification systems are therefore free of hyperparameters. Nevertheless, several reference set sizes (i.e. different values for K) are considered, in order to investigate its significance in terms of system proficiency.

The performance estimates obtained for the on-line systems developed in this study, when evaluated on the Philips database and MCYT-100, are presented in Tables 6.7 and 6.8 respectively.

From these tables it is clear that both the QDS and SVMS prove highly effective, when either data set is considered. As expected, an increase in system proficiency is witnessed (for both systems on both data sets) when the number of reference signatures is increased. Furthermore, there is no evidence to suggest that the QDS is impeded by the curse of dimensionality. This is a sensible result, since the on-line systems always utilise nineteen-dimensional dissimilarity vectors, whilst we showed in the previous section that the off-line QDS performed optimally for $T \approx 16$ on data sets of similar size. Nevertheless, as is the case for the off-line systems, the SVMS consistently outperforms the QDS. The superiority of the SVMS is especially pronounced when the Philips database is considered for performance estimation.

6.4 Comparison with previous work

From the performance estimates reported in the previous section, it is clear that, when either static or dynamic signatures are considered, both the QDS and the SVMS provide effective solutions to the problem of handwritten signature verification.

In order to place the performance of these systems into perspective, we now compare the results reported in this study with various historic and recent results reported in the literature. The comparisons presented in this section are of course limited to those systems

Table 6.9: Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on Dolfing’s data set, with those reported in this study.

System	Modelling strategy	EER (%)		
		$K = 5$	$K = 10$	$K = 15$
Coetzer <i>et al.</i> (2004)	WD	-	-	12.20
Swanepoel and Coetzer (2010)	WD	-	-	10.23
Panton and Coetzer (2010)	WD	-	-	8.89
QDS (This study)	WI	4.71	2.86	2.72
SVMS (This study)	WI	4.41	2.95	2.56

in the literature that have also been evaluated on the data sets considered in this study. Each of the systems that are considered here for the purpose of performance comparison is discussed in detail in Chapter 2.

It is worth noting that these existing systems predominantly utilise a *writer-dependent* approach to signature modelling and verification. However, recall from Section 1.2.6 that a writer-dependent system requires K genuine *training* signatures per writer, prior to model construction. When a writer-independent approach is employed, this is analogous to the requirement of K genuine *reference* signatures per writer, prior to model training. The writer-independent systems developed in this study are therefore deemed fit for comparison with writer-dependent systems proposed in the literature.

Although we explained earlier that the AUC represents a more comprehensive and reliable measure of system proficiency, we consider the EER for the purpose of performance comparison in this section, since the EER remains the most commonly reported metric in the literature.

6.4.1 Off-line verification systems

Dolfing’s data set

To the best of our knowledge, only the writer-dependent HMM-based systems proposed in Coetzer *et al.* (2004) and Panton and Coetzer (2010), as well as the ensemble-based system proposed in Swanepoel and Coetzer (2010), have previously been evaluated using Dolfing’s data set. A comparison of the EERs reported for these systems with the EERs reported in this study is presented in Table 6.9.

It is clear from Table 6.9 that, under similar operating conditions, the systems developed in this study outperform the systems proposed in Coetzer *et al.* (2004), Swanepoel and Coetzer (2010), as well as Panton and Coetzer (2010). Furthermore, the improvement witnessed in verification proficiency is substantial.

As previously mentioned, however, it is not considered reasonable to assume that as many as fifteen genuine reference signatures would be available per writer in a practical deployment scenario. It is therefore most promising to note that, even when the reference set size is reduced to $K = 5$, both the QDS and SVMS *still* significantly outperform *all* the

Table 6.10: Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on MCYT-75, with those reported in this study.

System	Modelling strategy	EER (%)	
		$K = 5$	$K = 10$
Fierrez-Aguilar <i>et al.</i> (2004)	WD	14.51	12.22
Alonso-Fernandez <i>et al.</i> (2007)	WD	17.76	14.44
Gilperez <i>et al.</i> (2008)	WD	10.18	6.44
Wen <i>et al.</i> (2009)	WD	15.02	-
Vargas <i>et al.</i> (2011)	WD	12.02	8.80
Ferrer <i>et al.</i> (2012)	WD	10.97	8.16
QDS (This study)	WI	11.09	9.68
SVMS (This study)	WI	10.37	9.48

existing systems listed in Table 6.9, despite the fact that they utilise a drastically reduced number of labelled positive samples. For $K = 5$, one does not expect a generative classifier (e.g. an HMM) to be able to construct a sufficiently representative writer-dependent signature model.

It is also interesting to note that the HMM-based systems proposed in Coetzer *et al.* (2004) and Panton and Coetzer (2010) both utilise the DRT for signature representation. In fact, the feature extraction process considered in Coetzer *et al.* (2004) is practically identical to the DRT-based feature extraction process utilised by the systems developed in this study. The fact that these sophisticated HMM-based systems are significantly outperformed by the systems developed in this study, even when a comparatively rudimentary classification technique such as QDA is utilised, suggests that the writer-independent approach to signature representation, when compared to its writer-dependent counterpart, provides a superior platform for signature modelling and verification.

MCYT-75

Unlike Dolfing's data set, MCYT-75 is both well-known and easily accessible to the global research community. As a result, this data set has been utilised for performance testing by numerous researchers the world over. Table 6.10 presents a performance comparison between several systems proposed in the literature and the systems developed in this study.

We conclude from Table 6.10 that both the QDS and SVMS developed in this study compare favourably with existing systems proposed in the literature. In fact, when only $K = 5$ genuine samples are available for model construction (i.e. in a realistic deployment scenario), the SVMS is *marginally* outperformed *only* by the system proposed in Gilperez *et al.* (2008).

However, as the number of available reference samples increases to $K = 10$, several existing systems outperform the systems developed in this study. The fact that the improvement in verification proficiency, resulting from the availability of additional reference samples, is relatively insignificant for both the QDS and SVMS when compared to those

reported for existing systems, suggests that the systems developed in this study may not optimally exploit the additional information provided by these extra samples. This possible deficiency in our proposed system design is probably associated with the model construction and/or verification stages and warrants further investigation. In the next chapter we discuss several proposed extensions to the system design presented in this study. These additional techniques, that have been identified as potential future work, are aimed towards improving the signature representation, model construction and verification stages. Once these proposed extensions are implemented, it should prove interesting to ascertain whether they are able to improve the proficiency of the QDS and SVMS to such an extent that their performance compares favourably with those of existing systems for *all* possible deployment scenarios.

It should be noted that the EERs reported in Table 6.10 for Fierrez-Aguilar *et al.* (2004) refer to the most proficient *individual* system proposed in their paper, that is the so-called local expert. As discussed in Section 2.2.1, this system may be fused with a global expert (which achieves EERs of 21.84% and 18.93% for $K = 5$ and $K = 10$ respectively), in order to obtain a superior combined system (which achieves EERs of 11.00% and 9.28% for $K = 5$ and $K = 10$ respectively). However, the local expert system is deemed most suitable for comparison with the systems developed in this study, since it is reasonable to expect that a similar fusion of the QDS and SVMS would also result in superior combined performance.

It is also worth mentioning that the existing systems listed in Table 6.10 represent only a small subset of published works that consider MCYT-75 for performance estimation. However, many of the works found in the literature (e.g. Prakash and Guru (2009); Azmi and Nasien (2014); Singh and Kaur (2014)) do not report an EER measure, but rather a combination of the *false acceptance rate* (FAR), *false rejection rate* (FRR) and *average error rate* (AER) metrics, as discussed in Section 1.2.7. Although these performance measures are inherently related to one another, it does not make sense to *directly* compare any of these metrics.

6.4.2 On-line verification systems

Philips database

An exhaustive search of the literature revealed that, to date, relatively few systems have been evaluated using the Philips signature database. In fact, only four works were found that report results which are deemed fit for comparison with those reported for the QDS and SVMS developed in this study. This performance comparison is presented in Table 6.11.

It is clear from Table 6.11 that, in a scenario where $K = 5$ reference samples are available per writer, the QDS outperforms existing systems that consider a similar number of training signatures. In contrast, when the number of available positive samples is increased to $K = 15$, the QDS still compares favourably with existing systems, but is ultimately outperformed by the systems proposed in Le Riche (2000) and Van *et al.* (2004). A most promising result, however, is the fact that the SVMS significantly outperforms *all* of the

Table 6.11: Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on the Philips database, with those reported in this study.

System	Modelling strategy	EER (%)		
		$K = 5$	$K = 10$	$K = 15$
Dolfing <i>et al.</i> (1998)	WD	-	-	1.90
Le Riche (2000)	WD	-	-	1.02
Van <i>et al.</i> (2004)	WD	3.54	-	0.95
Van <i>et al.</i> (2007)	WD	3.25	-	-
QDS (This study)	WI	2.18	2.04	1.65
SVMS (This study)	WI	1.19	0.85	0.65

existing systems listed in Table 6.11, regardless of the number of reference signatures available per writer.

Other works that consider the Philips database for performance estimation, but which are deemed not fit for comparison with the systems developed in this study, include the writer-dependent systems proposed in Fuentes *et al.* (2002) and Sindle (2003). The former evaluate their system using a combination of skilled *and* professional forgeries, which reportedly yields an FAR and FRR of 4.62% and 8.25% respectively. The latter reports EERs of 5.0% and 3.9% respectively when *either* over-the-shoulder *or* home-improved forgeries are considered for system evaluation. Both of the aforementioned systems require fifteen genuine training samples for model construction.

MCYT-100

Similar to its off-line counterpart, MCYT-100 has been considered extensively by a large cohort of researchers over a prolonged period of time for the purpose of system performance estimation. The results reported for several systems proposed in the literature are listed in Table 6.12, along with the results reported for the QDS and SVMS developed in this study.

It should be emphasised that the systems of which the performances are listed in Table 6.12 constitute a small subset of a much larger collection of works that also consider MCYT-100 for system evaluation. However, it is important to note that the works tabulated here constitute the *most proficient* systems currently found in the literature. Although numerous other works have also reported results for this data set, the systems proposed in e.g. Kahn *et al.* (2006), Nanni and Lumini (2008), Yanikoglu and Kholmatov (2009), Galbally *et al.* (2009), and Wibowo *et al.* (2013) achieve EERs in the range 7.2%–15.98% when five genuine samples are available for model construction, whereas the *least* proficient system listed in Table 6.12 achieves an EER of 5.4% under similar operating conditions.

When compared to previous systems also evaluated on MCYT-100, both of the systems developed in study compare favourably. The QDS outperforms several existing systems,

Table 6.12: Comparison of the EERs for several existing writer-dependent (WD) and/or writer-independent (WI) systems when evaluated on MCYT-100, with those reported in this study.

System	Modelling strategy	EER (%)		
		$K = 5$	$K = 10$	$K = 15$
Van <i>et al.</i> (2007)	WD	3.37	-	-
Lumini and Nanni (2009)	WD	5.4	-	-
Muramatsu and Matsumoto (2009)	WI	4.0	-	-
Montalvão <i>et al.</i> (2010)	WD	4.5	-	-
Sae-Bae and Memon (2013)	WD	4.02	2.72	-
QDS (This study)	WI	4.12	3.33	2.51
SVMS (This study)	WI	3.24	2.72	2.49

whilst the more proficient SVMS outperforms *all* previous systems (except for the system proposed in Sae-Bae and Memon (2013), for which there is no distinction in terms of the reported EER for $K = 10$).

It is worth noting that the system proposed in Muramatsu and Matsumoto (2009) considers the BIOMET and MCYT-100 databases (in their entirety) for the respective purposes of training and evaluation. It is unclear whether this protocol constitutes an advantage or a disadvantage in terms of system performance estimation. On the one hand, the data set considered for training is completely different to the data set considered for evaluation, which may result in an atypically poor system performance. On the other hand, this system considers a significantly larger number of training samples than would be possible within a protocol that incorporates cross-validation on a single data set. In addition, since the system utilises a writer-independent signature model, it should be able to generalise well and the availability of additional training samples may result in a superior model. It should therefore prove interesting to re-evaluate the systems developed in this study under similar experimental conditions to those considered in Muramatsu and Matsumoto (2009). Such a re-evaluation is, however, reserved for future work.

6.5 Contribution of this study

In order to ascertain whether or not the novel modelling techniques proposed in this study constitute significant contributions to the current state of the art, we re-evaluate each system for the set of different system configurations listed in Table 6.13. It should be clear from Table 6.13 that configuration C1 represents the current state of the art within the context of writer-independent *off-line* signature verification, whilst configuration C2 represents the current state of the art within the context of writer-independent *on-line* signature verification. Configuration C4 represents the system design proposed in *this study*.

Apart from the dichotomy transformation and/or dissimilarity normalisation strategy

Table 6.13: The set of system configurations considered for system re-evaluation.

Configuration	Dichotomy transformation	Normalisation strategy
C1	Euclidean	Global
C2	DTW	Global
C3	Euclidean	Writer-specific
C4	DTW	Writer-specific

utilised, the system design associated with each configuration is identical. Furthermore, the experimental protocol considered for each re-evaluation is also identical. By comparing the system performance estimates obtained for these different configurations, we are therefore able to quantify the impact on system proficiency resulting from the incorporation of the modelling techniques proposed in this study, namely the DTW-based dichotomy transformation for off-line signature representation and the writer-specific dissimilarity normalisation strategy.

In Section 6.3 we estimated system proficiency using two fundamentally different performance metrics, namely the AUC and EER measures. As mentioned in Section 1.2.7, however, the AUC is considered a more stable and comprehensive measure of system proficiency. For this reason, we only consider the AUC-based performance metric associated with each configuration-specific evaluation for comparison.

In order to gain an initial insight into configuration-specific system performance, graphical performance comparisons are presented in Figures 6.7–6.10. Each *notched box plot* represented in these figures is constructed from the entire set of AUC-based metrics obtained for all values of the system hyperparameters K and T (see Table 6.2).

In the case of the *off-line* verification systems, we note from Figures 6.7–6.8 that the inclusion of *either* the DTW-based dichotomy transformation *or* the writer-specific dissimilarity normalisation strategy leads to a significant increase in verification proficiency. It is also clear from these figures that the improvement in system proficiency is especially pronounced when these two techniques are incorporated simultaneously. When the system design proposed in this study (C4) is compared to the current state of the art (C1), we observe a notable improvement in terms of both *accuracy* (in the sense that the median AUC associated with C4 exceeds that of C1) and *consistency* (in the sense that the variability of the AUC-values, as indicated by the length of the box whiskers, is much smaller for C4 than for C1). In fact, both Figures 6.7 and 6.8 indicate that the *worst* expected performance associated with configuration C4 surpasses the *best* expected performance associated with configuration C1. Finally, the fact that the notches associated with boxes C4 and C1 do not overlap (in any of the four off-line system comparisons presented) suggests that the improvement witnessed is *statistically significant*. We formally address the issue of statistical significance later in this section.

In the case of the *on-line* verification systems, we note from Figures 6.9–6.10 that the inclusion of the writer-specific dissimilarity normalisation strategy leads to an increase in verification proficiency. When MCYT-100 is considered, the improvement witnessed

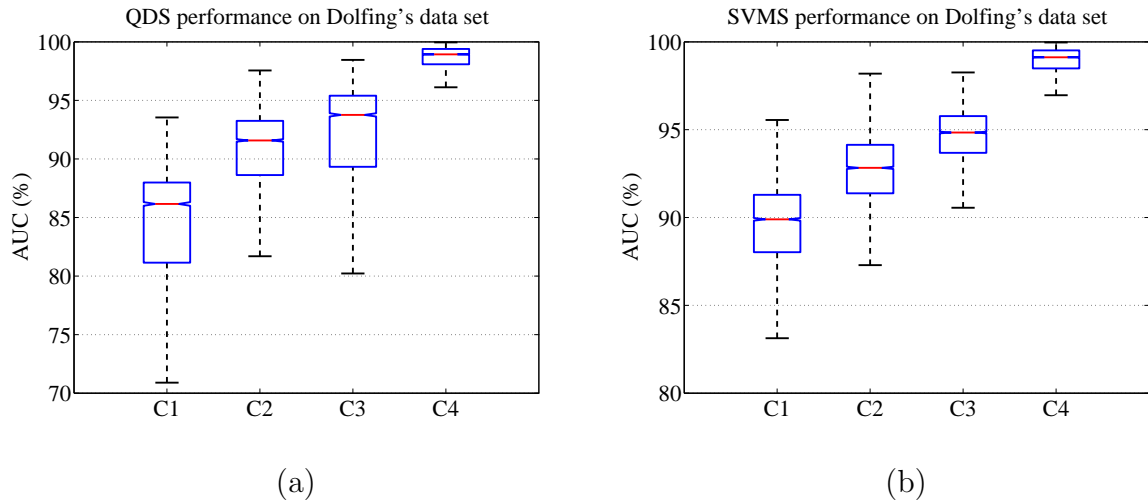


Figure 6.7: Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on Dolfig's data set for configurations C1–C4.

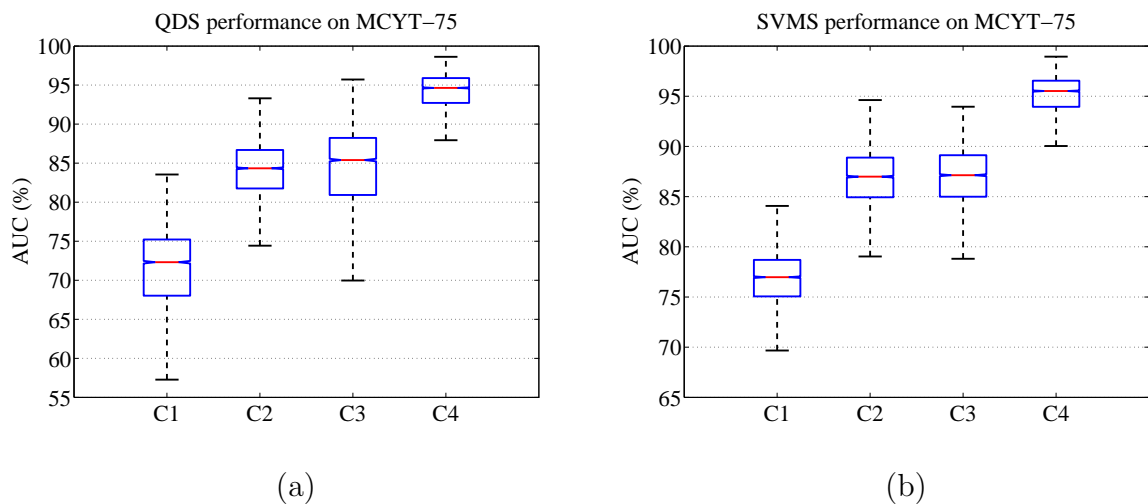


Figure 6.8: Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on MCYT-75 for configurations C1–C4.

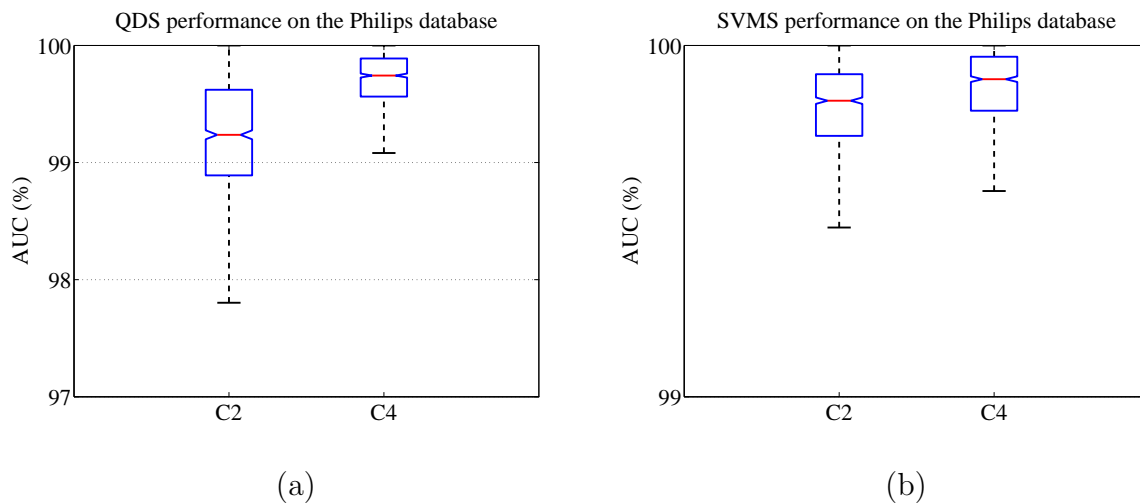


Figure 6.9: Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on the Philips database for configurations C2 and C4.

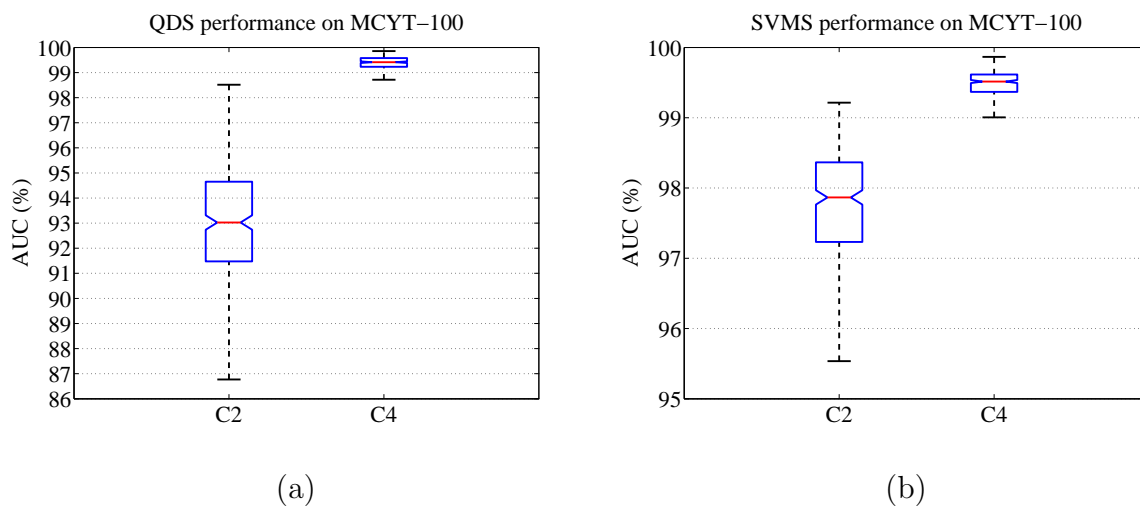


Figure 6.10: Comparison of the AUC-based performance metrics obtained for (a) the QDS and (b) the SVMS, when these systems are evaluated on MCYT-100 for configurations C2 and C4.

in terms of both accuracy and consistency is substantial. When the Philips database is considered, however, the improvement in proficiency is somewhat marginal. Nevertheless, it should be clear from Figure 6.9 that the state-of-the-art system performance on the Philips database is already exceptionally high, thereby rendering *any* improvement noteworthy. Furthermore, both Figures 6.9 and 6.10 indicate that a statistically significant improvement in verification proficiency is expected when the on-line system design proposed in this study is compared to that of the current state of the art.

Statistical significance

Although the graphical performance comparisons in Figures 6.7–6.10 indicate that the incorporation of the novel techniques proposed in this study leads to a significant improvement in system proficiency, it is advisable to confirm this assertion by means of a formal statistical test. The evidence resulting from such a *statistical hypothesis test* may be used to support or reject assertions made from experimental observations, by quantifying the likelihood that these observations are misleading due to *sampling error*.

In this study we employ the *two-sample t-test* for an $\alpha = 0.01$ *level of significance*⁴, as outlined in e.g. Cressie and Whitford (1986). This test compares the location parameters of two *independent, normally distributed* data samples A and B , in order to ascertain the likelihood that these samples are representative of populations with unequal means, and proceeds as follows:

1. Definitions

Let samples A and B refer to the sets of AUC-based performance metrics obtained by evaluating a system for two different configurations. The improvement expected in terms of system proficiency, as a result of utilising configuration A instead of configuration B , may be quantified by the *performance gradient* Δ_B^A , such that

$$\Delta_B^A = \mu(A) - \mu(B). \quad (6.1)$$

The performance gradient therefore represents the difference between the *sample means*. Furthermore, let $\mu_A^{(P)}$ and $\mu_B^{(P)}$ denote the *population means* associated with samples A and B respectively.

2. Hypothesis statement

Given the parameter definitions stated above, we wish to investigate the following hypotheses:

- Null Hypothesis H_0 : $\mu_A^{(P)} - \mu_B^{(P)} \leq 0$
- Alternative Hypothesis H_A : $\mu_A^{(P)} - \mu_B^{(P)} > 0$

⁴The level of significance may be interpreted as the probability of rejecting the *null hypothesis* given that it is true.

Since we specifically wish to ascertain whether $\mu_A^{(P)} > \mu_B^{(P)}$, that is whether the utilisation of system configuration A will in all likelihood result in superior proficiency when compared to system configuration B , the statistical test performed here is referred to as a *right-tailed test*.

3. Computations

The *test statistic* t_{stat} is determined as follows:

$$t_{\text{stat}} = \frac{\Delta_B^A}{S}, \quad (6.2)$$

$$S = \sqrt{\left(\frac{(n_A - 1)\sigma_A^2 + (n_B - 1)\sigma_B^2}{n_A + n_B - 2}\right) \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}, \quad (6.3)$$

where σ_X^2 and n_X denote the variance and size, respectively, of sample X .

Note that, in contrast to the pooled performance comparisons presented in Figures 6.7–6.10, we wish to conduct a separate hypothesis test for each combination of the system hyperparameter values considered, in order to determine whether or not the improvement witnessed in system proficiency is dependent on said values. Since each sample X therefore contains the set of $kn = 300$ performance metrics associated with specific values for K and T , we consequently also have $n_A = n_B = 300$.

4. Decision

At an $\alpha = 0.01$ level of significance, there is sufficient evidence to reject the null hypothesis if the test statistic t_{stat} exceeds the associated *critical value* t_{crit} , such that

$$\begin{aligned} t_{\text{stat}} &> t_{\text{crit}} \\ &= t_{(\alpha, n_A + n_B - 2)} \\ &= t_{(0.01, 598)} \\ &= 2.33 \end{aligned} \quad (6.4)$$

Consequently, if (and only if) $\Delta_B^A > 0$ and $t_{\text{stat}} > 2.33$, we can assert with 99% confidence that the utilisation of configuration A instead of configuration B should result in a superior verification system.

We now subject each of the novel concepts presented in this study to the above-mentioned statistical analysis, in order to confirm (or refute) our claim that each of these concepts may be considered a noteworthy contribution to the current state of the art.

6.5.1 A DTW-based dichotomy transformation for writer-independent off-line signature representation

As discussed in Chapter 2, a wide variety of writer-independent off-line signature verification systems have been proposed in recent years. In order to achieve the required dissimilarity-based signature representation, the systems proposed in the literature primarily utilise the Euclidean distance (or simple vector subtraction) to quantify the difference between a reference sample and any given training/questioned sample.

In this study we proposed that a fundamentally different distance measure, namely the DTW-based distance, be utilised for the aforementioned dissimilarity calculation. Although this method has proved successful within the context of *on-line* signature representation, where its use is necessitated by the inevitable variations in length of the time series data extracted from on-line signature samples, no *off-line* signature verification system currently found in the literature makes use of a DTW-based dichotomy transformation. We postulate that a DTW-based algorithm's ability to non-linearly align two feature vectors, prior to matching, allows for the construction of more robust dissimilarity vectors, since the adverse effects of intra-class variability are minimised. The strategy proposed in this study, for the purpose of obtaining a writer-independent off-line signature representation, may therefore be considered novel.

The impact on system proficiency, as a direct result of utilising a DTW-based distance measure instead of the Euclidean distance for the purpose of dissimilarity vector construction (that is a comparison between the performance estimates associated with system configurations C2 and C1), is presented in Tables 6.14–6.15. These results clearly indicate that the proposed DTW-based approach consistently results in superior performance, regardless of the classification technique, system hyperparameter values or data set considered. In the case where MCYT-75 is considered for system evaluation, this improvement in performance is especially pronounced. More importantly, however, the associated *t*-statistics indicate that there is sufficient evidence to assert, with 99% confidence, that the aforementioned improvement is statistically significant - for *all* cases investigated.

It should be clear that the results presented in Tables 6.14–6.15 confirm the superiority of the DTW-based approach, but *only* when utilised in conjunction with a DRT-based feature extraction process. Nevertheless, it is reasonable to expect that this approach should also prove effective when coupled with several other feature extraction techniques. However, it is important to realise that, in certain instances, the use of a DTW-based dichotomy transformation may in fact be entirely inappropriate. For example, if one were to utilise *local binary pattern* (LBP) histograms for the initial signature representation (see e.g. Nicolaou *et al.* (2013)), it would not be sensible to consider non-linear feature correspondences during dissimilarity calculation.

Ultimately, given the results presented in this section, we assert that the DTW-based dichotomy transformation proposed in this study constitutes a notable contribution to the field of writer-independent off-line signature verification.

Table 6.14: Performance gradients (and corresponding test statistics) obtained between configurations C2 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.

QDS				
Δ_{C1}^{C2} (t_{stat})		K		
		5	10	15
T	2	6.11 (43.13)	7.15 (50.88)	7.27 (51.35)
	4	3.57 (26.27)	4.30 (31.35)	4.50 (32.35)
	8	4.71 (34.09)	5.30 (38.97)	5.49 (39.69)
	16	5.86 (42.11)	6.74 (49.91)	6.98 (50.96)
	32	4.92 (31.59)	5.45 (36.33)	5.58 (37.72)
	64	5.20 (23.37)	5.47 (24.96)	5.42 (24.90)
	128	8.76 (31.49)	9.59 (33.94)	9.93 (34.55)
	256	7.83 (27.17)	8.99 (30.27)	9.32 (30.79)
SVMS				
Δ_{C1}^{C2} (t_{stat})		K		
		5	10	15
T	2	6.63 (39.88)	7.37 (45.78)	7.63 (47.19)
	4	3.05 (18.67)	4.00 (25.50)	4.36 (27.29)
	8	2.50 (14.81)	2.82 (17.25)	3.22 (19.73)
	16	2.18 (14.37)	2.68 (18.02)	2.83 (18.91)
	32	2.03 (13.49)	2.50 (17.40)	2.65 (18.19)
	64	2.13 (14.35)	2.58 (18.14)	2.71 (18.94)
	128	2.57 (17.51)	2.79 (19.84)	2.88 (20.20)
	256	2.67 (18.51)	2.83 (20.33)	2.89 (20.51)

Table 6.15: Performance gradients (and corresponding test statistics) obtained between configurations C2 and C1 when the QDS and SVMS are evaluated on MCYT-75.

QDS				SVMS			
Δ_{C1}^{C2} (t_{stat})		K		Δ_{C1}^{C2} (t_{stat})		K	
		5	10			5	10
T	2	12.01 (66.57)	11.89 (60.75)	T	2	11.78 (61.22)	10.72 (53.49)
	4	8.48 (45.05)	8.35 (41.34)		4	8.78 (45.17)	8.05 (39.72)
	8	9.62 (51.63)	10.04 (50.90)		8	9.58 (50.51)	9.11 (46.66)
	16	12.19 (64.05)	12.84 (66.16)		16	9.68 (54.80)	9.53 (51.68)
	32	12.11 (57.77)	13.01 (62.77)		32	10.44 (57.81)	10.49 (56.36)
	64	12.51 (58.51)	12.72 (56.98)		64	10.61 (60.09)	10.25 (55.60)
	128	14.91 (61.18)	15.70 (60.09)		128	10.54 (60.77)	10.15 (56.31)
	256	19.20 (70.71)	19.28 (66.94)		256	10.64 (60.44)	10.17 (55.60)

6.5.2 A writer-specific dissimilarity normalisation strategy for writer-independent handwritten signature modelling

In this study we proposed the utilisation of a novel dissimilarity normalisation technique that exploits *writer-specific* information within a *writer-independent* signature modelling framework. In Section 5.2.2 we showed that, as opposed to the traditional global normalisation strategy, the proposed writer-specific approach is able to significantly improve inter-class separability in dissimilarity space and therefore also improve the efficacy of the resulting writer-independent signature model.

It should be clear from the literature survey presented in Chapter 2 that the incorporation of writer-specific information into a writer-independent signature modelling framework does not in itself constitute a novel concept, since the benefits of such an approach was previously investigated by Muramatsu and Matsumoto (2009). Recall from Section 2.3.2 that the aforementioned system first estimates a mean dissimilarity vector from the entire set of reference signatures associated with a specific writer. This is of course similar to the writer-specific mean dissimilarity vector $\boldsymbol{\mu}^{(\omega)}$ estimated by the systems developed in this study, as explained in Section 5.2.2. However, Muramatsu and Matsumoto (2009) propose that this mean vector should subsequently be *concatenated* to any dissimilarity vector belonging to (or claimed to belong to) said writer, prior to model construction (or verification), in order to *personalise* the dissimilarity vector. The utilisation of a writer-specific dissimilarity *normalisation function*, as proposed in this study, is fundamentally different from this previously proposed strategy and may therefore be considered novel. Furthermore, we assert that the dissimilarity normalisation strategy proposed in this study is (from a statistical perspective) more elegant and intuitive than the existing dissimilarity vector concatenation-based strategy. Finally, it is worth noting that the aforementioned concatenation process invariably *doubles the dimension* of any dissimilarity vector submitted for personalisation. In Section 6.3.1 we showed that a substantial increase in the model dimension may have a severely detrimental effect on the proficiency of a system that utilises a generative classifier for signature modelling and verification. The concatenation-based approach is therefore only fit for incorporation into verification systems that either consider a relatively small feature set for signature representation or employ a classification technique that is adept at constructing high-dimensional signature models. In contrast, the strategy proposed in this study does not alter the dimension of any dissimilarity vector submitted for normalisation and is therefore fit for incorporation into *any* writer-independent verification system.

Although our proposed approach to incorporating writer-specific information into the writer-independent signature modelling framework is certainly novel, it is equally important to ascertain whether or not this approach is in fact effective. Since the utilisation of this dissimilarity normalisation function is equally applicable to off-line and on-line signature modelling, we investigate its impact on system proficiency for both of these scenarios.

The impact on system proficiency, as a direct result of utilising the proposed writer-specific dissimilarity normalisation strategy instead of a global normalisation strategy, within the context of *off-line* signature modelling (that is a comparison between the performance estimates associated with system configurations C3 and C1), is presented in

Tables 6.16–6.17. A similar analysis is performed within the context of *on-line* signature modelling (that is a comparison between the performance estimates associated with system configurations C4 and C2) and is presented in Tables 6.18–6.19.

We find that, for both the off-line and on-line scenarios, the utilisation of a writer-specific approach to dissimilarity normalisation consistently results in superior performance, regardless of the classification technique, system hyperparameter values or data set considered. The observed improvement in terms of verification proficiency is substantial when MCYT-75 is considered for system evaluation, whilst the results achieved on the Philips database indicate a somewhat marginal improvement. Ultimately, although the magnitude of the resulting performance improvement appears to be dependent on the data set considered, there is no denying that an improvement is consistently observed. More importantly, there is sufficient evidence to assert (with 99% confidence) that the witnessed improvement is statistically significant - for *all* cases investigated.

The results presented in this section, coupled with the fact that the strategy proposed in this study is generic (in the sense that it may easily be incorporated into *any* existing writer-independent verification system), leads us to assert that the writer-specific dissimilarity normalisation strategy proposed in this study constitutes a significant contribution to the field of writer-independent handwritten signature verification – that is for both the off-line and on-line scenarios.

6.5.3 A writer-independent off-line signature verification system that is both novel and proficient

The off-line systems developed in this study utilise two novel signature modelling techniques, namely the DTW-based dichotomy transformation and the writer-specific dissimilarity normalisation strategy proposed in this study. The off-line QDS and SVMS developed in this study may therefore also be considered novel.

In Sections 6.5.1–6.5.2 we showed that, within the context of off-line signature modelling, the incorporation of *either* of the aforementioned techniques leads to a statistically significant improvement in verification proficiency. It is therefore not strictly necessary to perform a similar analysis for the *simultaneous* incorporation of these two techniques. Nevertheless, we provide such an analysis (that is a comparison between the performance estimates associated with system configurations C4 and C1) in Tables 6.20–6.21, in order to emphasise the extent of the resulting improvement in performance.

It is clear from Tables 6.20–6.21 that, regardless of the classification technique, system hyperparameter values or data set considered, the simultaneous incorporation of the novel techniques proposed in this study leads to a *substantial* increase in system proficiency. As expected, the associated *t*-statistics confirm that there is overwhelming evidence to assert that this improvement is statistically significant – for *all* cases considered. Furthermore, in Section 6.4.1 we showed that, when system performance estimation is conducted under similar experimental conditions, both the QDS and SVMS proposed in this study either outperform or compare favourably with existing systems proposed in the literature.

Table 6.16: Performance gradients (and corresponding test statistics) obtained between configurations C3 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.

QDS				
Δ_{C1}^{C3} (t_{stat})		K		
		5	10	15
T	2	6.83 (50.10)	7.93 (59.29)	8.21 (65.13)
	4	6.25 (52.63)	7.17 (60.41)	7.59 (64.57)
	8	7.54 (64.94)	8.27 (73.00)	8.68 (76.98)
	16	7.57 (61.57)	8.60 (73.20)	9.09 (79.75)
	32	6.70 (49.07)	7.87 (61.55)	8.16 (66.67)
	64	6.71 (34.30)	7.84 (39.76)	7.84 (39.94)
	128	6.30 (24.76)	8.86 (33.10)	9.12 (33.16)
	256	1.52 (6.14)	4.79 (18.55)	4.86 (17.93)
SVMS				
Δ_{C1}^{C3} (t_{stat})		K		
		5	10	15
T	2	5.99 (37.56)	6.63 (42.24)	6.75 (43.71)
	4	5.90 (44.86)	6.40 (48.66)	6.73 (51.10)
	8	5.97 (44.21)	6.16 (47.56)	6.42 (48.71)
	16	4.74 (37.93)	5.25 (42.11)	5.75 (46.49)
	32	4.18 (33.95)	4.72 (39.27)	5.35 (45.16)
	64	4.01 (33.04)	4.64 (39.26)	5.28 (45.56)
	128	3.90 (32.06)	4.63 (39.90)	5.17 (45.38)
	256	3.86 (31.82)	4.61 (40.03)	5.11 (45.35)

Table 6.17: Performance gradients (and corresponding test statistics) obtained between configurations C3 and C1 when the QDS and SVMS are evaluated on MCYT-75.

QDS				SVMS			
Δ_{C1}^{C3} (t_{stat})		K		Δ_{C1}^{C3} (t_{stat})		K	
		5	10			5	10
T	2	10.63 (57.28)	12.42 (62.92)	T	2	8.18 (42.88)	8.22 (39.96)
	4	10.42 (59.90)	11.84 (65.80)		4	9.22 (54.57)	9.36 (54.04)
	8	11.29 (69.93)	13.08 (77.95)		8	10.24 (65.93)	11.28 (73.02)
	16	12.38 (73.02)	14.98 (90.03)		16	9.58 (62.93)	12.12 (84.66)
	32	12.67 (70.68)	15.38 (89.92)		32	9.54 (59.29)	11.58 (75.86)
	64	13.10 (76.43)	16.07 (90.94)		64	9.47 (60.04)	11.01 (73.12)
	128	11.13 (56.27)	15.45 (73.85)		128	9.39 (59.43)	10.85 (71.60)
	256	7.02 (30.68)	10.01 (38.42)		256	9.16 (58.08)	10.67 (70.33)

Table 6.18: Performance gradients (and corresponding test statistics) obtained between configurations C4 and C2 when the QDS and SVMS are evaluated on the Philips database.

Δ_{C2}^{C4} (t_{stat})	K		
	5	10	15
QDS	0.31 (8.81)	0.48 (14.40)	0.68 (19.85)
SVMS	0.07 (4.67)	0.07 (5.45)	0.10 (7.71)

Table 6.19: Performance gradients (and corresponding test statistics) obtained between configurations C4 and C2 when the QDS and SVMS are evaluated on MCYT-100.

Δ_{C2}^{C4} (t_{stat})	K		
	5	10	15
QDS	6.37 (49.30)	6.42 (48.22)	6.37 (48.68)
SVMS	1.80 (36.53)	1.74 (36.35)	1.70 (36.38)

Consequently, we assert that the off-line system design proposed in this study constitutes a notable contribution to the field of off-line signature verification.

6.5.4 A writer-independent on-line signature verification system that is both novel and proficient

As is the case with the off-line systems discussed in the previous section, both the on-line QDS and SVMS developed in this study employ the proposed writer-specific dissimilarity normalisation strategy, thereby also ensuring their novelty.

In Section 6.5.2 we showed that, within the context of on-line signature modelling, the incorporation of this normalisation strategy leads to a statistically significant improvement in verification proficiency and, since this is the only novel technique incorporated into the on-line verification systems developed in this study, no further analysis is required. We also showed in Section 6.4.2 that both the QDS and SVMS proposed in this study either outperform or compare favourably with existing systems proposed in the literature. In fact, it appears to be entirely plausible that, for realistic deployment scenarios, where limited genuine reference samples are available for each writer enrolled into the system, the SVMS represents one of the most proficient signature verification systems evaluated to date on both the Philips database and MCYT-100.

We therefore assert that the on-line system design proposed in this study constitutes a notable contribution to the field of on-line signature verification.

6.6 Concluding remarks

In this chapter we conducted a rigorous experimental evaluation of the off-line and on-line signature verification systems developed in this study. Several large, well-known signature

Table 6.20: Performance gradients (and corresponding test statistics) obtained between configurations C4 and C1 when the QDS and SVMS are evaluated on Dolfig's data set.

QDS				
Δ_{C1}^{C4} (t_{stat})		K		
		5	10	15
T	2	12.19 (107.45)	13.37 (120.91)	13.48 (124.09)
	4	10.13 (99.71)	10.80 (105.67)	10.92 (106.20)
	8	11.26 (112.11)	11.70 (118.16)	11.72 (118.26)
	16	11.65 (110.37)	12.24 (120.25)	12.25 (120.95)
	32	11.44 (94.12)	11.73 (102.00)	11.59 (104.70)
	64	14.25 (82.95)	14.20 (83.04)	13.88 (82.58)
	128	20.58 (105.21)	21.66 (107.71)	21.83 (105.18)
	256	20.45 (99.48)	23.07 (114.76)	23.63 (114.95)
SVMS				
Δ_{C1}^{C4} (t_{stat})		K		
		5	10	15
T	2	11.79 (85.71)	12.79 (96.77)	12.89 (97.47)
	4	9.60 (83.55)	10.36 (92.31)	10.47 (91.39)
	8	9.66 (80.49)	10.07 (87.85)	10.19 (89.37)
	16	8.91 (81.06)	9.20 (85.14)	9.18 (85.10)
	32	8.42 (78.68)	8.71 (86.09)	8.66 (84.80)
	64	8.44 (81.20)	8.57 (85.63)	8.46 (84.28)
	128	8.55 (83.62)	8.57 (87.67)	8.37 (85.28)
	256	8.52 (83.77)	8.48 (87.27)	8.27 (84.85)

Table 6.21: Performance gradients (and corresponding test statistics) obtained between configurations C4 and C1 when the QDS and SVMS are evaluated on MCYT-75.

QDS				SVMS			
Δ_{C1}^{C4} (t_{stat})		K		Δ_{C1}^{C4} (t_{stat})		K	
		5	10			5	10
T	2	20.50 (121.67)	21.31 (126.85)	T	2	17.29 (97.10)	17.08 (97.97)
	4	18.95 (122.66)	19.31 (122.92)		4	17.01 (110.73)	16.60 (109.77)
	8	18.98 (127.97)	19.57 (130.02)		8	18.61 (127.63)	18.09 (127.03)
	16	20.56 (135.05)	20.90 (137.59)		16	18.98 (141.61)	18.77 (142.59)
	32	22.09 (138.77)	22.09 (141.33)		32	18.90 (135.33)	18.70 (137.38)
	64	23.76 (158.61)	24.17 (155.47)		64	18.57 (137.40)	18.28 (135.83)
	128	26.32 (158.04)	28.22 (157.85)		128	18.57 (137.93)	18.04 (133.87)
	256	26.81 (140.34)	30.19 (146.95)		256	18.42 (137.12)	17.92 (132.95)

corpora were considered for system proficiency testing, whilst the experimental protocol ensured that the performance estimates reported in this chapter are both comprehensive and unbiased.

The results reported in this chapter indicate that, under optimal operating conditions, the SVMS consistently outperforms the QDS. This is the case for both the off-line and on-line verification scenarios investigated. Furthermore, the proficiency of the systems developed in this study were placed into context by means of a performance comparison with existing systems proposed in the literature. When compared to existing systems that were also evaluated on Dolfing's data set, the Philips database and the MCYT-Signature-100 subcorpus, we found that the SVMS outperforms (to the best of our knowledge) *all* previously proposed systems, whilst it outperforms *most* existing systems also evaluated on the MCYT-SignatureOff-75 subcorpus. The QDS compares favourably with previously proposed systems and outperforms several existing systems, although we find that this system is, unlike the SVMS, susceptible to the curse of dimensionality.

Finally, we presented an extensive analysis of the contributions made by the novel techniques proposed in this study. We showed that the utilisation of a DTW-based dichotomy transformation or a writer-specific dissimilarity normalisation strategy consistently leads to a substantial improvement in system proficiency, especially when both methods are incorporated simultaneously. Most importantly, a formal statistical significance test confirmed that both of the novel techniques proposed in this study constitute noteworthy contributions to the current state of the art.

In the next chapter, we conclude this study by highlighting selected key aspects in a concise overview. We also present brief discussions on a selection of strategies for potentially improving the proposed systems in the future.

Chapter 7

Conclusion and Future Work

“The important thing is to not stop questioning.”
- Albert Einstein (1879–1955)

7.1 Conclusion

The purpose of this study was to design and develop an effective, writer-independent hand-written signature verification system. Two fundamentally different verification scenarios, namely those concerned with off-line and on-line signatures, were investigated.

The systems developed in this study utilise various existing techniques documented in the literature, as well as several novel techniques as proposed in this study. The most notable of these novel techniques include a *dynamic time warping* (DTW) based dichotomy transformation for the purpose of off-line signature representation, specifically proposed for the conversion of writer-dependent *discrete Radon transform* (DRT) based feature vectors into writer-independent dissimilarity vectors, as well as a writer-specific dissimilarity normalisation strategy, which is considered generic in the sense that it is equally applicable to both off-line and on-line signature modelling.

The DTW-based dichotomy transformation provides an effective solution to the problem of intra-class variability, whilst the writer-specific dissimilarity normalisation strategy considerably improves inter-class separability in dissimilarity space. Both of these issues are of critical importance to the successful development of any writer-independent signature verification system. Furthermore, when compared to state-of-the-art writer-independent signature verification systems, the novel techniques proposed in this study make no additional assumptions regarding potential deployment conditions, and may therefore be easily incorporated into any existing system documented in the literature.

In order to quantify the significance of the proposed techniques, the systems developed in this study were subjected to a rigorous experimental evaluation protocol using several large signature corpora. The results obtained during this evaluation indicate that both the off-line and on-line verification systems developed in this study provide highly effective solutions to the problem of automated handwritten signature verification. However, we

also found that the SVMS, that is the system that utilises a support vector machine with a radial basis function kernel, unquestionably represents the most proficient system developed in this study. This is the case when either off-line or on-line signature verification is performed.

A comparison between the results reported here with those reported in the literature (i.e. existing systems evaluated on the *same* data sets) confirm the efficacy the systems proposed in this study. In addition, it is shown that the novel techniques proposed in this study are able to substantially improve system proficiency, especially when implemented simultaneously. Furthermore, we confirmed that this improvement in system performance is statistically significant. The proposed techniques may therefore not only be considered novel, but also constitute noteworthy contributions to the field of handwritten signature verification.

In terms of the primary objectives stated in Section 1.3, this study is therefore deemed successful. Furthermore, the success of this study is confirmed by the fact that it has resulted in several peer-reviewed publications (Swanepoel and Coetzer (2012, 2013, 2014)), whilst the initial *proof of concept* (Swanepoel and Coetzer (2012)) also received the *International Graphonomics Society Best Student Paper Award* at the *2012 International Conference on Frontiers in Handwriting Recognition*. However, despite these successes, there remains considerable room for potential improvement regarding various stages of the proposed system design. In the next section, we present several possible avenues for future research that may prove beneficial to the systems developed in this study.

7.2 Future work

During the course of this study, several concepts were encountered that may warrant further investigation. These concepts are, however, not included in this dissertation – either due to being considered outside the scope of this study, or simply due to time constraints. In this section, several of these concepts are briefly discussed as a possible continuation of this research.

7.2.1 Improved image processing for off-line signature representation

In Section 3.2.1 we discussed the importance of incorporating effective image pre-processing techniques prior to off-line signature feature extraction. We also discussed several techniques aimed specifically towards image binarisation, noise reduction and signature segmentation. As mentioned in Section 3.5, however, none of the image processing methods utilised by the systems developed in this study are necessarily deemed *optimal* for their respective tasks, since document image processing is not considered a primary objective of this study. It should therefore prove useful to investigate existing state-of-the-art techniques (and possibly develop novel algorithms), in order to optimally address the various

issues related to document image processing, and in so doing maximise the efficacy of the off-line feature extraction process.

Consider, for example, the issue of document image binarisation. In this study we implemented Otsu's method, due to its ease of use and the fact that it has proven generally effective throughout the literature (see e.g. Freire *et al.* (2007); Alonso-Fernandez *et al.* (2007); Prakash and Guru (2009)). However, since this method employs *global* thresholding, its success may be severely impeded when confronted with a degraded¹ document image. A variety of state-of-the-art *adaptive* image binarisation algorithms are detailed in e.g. Singh *et al.* (2011), Zhang and Wu (2011), as well as Lazzara and Thierry (2014). Since these algorithms perform *local* thresholding within a dynamically adjustable window, they are able to produce significantly more accurate pen stroke information within the resulting binary image representation.

Another key issue in document image analysis is that of noise reduction. Again, due to its straightforward implementation and historic success, we implemented the median filter in this study. However, alternative algorithms such as the *adaptive median filter* (AMF) or other morphological operators may prove superior, especially when the erroneous pixels represented in the binary image are not limited to impulse noise. The AMF, for instance, is particularly adept at removing relatively high density noise, whilst also preserving the shape of the handwriting.

Various additional issues may also be addressed in potential future research, such as *stroke repair*² and *stroke width normalisation*, so that the resulting binary image representation may resemble an *ideal* off-line representation (similar to the samples from Dolfing's data set as presented in Figure 6.1).

The incorporation of these specialised image pre-processing algorithms, into the systems developed in this study, should undoubtedly improve the quality of the features subsequently extracted from static signature images and therefore also improve overall modelling and verification proficiency.

7.2.2 Feature weighted signature modelling

The *on-line* verification systems developed in this study consider several *different* spatial and temporal features for signature representation. Since each one of these features measures a different aspect of signature production, it is reasonable to expect that not all the features considered would provide an equally discriminative platform for signature modelling. For instance, we noted in Section 3.3.2 that the discriminative superiority of temporal features, when compared to their spatial counterparts, is a well-established fact.

¹Within the context of document image processing, document degradation may refer to e.g. the presence of relatively large areas of ink residue, dust and/or other foreign materials captured during the digitisation process. Other issues that may impede document image analysis include poorly contrasted images, as well as relatively poorly defined pen strokes due to e.g. an excessive variation in pen pressure during signature production.

²It should be clear that *optimal* performance of the preceding algorithms, aimed at image binarisation and noise reduction, would obviate the need for an additional technique aimed toward pen stroke repair.

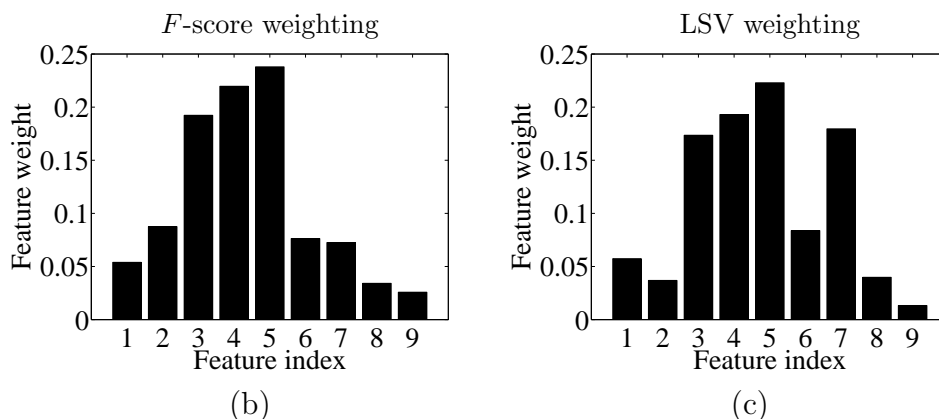


Figure 7.1: Average feature weights reported in Swanepoel and Coetzer (2014), as determined using (a) the F -score and (b) the linear support vector weighting methods. The feature indices correspond to the columns of the feature set considered, that is $\mathbf{X} = [\mathbf{p}, \mathbf{x}, \mathbf{y}, \dot{\mathbf{x}}, \dot{\mathbf{y}}, \ddot{\mathbf{x}}, \ddot{\mathbf{y}}, \theta_x, \theta_y]$. Although both strategies clearly identify the vertical position \mathbf{y} , horizontal velocity $\dot{\mathbf{x}}$ and vertical velocity $\dot{\mathbf{y}}$ (indices 3–5) as highly discriminative, it should be noted that substantially different weights are associated with the horizontal position \mathbf{x} and the vertical acceleration $\ddot{\mathbf{y}}$ (indices 2 and 7).

This expected variability in the discriminative potential associated with different features is, however, not exploited by the systems developed in this study.

This potential shortcoming in the signature modelling framework proposed in this study may of course be addressed through the incorporation of *feature weighting* into the model construction process. An investigation into the potential gain in system proficiency, as a result of utilising a feature weighted signature modelling framework, is therefore deemed warranted. Furthermore, the concept of feature weighting may easily be incorporated into the systems developed in this study. For instance, when an SVM is employed for signature modelling, Chang and Lin (2008) explain that feature weighting may be achieved by simply introducing a *weight factor* into the associated kernel function.

An initial investigation into the feasibility and significance of feature weighted *on-line* signature modelling was launched during the latter stages of this study and may be found in Swanepoel and Coetzer (2014). In this paper, a set of 9 features is considered for signature representation, whilst two fundamentally different weighting strategies, namely the F -score (Duda *et al.* (1999)) and the *linear support vector* weighting method (Chang and Lin (2008)), are investigated. The impact of incorporating the feature weights yielded by these two strategies, into an SVM-based on-line signature model, is determined experimentally and these initial findings are presented as a *proof of concept*.

The findings presented in Swanepoel and Coetzer (2014) yields two notable results. Firstly, it is shown that the utilisation of different weighting strategies may lead to substantially different feature weight values, as is evident from Figure 7.1. This result confirms that the selection of a suitable weighting strategy is not a trivial task, but in fact demands serious consideration. More importantly, however, it is shown that the proposed *feature*

weighted SVM-based model consistently outperforms its *conventional* SVM-based counterpart. This result confirms that the concept of feature weighting, within the context of writer-independent signature modelling, represents an avenue of research worth further investigation.

Although consistent, the improvement in verification proficiency reported in Swanepoel and Coetzer (2014) is relatively modest. This paper does, however, identify several topics deemed warranted for continued investigation. For instance, in order to maximally exploit the potential of the proposed feature weighted approach, the utilisation of an expanded³ feature set is suggested. In addition, an investigation into alternative, more sophisticated feature weighting strategies is also suggested. Candidate strategies include, amongst others, the FSDD feature ranking algorithm proposed in Liang *et al.* (2008), as well as the *receiver operating characteristic* (ROC) based approach proposed in Zhang *et al.* (2009). A continuation of the initial investigation presented in Swanepoel and Coetzer (2014) is currently underway.

It should be clear from the discussion above that, within the context of this study, the incorporation of feature weighting into the modelling framework is *only* applicable to the *on-line* verification systems, since only these systems consider a set of independent, fundamentally different features for signature representation. This is not the case for the off-line systems, where each feature represents a DRT-based projection profile associated with a different projection angle. It is not reasonable to expect that, specifically within a *writer-independent* framework, the projection profiles associated with a specific set of predetermined angles will *always* prove more discriminative than others. It is of course not impossible, nor inadvisable, to incorporate the concept of feature weighting into an off-line signature modelling framework. Such an endeavour will, however, necessitate the use of an alternative feature extraction technique, since the DRT-based method utilised in this study is not deemed suitable for this purpose.

7.2.3 Confidence weighted score fusion

In the previous section we discussed the expected benefits of incorporating *feature* weighting into the signature modelling framework. This concept of assigning varying levels of importance to different data sources is, however, by no means limited to features.

Consider, for example, the signature verification protocol utilised by the systems developed in this study. In Section 5.3 we explained that, when presented with a questioned signature sample, a trained model yields as output a set of *signed distance measures* relative to the corresponding decision boundary, where each measure is obtained from a comparison between the questioned sample and a known genuine reference sample. In addition, we explained how such a distance measure may subsequently be converted into a *confidence score* $s \in [0, 1]$, where confidence values of 1, 0 and 0.5 are associated with certain

³The set of nineteen features considered by the on-line signature verification systems developed in this dissertation already constitutes a noteworthy feature set expansion, when compared to the set of nine features considered in Swanepoel and Coetzer (2014). Of course, many additional features are available for potential incorporation into an even further expanded feature set.

acceptance, certain rejection and total ambiguity, respectively.

Furthermore, we explained that the systems developed in this study combine the set of model outputs by means of score averaging, as opposed to the more commonly used majority voting strategy (Santos *et al.* (2004)). This approach is able to incorporate additional information relating to classifier certainty, into the decision-making process, thereby yielding a more comprehensive and reliable result. The available information regarding classifier certainty may, however, be exploited even further if fusion is achieved by employing *weighted* score averaging, where the weight associated with each classification event is related to the confidence score obtained from said classification.

The concept of confidence weighted information fusion has previously been proposed in the literature, predominantly in the field of *multi-sensor fusion* (Elmenreich (2007); Apartsin (2012)), that is the fusion of measurement samples from multiple sensors, in order to obtain a dependable estimation of the variable/entity considered for measurement. However, to the best of our knowledge, confidence weighted *classifier score fusion* has not been investigated within the context of *pattern authentication*.

It is therefore our intention to develop a novel confidence weighted score fusion protocol. Furthermore, we propose that the weight associated with each classification be determined directly from its corresponding confidence score s , by means of a *confidence weighting function* $f(s) \in [0, 1]$. To the best of our knowledge, the utilisation of such a function, for the purpose of determining a suitable confidence weighting value, has not yet been proposed in the literature. Although the weighting function may be chosen arbitrarily, any suitable function would in all likelihood possess the following important properties: (1) In order to remain unbiased toward either acceptance or rejection, the proposed function should be symmetric with respect to the line $s = 0.5$; (2) In order to yield sensible weight values, $f(s)$ should be monotonically *decreasing* in the interval $s \in [0, 0.5]$ and monotonically *increasing* in the interval $s \in [0.5, 1]$. Such a function would therefore maximise the contribution of any classification made with relatively high certainty, whilst classifications associated with total ambiguity would, in essence, be discarded.

Several functions that possess these desired properties are presented in Figure 7.2. Although countless other functions may exist that also satisfy the aforementioned requirements, the candidate functions illustrated in Figure 7.2 are deemed fit for investigation within an initial study.

It is clear from the discussion presented in this section that the implementation of the proposed classifier score fusion strategy, that is the substitution of the currently utilised score averaging strategy with a weighted average that incorporates a weighting function, does not represent a particularly daunting research challenge. However, it should also be clear that the success of the proposed strategy is entirely dependent on the optimality of the confidence scores yielded by each classification. As explained in Section 5.3, the systems developed in this study convert model outputs into confidence scores by means of the conventional logistic function. Although this method has proved successful in this study, it is by no means considered optimal. Furthermore, the identification of an optimal weighting function remains a problem that demands further investigation.

An initial investigation into the feasibility and significance of an efficient confidence

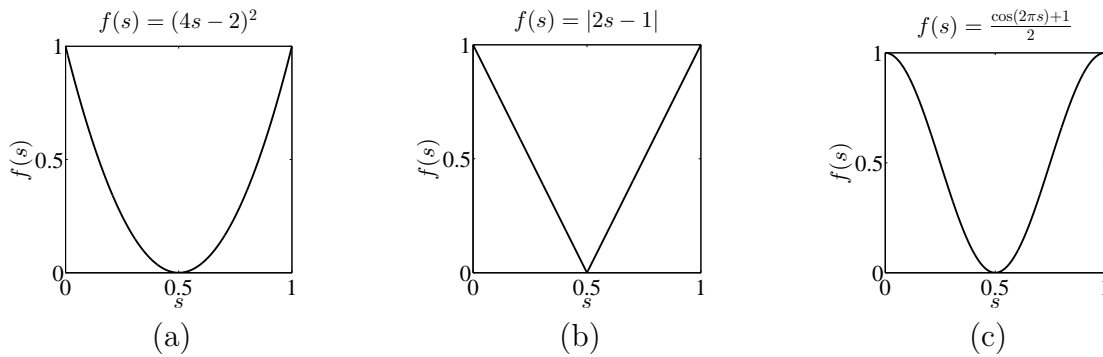


Figure 7.2: Confidence weighting function candidates. It should be clear that these functions may intuitively be associated with a (a) conservative, (b) neutral, and (c) liberal approach to weight assignment, respectively.

weighted classifier score fusion protocol, within the context of handwritten signature verification, is currently underway.

7.2.4 Other applications

Although we demonstrated that the pattern representation, modelling and verification techniques presented in this study are well suited for the development of a proficient handwritten signature verification system, the application of these methods is in no way limited to this specific field of research. It is in fact entirely reasonable to expect that the systems developed in this study may, subject to relatively minor modifications, provide effective solutions to a variety of other problems in the wider field of pattern recognition.

Consider, for example, the problem of *text-dependent speaker verification*, that is the automatic authentication of a person's claimed identity from an audio sample representative of a *specific* vocal utterance. Although the acoustic information represented in an audio sample differs significantly from the pen stroke information associated with a handwritten signature sample, the underlying pattern recognition problem is similar to that of *on-line* signature verification⁴. The on-line systems developed in this study may therefore, subject to the development of a modified feature extraction process aimed toward acoustic signal analysis, provide an effective tool for the modelling and verification of *speaker-independent* vocal utterances. Furthermore, if several independent features are considered for pattern representation, the utilisation of a *weighted* SVM may prove particularly effective.

Another potential field of application is that of *handwritten word spotting*, that is the automatic retrieval of a specific *target word* within a digitised handwritten manuscript. Following the successful segmentation of each word represented within a given manuscript, the problem of either accepting or rejecting each of these questioned words as the target word constitutes a verification task similar to that of *off-line* signature verification. Fur-

⁴Despite the fact that the source data associated with text-dependent speaker verification and on-line signature verification represent fundamentally different information, both types of data comprise time series information of variable length, and both are associated with a specific predefined gesture.

thermore, in order to construct a suitable model for the target word, one or more known samples of said word is required. The modelling of a specific target word, produced by a specific writer, may therefore be considered analogous to the construction of a writer-dependent handwritten signature model. It may therefore prove interesting to deploy the off-line systems developed in this study, without any system design modifications, within the context of *target-independent* handwritten word spotting. It should be clear, however, that if no design modifications whatsoever are made to the off-line signature verification systems developed in this study, the resulting word spotting systems may prove proficient in *either* the retrieval of *arbitrary words* produced by a *specific writer*, or the retrieval of a *specific word* produced by *arbitrary writers*, but not both. If a word spotting system is to be developed that is both writer-independent *and* target-independent, it is entirely possible that the model construction process proposed in this study may require several modifications, specifically aimed toward addressing the shape variations expected for a specific word – when this word may be produced by one of several potential writers.

Finally, we are confident that the applications described above represent only a small portion of a wide variety of additional applications that will undoubtedly be identified in the future.

Bibliography

- Ahmed, S., Malik, M., Liwicki, M. and Dengel, A. (2012). Signature segmentation from document images. *International Conference on Frontiers in Handwriting Recognition*, pp. 425–429.
- Alonso-Fernandez, F., Fairhurst, M., Fierrez, J. and Ortega-Garcia, J. (2007). Impact of signature legibility and signature type in off-line signature verification. *Biometrics Symposium*, pp. 1–6.
- Apartsin, A. (2012). *Fusion of Biased Estimators Using Machine Learning for Time of Flight Estimation in the Presence of Outliers*. Ph.D. thesis, Tel Aviv University.
- Azmi, A. and Nasien, D. (2014). Freeman chain code (FCC) representation in signature fraud detection based on nearest neighbour and artificial neural network (ANN) classifiers. *International Journal of Image Processing*, vol. 8, no. 6, pp. 434–454.
- Batista, L., Granger, E. and Sabourin, R. (2010). Applying dissimilarity representation to off-line signature verification. *International Conference on Pattern Recognition*, pp. 1293–1297.
- Ben-Hur, A. and Weston, J. (2010). A users guide to support vector machines. In: *Data mining techniques for the life sciences*, pp. 223–239. Springer.
- Beylkin, G. (1987). Discrete radon transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 2, pp. 162–172.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bouveyron, C., Girard, S. and Schmid, C. (2007). High-dimensional discriminant analysis. *Communications in Statistics - Theory and Methods*, vol. 36, no. 14, pp. 2607–2623.
- Brockly, M., Elliott, S., Burdine, J., Frost, M., Riedle, M. and Guest, R. (2014). An investigation into biometric signature capture device performance and user acceptance. *International Carnahan Conference on Security Technology*, pp. 1–5.
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167.

- Business Day (2009 February). Beware fraud if you pay by cheque. Y. Gerakaris.
- Chang, Y. and Lin, C. (2008). Feature ranking using linear SVM. *Causation and Prediction Challenge: Challenges in Machine Learning*, vol. 2, pp. 47–57.
- Cheque & Credit Clearing Company (2014 March). Cheque fraud.
- Cherkassky, V. and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, vol. 17, no. 1, pp. 113–126.
- Coetzer, J. (2005). *Off-line Signature Verification*. Ph.D. thesis, Stellenbosch University.
- Coetzer, J., Herbst, B. and du Preez, J. (2004). Offline signature verification using the discrete Radon transform and a hidden Markov model. *European Association for Signal Processing Journal on Applied Signal Processing*, vol. 4, pp. 559–571.
- Coetzer, J., Herbst, B. and du Preez, J. (2006). Off-line signature verification: A comparison between human and machine performance. *International Workshop on Frontiers in Handwriting Recognition*.
- Coetzer, J., Swanepoel, J. and Sabourin, R. (2012). Efficient cost-sensitive human-machine collaboration for offline signature verification. *IS&T/SPIE Electronic Imaging*, pp. 82970J-1–82970J-8.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, vol. 20, no. 3, pp. 273–297.
- Cressie, N. and Whitford, H. (1986). How to use the two sample *t*-test. *Biometrical Journal*, vol. 28, no. 2, pp. 131–148.
- Dolfing, J., Aarts, E. and van Oosterhout, J. (1998). On-line signature verification with hidden Markov models. *International Conference on Pattern Recognition*, vol. 2, pp. 1309–1312.
- Duda, R., Hart, P. and Stork, D. (1999). *Pattern classification*. John Wiley & Sons.
- Dumas, B., Pugin, C., Hennebert, J., Petrovska-Delacrétaz, D., Humm, A., Evéquoq, F., Ingold, R. and Von Rotz, D. (2005). Myidea-multimodal biometrics database, description of acquisition protocols. *European Co-Operation in the Field of Scientific and Technical Research: Biometrics on the Internet*, pp. 59–62.
- El-Henawy, I., Rashad, M., Nomir, O. and Ahmed, K. (2013). Online signature verification: State of the art. *International Journal of Computers & Technology*, vol. 4, no. 2, pp. 664–678.
- Elmenreich, W. (2007). Fusion of continuous-valued sensor measurements using confidence-weighted averaging. *Journal of Vibration and Control*, vol. 13, no. 9–10, pp. 1303–1312.

- Eskander, G., Sabourin, R. and Granger, E. (2012). Adaptation of writer-independent systems for offline signature verification. *International Conference on Frontiers in Handwriting Recognition*.
- Fang, B., Leung, C., Tang, Y., Tse, K., Kwok, P. and Wong, Y. (2003). Off-line signature verification by the tracking of feature and stroke positions. *Pattern Recognition*, vol. 36, no. 1, pp. 91–101.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, vol. 27, pp. 861–874.
- Ferrer, M., Vargas, F., Morales, A. and Ordóñez, A. (2012). Robustness of off-line signature verification based on gray level features. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 966–977.
- Fierrez-Aguilar, J., Alonso-Hermira, N., Moreno-Marquez, G. and Ortega-Garcia, J. (2004). An off-line signature verification system based on fusion of local and global information. In: Maltoni, D. and Jain, A. (eds.), *Biometric Authentication*, pp. 295–306. Springer Berlin Heidelberg.
- Fierrez-Aguilar, J., Krawczyk, S., Ortega-Garcia, J. and Jain, A. (2005). Fusion of local and regional approaches for on-line signature verification. In: *Advances in biometric person authentication*, pp. 188–196. Springer.
- Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188.
- Freire, M., Fierrez-Aguilar, J., Martinez-Diaz, M. and Ortega-Garcia, J. (2007). On the applicability of off-line signatures to the fuzzy vault construction. *International Conference on Document Analysis and Recognition*, pp. 1173–1177.
- Friedman, J. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175.
- Fuentes, M., Garcia-Salicetti, S. and Dorizzi, B. (2002). On line signature verification: Fusion of a hidden Markov model and a neural network via a support vector machine. *International Workshop on Frontiers in Handwriting Recognition*, pp. 253–258.
- Galbally, J., Fierrez, J., Martinez-Diaz, M. and Ortega-Garcia, J. (2009). Improving the enrollment in dynamic signature verification with synthetic samples. *International Conference on Document Analysis and Recognition*, pp. 1295–1299.
- Garcia-Salicetti, S., Alonso-Fernandez, J.F.-A.F., Vielhauer, C., Guest, R., Allano, L., Doan Trung, T., Scheidat, T., Ly Van, B., Dittman, J., Dorizzi, B., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bacile di Castiglione, M. and Fairhurst, M. (2007). Biosecure reference systems for on-line signature verification: A study of complementarity. *Annals of Telecommunications: Special Issue on Multimodal Biometrics*, pp. 36–61.

- Gilperez, A., Alonso-Fernandez, F., Pecharroman, S., Fierrez, J. and Ortega-Garcia, J. (2008). Off-line signature verification using contour features. *International Conference on Frontiers in Handwriting Recognition*.
- Gonzales, R. and Woods, R. (2002). *Digital Image Processing*. 2nd edn. Prentice-Hall Inc.
- Guo, X., Yang, J. and Liang, C.W.Y. (2008). A novel LS-SVMs hyper-parameter selection based on particle swarm optimization. *Neurocomputing*, vol. 71, no. 16, pp. 3211–3215.
- Guyon, I., Vapnik, V., Boser, B. and Solla, A. (1992). Structural risk minimization for character recognition. *Advances in Neural Information Processing Systems*, vol. 4, pp. 471–479.
- Henniger, O. and Muller, S. (2007). Effects of time normalization on the accuracy of dynamic time warping. *International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–6.
- Ho, T. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844.
- Hou, W., Ye, X. and Wang, K. (2004). A survey of off-line signature verification. *International Conference on Intelligent Mechatronics and Automation*, pp. 536–541.
- Houmani, N., Garcia-Salicetti, S. and Dorizzi, B. (2012). On measuring forgery quality in online signatures. *Pattern Recognition*, vol. 45, no. 3, pp. 1004–1018.
- Hsu, C.-W., Chang, C.-C. and Lin, C.-J. (2003). A practical guide to support vector classification. Tech. Rep., National Taiwan University.
- Huang, C.-L. and Wang, C.-J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert systems with applications*, vol. 31, no. 2, pp. 231–240.
- Ibrahim, M., Alimgeer, K., Kahn, M. and Taj, I. (2007). Creation and selection of most stable discriminating features for on-line signature verification. *IEEE International Conference on Machine Vision*, pp. 97–101.
- Impedovo, D. and Pirlo, G. (2008). Automatic signature verification: the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 5, pp. 609–635.
- Impedovo, D., Pirlo, G. and Plamondon, R. (2012). Handwritten signature verification: New advancements and open issues. *International Conference on Frontiers in Handwriting Recognition*, pp. 365–370.
- Jain, A., Duin, R. and Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37.

- Jain, A., Nandakumar, K. and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, vol. 38, pp. 2270–2285.
- Jayadevan, R., Kohe, S. and Patil, P. (2009). Dynamic time warping based static hand printed signature verification. *Pattern Recognition Research*, vol. 4, no. 1, pp. 52–65.
- Kahn, M., Kahn, M. and Kahn, M. (2006). Velocity-image model for online signature verification. *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3540–3549.
- Kalera, M., Srihari, S. and Xu, A. (2004). Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 7, pp. 1339–1360.
- Kan, K. and Shelton, C. (2008). Catenary support vector machines. In: *Machine Learning and Knowledge Discovery in Databases*, pp. 597–610. Springer.
- Keogh, E. and Pazzani, M. (2001). Derivative dynamic time warping. *SIAM International Conference on Data Mining*, pp. 1–11.
- Ketabdar, H., Richiardi, J. and Drygajlo, A. (2005). Global feature selection for on-line signature verification. *IGS Conference*.
- Klecka, W. (1980). *Discriminant analysis*. Sage.
- Kumar, R., Sharma, J. and Chanda, B. (2011). Writer-independent off-line signature verification using surroundedness feature. *Pattern Recognition Letters*, vol. 33, no. 3, pp. 301–308.
- Lazzara, G. and Thierry, G. (2014). Efficient multiscale Sauvola’s binarization. *International Journal on Document Analysis and Recognition*, vol. 17, no. 2, pp. 105–123.
- Le Riche, P. (2000). *Handwritten signature verification: a hidden Markov model approach*. Master’s thesis, Stellenbosch University.
- Leclerc, F. and Plamondon, R. (1994). Automatic signature verification: the state of the art – 1989–1993. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 8, no. 3, pp. 643–660.
- Liang, J., Yang, S. and Winstanley, A. (2008). Invariant optimal feature selection: A distance discriminant and feature ranking based solution. *Pattern Recognition*, vol. 41, no. 5, pp. 1429–1439.
- Lin, S.-W., Lee, Z.-J., Chen, S.-C. and Tseng, T.-Y. (2008). Parameter determination of support vector machine and feature selection using simulated annealing approach. *Applied Soft Computing*, vol. 8, no. 4, pp. 1505–1512.

- Lumini, A. and Nanni, L. (2009). Ensemble of on-line signature matchers based on Over-Complete feature generation. *Expert Systems with Applications*, vol. 36, no. 3, pp. 5291–5296.
- Mandal, R., Roy, P. and Pal, U. (2011). Signature segmentation from machine printed documents using conditional random field. *International Conference on Document Analysis and Recognition*, pp. 1170–1174.
- Mattera, D. and Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In: *Advances in kernel methods*, pp. 211–241. MIT Press.
- McLachlan, G. (2004). *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons.
- Min, J. and Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, vol. 28, no. 4, pp. 603–614.
- Momma, M. and Bennett, K. (2002). A pattern search method for model selection of support vector regression. *SIAM International Conference on Data Mining*.
- Montalvão, J., Houmani, N. and Dorizzi, B. (2010). Comparing GMM and Parzen in automatic signature recognition – a step backward or forward? *Congresso Brasileiro de Automática*, pp. 4463–4468.
- Muramatsu, D. and Matsumoto, T. (2009). Online signature verification algorithm with a user-specific global-parameter fusion model. *IEEE International Conference on Systems, Man and Cybernetics*, pp. 486–491.
- Nanni, L. and Lumini, A. (2005). Ensemble of Parzen window classifiers for on-line signature verification. *Neurocomputing*, vol. 68, pp. 217–224.
- Nanni, L. and Lumini, A. (2008). A novel local on-line signature verification system. *Pattern Recognition Letters*, vol. 29, pp. 559–568.
- Nguyen, V. (2012). *Off-line Signature Verification using Novel Feature Extraction Techniques and Trajectory Recovery*. Ph.D. thesis, Griffith University.
- Nicolaou, A., Liwicki, M. and Ingolf, R. (2013). Oriented local binary patterns for writer identification. *International Workshop on Automated Forensic Handwriting Analysis*, pp. 15–20.
- Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.-J., Vivaracho, C., Escudero, D. and Moro, Q.-I. (2003). MCYT baseline corpus: a bimodal biometric database. *IEE Proceedings on Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401.

- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66.
- Pal, S., Blumenstein, M. and Pal, U. (2011). Automatic off-line signature verification systems: A review. *International Journal of Computer Applications*, vol. 14, pp. 20–27.
- Panton, M. and Coetzer, J. (2010). Off-line signature verification using ensembles of local Radon transform-based HMMs. *Annual Symposium of the Pattern Recognition Association of South Africa*, pp. 201–206.
- Pekalska, E., Paclik, P. and Duin, R. (2002). A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research*, vol. 2, pp. 175–211.
- Piyush Shanker, A. and Rajagopalan, A. (2007). Off-line signature verification using DTW. *Pattern Recognition Letters*, vol. 28, no. 12, pp. 1407–1414.
- Plamondon, R. and Lorette, G. (1989). Automatic signature verification and writer identification – the state of the art. *Pattern Recognition*, vol. 22, no. 2, pp. 107–131.
- Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Microsoft Research*.
- Prakash, H. and Guru, D. (2009). Geometric centroids and their relative distances for off-line signature verification. *International Conference on Document Analysis and Recognition*, pp. 121–125.
- Pudil, P., Novovičová, J. and Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125.
- Rabiner, L. and Schmidt, C. (1980). Application of dynamic time warping to connected digit recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 377–388.
- Radon, J. (1917). Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, vol. 69, no. 262–277.
- Radon, J. (1986). On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, vol. 5, no. 4, pp. 170–176. Translated by P. Parks from the original German text.
- Sabourin, R., Plamondon, R. and Lorette, G. (1992). Off-line identification with handwritten signature images: Survey and perspectives. In: *Structured Document Image Analysis*, pp. 219–234. Springer.

- Sae-Bae, N. and Memon, N. (2013). A simple and effective method for online signature verification. *International Conference of the Biometrics Special Interest Group*, pp. 147–158.
- Santos, C., Justino, E., Bortolozzi, F. and Sabourin, R. (2004). An off-line signature verification system based on the questioned document expert's approach and a neural network classifier. *International Workshop on Frontiers in Handwriting Recognition*, pp. 498–502.
- Schmidt, C. (1994). Signature verification using time-delay neural networks. *Midwest Symposium on Circuits and Systems*, vol. 37, no. 2, pp. 1395–1398.
- Sindle, C. (2003). *Handwritten Signature Verification using Hidden Markov Models*. Master's thesis, Stellenbosch University.
- Singh, S. and Kaur, A. (2014). Off-line signature verification using sub uniform local binary patterns and support vector machine. *International Conference on Chemical Engineering & Advanced Computational Technologies*, pp. 93–100.
- Singh, T., Roy, S., Singh, O., Sinam, T. and Singh, K. (2011). A new local adaptive thresholding technique in binarization. *International Journal of Computer Science Issues*, vol. 8, no. 2, pp. 271–277.
- Snelick, R., Uludag, U., Mink, A., Indovina, M. and Jain, A. (2005). Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 450–455.
- Staelin, C. (2003). Parameter selection for support vector machines. Tech. Rep. HPL-2002-354R1, Hewlett-Packard Company.
- Suykens, J. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300.
- Swanepoel, J. and Coetzer, J. (2010). Off-line signature verification using flexible grid features and classifier fusion. *International Conference on Frontiers in Handwriting Recognition*, pp. 297–302.
- Swanepoel, J. and Coetzer, J. (2012). Writer-specific dissimilarity normalisation for improved writer-independent off-line signature verification. *International Conference on Frontiers in Handwriting Recognition*, pp. 391–396.
- Swanepoel, J. and Coetzer, J. (2013). A robust dissimilarity representation for writer-independent signature modelling. *IET Biometrics*, vol. 2, no. 4, pp. 159–168.
- Swanepoel, J. and Coetzer, J. (2014). Feature weighted support vector machines for writer-independent on-line signature verification. *International Conference on Frontiers in Handwriting Recognition*, pp. 434–439.

- Toft, P. (1996). *The Radon Transform: Theory and Implementation*. Ph.D. thesis, Technical University of Denmark.
- Ulusoy, I. and Bishop, C. (2005). Generative versus discriminative methods for object recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 258–265.
- Van, B., Garcia-Salicetti, S. and Dorizzi, B. (2004). Fusion of HMMs likelihood and Viterbi path for on-line signature verification. In: *Biometric Authentication*, pp. 318–331. Springer.
- Van, B., Garcia-Salicetti, S. and Dorizzi, B. (2007). On using the Viterbi path along with HMM likelihood information for online signature verification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 5, pp. 1237–1247.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.
- Vargas, F., Ferrer, M., Travieso, C. and Alonso, J. (2007). Off-line handwritten signature GPDS-960 corpus. *International Conference on Document Analysis and Recognition*, pp. 764–768.
- Vargas, J., Ferrer, M., Travieso, C. and Alonso, J. (2011). Off-line signature verification based on gray level information using texture features. *Pattern Recognition*, vol. 44, no. 2, pp. 375–385.
- Watanabe, S. (1985). *Pattern recognition: Human and Mechanical*. John Wiley & Sons, Inc.
- Wen, J., Fang, B., Tang, Y. and Zhang, T. (2009). Model-based signature verification with rotation invariant features. *Pattern Recognition*, vol. 42, no. 7, pp. 1458–1466.
- Wibowo, C., Thumwarin, P. and Matsuura, T. (2013). On-line signature verification based on angles of pen-motion. *International Conference on Simulation Technology*.
- Yanikoglu, B. and Kholmatov, A. (2009). Online signature verification using fourier descriptors. *European Association for Signal Processing Journal on Advances in Signal Processing*.
- Yeung, D.-Y., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T. and Rigoll, G. (2004). SVC2004: First international signature verification competition. *International Conference on Biometric Authentication*, pp. 16–22.
- Zhang, S., Maruf Hossain, M., Rafiul Hassan, M., Bailey, J. and Ramamohanarao, K. (2009). Feature weighted SVMs using receiver operating characteristics. *SIAM International Conference on Data Mining*, pp. 497–508.

- Zhang, Y. and Wu, L. (2011). Fast document image binarization based on an improved adaptive Otsu's method and destination word accumulation. *Journal of Computational Information Systems*, vol. 7, no. 6, pp. 1886–1892.
- Zhu, G., Zheng, Y., Doermann, D. and Jaeger, S. (2007). Multi-scale structural saliency for signature detection. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.

Appendix A

Dynamic Time Warping

As mentioned in Section 3.4, there are currently several variants of the DTW-algorithm described in the literature. In this appendix, we discuss the *specific* algorithm utilised by the systems developed in this study. For a graphical conceptualisation of this algorithm, the reader is referred to Figure 3.10 in Section 3.4.

A.1 Algorithm

In order to obtain a DTW-based distance between two vectors \mathbf{x}_q and \mathbf{x}_k , which is denoted by $\mathcal{D}(\mathbf{x}_k, \mathbf{x}_q)$, a *distance grid* is first constructed where each node (i, j) relates element i of \mathbf{x}_q to element j of \mathbf{x}_k . For each such node, the distance

$$D_{\text{node}}(i, j) = (\mathbf{x}_q(i) - \mathbf{x}_k(j))^2 \quad (\text{A.1})$$

is computed, which is said to reflect the *node-based* cost associated with the elements $\mathbf{x}_q(i)$ and $\mathbf{x}_k(j)$. The *optimal path* through this distance grid, that terminates at node (i, j) , is subsequently defined as the path $(i_0, j_0), (i_1, j_1), \dots, (i_K, j_K)$ for which the *total* node-based cost

$$D_{\text{node}}^{(\text{compl})}(i, j) = \sum_{k=0}^K D_{\text{node}}(i_k, j_k) \quad (\text{A.2})$$

is minimised. Furthermore, several constraints are imposed on the solution space, in order to ensure that the resulting optimal path is in fact a valid path.

Firstly, it is required that the optimal path is *complete*. In other words, for two feature vectors $\mathbf{x}_q \in \mathfrak{R}^m$ and $\mathbf{x}_k \in \mathfrak{R}^n$, $\mathbf{x}_q(1)$ must always be related to $\mathbf{x}_k(1)$, whilst $\mathbf{x}_q(m)$ must always be related to $\mathbf{x}_k(n)$. This is achieved by the requirement

$$(i_0, j_0) = (1, 1), \quad (\text{A.3})$$

$$(i_K, j_K) = (m, n). \quad (\text{A.4})$$

Note that it is clearly not required that \mathbf{x}_q and \mathbf{x}_k have the same dimension.

Secondly, it is required that the optimal path be *monotonically increasing*. In other words, a node (i, j) may therefore only be considered for inclusion into the optimal path if it is preceded by the node $(i - 1, j - 1)$, $(i, j - 1)$ or $(i - 1, j)$. Consequently, it follows that

$$i_k \geq i_{k-1}, \quad k = 2, 3, \dots, K, \quad (\text{A.5})$$

$$j_k \geq j_{k-1}, \quad k = 2, 3, \dots, K. \quad (\text{A.6})$$

As a result, any node that fails to satisfy (A.5) and (A.6) does not form part of a valid path and can therefore not form part of the optimal path. Note that, since it is not demanded that the optimal path be *strictly increasing*, but simply *non-decreasing*, this constraint facilitates the consideration of *one-to-many* element correspondences.

Finally, when constructing the optimal path, it is required that each included node is located within a predetermined *bandwidth* β around the diagonal, such that

$$|j_k - i_k| \leq \beta, \quad k = 0, 1, \dots, K. \quad (\text{A.7})$$

This bandwidth may be chosen arbitrarily. However, since this value has a major influence on the practicality of the obtained solution, careful consideration is required in selecting a sensible value. Firstly, it ensures that components with exceedingly different indices are not related, which would of course not be sensible. Secondly, it limits the computational cost associated with the DTW-algorithm. It should be clear that, for feature vectors of a very high dimension, the construction of a complete cost grid could become computationally exhaustive. Note that when $\beta = 0$, the optimal path is restricted to the diagonal $i = j$, which is equivalent to computing the Euclidean distance between \mathbf{x}_k and \mathbf{x}_q .

Before we discuss the DTW-based algorithm for obtaining the *complete optimal path* and its associated *complete optimal cost*, it is necessary to define two key concepts. Let $D_{\text{node}}^{(\text{part})}(i, j)$ denote the *partial optimal path* that terminates at node (i, j) . Also, let $\leftarrow (i, j)$ denote the *optimal preceding node* for node (i, j) . A preceding node is deemed optimal if the partial optimal path that passes through it and terminates at (i, j) minimises the *partial optimal cost*.

The complete procedure for finding the optimal path and corresponding DTW-based distance by means of the DTW-algorithm can now be stated as follows:

- **Initialisation:**

$$D_{\text{node}}^{(\text{part})}(1, 1) = D_{\text{node}}(1, 1). \quad (\text{A.8})$$

- **Recursion:**

All nodes within the allotted bandwidth are considered in a left-to-right, bottom-to-top fashion. For each node considered, $\leftarrow (i, j)$ is computed as

$$D_1 = D_{\text{node}}^{(\text{part})}(i - 1, j - 1), \quad (\text{A.9})$$

$$D_2 = D_{\text{node}}^{(\text{part})}(i, j - 1), \quad (\text{A.10})$$

$$D_3 = D_{\text{node}}^{(\text{part})}(i - 1, j), \quad (\text{A.11})$$

$$\leftarrow (i, j) = \operatorname{argmin}\{D_1, D_2, D_3\}. \quad (\text{A.12})$$

If the minimum value for D_i is shared by more than one of the preceding nodes, $\leftarrow (i, j)$ is selected in the following order of preference: $(i - 1, j - 1)$, then $(i, j - 1)$, else $(i - 1, j)$. Subsequently, $D_{\text{node}}^{(\text{part})}(i, j)$ is computed as

$$D_{\text{node}}^{(\text{part})}(i, j) = D_{\text{node}}(i, j) + D_{\text{node}}^{(\text{part})}(\leftarrow (i, j)). \quad (\text{A.13})$$

- **Path backtracking:**

As ensured by (A.3) and (A.4), nodes (i_0, j_0) and (i_K, j_K) of the optimal path are reserved by $(1, 1)$ and (m, n) respectively. The remainder of the optimal path may be obtained through backtracking from node (i_K, j_K) by iteratively letting

$$(i_k, j_k) = \leftarrow (i_{k+1}, j_{k+1}). \quad (\text{A.14})$$

- **Termination:**

$$\mathcal{D}(\mathbf{x}_k, \mathbf{x}_q) = D_{\text{node}}^{(\text{compl})}(\mathbf{x}_q, \mathbf{x}_k) \quad (\text{A.15})$$

$$= D_{\text{node}}^{(\text{part})}(m, n). \quad (\text{A.16})$$

In short, the DTW-algorithm therefore computes the Euclidean distance between two vectors whose elements are first non-linearly and optimally aligned through the use of dynamic programming.