



Kent Academic Repository

Kaplani, Eleni (2003) *Human and computer-based verification of handwritten signatures*. Doctor of Philosophy (PhD) thesis, University of Kent.

Downloaded from

<https://kar.kent.ac.uk/94454/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.22024/UniKent/01.02.94454>

This document version

UNSPECIFIED

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

This thesis has been digitised by EThOS, the British Library digitisation service, for purposes of preservation and dissemination. It was uploaded to KAR on 25 April 2022 in order to hold its content and record within University of Kent systems. It is available Open Access using a Creative Commons Attribution, Non-commercial, No Derivatives (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) licence so that the thesis and its author, can benefit from opportunities for increased readership and citation. This was done in line with University of Kent policies (<https://www.kent.ac.uk/is/strategy/docs/Kent%20Open%20Access%20policy.pdf>). If you ...

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Human and Computer –Based Verification of Handwritten Signatures

A Thesis Submitted to The University of Kent at Canterbury
For The Degree Of Doctor of Philosophy
In Electronic Engineering

By

Eleni Kaplani

May 2003

στους γονείς μου,

to my parents,

Acknowledgements

Now it is time for me to look back into these three and a half years in appraisal of their worth, for which I must say the outcome is invaluable. It has been a wonderful journey, with its highs and lows, in a constant quest for knowledge. I am left fascinated by a realisation of the extent of the power of thought, the delicate process of it maturing and the final stage of enlightenment that succeeds. I was taught that there is always an answer to every troubling question, and the secret has been search, endeavour and devotion. Nevertheless, there are also a number of people who contributed their part into making this thesis a reality, whom I would like to hereby thank.

First of all, I would like to express my deep appreciation and gratitude to my supervisor, Prof. Mike Fairhurst, for warmly accepting me for my postgraduate studies, for his valuable guidance and share of expertise over these past years, and for offering continuous support and encouragement. Not to forget his friendly advice in moments of trouble. For all these I am very grateful. His words, that *each step of the way was necessary for me to get here*, will always accompany me in life.

Further, I would like to thank all the people in the Digital Systems Research Group, current and past members, including: Dr. Chris Allgrove, Dr. Will Cobbah, Mrs Julia Chapran, Dr. Richard Guest, Dr. Sam Chindaro, Dr. Kostas Sirlantzis and Dr. Sanaul Hoque, for friendly conversations and for baring several of my human perception experiments, and finally, our lab technician Mr. Brian Havell. Also, many thanks to Mr. Jose Casal-Gimenez, Mr. Emmanuel Ogakwu, Ms. Sue Rumbelow, and to all the people – UKC and non-UKC members – who participated in my experiments, giving me access to the fascinating world of human perception.

I would also like to thank Mr. Arda Can Kumbaracibasi for his understanding and support and for adding some colour in my everyday life here, and Miss Victoria Kaplani for cheerful conversations and encouragement.

Last but certainly not least, I would like to express my dearest thanks and gratitude to my parents, Prof. Socrates Kaplanis and Mrs Ioanna Sivropoulou, to whom I owe what I have achieved, for all they have offered me during this on-going quest for knowledge, for supporting me, in every possible way, believing in me and for teaching me and introducing me to the world of insight. I would further like to thank them for helpful suggestions they have given me, and especially my father for valuable conversations and for his scientific sharing of knowledge. *I will always be grateful – θα σας είμαι πάντα ευγνώμων.*

Elina Kaplani

Canterbury, May 2003

Abstract

The work reported in this thesis is set within the broad field of biometrics and, especially, is concerned with the link between human and machine –based biometric processing. The use of the handwritten signature as a biometric for authenticating the identity of individuals is investigated and aspects of the handwritten signature known to influence its susceptibility to fraudulent penetration or a false rejection of its original are particularly assessed. Two approaches to handwritten signature verification are considered: through an automatic signature verification system and by means of human visual inspection. A strong emphasis is given in the processing of static signature information.

Experimental studies on human perceptual judgements regarding the complexity of signatures and their intra-class variability directed the development and assessment of automatic algorithmic solutions that reflect human perceptual criteria. A novel method is developed to quantitatively estimate the degree of a signature's complexity based on the static signature image alone. Valuable insight is gained, through a number of experiments on human visual inspection of handwritten signatures, into strategies employed by humans in analysing signatures and the key factors that affect human performance in correctly verifying signatures' authenticity, allowing predictions about the likelihood of a signature's susceptibility to forgery and proposing practical scenarios for improvement.

Finally, the integration of knowledge gained from the investigation of the advanced human perceptual capabilities in inspecting handwritten signature samples with automatic signature verification processing is investigated and a number of possible schemes are proposed. The possibility of improvement is, thus, offered with respect to both automatic identity authentication and also regarding human related procedures in inspecting and authenticating handwritten signature samples, which remains an important aspect of signature checking in common use.

Table of Contents

List of Tables.....	ix
List of Figures	xii
Chapter 1 - Introduction.....	1
1.1 Biometrics	2
1.2 Handwritten Signature	4
1.3 Signature Verification	6
1.4 Aim and Structure of Thesis	9
Chapter 2 - Literature Review	12
2.1 Automatic Signature Verification	12
2.1.1 Static Verification	14
2.1.2 Dynamic Verification	27
2.2 Visual Inspection of Signatures	44
2.3 Conclusions	45
Chapter 3 - Data Collection	47
3.1 Introduction	47
3.2 Data Acquisition.....	48
3.2.1 Digitiser	48
3.2.2 Data Format	49
3.2.3 Pen-Up Information	50
3.3 Signature Acquisition.....	52
3.3.1 Experimental Set Up.....	52
3.3.2 Experimental Procedure.....	53
3.4 Data Preprocessing	55
3.5 Conclusions	61

Chapter 4 - Automatic Handwritten Signature Verification 63

4.1 Introduction 64

4.2 Automatic Signature Verification 66

4.2.1 Signature Database 66

4.2.2 System Design 67

4.3 Static Verification 72

4.3.1 Features 73

4.3.2 System Performance 78

4.3.3 Optimisation..... 82

4.4. Dynamic Verification..... 91

4.4.1 Features 92

4.4.2 System Performance 98

4.4.3 Optimisation..... 100

4.5 Conclusions 109

Chapter 5 - Handwritten Signature Complexity 113

5.1 Introduction 113

5.2 Perceived Signature Complexity..... 115

5.2.1 Experimental Procedure..... 116

5.2.2 Analysis of Results 117

5.2.3 Perception Remarks 124

5.3 Automatic Complexity Estimation..... 125

5.3.1 Methodological Approach 126

5.3.2 Analysis of Results 132

5.4 Conclusions 135

Chapter 6 - Handwritten Signature Variability 137

6.1 Introduction 137

6.2 Perceived Signature Variability 139

6.2.1 Experimental Procedure..... 140

6.2.2 Analysis of Results 140

6.3 Objective Signature Variability..... 148

6.3.1 Methods Employed 148

6.3.2 Analysis of Results	155
6.4 Conclusions	164
Chapter 7 - Human Visual Inspection of Handwritten Signatures	166
7.1 Introduction	166
7.2 Human Verification of Signatures	171
7.2.1 Experiment 1	171
7.2.2 Experiment 2.....	178
7.2.3 Experiment 3.....	182
7.2.4 Comparative Analysis.....	186
7.3 Complexity and Variability vs Verification Performance	199
7.4 Confidence in Verifying Signatures.....	207
7.5 Conclusions	218
Chapter 8 - Combined Scheme and Conclusions.....	221
8.1 Human Inspection vs Machine Verification of Signatures	221
8.2 Human – Machine Integration Scenarios	225
8.3 Summary of Thesis	232
8.4 Discussion	236
8.5 Novelties and Contributions.....	238
8.6 Suggestions for Future Work	239
8.7 Conclusions	240
References	243
Appendixes	253
Appendix A - Human Perception Experiments.....	A-i
Appendix B - Automatic Signature Verification	B-i
Appendix C - Handwritten Signature Variability	C-i
Appendix D - Human Visual Inspection of Signatures.....	D-i

List of Tables

Table 2.1: Static signature verification systems.....	28
Table 2.1: Static signature verification systems (continued).....	29
Table 2.2: Dynamic signature verification systems.	43
Table 4.1: FR factor for the seven central moments based on targets' reference set.	77
Table 4.2: FR factor for the seven normalised central moments based on targets' reference set.....	77
Table 4.3: FR factor for the seven Hu moments based on targets' reference set.	77
Table 4.4: FR factor for the twelve normalised central moments based on targets' reference set.	78
Table 4.5: Static verification –Equal Error Rates for the different feature sets.....	81
Table 4.6: Static Verification –Error rates for the target signatures.....	83
Table 4.7: Static verification –Intra-class distance measures for the targets' reference set.....	87
Table 4.8: Static verification –Error rates for the target signatures using individual thresholds.....	87
Table 4.9: Static verification –Error rates for different reference sets.	89
Table 4.10: Dynamic verification –Extracted features.....	96
Table 4.11: FR factor for the 40 features and the target signatures.....	97
Table 4.12: Dynamic verification –Selected features.....	98
Table 4.13: Dynamic verification –Error rates for the target signatures.	100
Table 4.14: Dynamic verification –Intra-class distance measures for the targets' reference set.....	105
Table 4.15: Dynamic verification –Error rates for different reference sets.....	106
Table 5.1: Perceived complexity statistics for the target signatures.....	118
Table 5.2: Perceived complexity statistics for Target 5 including and excluding the outlier.....	123
Table 5.3: Human and computer estimates of target signature complexity.	132
Table 6.1: Relative variability ranking statistics for the target signatures.	145
Table 6.2: Variability ranking order (A) of the target signatures.	145
Table 6.3: Variability ranking order (B) of the target signatures.	147
Table 6.4: Variability ranking order (C) of the target signatures.	149
Table 6.5: Baseline angle measurements (in degrees), for the five samples of the target signatures...	154
Table 6.6: Intra-class similarity of target signatures.	156
Table 6.7: Inter-class similarity of target signatures.	156
Table 6.8: Variability ranking order (D) of the target signatures.	157
Table 6.9: Intra-class similarity values for different degrees of smearing.	159
Table 6.10: Non-parametric correlation between subjective and objective similarity rank orders.	163
Table 6.11: Pearson correlations between subjective/objective complexity and objective similarity measures.....	163

Table 7.1: Experiment 1 – Average human performance in verifying signatures. 173

Table 7.2: Experiment 1 – Statistics of the correct classification and total error. 174

Table 7.3: Experiment 1 – Statistics of the FRR and FAR. 176

Table 7.4: Experiment 2 – Average human performance in verifying signatures. 179

Table 7.5: Experiment 2 – Statistics of the correct classification and total error. 180

Table 7.6: Experiment 2 – Statistics of the FRR and FAR. 181

Table 7.7: Experiment 3 – Average human performance in verifying signatures. 183

Table 7.8: Experiment 3 – Statistics of the correct classification and total error. 184

Table 7.9: Experiment 3 – Statistics of the FRR and FAR. 186

Table 7.10: Correct Classification (%) statistics for the three experiments. 187

Table 7.11: Multiple Comparisons of the correct classification means for the three experiments. 189

Table 7.12: Two-independent sample t-Test assuming equal variances for the correct classification
population means of Experiments 2 and 3. 190

Table 7.13: FRR (%) statistics for the three experiments. 191

Table 7.14: Multiple Comparisons of the FRR means for the three experiments. 192

Table 7.15: Two-independent sample t-Test assuming equal variances for the FRR population means of
Experiments 2 and 3. 193

Table 7.16: FAR (%) statistics for the three experiments. 193

Table 7.17: Kruskal-Wallis One-way ANOVA for comparison of the three experiment FAR probability
distributions. 195

Table 7.18: Two-independent sample t-Test assuming unequal variances for the FAR population means
of Experiments 2 and 3. 195

Table 7.19: Mann-Whitney U test for comparison of the FAR populations of Experiment 2 and 3. 196

Table 7.20: Experiment 1 – Average error rates for the target signatures. 200

Table 7.21: Spearman’s rank correlation coefficient with error rates from Experiment 1. 200

Table 7.22: Effect of variability in targets with similar complexity values. 201

Table 7.23: Targets with similar perceived intra-class variability in Experiment 1. 202

Table 7.24: Experiment 3 – Average error rates for the target signatures. 202

Table 7.25: Spearman’s rank correlation coefficient with error rates from Experiment 3. 203

Table 7.26: Targets with similar perceived intra-class variability in Experiment 3. 204

Table 7.27: Percentage of reductions affected both by the complexity and the intrinsic variability of the
targets. 205

Table 7.28: Degree of confidence in classification decisions of Experiment 1. 208

Table 7.29: Degree of human confidence with respect to the target signatures. 209

Table 7.30: Degree of confidence for genuine and forged signatures. 209

Table 8.1: Error rates for individual threshold selection Scheme 1 and different α ’s. 227

Table 8.2: Error rates for individual threshold selection Scheme 2 and different β ’s. 229

Table 8.3: Error rates for individual threshold selection Scheme 3. 230

Table C.1a: Intra-class similarity of target signatures, using the method proposed in [57]. C-i

Table C.1b: Inter-class similarity of target signatures, using the method proposed in [57]. C-i

Table C.2a: Intra-class similarity of target signatures, using the method proposed in [140].	C-ii
Table C.2b: Inter-class similarity of target signatures, using the method proposed in [140].	C-ii
Table C.3a: Intra-class similarity of target signatures, using the method proposed in [121] without weights.	C-iii
Table C.3b: Inter-class similarity of target signatures, using the method proposed in [121] without weights.	C-iii
Table C.4a: Intra-class similarity of target signatures, using baseline adjustment.	C-iv
Table C.4b: Inter-class similarity of target signatures, using baseline adjustment.	C-iv
Table D.1: Mean confidence values for the target signatures.	D-ii

List of Figures

Figure 1.1: Biometric modalities.....	2
Figure 1.2: Signing process.....	4
Figure 2.1: Type I/ Type II error curves vs the decision threshold.	13
Figure 3.1: The WACOM ArtPad II tablet with Ultrapen.....	48
Figure 3.2: Transmission of information between pen and tablet [7].....	49
Figure 3.3: Sample signature file.....	49
Figure 3.4: (a) Pen-down points and (b) reconstructed signature line of a sample.	50
Figure 3.5: (a) Pen-up points and (b) pen-up and pen-down information of a sample.....	51
Figure 3.6: Pen-up and pen-down information for (a) and (b) genuine samples, and (c) and (d) forgeries.	52
Figure 3.7: Digitiser experimental set up and signing process.....	53
Figure 3.8: Scanned images of (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 signature.....	54
Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5.	56
Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).	57
Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).	58
Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).	59
Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).	60
Figure 4.1: Schematic of handwritten signature verifier.	67
Figure 4.2: C++ Code for the calculation of the vertical and horizontal projections.....	73
Figure 4.3: Static verification –Error curves vs threshold.....	79
Figure 4.4: Static verification –Error tradeoff curve.	80
Figure 4.5: Static verification –Error tradeoff curves for the different feature sets.	81
Figure 4.6: Static verification –FAR vs threshold for random forgeries.....	82
Figure 4.7: Static verification –Error curves vs threshold for Target 1.....	83
Figure 4.8: Static verification –Error curves vs threshold for Target 2.....	84
Figure 4.9: Static verification –Error curves vs threshold for Target 3.....	84
Figure 4.10: Static verification –Error curves vs threshold for Target 4.....	85

Figure 4.11: Static verification –Error curves vs threshold for Target 5.....	85
Figure 4.12: Static verification –EER vs number of reference samples.....	89
Figure 4.13: Static verification –Error tradeoff curves for different reference sets.	90
Figure 4.14: Static verification –FAR curves for random forgeries and different reference sets.....	91
Figure 4.15: Velocity components as a function of time $u_x(t)$, $u_y(t)$, and total velocity signal $u(t)$	94
Figure 4.16: Acceleration components as a function of time $\alpha_x(t)$ and $\alpha_y(t)$, and the total acceleration $\alpha(t)$	94
Figure 4.17: Dynamic verification –Error curves vs threshold.	99
Figure 4.18: Dynamic verification –Error tradeoff curve.	99
Figure 4.19: Dynamic verification –FAR vs threshold for random forgeries.	100
Figure 4.20: Dynamic verification –Error curves vs threshold for Target 1.	101
Figure 4.21: Dynamic verification –Error curves vs threshold for Target 2.	101
Figure 4.22: Dynamic verification –Error curves vs threshold for Target 3.	102
Figure 4.23: Dynamic verification –Error curves vs threshold for Target 4.	102
Figure 4.24: Dynamic verification –Error curves vs threshold for Target 5.	103
Figure 4.25: Dynamic verification –EER vs number of reference samples.	106
Figure 4.26: Dynamic verification –Error tradeoff curves for different reference sets.	107
Figure 4.27: Dynamic verification –FAR vs threshold for random forgeries and different reference sets.	108
Figure 4.28: Dynamic verification –EER vs number of features.	109
Figure 5.1: Signature images of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, and (e) Target 5.	117
Figure 5.2: Frequency distributions of the perceived degree of complexity for target signatures: (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.	119
Figure 5.3: Box-plots of the perceived complexity judgments for the target signatures.	120
Figure 5.4: Perceived complexity z-scores ($z = (x - \bar{x}) / s$) of all subjects for Target 5.	122
Figure 5.5: Perceived complexity frequency distribution for Target 5 (a) including and (b) excluding the outlier.	122
Figure 5.6: Vertical projections of Target: (a) 1, (c) 2, (e) 3, (g) 4, (i) 5, and horizontal projections of Target: (b) 1, (d) 2, (f) 3, (h) 4, (j) 5, along with their associated Entropy value.	129
Figure 5.7: Linear regression between human and computer estimation of signature complexity.	133
Figure 5.8: (a) Standardized residual plot, and (b) its frequency distribution.	134
Figure 6.1: Genuine samples of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, (e) Target 5.	141
Figure 6.1: Genuine samples of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, (e) Target 5 (continued).	142
Figure 6.2: Frequency graphs of the perceived relative variability ranking position for the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5.	144
Figure 6.3: Variability ranking box-plots for the target signatures.	146

Figure 6.4: Polar quantization of shape.....150

Figure 6.5: Baseline detection for (a) Target 2 (Sample 4), (b) Target 1 (Sample 5), and (c) Target 5 (Sample 2)..... 153

Figure 6.6: Effect of smearing on the intrinsic similarity values. 159

Figure 6.7: Effect of smearing on original signatures (a), for thresholds of (b) 5, (c) 10, (d) 20 and (e) 30 pixels..... 161

Figure 7.1: Experiment 1 – (a) frequency distribution and (b) normal probability plot of the participants’ correct classification performance..... 173

Figure 7.2: Experiment 1 – (a) frequency distribution and (b) normal probability plot of the FRR. ... 175

Figure 7.3: Experiment 1 – (a) frequency distribution, and (b) normal probability plot of the FAR. .. 175

Figure 7.4: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the participants’ correct classification performance..... 179

Figure 7.5: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the FRR. ... 181

Figure 7.6: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the FAR. 181

Figure 7.7: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the participants’ correct classification performance..... 183

Figure 7.8: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the FRR. ... 184

Figure 7.9: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the FAR. 185

Figure 7.10: Box-plots of the correct classification performance in the three experiments. 187

Figure 7.11: Box-plots of the FRR in the three experiments. 191

Figure 7.12: Box-plots of the FAR in the three experiments. 194

Figure 7.13: Mean error rate performance vs number of reference samples..... 196

Figure 7.14: Degree of human confidence in verifying signatures. 208

Figure 7.15: Degree of confidence in verifying (a) Genuine signatures, (b) Forgeries..... 210

Figure 7.16: Rejection rate vs confidence threshold. 211

Figure 7.17: Implementation of the degree of confidence in the verification scheme. 211

Figure 7.18: Verification error and relative error vs confidence threshold. 212

Figure 7.19: Total error vs confidence lower threshold. 213

Figure 7.20: Rejection rates of genuine signatures (r_g) and forgeries (r_f) vs confidence threshold..... 214

Figure 7.21: First stage FRR_v and the relative FRR_v' vs confidence threshold..... 215

Figure 7.22: Total FRR of the system (with $FRR_f=26.43$) vs confidence threshold. 216

Figure 7.23: First stage FAR_v and the relative FAR_v' vs confidence threshold. 216

Figure 7.24: Total FAR of the system (with $FAR_f=10$) vs confidence threshold..... 217

Figure 8.1: Human inspection of signatures –Error rates vs number of reference samples. 222

Figure 8.2: Static signature verification –Error rates vs number of reference samples..... 223

Figure A.1: Interface of the perceived signature complexity experiment. A-i

Figure A.2: Interface of the human perception Experiment 1..... A-ii

Figure A.3: Interface of the human perception Experiment 2..... A-ii

Figure A.4: Interface of the human perception Experiment 3..... A-iii

Figure B.1: Static verification performance for the set of seven central moments (a) error rates vs threshold and (b) error tradeoff curve. B-i

Figure B.2: Static verification performance for the set of seven normalised central moments (a) error rates vs threshold and (b) error tradeoff curve. B-ii

Figure B.3: Static verification performance for the set of seven Hu moments (a) error rates vs threshold and (b) error tradeoff curve. B-iii

Figure B.4: Static verification performance for the set of seven Global features (a) error rates vs threshold and (b) error tradeoff curve. B-iv

Figure B.5: Static verification performance for the Limited Reference Set (a) error rates vs threshold and (b) error tradeoff curves, using skilled forgeries.B-v

Figure B.6: Static verification performance for the Small Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries. B-vi

Figure B.7: Static verification performance for the Large Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries. B-vii

Figure B.8: Static verification –FAR vs threshold for random forgeries and the Limited Reference Set. B-viii

Figure B.9: Static verification –FAR vs threshold for random forgeries and the Small Reference Set. B-viii

Figure B.10: Static verification –FAR vs threshold for random forgeries and the Large Reference Set. B-ix

Figure B.11: Dynamic verification performance for the Limited Reference Set (a) error rates vs threshold and (b) error tradeoff curves, using skilled forgeries.B-x

Figure B.12: Dynamic verification performance for the Small Reference Set (a) error rates vs threshold and (b) error tradeoff curves, using skilled forgeries. B-xi

Figure B.13: Dynamic verification performance for the Large Reference Set (a) error rates vs threshold and (b) error tradeoff curves, using skilled forgeries. B-xii

Figure B.14: Dynamic verification –FAR vs threshold for random forgeries and the Limited Reference Set. B-xiii

Figure B.15: Dynamic verification –FAR vs threshold for random forgeries and the Small Reference Set. B-xiii

Figure B.16: Dynamic verification –FAR vs threshold for random forgeries and the Large Reference Set. B-xiv

Figure D.1: Frequency distribution of human confidence in verifying the authenticity of target signatures: (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5. D-i

Figure D.2: Rejection rate vs confidence threshold for Experiment 2. D-ii

Figure D.3: Error rates vs confidence threshold for Experiment 2. D-ii

Figure D.4: Rejection rate vs confidence threshold for Experiment 3. D-iii

Figure D.5: Error rates vs confidence threshold for Experiment 3. D-iii

“...τὴν τοῦ ἀγαθοῦ ἰδέαν φάθι εἶναι αἰτίαν δ’ ἐπιστήμης καὶ ἀληθείας...”

Πλάτων (428 – 347 π.Χ.)

ΠΛΑΤΩΝΟΣ ΠΟΛΙΤΕΙΑ (ΣΤ') [508e]

“...the concept of good is the cause of science and truth...”

Plato (428 – 347 B.C.)

THE REPUBLIC OF PLATO (6) [508e]

Chapter 1

Introduction

This thesis describes an investigation into aspects of the handwritten signature as a means of verifying the identity of an individual. The work is thus set within the broad field of biometrics and, especially, is concerned with the link between human and machine –based biometric processing. This chapter sets the scene through an introduction to biometrics technology and the examination of the advantages and disadvantages of the handwritten signature over other biometrics. Several concepts concerning the complex process of signing and forging will be discussed, while some issues related to the inherent variability of signature samples, the different styles and variations encountered and the existence of forgeries will be further analysed, especially with respect to potential difficulties presented to automatic verification systems. An introduction to a typical scenario of human inspection of signatures and the automatic signature verification process will define some important problems and highlight the need for improvement. Finally, the approach followed in this thesis along with a structural layout is given, proposing an enhanced strategy offering solutions for improvement both in human checking procedures as well as in the development of automatic signature verification systems.

1.1 Biometrics

In a fast moving world, where automation, electronic transactions and e-commerce are increasingly part of our lives, the battle against plastic card fraud is more essential than ever [6]. The use of PIN numbers and passwords has raised security issues, since they can be lost or stolen, while easily forgotten. Biometric technology, on the other hand, is a promising alternative.

Biometrics are personal attributions that uniquely characterize individuals and can, therefore, be used for verification of their identity. The biometric modalities (Figure 1.1) can be classified into two categories:

- Physiological biometrics
- Behavioural biometrics

The physiological biometrics, which are unique physical characteristics of a person, involve technologies such as the following: fingerprint verification [69], hand geometry verification [70], iris [44], face [38], ear [66] recognition, etc. The behavioural biometrics, which are based on behavioural traits of individuals, are signified by the following: handwritten signature verification [51], speaker verification [34], gait recognition [150], etc.

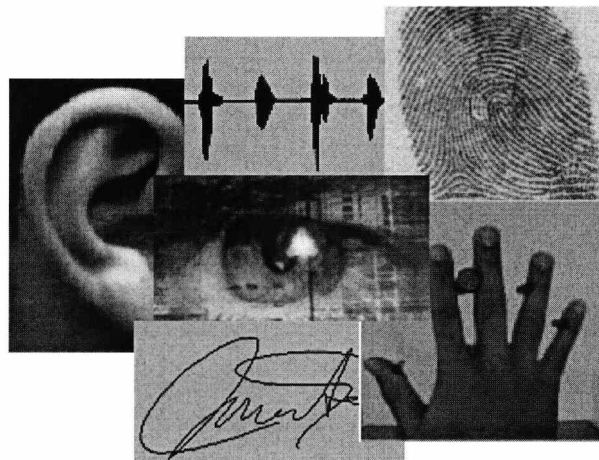


Figure 1.1: Biometric modalities.

Although both recognition and verification tasks have been studied with respect to most biometrics, the two processes are very different. The *recognition* process involves establishing a person's identity through searching an entire database, whereas the *verification* process involves the authentication of a person's claimed

identity by performing a one-to-one comparison. Plamondon and Lorette [112] discriminate between the *recognition* task and the *identification* task in handwriting in the context that the former is based on processing the ‘semantic’ information of handwriting, while the latter on the ‘singular’ part.

Apart from voice prints the rest of the biometric modalities mentioned here are visual biometrics and, therefore, the associated recognition methods generally involve common visual pattern classification and recognition techniques. However, signal comparison methods initially developed for speech recognition, have also been applied for the verification of handwritten signatures.

Some of the criteria that need to be considered in the evaluation of the suitability of a biometric technology for a particular application involve the following:

- Accuracy
- Reliability
- Robustness
- Public acceptance
- Ease of use
- Speed of verification
- Cost of implementation
- Data storage requirements

Since a biometric technology fitting all the criteria remains an ideal case, in reality different biometrics are more or less suitable for different applications. For example, while iris scanning is highly suitable for military applications [126], where high accuracy is a priority, it would not necessarily be the right choice for everyday point-of-sales applications, as the cost of implementation would be high, and most importantly, as the introduction of such a technology might be considered invasive by the general public. Fingerprint verification, which is also highly accurate, may be found intrusive by the public, due to its criminal connotations. A public survey [1] showed that while automatic signature verification was found by the general public to be an acceptable method of personal identity authentication in point-of-sales applications, fingerprint checking was not.

1.2 Handwritten Signature

The handwritten signature has a number of advantages over the other biometrics that make it highly suitable for point-of-sales (POS) and similar applications. First of all, authentication based on signature checking is already familiar to the public as a result of its widespread use. In fact the handwritten signature has long been established as a common means for providing proof of identity, as related to bank transactions, certification of documents, wills, etc. Furthermore, the automatic handwritten signature verification has relatively low cost of implementation and low storage requirements for a template. However, while the handwritten signature generally receives high acceptability ratings from the public, it also stands for a challenge to potential forgers. Hence, it is the unknown class of forgeries which pose the greatest threat to automatic signature verification systems.

Signing process

The act of signing (Figure 1.2) is a complex process that requires the neuromuscular systems coordination in the arm, hand and wrist, for the production of the necessary movements. Several models have been proposed for the generation of



Figure 1.2: Signing process.

handwriting [110;111;115]. As related to the biomechanical processes involved in fast handwriting, it may be quoted from Plamondon and Maarse [115] that “...*some central nervous system mechanisms within the brain fire, with a predetermined intensity and duration, the nerve network which activates the proper muscles in a predetermined order. The motion of the pen on the paper, resulting from muscle contraction/relaxation, leaves a partial trace of the pen-tip trajectory*”. Furthermore, it is interesting to mention that, with respect to the variables controlled by the nervous system during handwriting the authors tested the velocity, acceleration and length, and suggested the velocity as a controlled variable. Schomaker and Plamondon [132] suggested that trajectory control can be separate from pen-force control, as their test results for the majority of the subjects showed low correlation between kinematics and pen pressure in writing predetermined patterns and cursive script. This was explained as a possibility for a separate control component that regulates the pen force.

Signature samples

Due to the nature of the signing process, no individual can exactly replicate his signature. Thus, a person's signature varies from sample to sample. This variation is highly dependent on the signing conditions, i.e. the position of signing as well as the writing instrument and writing surface, the psychological and physical state of the writer at the time of signing, the time interval between the submitted samples, the signer's age, occupation, etc. While some signers are more stable than others, some may exhibit a relatively high degree of inconsistency in their signature samples. The intrinsic variability of signatures is a delicate factor that should be taken into consideration in the design of automatic signature verification systems as it highly affects their performance.

The nationality of the signer is another important issue as different countries and cultures exhibit different signing styles. Where Americans generally tend to sign their names, the European signatures have more of a resemblance for shapes, usually displaying a number of loops, lines and curves surrounding the main body of the signature, while the handwritten version of the latter is sometimes even absent. Furthermore, Arabic signatures are written in the opposite direction from right-to-left, while Japanese people do not have habits of signing for the authentication of their identity [152] and, thus, they are not as accustomed to the signing process itself. Therefore, for an automatic signature verification system to be universally successful, it should incorporate design considerations to cover a wide range of signature styles encountered in the literate world.

Forgeries

Perhaps the highest threat imposed on signature verification systems is the possibility of forgeries. Although, according to experts, the vast majority of forgeries are usually poorly performed, with some exhibiting no resemblance to the original signature at all, there are still a number of highly skilled forgeries that are difficult to detect. In fact, the extent to which highly skilled forgeries are produced would be difficult to estimate since any number of them could have passed undetected. Since the task of discovering professional forgeries is a very difficult one, it would be a good

accomplishment for an automatic signature verification system if it succeeded in detecting imitations performed by amateur forgers.

It is generally accepted that a forger concentrates on the spatial characteristics of the signature to be imitated, missing out the dynamics, while if an attempt is made to capture the dynamics, the resulting forgery is expected to miss out on important details of the original signature shape. Since the task of imitating a signature both in the static and dynamic domain is a relatively difficult one, it is considered vital for an automatic signature verification system to evaluate both static and dynamic features of the signatures to be tested. In their experimental study involving copying both handwritten signatures and script, Van Gemmert and Van Galen [142], suggested that kinematic measures are preferable to spatial characteristics and that static script alone cannot be used to distinguish forgers from authentic writers. Furthermore, it was reported that the imitations were achieved with significantly increased movement times, more dysfluencies, and more intermittent and smaller force pulses, while also indicative of a higher degree of limb stiffness. With respect to the latter, the authors also concluded that *“psychomotor and biomechanical strategies used to forge another person’s script are similar to strategies subjects more generally use during stressful conditions”*.

1.3 Signature Verification

Human inspection

Currently, the majority of the verification tasks required for the examination of the authenticity of handwritten signatures are carried out by human inspectors, whose level of expertise depends on the nature of the application in question. Highly specialised cases such as the examination of important documents may be carried out by forensic experts, whereas everyday signature checking encountered in public services, or sales departments usually involve non-experts having undergone preliminary training or in many situations no training at all. Furthermore, everyday experience may highlight that the signature checking carried out in most point-of-sales applications today is minimal. Although it may be argued that a noisy background and a busy situation may impair the concentration levels of a counter-

clerk, the associated human perceptual capabilities have not been particularly investigated.

In spite of the large extent of the handwritten signature verification literature, only very few studies exist investigating human performance in inspecting the authenticity of signatures. Randolph and Krishnan [119] have cited previous studies stating that *“the majority of forgeries accepted by bank tellers are totally unskilled”*, and that *“under test conditions, tellers accept 10-50% of forgeries as valid signatures and accept nearly 100% of forgeries in their actual jobs”*. Although it is recognised that both handwritten signatures and visual inspection have long been in common use, both of the studies cited date back to 1977. Of course, if nowadays bank tellers are trained to perform better, it could still be argued that shop-clerks might be not, and although the handwritten signature is widely used today for most electronic transactions ensuring a high level of familiarisation with its use in human checking procedures, the previously mentioned rates may still reflect the situation today. With respect to the results brought by document experts a high level of performance is generally anticipated. However, due to the relative confidentiality of the matter and the separate research area involved, detailed information about this is not readily available.

Automatic verification

A key solution to the high levels of fraudulent activity mostly related to electronic transactions is the implementation of automatic signature verification (ASV) systems. Although many commercial ASV systems have been introduced to the market [2-5], automated signature checking has not been yet established as a secure and completely reliable biometric technology. The design of a highly accurate and reliable ASV system is a difficult task. This is confirmed by the numerous research studies concentrated in constantly offering improvements in the field. In fact, for reasons earlier mentioned, such as related to the unknown class of forgeries and the unsuspected variability of the genuine class, the separation of the two classes is a delicate issue. Although in an ideal case the separation of the two classes would be possible with the use of an optimum feature set, in practice the two classes overlap. As a result, the error curves associated with the two classes – i.e. the false rejection of genuine signatures (Type I error) and the false acceptance of forgeries (Type II error)

– intersect. Hence, a decrease in one error rate generally leads to an increase in the other. Thus, the selection of a decision threshold for the pattern classification task depends on the requirements of the application in question for the minimisation of the appropriate error rate. Usually, point-of-sales applications require primarily a very small Type I error in order to avoid potential nuisance caused to the customers, while a relatively small Type II error would offer a significant improvement to the large number of forgeries penetrating the system as a result of the minimal checking carried out today. Lee et al. [79] suggested that a good performance for a POS application would have the Type II error below 25% while the Type I error would be near zero.

The performance of verification systems is usually evaluated through the Type I/II error rates. As earlier explained, zero error rates cannot be achieved in a practical situation and thus, a 100% accuracy is not a realistic achievement for a handwritten signature verification system. As a result, numerous research studies have concentrated on minimising the error rates improving the performance of ASV systems. However, the lack of a common database to test the different systems, in conjunction with the different methods and procedures employed in different studies, constitutes a real difficulty in comparatively assessing those systems, leading to relatively slow improvements in the field of ASV.

Automatic signature verification may be approached in two ways: through a static analysis of the signature image acquired after the signature was written, or through an analysis of the signature data captured dynamically during the signing process. The static (off-line) verification of signatures is generally weaker than the dynamic (on-line) verification, since the latter evaluates kinematic information, which is well hidden from potential forgers. On the other hand, static signature verification is essential for particular applications where only the image of the signatures is available. Moreover, these situations are the ones most likely to be encountered in a wide range of applications involving document processing, cheque processing, etc. Therefore, improvements in the automatic verification of signatures should be looked at both from a static and a dynamic point of view.

1.4 Aim and Structure of Thesis

This thesis describes research which is aiming at an improvement in signature verification processes so as to offer support and security in areas where authorisation of individuals' identity is needed. The procedure followed involves the examination of handwritten signature verification both in terms of an automatic approach but also through investigation of human inspection processes. It is anticipated that knowledge gained from human perceptual judgements can offer an improvement in the static verification of signatures, and an attempt to combine related findings obtained will be made. Finally, the general approach used involves the consideration of information acquired from different disciplines, including: cognitive science, forensic science, document processing, handwriting analysis, shape description, pattern recognition, and automatic signature verification, as it is believed that marrying knowledge from different disciplines and following a more universal approach is especially promising.

An overview of the structure of the thesis is given below. The following paragraphs include a brief description of the contents of each chapter and the approach followed:

Chapter 1 sets the scene through an introduction to biometrics technology and the pros and cons of the handwritten signature as a means of authenticating individuals' identity. Several aspects concerning the complex process of signing and forging are discussed, while issues related to inherent signature characteristics and the existence of forgeries, presenting potential difficulties to automatic signature verification systems, are further analysed. The current situation of human checking and existing automatic signature verification systems are discussed, highlighting the need for improvement.

Chapter 2 gives an overview of the several methods implemented and the systems developed in a number of research studies in the field, both in a static and a dynamic signature verification approach. The visual inspection of handwritten signatures as performed by human experts and non-experts is also briefly discussed, based on the very limited number of relative studies available.

Chapter 3 describes the data acquisition process, the means and methods used for the construction of a signature database for its use in the experiments undertaken and applications developed in this thesis. The signature data were obtained through an experiment with human subjects, collecting a large number of genuine signatures and forgeries. The signing process was captured with a digitising tablet and further issues concerning the format of the data and the digitised signature images are also discussed.

Chapter 4 analytically describes the design and development of an automatic signature verification system as implemented in this thesis. Both static and dynamic verification processes are developed, generally based on a common verification design. The construction of the verifier is guided step by step through the system design, presenting the data sets, the enrolment stage, the feature selection, the classification and the threshold selection procedures used, while the feature extraction is discussed as related to the separate parts of the static and dynamic approach. The performance of the system under both static and dynamic conditions is extensively evaluated, while some optimisation techniques are discussed and assessed.

Chapter 5 examines the complexity of signatures from the point of view of static signature image analysis. An experimental study is carried out in order to assess the perceptual viewpoints and judgements of human subjects with respect to the degree of complexity inherent in five different signatures. Furthermore, a novel method is developed to quantitatively estimate the degree of a signature's complexity reflecting human perceptual criteria.

Chapter 6 is engaged with the evaluation of the intrinsic variability of handwritten signatures from a static point of view. The intra-class dissimilarity of signatures is assessed both in a subjective and an objective approach. A human perception experiment is the source of the subjective signature variability results, while for the objective measurement several methods are considered and the one closest to the human perception results is further evaluated.

Chapter 7 includes an analysis of the visual inspection of handwritten signatures based on data obtained from three experiments on human judgements. The human

performance in correctly verifying the authenticity of signatures and identifying forgeries is extensively evaluated, while its relation to the complexity of signatures and their intrinsic variability is also examined, assessing the likelihood of a signature's susceptibility to forgery. The human confidence in making verification decisions of this nature is analysed and possible system rejection scenarios are proposed to assist the human checking procedures carried out today.

Chapter 8 provides a comparative analysis between the human performance in inspecting signatures' authenticity and the automatic signature verification system developed. Several combined schemes are proposed for the integration of knowledge and implementations developed in all main chapters offering an improvement in the static verification of signatures. Furthermore, the chapter provides a summary of the work carried out in the thesis and the contributions made to the field. Some general issues are discussed and suggestions are given for future work, while final conclusions are drawn setting out a clear message for the future.

The various applications developed in this research were designed and constructed using object-oriented programming, specifically using Borland C++ Builder 3 and Borland Database tools.

Chapter 2

Literature Review

A great number of commercial products, patents and research systems have been developed for the automatic verification of handwritten signatures. In this overview methods and systems developed in research studies are included and briefly reviewed in the context of the section within which they are described. Automatic signature verification is divided into static and dynamic approaches, which in turn are partitioned into the major stages involved in building an automatic signature verification system. Human performance in visually inspecting the authenticity of handwritten signatures is also reviewed, from the point of view of expert examiners and non-experts, although available studies in this field are very limited.

2.1 Automatic Signature Verification

The design considerations for an automatic signature verification system include the following stages: preprocessing, feature extraction and feature selection, classification, and performance evaluation. The construction of a signature database is the first requirement and its size and quality are major factors that affect the design and the performance of a system. Usually, a number of genuine signatures and

forgeries are initially collected. Three types of forgeries have been defined [112]: *random* where the submitted forgeries do not resemble to any intended degree the original signature as it is usually the forger’s own signature, *simple* where no attempt is made by the forger to imitate the original signature as only information about the name of the original signer is available, and *skilled* where the forger is given time to practice the original target signature and sometimes information about its dynamics is also provided. It immediately becomes apparent that a comparison of different systems based on different databases and even different types of forgeries collected is already a difficult task. Moreover, different techniques are implemented depending on the data acquisition process, and the discretion of the researcher/system developer. The performance of a system is usually assessed in terms of the attained error rates: the False Rejection Rate (FRR) or Type I error, which is the percentage of genuine signatures falsely rejected by the system as attempted forgeries, and the False Acceptance Rate (FAR) or Type II error, which is the percentage of forgeries falsely accepted as genuine samples. Alternatively, the Equal Error Rate (EER) is often used, which corresponds to the point where the two error curves intersect (see Figure 2.1) and hence, it is defined by an equality of the two types of error. An extensive analysis of the several methods and approaches used for the automatic handwritten signature verification and a report on the performances obtained in the different research studies will follow, both in a static and a dynamic signature verification context.

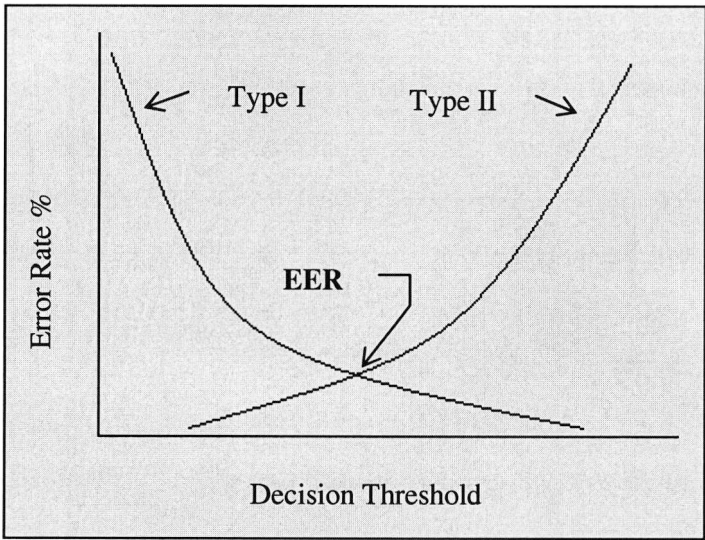


Figure 2.1: Type I/ Type II error curves vs the decision threshold.

2.1.1 Static Verification

The static verification of signatures involves an analysis of the 2-D signature image. The signatures may be captured with the help of a scanner or a camera, usually a TV vidicon camera or a CCD matrix camera, while the image may be binary or grey-scale with up to 256 levels of quantization. There is also the option of dynamically acquiring the signature data with a digitising tablet and then using solely the x- and y-coordinate pairs, assessing them from left-to-right and top-to-bottom thus, simulating 'reading' a digital image in an off-line approach. A number of off-line techniques are reviewed by Plamondon et al. [114], Plamondon and Lorette [112], and Leclerc and Plamondon [78].

Preprocessing

The preprocessing stage may involve several or all of the following procedures: separation of the signature from the background, noise reduction, data area cropping, size normalisation, thinning/skeletonization, and segmentation of the signature. The precise requirements depend on the signature acquisition method (camera-based acquisition requires binarization and filtering, while an image obtained with a scanner is automatically thresholded [101]), the background pattern (a signature written on a bank cheque would require the extraction of the background as opposed to a signature written on white paper), and the type of features implemented (some particular features may require the segmentation of the signature), etc.

To separate the ink from the background, Nagel and Rosenfeld [97], used thresholding of the grey levels in the image. In order to precisely locate the best threshold between the large and small peaks in the grey-level histogram, the histogram was filtered using a Laplacian^[56] operator selecting pixels with values in the upper 15% of its range. For the reduction of noise, Bajaj and Chaudhary [20] used a modified median filter^[56] with a 7x7 mask, according to which, if the number of black-pixels in the pixel-neighbourhood was greater than a threshold, then the pixel under consideration was set to 1. However, a lower threshold to that of the standard median filter was found more adequate. Baltzakis and Papamarkos [21] proposed the use of a noise reduction filter on the binary scanned images with a 3x3 mask and a decision rule according to which, if the number of the 8-neighbours of a pixel having

the same colour with the central pixel was less than two, then the colour of the central pixel was reversed. Furthermore, the cropping of the signature area was achieved with the help of the vertical and horizontal image projections, while for the width normalisation of the signature images the width was adjusted to the default value, sustaining the height-to-width ratio unchanged. A simplified skeletonization algorithm was also proposed. Ramesh and Murty [118] suggested that the signatures are normalised in terms of orientation and that the images are resized, for the extraction of certain features that cannot be size-normalised. As the resizing of the signature images requires additional averaging or interpolation methods that are bound to distort the images, a better solution was found through an algorithm proposed to minimise the number of breaks formed between the lines in the image, without the need for averaging.

Ammar et al. [12] proposed a background equalization and reduction method, followed by noise reduction achieved by averaging and then thresholding for the signature to be separated from the background. Furthermore, the authors proposed the extraction of *high pressure regions* (i.e. signature regions with higher grey levels than a predetermined threshold) and a selection procedure for this threshold. In addition, the proposed method involved the evaluation of a *pressure factor*, defined as the ratio of the area of high pressure regions to the signature area. In another study [14] the authors proposed a *middle zone demarcation* method, based on the y-axis projection of the signatures, for their segmentation into upper, middle and lower zones, as well as a segmentation procedure of the signatures into elements, with the help of the signatures x-axis projection, for the extraction of features and relations between signature properties for their global and local description. Nagel and Rosenfeld [97], introduced different types of y-axis projections, designed to be insensitive to long lines and sensitive to the writing (by counting the number of ‘runs’ of ink points on each row), or sensitive to long lines and insensitive to the writing (by measuring the length of the longest ‘run’ in each row), in order to locate the signature line and segment the writing into upper, middle and lower zones. A horizontal segmentation of the writing was also proposed at the points where the long strokes cross into the upper and lower zones.

The segmentation of the signature into arbitrarily-shaped primitives was proposed by Sabourin et al. [124]. The primitive extraction process involved a *region-growing-with-merging* process, which merges pixels into atomic regions characterized by homogeneity of the orientation of the gradient, and a *high-level merging* process, which initially removes spurious noise, characterised by sparse atomic regions in the image area, followed by a hierarchical merging of collinear atomic regions. The gradient computation process uses the Sobel^[56] operator applied to the grey-level image, and the primitive extraction process uses the statistics of directional data. Lee and Pan [80] proposed an approach for the off-line tracing of signatures involving a set of heuristic rules, based on the way humans normally trace signatures. The heuristic rules involve pixel processing, locally assessing the neighbour pixels of each point, and stroke processing, globally reordering the sequence of strokes. Furthermore, a *critical-point segmentation* method was proposed to extract local feature points for signature representation, as well as a *critical-point normalisation* method. Note that, in order to ensure proper tracing a thinning process was initially applied to the binary signature images, producing skeletons of 1-pixel width, while undesirable artifacts produced as a result of the thinning process were removed. Noise reduction was also performed to remove the *salt-and-pepper noise* associated with the thresholding applied for the binarization of the signature images, while the resulting broken strokes were restored.

Feature extraction and selection

The feature extraction procedure for the off-line representation and verification of signatures in the various research studies involves the use of some of the following feature types: global geometric features, local shape features, moment-based features, envelope features, wavelet features, grid features, texture features, etc.

Nagel and Rosenfeld [97] used global and local features extracted from signatures that were segmented into vertical and horizontal zones. The two global features were the ratio of signature width to short-stroke height and to long-stroke height, while the two types of local features used were ratios of the height of a long stroke to the height of the immediately preceding short stroke and slope features for appropriate long letters. The identification of the tall letters was carried out with respect to the known signature spelling. Other studies related to projections and zone features of global and

local descriptions are those by Huang and Yan [65] and Ammar et al. [14], which also include features related to grey-level intensities in the signature image, which will be discussed later.

Baltzakis and Papamarkos [21] extracted the following global features: the signature height, image area, pure width, pure height, baseline shift, vertical and horizontal centres of signature, maximum vertical and horizontal projections, vertical and horizontal projection peaks, global and local slant angles, number of edge and cross points and number of closed loops. In addition grid information features and texture features were used, providing information about the overall appearance of the signatures. The grid feature vector holds the number of foreground pixels contained in each of the 96 rectangular regions into which the skeletonised signature image was segmented. The texture feature group contained information about the transition of black and white pixels in the four different directions, for each one of the six rectangular areas into which the signature was segmented in this second scheme. Local grid features were also used by Qi and Hunt [117] along with 7 global geometric features, which included the height and width of a signature image, the pure width, the slant angle of the signature, the vertical centre of gravity of black pixels, the maximum horizontal projection, the area of black pixels and the baseline shift of the signature image. Ismail and Gad [67] proposed a combination of global and local features for the recognition and verification of Arabic signatures. Five types of computationally exhaustive local features – central line features, corner line features, central circle features, corner curve features and critical point features – were evaluated and the critical point features were found most effective.

Bajaj and Chaudhury [20] introduced the use of normalised projection moments for the classification of signature images. The proposed moment measures are related to the horizontal and vertical projections of the signature image, and include: kurtosis and skewness measures, relative kurtosis and skewness measures, and relative vertical and horizontal projection measures. In addition, the authors used envelope characteristics of the signatures. The envelope curve is constructed from the external points of the signature image. The points lying on the upper side of the principal axis of the signature form the upper envelope and those on the lower side similarly form the lower envelope. In order to extract the shape information from these curves a grid

is usually overlaid on the corresponding envelopes and numerical values are assigned to the grid-regions according to characteristics of the envelope curve encountered in the particular region, such as if the envelope curve passes through that region, if there is a prominent peak or valley, and so on. Ramesh and Murty [118], in addition to grid features extracted from the envelope curve, proposed features related to the degree of convolution of the envelopes – these include the number of turns in the upper and the lower envelope, and the number of gaps in the envelopes – which were claimed to be quite capable of detecting many types of simple forgeries. Along with the envelope characteristics, the authors also used a set of global geometric features, moment-based features associated with the vertical and horizontal projections of the signature image identical to those earlier described, and wavelet-based representations. With respect to the latter, a tree-structured wavelet^[56] transform was performed, decomposing a signature image into four subimages and calculating the energy at its leaves, while, depending on the energy value at the leaf, the decomposition procedure was or was not further applied to the subimage. At the end of the procedure, energy maps were obtained, and the prominent energy maps of signatures were compared for verification.

Other methods include contour features extracted from each of the signature's connected components as proposed by Congedo et al. [37], the representation of the signatures by their sinogram (Radon transform) introduced by Herbst and Coetzer [63], which is information preserving (i.e. a signature can be reconstructed from its descriptor by using an inverse transformation), and the use of Shape Matrices for the representation of signatures, proposed by Sabourin et al. [121], which is also an information preserving shape descriptor. Furthermore, Sabourin et al. [122] used granulometric size distributions for the extraction of local shape descriptors for the signature images, while Pavlidis et al. [105;106] used revolving active deformable models to capture the unique characteristics of a signature's silhouette, addressing the problem of signature recognition and identification.

Pseudodynamic features, which incorporate dynamic information but are yet extracted from the static signature's image and thus are considered more effective against forgery detection, have also been used. These may include pressure information, inferred indirectly from the signature's grey-levels image, or timing information, as

related to the sequence of strokes recovered by tracing the signature line. As earlier described, Ammar et al. [12] introduced the use of the *high pressure regions* and the *pressure factor*, associated with high grey-level intensities, and reported on their effectiveness, while Lee and Pan [80] proposed a multiresolution *critical-point segmentation* method for the extraction of critical points according to the tracing sequence, and a *critical-point normalisation* method for the invariant – under translation, rotation, and scaling – representation of signatures. Ammar et al. [13] extracted a number of static parameters consisting of zone and slant features (negatively, vertically, positively and horizontally slanted pixels), as well as the width, length and area of the signature. A feature selection procedure was proposed to test different subsets and select the one which produced the best verification result. The best subset included three global features: the percentage of vertically and positively slanted pixels and the global baseline, while the addition of a density feature, the *pressure factor*, led to a further improvement in the performance.

Classification process

Signature verification is represented by a two-class problem, described by the ‘genuine’ and the ‘forgery’ class. The classification process, which deals with the problem of assigning a test sample (unknown signature) to one of the two classes, requires the calculation of its distance from the known class(es) – subclass(es) of the vicinity of the respective set(s). The simplest and most popular measure of the distance between a test sample and the set of known genuine signatures (genuine reference set) is the Euclidean distance measure. Its basic form is given by equation (1), where the value t_i of the i th feature of the test sample is compared to the mean m_i of the i th feature computed on the genuine reference set, for a number of n features selected to represent the signature. More advanced measurements based on the Euclidean distance metric are given by equations (2) and (3), where the difference between the test sample and the reference set for a particular feature, is weighted by the standard deviation s_i of that feature computed on the reference set.

$$dist = \sqrt{\sum_{i=1}^n (t_i - m_i)^2} \quad (1)$$

$$dist = \sqrt{\sum_{i=1}^n \left(\frac{t_i - m_i}{s_i} \right)^2} \quad (2)$$

$$dist = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{t_i - m_i}{s_i} \right)^2} \quad (3)$$

For the classification decision, if the measured distance for a test signature is greater than a predefined threshold T then the signature is classified as a ‘forgery’, otherwise as a ‘genuine’ sample.

Nagel and Rosenfeld [97] used the Euclidean distance classifier – specified by the distance measure of equation (2) – with a threshold chosen to minimise the number of false rejections. The threshold selection method proposed is based on a given probability that each feature value has of belonging to the normal distribution with mean m_i and standard deviation s_i . The threshold T was defined as $z\sqrt{n}$, where a value of $z=2.575$ corresponds to a probability of 99%. Ammar et al. [13] used the Euclidean distance classifier with the above threshold selection method but reported that a smaller value of z was needed in order to achieve acceptable error rates for their data. In particular, a value of $z=1.51$ was considered most appropriate for the verification of skilled forgeries. However, it seems that their argument may not be entirely correct, as it appears to be that the threshold of $T = z\sqrt{n}$ was used with the Euclidean distance measure described by equation (3). On the other hand, for this distance form the n factor in the determination of T is cancelled out and, hence, the threshold should only be determined by the value of z . If that was the case, then the redundant factor \sqrt{n} could be the reason why a much smaller value of z was needed for a more appropriate threshold definition.

Several verification criteria have been used by Ramesh and Murty [118], including the Euclidean distance classifier with the above threshold selection method, which was denoted as the *Confidence Interval approach*. This approach can be employed where no forgeries are available for training. Another approach suitable for this case is the *N-dimensional boundary*. According to the authors, this method involves a set of thresholds $\{T_1, T_2, \dots, T_N\}$ corresponding to each of the features, rather than a single threshold T . This solves the case where a forgery is far from the genuine set with respect to one feature but close to it with respect to all other features. The set of thresholds is represented by a set of lower and upper limits. For a given feature the limits are defined by the minimum (f_{min}) and maximum (f_{max}) feature values of all the genuine training samples, a corresponding deviation measure (u) and a factor of deviation (k_u) – i.e. the lower and upper limits are given by $f_{min} - k_u u$ and $f_{max} + k_u u$.

respectively. For the case where both genuine and forgery samples are available two other approaches were proposed: the *Min-Max approach* and the *Neighbourhood approach*. The former involves the calculation of the maximum distance between a forged training sample i and all the genuine samples, say d_{\max}^i . The forgery sample closest to the genuine training set is identified by selecting the minimum of all d_{\max}^i values (let this be d_m). The threshold is defined as $T=k_m d_m$ where k_m is a constant. Thus, if the maximum of the distances between a test signature and all the genuine training samples is greater than T the signature is labelled a 'forgery', otherwise it is considered a 'genuine' sample. The *Neighbourhood approach* involves the use of the Nearest-Neighbour Classifier (NNC) or the K-Nearest-Neighbour (KNNC). The decision is based on which class, the genuine or the forgery, is closer to the test signature. The NNC decision is based on the class label of the training sample which is closest to the test signature, while the KNNC considers K nearest training samples and the decision is formed on the class label which has the majority in the K neighbours. The authors explain how the KNNC has a better performance than the NNC with respect to genuine signatures, but performs worse with respect to forgeries. As a result a modified KNNC was proposed, implementing distance measures between the test signature and each nearest neighbour, associating the latter with an importance measure. The sum of the importance measures for each of the two classes is considered and compared, and hence, the decision is made according to which is the largest. Finally, a *Hybrid approach* was also proposed integrating different classifiers. According to this approach, if the modified KNNC determines a signature as genuine then the case is passed to any of the other previously mentioned classifiers for a more accurate decision, while if the signature is concluded as forgery no further action is taken. The authors performed an extensive comparative study for the evaluation of the performance of the different classifiers under different types of features and concluded the following. The Confidence Interval classifier generally performed better than the Min-Max classifier. While it was suggested that a good verification system should not depend on a neighbourhood classifier alone, two hybrid approaches combining the modified KNNC with the Confidence Interval classifier and with the N-dimensional boundary were proposed and the results showed that an integration of the Neighbourhood method with the threshold-based techniques performed well, particularly for forgeries.

A different classification process, which will be discussed more analytically in relation to the dynamic approach, and which relates to signal comparison and matching is evaluated through dynamic programming. Although Dynamic Programming Matching is usually associated with on-line techniques, it has been used in a number of off-line related studies. Nouboud [101] used dynamic programming for the comparison of characteristic curves, Qi and Hunt [117] used a dynamic programming procedure for the matching of grid features, and Sabourin et al. [125] for the matching of primitive sets. Furthermore, Herbst and Coetzer [63] used a dynamic programming algorithm for the warping of signature projections (rows of their sinograms), and Yoshimura and Yoshimura [153] for the matching of the x-axis projections.

Finally, several studies including those by Bajaj and Chaudhury [20], Baltzakis and Papamarkos [21], Huang and Yan [65], and Cardot et al. [33] have reported on the use of Neural Network classifiers for the off-line verification of signatures. As an investigation of the implementation of Neural Networks is out of the scope of this review, the reader is referred to Gonzalez and Woods [56] for a brief description of the principles of neuromorphic systems and to Leclerc and Plamondon [78] for a comprehensive review of studies in the field of automatic signature verification using Neural Networks. The current report will be restricted to a definition of the term only as phrased by Gurney [60]: *“A Neural Network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns”*.

Performance evaluation of systems

So far it has been shown that many different techniques have been used in a range of research studies: different types of preprocessing have been applied, features extracted or classification methods used. Moreover, each study makes use of its own algorithmic implementation and database, some of which may include only random or simple forgeries while others may use skilled forgeries. Therefore, it is obvious that a comparison between the verification performance of different systems is not possible

in any meaningful way. For this reason, a report on the attained error rates achieved in different studies will be included here only as an attempt to give the range of attainable performance in the static verification of signatures.

Nagel and Rosenfeld [97] used local and global features including ratios of stroke heights and slope features for long letters, with the Euclidean distance classifier and the common threshold procedure discussed earlier, and reported an 8% FRR and a 0% FAR using simple forgeries in a single Jackknife experiment and a 12% FRR and a 0% FAR in a double Jackknife test. Note that the Jackknife procedure is used when the true samples available for training are very few. According to this method, one (single Jackknife) or two (double Jackknife) genuine samples are left out at a time for testing, while the remaining true samples are used for training. However, only a small database of 11 genuine signatures and 14 simple forgeries was used.

The study of Ramesh and Murty [118] can give a comparative view about the performances of different classifiers and different types of features. With a signature database of 650 signatures, 15 out of the 20 genuine samples from each of the 15 signers and 13 out of the 23 corresponding forgeries were used for training. The authors used global geometric features, moment-based features, envelope features, and wavelet features, while the decisions of the four subsystems were combined to form a final decision in a combination system. Note that weighted feature vectors were implemented, with the weights determined by Genetic Algorithms, and this approach was found to represent the patterns better than the raw feature vectors. While the single verification performance of each type of features was comparatively assessed also for different classifiers, the combination system gave generally a relatively good performance. As an indication of the performance attained for different classifiers and the combination system the following rates were attained. In the region where the Confidence Interval classifier produced less than 15% errors for genuine signatures, the errors varied from 25% up to 80% for forgeries. In the corresponding region for the Min-Max classifier the minimum error was 40% for forgeries. Furthermore, with the N-dimensional boundary classifier, in the region where this classifier produced less than 15% error for genuine cases, the minimum error was 10% for simple forgeries and 40% for skilled forgeries. For the modified-KNNC verification accuracies of 95.6% for genuine signatures and 86.7% for forged

signatures were achieved. For the first hybrid system (modified KNNC and Confidence Interval classifier), focusing on the low threshold region, a 95% accuracy for simple forgeries, more than 90% for genuine signatures, and about 70% for skilled forgeries were reported, but at the cost of a rejection rate as high as 20%. A slightly better performance is reported for the second hybrid system (modified KNNC and N-dimensional boundary) with a verification accuracy of 90% for genuine cases, over 98% for simple forgeries and around 70-80% for skilled forgeries, at a cost of 15-20% rejections. Although the error rates reported in this study may appear large with respect to the optimisation methods employed, they nevertheless seem realistic for the problem of static signature verification.

Another comparative study of the performance of different feature representations was presented by Qi and Hunt [117]. The database used comprised 450 signatures of which 300 were genuine and the rest were forgery samples. For the training of the system 15 genuine samples were used from each signer, while only 5 separate samples were used for testing. For the simple forgeries the subjects were given only the printed name of the person, while the skilled forgeries were produced by people skilled in hand-drawing or simulating signatures. The authors implemented two types of features, 7 global geometric features and local grid features. For the geometric features the Euclidean distance measure was used, while for the grid feature space a dynamic programming algorithm was proposed. The verification of a test signature was made by calculating the minimum distance between the test sample and the templates. The measured distance was compared to a threshold which was set to minimise the total error rate. Although it was recognised that the determination of threshold by minimising the total verification error tends to underestimate the mis-verification rate, this threshold selection approach was adopted for a comparison of the different feature representations separately and collectively for signature verification. The combination of the two feature representations needed a special arrangement and thus, for the verification function, the weighted average of minimum distances was used, where the weight vector was derived so as to reflect the discriminating power of each scale. The error rates obtained when using skilled forgeries were $FRR=8.8\%$ and $FAR=15.7\%$ for the geometric feature implementation, and similar results were obtained for the grid feature implementation ($FRR=8\%$ and $FAR=16\%$). When using the combination of both geometric and grid features the

attained error rates were much lower i.e. $FRR=3\%$ and $FAR=8\%$. For the case where only simple forgeries were used the error rates were much lower for the single-verification functions, while reaching 0% when the combined system – with both geometric and grid features – was used.

Herbst and Coetzer [63] proposed the use of the Radon transform with a dynamic programming algorithm. The Radon transform of a function consists of projections of that function obtained at different angles. It is often displayed as a grey-scale image which is called a sinogram. A dynamic programming algorithm was proposed to warp a projection of a test signature onto the corresponding projection of a reference signature. The database consisted of 966 signatures with both skilled and simple forgeries. The training set was formed from 10 genuine samples per writer, while the test set from 20 genuine samples, 6 skilled and 6 simple forgeries. The reference signature was the best representative selected from the 10 sample set, i.e. the one having the lowest average dissimilarity with respect to the other nine samples. The verification results reported include an EER of about 23% when using only skilled forgeries and an EER of about 10% when using only simple forgeries. For a small FRR of about 5.2% an FAR of about 41.3% was obtained when all forgeries were considered.

Smaller error rates were reported by Nouboud [101] where characteristic curves from signatures were extracted and compared using dynamic programming. The reported database contained 1000 signatures, written on plain white paper, from 100 people, with 6 genuine samples per person retained as references and 4 used for testing. The error rates obtained were an FRR of under 2% and an FAR of about 8%. In another experiment with bank cheques the author used both characteristic curves and geometric parameters. The latter included 39 features, such as inertia moments, slopes of the signature outlines, density of the lines, perimeter, and surface, from which the most stable and discriminating parameters were selected, although no information is given on how this was carried out. The resulting error rates were an $FRR=2.1\%$ and an $FAR=1.8\%$. Another experiment is also reported where image, characteristic curves and geometric parameters were used with neural networks, while based on the previous database. The error rates reported in this case were $FRR=3.3\%$ and $FAR=0.6\%$. It is also noted that the experiments were carried out on European

signatures. While very low error rates are reported in this study, the information provided about the implementations used is limited and, hence, does not allow a full assessment of its performance here.

Ammar et al. [12] used 7 global pressure features: the pressure factor, the maximum grey level, the range of grey levels (max – min value), the threshold separating high pressure regions from other image parts, the area of the signature, and the positions at which the vertical projection of the binary image and that of the high pressure image have a peak frequency. These features were used with a weighted distance measure proposed by the authors and two threshold selection procedures – a simple and a revised threshold – and reported an FRR=10% and an FAR=6.5% with the simple threshold, compared with an FRR=4% and an FAR=6.5% with the revised method. While using the leave-one-out (single Jackknife) method for verification, the authors tested the performance of their system also with the Euclidean distance measure and reported an FRR=9% and an FAR=7.5%. The reported database included 200 genuine signatures from 20 people and 200 skilled forgeries written by 10 different forgers.

Sabourin et al. [124] proposed a different approach to signature image representation. The signature was segmented into arbitrarily-shaped primitives, while for the primitive extraction process information about directional data in the gradient space was used. The comparison process was responsible for the structural match between a reference primitive set and a test primitive set. More specifically, the comparison process included the *Local Interpretation* of test primitives (LIP) process, which is related to the final merging of test primitives and permits the labelling of all primitives from the test primitive set, and the *Global Interpretation of the Scene* (GIS) process, which proceeds with the structural match of a reference primitive set and a test primitive set, and results in a static similarity measure. With a database of 800 signatures, the attained results showed an FRR=1.50% and an FAR=1.37% with random forgeries, for the best case when using a minimum distance classifier and two reference signatures.

It is apparent that the verification results reported in all of these studies are very different, ranging from error rates of near zero with random and simple forgeries up to about 50% with skilled ones. This shows that the evaluation of a system cannot be

carried out through the attained error rates per se, but through an evaluation of the different parameters (types of forgeries, size of database, number of reference samples used for training, etc.) and implementation strategies (preprocessing, feature selection, distance measure and threshold selection, etc.) used. A comparison of the performance of the various systems discussed here is provided by Table 2.1. It seems that systems based on pseudodynamic features, as the case of the last two studies reviewed, perform generally better than systems based on other types of static representations. It may also be said that systems using a combination of different types of representations in the feature space show an improved performance compared to that resulting from the single representations alone. Although the combination systems may be computationally expensive, the improvement generally offered in the verification performance seems to be worthwhile. Finally, it was observed that the performance of off-line systems with skilled forgeries is generally poor, which sets a further challenge to the static verification approach.

2.1.2 Dynamic Verification

Dynamic signature verification techniques rely primarily on the on-line acquisition of the signature data, most often by means of a digitiser or an instrumented pen. Other more unusual methods include a camera-based acquisition system incorporating a visual tracker of the pen-tip position in the writing surface [95] and a glove-based acquisition device in conjunction with a virtual-reality-based environment to support the signing process [141]. A standard digitising tablet captures the x- and y-coordinate pairs of the pen-tip position on the image plane at a constant sampling rate (ranges between 60Hz and 200Hz have been reported [87]), but more recent tablets can capture the position of the pen in the air when it lies within a small distance from the surface of the tablet, the pen-tip pressure, and, in some cases, the pen azimuth and altitude angles, etc. Furthermore, some instrumented pens are designed to capture the acceleration of the pen on the writing surface. From the position coordinates, the pressure and time information a number of signals may be derived such as position signals, velocity and acceleration signals, pressure signal, etc. It is, therefore, apparent that on-line systems, being able to capture information about the dynamics of a

Table 2.1: Static signature verification systems.

Research Study	Signature Database	Feature Extraction	Classification Process	System Performance
Nagel and Rosenfeld [97]	11 genuine signatures (6 + 5) samples - 2 writers 14 simple forgeries (9 + 5) corresponding forgeries different forgers	- 2 global features - 2 types of local features	weighted Euclidean distance	8% FRR average result of 0% FAR single Jackknife test
			Threshold set to minimise the FRR	12% FRR average result of 0% FAR double Jackknife test
Ramesh and Murty [118]	650 signatures 20 genuine samples/ writer 15 writers 23 (simple+skilled) forgeries/ sign. 5 forgers	- global geometric features - moment-based representations - envelope characteristics - wavelet features	1. Confidence Interval Approach (weighted Euclidean distance) 2. N-dimensional boundary 3. Min-Max Approach	⇒ combination system: 15% FRR 25% FAR simple, 50% FAR skilled ⇒ combination system: 15% FRR 10% FAR simple, 40% FAR skilled ⇒ combination system: 15% FRR 40% FAR simple, 65% FAR skilled
	<u>training:</u> 15 genuine samples/writer 13 forgeries/ signature	weighted feature vector Genetic Algorithms	4. Neighbourhood Approach ? NNC, KNNC ? modified KNNC 5. Hybrid Approach 1. modified KNNC + Confidence Interval Classifier 2. modified KNNC + N-dimensional boundary	⇒ 4.4% FRR, 13.3% FAR ⇒ < 10% FRR, 20% rejection <5% FAR simple, 30% FAR skilled ⇒ 10% FRR, 15-20% rejection 2% FAR simple, 20-30% FAR skilled
Qi and Hunt [117]	450 signatures (300 G +150 F) 20 genuine samples/ writer 15 writers 5 simple forgeries/ sign. (5 forgers) 5 skilled forgeries/ sign. (5 different forgers)	- global features - local grid features	1. Euclidean distance 2. Dynamic Programming	for Geometric feature representation: 8.8% FRR, 15.7% FAR skilled forgeries for Grid feature representation: 8% FRR, 16% FAR skilled forgeries
	<u>training:</u> 15 genuine samples/ writer		Threshold set to minimise the total error rate of each verification function	combination system: 3% FRR, 8% FAR skilled forgeries 0% FRR, 0% FAR simple forgeries

Table 2.1: Static signature verification systems (continued).

Research Study	Signature Database	Feature Extraction	Classification Process	System Performance
Herbst and Coetzer [63]	966 signatures simple and skilled forgeries <u>training:</u> 10 genuine samples/ writer	sinogram of signature	Dynamic Programming	23% EER for skilled forgeries 10% EER for simple forgeries 5.2% FRR 41.3% FAR all forgeries
Nouboud [101]	1000 signatures 100 signers <u>training:</u> 6 genuine samples/ writer	characteristic curves	Dynamic Programming	2% FRR 8% FAR
Ammar, Yoshida and Fukumura [12]	200 genuine samples 20 writers 200 skilled forgeries 10 forgers	7 global pressure features	weighted distance Euclidean distance	with simple threshold 10% FRR 6.5% FAR with revised threshold 4% FRR 6.5% FAR for single Jackknife test 9% FRR 7.5% FAR
Sabourin, Plamondon and Beaumier [124]	800 signatures 20 writers 40 genuine samples/ writer <u>training:</u> 248 signatures - 8 writers 2 genuine samples/ writer	arbitrarily shaped primitives	structural matching minimum distance classifier	1.50% FRR 1.37% FAR random forgeries

signature – which are both regarded to be characteristic for individuals but also hides important information from potential forgers – are more promising than their off-line equivalents. Hence, a much greater number of studies in the field of ASV are engaged in the dynamic rather than the static approach. Studies providing a review of on-line signature verification methods and systems include those by Plamondon and Lorette [112], Leclerc and Plamondon [78], Lorette and Plamondon [87], Plamondon and Lorette [113], and Gupta and McCabe [58].

Preprocessing

The preprocessing involved in the on-line approach is generally less extensive than for its static counterpart, since the signature data obtained from the digitising tablet is much less noisy or redundant. Depending on the capturing device, if the quality of the signal is good then minimal preprocessing is necessary. In fact, many relevant research studies omit this stage altogether. The preprocessing stage may involve smoothing, resampling, normalisation, extraction of pen-up/pen-down signal, detection of gaps, and segmentation of the signature.

Herbst and Richards [64] suggested smoothing the data points representing the signature with cubic smoothing splines, in order to remove discretization errors caused by the digitising tablet, while Plamondon et al. [116] proposed filtering the signature coordinates with a third-order linear low-pass filter having a bandwidth of about 10Hz. Furthermore, Hastie et al. [62] smoothed each coordinate of the signature separately using a cubic smoothing spline. It was explained that very little smoothing was performed, enough to eliminate the small amount of measurement error, but mainly to have a differentiable-function representation of the entire curve. Segmentation of the signature was also performed, at regions of low pen speed. The lower speed threshold was chosen to be 15% of the mean speed, although a modification of this was proposed which results in thresholds between 15-18% leading to a more stable segmentation process. Jain et al. [68] used a Gaussian filter to smooth the signature and proposed resampling the signature uniformly with equidistant spacing. The critical points – start and endpoints of a stroke and points of trajectory change – were extracted before preprocessing and retained throughout the resampling and smoothing process. In a different approach, Lee et al. [79] minimised the effect of the spatial resolution of the tablet by eliminating one of any two

consecutive sample points that were separated by only one or two basic spatial resolution units. The authors also eliminated the problem of signature rotation and shifting by providing a horizontal line on paper taped over the tablet surface as reference for writing, while a normalised feature set was also proposed. Using normalised features is generally a better response to normalisation issues, as resizing can affect the image quality, but this may not always be appropriate. Normalised features and detection of gaps in the signature are issues further discussed by Penagos et al. [107].

Brault and Plamondon [28] proposed a method to segment a signature at its perceptually important points. According to this method, a vertex is constructed for every point of the signature, centred on that point, with the help of neighbouring points as specified by prescribed geometric conditions. A function is then determined to weight the perceptual importance of every point, while the segmentation points are selected to correspond to the local maxima of that function. Furthermore, it is suggested that the sampled signature data be preprocessed with a moving average filter – a Gaussian kernel – to remove the noise that comes with the acquisition process from the digitiser. In another study, Plamondon [108] proposed a general segmentation framework for the partitioning of handwriting, including handprinted characters, cursive script and signatures, based on a handwriting generation model and an analysis-by-synthesis experiment. According to this view, a handwritten trace can be hierarchically segmented into components, strings, and curvilinear and angular strokes. It is highlighted that the strokes, which are the fundamental units of a handwriting signal, have to be superimposed to generate a smooth handwritten trace, and are thus partially hidden in the signal.

It should be noted that different definitions for a ‘stroke’ exist in the literature of ASV. Schmidt and Kraiss [128] define a natural stroke as a writing segment between a pen-down and a pen-up movement, which can be derived from the writing pressure signal of the signature, and further propose a segmentation algorithm to partition a stroke into characteristic segments and identify significant points. In the same context, Dimauro et al. [46] defined a stroke as the writing included between a pen-down and a pen-up movement of the instrumented pen. A distinction was also made between pen-down and pen-up strokes. The latter, corresponding to invisible parts of the signature,

show greater variability than the pen-down counterparts and, hence, were not considered for the verification process. In a different approach, Schomaker [130] defines a stroke as a trajectory between two velocity minima. It is also illustrated that the points corresponding to a dip in the velocity are characterised by high curvature.

Feature extraction and selection

The feature extraction stage involves the representation of signatures with functions or parameters. The former approach is easier at this stage, since it involves the extraction of signals which can be either directly measured in the acquisition process with special pens or derived from other signals – as, for instance, the velocity and acceleration signals may be respectively obtained from the first and second time derivatives of the position signals. The parameter approach involves the extraction of specific parameters from the signals such as maximum, minimum, or mean values, etc, but special care must be taken in finding appropriate parameters, which are ideally both stable with respect to the genuine class and discriminatory against forgeries.

The signals involved in the function-based approach are the position, velocity, acceleration, pressure, etc., while an implementation may use one or more signals together. Sato and Kogure [127] proposed the use of the coordinates in a complex function defined as: $z(t) = x(t) + iy(t)$, which was normalised with respect to the time duration, location, rotation, constant movement (a result of signing from left to right) and size, and the use of the pressure signal, which was also normalised with respect to the time duration and amplitude. Herbst and Richards [64] proposed the use of two functions: the rhythm (speed signal) and the curvature of a signature. Liu et al. [84] used the pressure waveform and the two orthogonal acceleration components obtained with two accelerometers, while Tanabe et al. [138] proposed the use of the pressure function obtained with a new digital pen device sensing the z-axis component of writing pressure. Plamondon [109] suggested the use of the velocity domain, and particularly proposed the use of curvilinear and angular velocities, for the representation of signatures after evaluating the results of a number of studies on different signal representations. Moreover, a comparative study on the position, velocity and acceleration signals, reviewed – while also a joint study – by the author

concluded that the velocity domain was the best representation space for a 2D signature verification system.

The parameter approach usually involves the extraction of a number of dynamic parameters such as time features, statistical values (mean, maximum, minimum, etc.) computed on the velocity, acceleration and pressure signals, while sometimes shape related features, such as length, curvature, width/height ratio, are also incorporated. The extracted parameters may be global, such as total signing time, number of components, number of zero crossings, mean velocity, mean acceleration, or local described by measurements made on segments, components, or strokes – examples include starting direction, starting moment, etc. It is apparent that a great number of parameters may be extracted. Lee et al. [79] extracted a 42-parameter feature set, which was advanced to a set of 49 normalised features, Achemlal et al. [8] proposed a 40-parameter set, Nelson et al. [100] a 25 sized feature set, Crane and Ostrem [43] a set of 44 features, while Bauer and Wirtz [23] extracted as many as 300 parameters. While the feature extraction process at this stage is an easy task, the selection of appropriate features to be used in the verification process is a more challenging issue.

Bauer and Wirtz [23] performed an elimination of highly correlated features, elimination of uncharacteristic parameters by Fisher's F test and personalised parameter selection using a classifier and the lowest error rate criterion. Lee et al. [79] proposed algorithms for selecting features according to the availability or otherwise of forgery training data based on distance measures, and an alternative to optimum individualised subsets being the determination of the *common* feature set with features that are good for most people in the population. The criterion used by Crane and Ostrem [43] for the selection of features involved the direct minimization of the Type I/ Type II error rate, while the use of personalised feature sets was also proposed. Nelson et al. [100] selected the subset of features which showed a reasonable separation in their genuine/forgery distributions, while a further subset of features for each signer was selected that had the least variability with respect to the training set. The approach proposed by Achemlal et al. [8] for personalised feature selection was the maximisation of the difference between the two thresholds: the one under which all forgeries lie and the one above which all genuine signatures lie. In a different approach Penagos et al. [107] proposed user-dependent weights for each parameter –

i.e. the parameters found to be more constant were assigned more weight than those which varied among the reference samples. Other studies involved with discriminative feature selection, personalised parameter selection or personalised parameter weights include respectively those reported in [75;99;120]. As the feature selection procedure when based on the error rate minimisation criterion can be computationally exhaustive, requiring a vast amount of time to complete the search especially when it involves a large list of parameters, Brittan and Fairhurst [30] and Fairhurst and Brittan [50] proposed the use of parallel processing for optimisation in individualised feature selection. According to this parallel implementation, the evaluation algorithm is replicated across a number of processors which work in parallel. Each section, given an active feature vector, adds or deletes its own features assessing their relative merits, and then returns the results to the master processor for global update and assessment.

The combination of both a parameter-based and a function-based representation in a single verification system has also been proposed. Nelson and Kishon [99] proposed a fast first stage verification procedure using dynamic features such as duration, length, rms (root-mean-square) measures of speed, centripetal and tangential acceleration, etc., to screen out only crude forgeries and signing mistakes. Signatures passing the first stage would enter the second stage, which was based on an elastic matching of speed functions and a 2D shape analysis procedure. Plamondon et al. [116] proposed a multi-level signature verification system using three different representations: one based on global parameters and two based on functions. The first level makes use of simple dynamic parameters such as the total pen-down time, percentage of pen-up duration, percentage of time that the pen is stopped, etc. The second level calculates the curvilinear and angular velocity signals, while segmentation of these signals is also performed. Finally, the third level is involved with the signature's trajectory information, and is concerned with the normalization of the signature image.

Classification process

In the parameter approach the classification process generally involves the evaluation of a distance measure to compare the feature vectors of the test signature and the reference set. Some of the possible classification procedures have been described in the earlier section describing the static approach. A widely used distance measure is

the Euclidean Distance Model. Nelson and Kishon [99] used the Euclidean distance measure in a first stage screening procedure to quickly reject crude forgeries or errors in signing. The threshold was selected as 10% greater than the largest distance occurring in the reference set for a particular signer. Crane and Ostrem [43] proposed the use of the weighted Euclidean distance metric and a pre-defined threshold selected using the Type I/ Type II error curves to obtain the optimum trade-off between the two types of errors according to the interests of particular applications. Bauer and Wirtz [23] proposed a number of different Norms, including the Euclidean distance metric named *Pseudo-Euclidean Norm* which did not consider parameter selection, and different norms that incorporated in different ways personalised parameter selection information based on the error rate of every parameter taken alone. Achemlal et al. [8] proposed the use of the linear classifier, as given by equation (4), with a pre-selected threshold defined by $1.1 \cdot S_v$ where S_v denotes the threshold above which all true signatures are accepted.

$$dist = \frac{1}{n} \sum_{i=1}^n \left| \frac{t_i - r_i}{r_i} \right| \quad (4)$$

Here, n is denoted the number of features used, t_i the value of the i th feature of the test sample and r_i the reference mean for the i th feature. Another classification approach is the Majority Decision Rule which is given by equation (5) as defined in [79]:

$$N_a = \left| \left\{ i : \frac{|t_i - m_i|}{s_i} \leq a \right\} \right| \quad (5)$$

According to the Majority Decision Rule if $N_a \geq n/2$ then the test signature is declared genuine, otherwise it is classified as a forgery. As earlier, t_i is the value of feature i of the test signature, while m_i and s_i are the mean and standard deviation respectively of the reference set for a particular signer, and a is a fixed threshold. Thus, the Majority classifier weighs the features equally and not based on their distance from the reference and, hence, considers only the number of features that pass the criterion. Lee et al. [79] proposed two modified versions of the Majority Decision Rule, the *Presoft Majority Decision* and the *Prehard Majority Decision* to work in conjunction with time-normalised features, according to which a signature is penalised if its writing time is highly deviant from the average reference time. The results showed an improvement in performance with the use of modified Majority Decision Rules. In a comparative study, Nelson et al. [100] evaluated the performance

of different statistical methods for the on-line verification of signatures: the Euclidean Distance Model, the Mahalanobis Distance Model, the Quadratic Discriminant Model –which uses statistical properties of both forgeries and genuine signatures, and the Majority Vote Scheme. The results showed that the Euclidean Distance Model performed better than the Mahalanobis distance and the Majority Vote, and was found an adequate method for applications that require a very small FRR at the expense of a higher FAR. The more complex Quadratic Discriminant Model using forgery data performed much better than all the other methods and appeared a promising choice for applications that require a much smaller FAR.

The comparison between a test and a reference signature in the function-based approach involves the matching of the two respective signals. Because of the variations encountered among signature samples in the time domain, even for the same signer a direct alignment of points cannot be performed. A number of signal matching techniques have therefore been proposed.

Liu et al. [84] proposed a Regional Correlation algorithm to compare two waveforms on the basis of heuristically determined regions derived initially from pen-paper contact. The segments, which were modified based on the duration of the interval and discrepancies between the test sample and reference, were then shifted in the correlation analysis through 20% of the reference segment's duration to find the best match. The authors used the pressure and acceleration signals and proposed a verification measure which combined, by averaging, both the pressure correlation and the acceleration correlation measures. Another signal matching technique which has long been used for speech recognition is Dynamic Time Warping. Yasuhara and Oka [149] explain how linear time alignment, which performs linear expansion or compression of the time axis so that the two signatures under comparison have the same time length, do not perform very well due to local variations in the signatures. For this reason non-linear time alignment is proposed, which allows for the local fine alignment of time. In this context, the time-registration algorithm searches for a path for which the Euclidean distance between the template and the input signature becomes minimum along the entire path within a limited time warping range. A dynamic programming algorithm to perform the recursive search is developed. The classification decision compares the final distance between the template and the input

signature to a predefined threshold and if it is greater than the threshold then the signature is rejected as forgery otherwise accepted as genuine. Finally, the authors proposed the use of a force function as a signature representation signal. The Dynamic Time Warping method has been widely used in the literature of ASV with different signals such as position, velocity, pressure, force, etc., while embedded in various implementation strategies. Some of these studies are included here [62;64;89;95;96;127;138;147]. Another signal comparison approach is Skeletal Tree Matching, which evaluates the distance between the corresponding tree representations. A waveform is represented by a tree, which includes information about the succession of peaks and valleys in the waveform. The magnitude of a peak/valley is reflected by the depth of a leaf in the skeletal tree representation. Gupta and Joyce [59] proposed a modification providing a single representation for the two position profiles. Thus, the signature shape was represented by a string interleaving the x and y profile representations. Furthermore, information about the peak/valley magnitudes and the time duration between consecutive peaks/valleys were also incorporated in the representation. The comparison between the test and reference signature representations was achieved with a string matching algorithm.

Parizeau and Plamondon [102] compared the performance of the three signal matching algorithms: Regional Correlation, Dynamic Time Warping and Skeletal Tree Matching, over the same database containing signatures, handwritten passwords and initials and tested against random forgeries, while using three different types of signals: position, velocity and acceleration. The verification errors – evaluated by searching for the threshold which minimized the total error – showed that no algorithm outperformed the others in all cases. With respect to signatures significant differences between the algorithms were observed only for the following signals: $x(t)$, $u_y(t)$ and $a_x(t)$ (i.e. x-position, y-velocity and x-acceleration component) whereas the Regional Correlation performed best for the $x(t)$ and $a_x(t)$ signals and the Skeletal Tree Matching was the best for the $u_y(t)$. It was also concluded that all three algorithms were less discriminant with the acceleration signals. With respect to computation time, Regional Correlation was the fastest algorithm, with Dynamic Time Warping coming second, while Skeletal Tree Matching was very slow. Finally, it was suggested that the Dynamic Time Warping algorithm is used in the velocity domain as it is less parameter-sensitive than Regional Correlation. It should be noted

here that parallelism in Dynamic Time Warping has been proposed by Bae and Fairhurst [19] for automatic signature verification reducing the computational complexity inherently involved.

It should be mentioned that in a function-based approach the reference set representation is not as simple as selecting the mean feature vector in relation to the parameter approach, and hence a number of studies are concerned with finding the best reference set representation. Kim et al. [74] proposed the selection of a prototype signature to represent the reference set. For the prototype determination the difference between each of the reference samples with the others from the set was calculated using Dynamic Programming Matching. The reference sample with the smallest average dissimilarity among the other reference samples was chosen as the prototype signature. Multiple prototype signatures were also suggested to accommodate a range of variations in variable signatures. In a different approach, Wirtz [148] proposed the construction of an average prototype signature by performing time- and position-based averaging of representative input signatures for a stroke-based verification approach. Furthermore, Congedo et al. [36] proposed the selection of three best signatures from a seven reference sample set based on a local stability index determined by the frequency of *direct matching points* occurring when one genuine signature is matched with another through an elastic matching procedure.

The threshold selection procedures used in the function-based approach are also more interesting. Kim et al. [74] proposed two methods for the determination of the decision threshold for a particular signer i : one based on the mean difference of the prototype from the rest of the reference samples $\theta_i = c \cdot \mu_i$, and the other as a function of both the mean and the respective standard deviation value $\theta_i = \mu_i + c \cdot \sigma_i$, where c is a constant. Jain et al. [68] suggested the selection of a common threshold based on the minimum error criterion evaluated on the training data. The use of writer-dependent thresholds was also proposed and the selected values were based on the common threshold and the addition of a writer-specific offset. For the determination of the latter, three approaches were investigated: the use of the minimum distance between all the references, the average distance, and the maximum distance, where the distance was measured after matching pairs of strings using dynamic

programming. The results showed that the minimum value gave the lowest error rates. Yoshimura and Yoshimura [151] selected three representatives from the reference set of a writer and calculated the dissimilarity between each of the remaining reference samples z_j from each of the three representatives z_i , using an Elastic Matching procedure. The proposed threshold was determined by: $T_h = c \cdot \max[\min\{D(z_i, z_j)\}]$, where the minimum was taken over the distance from z_i 's and the maximum over the distance values corresponding to each of the z_j 's, and c was a fixed constant.

Although perhaps to a lesser extent – compared to the relative literature volume – than that observed in the static verification approach, Neural Networks have also been used as a classification approach for the dynamic verification of signatures [78;81;90;129].

Performance evaluation of systems

Similarly to the static systems evaluation, a brief review of the verification performance attained in some of the research studies will be included here in an attempt to give an indication of the rates that are likely to be obtained in a dynamic approach. However, it is reiterated here that any attempt to directly compare the several systems would be meaningless.

Achemlal et al. [8] proposed an initial set of 40 parameters from which 10 were selected based on the maximum discrimination between true signatures and real forgeries for each signer. The training of the system was based upon 10 genuine and 10 forged signatures for every signer, while 60 genuine signers and one forger took part in the experiment. For the comparison process the Linear classifier and the pre-selected threshold earlier described were used. Although not very clear in the paper, it seems that the FRR was computed on a new set of genuine signatures: 1 sample per writer, and the FAR from skilled forgeries performed by one forger having up to 3 attempts per signature. The error rates obtained were 11% for the FRR and 8% for the FAR.

Using a large database of more than 10,000 signatures, Lee et al. [79] proposed a 49 normalised feature set from which 15 individualised parameters were selected. Testing several classifiers, the best performance was obtained with a modified version

of the Majority classifier. The *Prehard Majority Classifier* declared a signature as a 'forgery' if its writing time was highly deviant from the average writing time of the corresponding reference set, in particular if their absolute difference was greater than 20% of the corresponding average writing time. An EER of 2.5% was obtained, or for POS applications (according to the authors, the performance requirements for POS applications are specified by a Type I error of near zero and a Type II error $\leq 25\%$) a zero FRR and a 7% FAR were achieved, with a forgery database consisting of simple, statically skilled and timed forgeries (where in addition to the static image the forger is provided with information about the average genuine writing time).

Crane and Ostrem [43] collected a database of 5,220 genuine signatures obtained from 58 subjects and 648 forgeries from 12 trained forgers. The signatures were collected using a three-axis force-sensitive pen –i.e. the instantaneous force on the pen tip in three orthogonal directions P, X and Y, was captured. The best individual feature set was selected from a 44 parameter set based on the minimum error rate criterion and 10 to 12 signatures per signer were used for the template construction. For the classification process the Euclidean distance metric was used with a predefined threshold selected using the Type I/ Type II error curves to obtain the optimum trade-off between the two errors. An EER of 1.5% was obtained with random forgeries and a 2.25% EER with skilled forgeries, while for the verification up to 3 attempts were allowed. Lower error rates were achieved when some signers were excluded for failing an enrolment criterion.

Sato and Kogure [127] based their experiment on a database of 110 genuine signatures from 11 subjects (10 Japanese and 1 Chinese) and 330 skilled forgeries performed by 3 forgers. *Pseudo-distance* measures were derived from a normalised coordinate function and the writing pressure. Adopting three principal features, the shape, motion and writing pressure, and with the use of Dynamic Programming Matching, three difference measures called pseudo-distances were defined with respect to the corresponding features, and an efficient different measure was composed from these pseudo-distances. An FRR of only 1.8% and an FAR of 0% were obtained.

Unlike most research studies that are evaluated in favourable conditions as described by laboratory tests, Liu et al. [84] performed a large-scale field test over a period of six months involving 248 subjects and a total of more than 6,000 signatures. The authors used pressure and acceleration signals derived from two accelerometers, and a Regional Correlation algorithm for the comparison of corresponding sample and reference signals. A verification measure was used combining both pressure and acceleration correlation results. The error rates obtained were an FRR of 1.7% and an FAR of 0.4% with skilled forgeries, compared with 0.02% with random forgeries. Adaptation procedures were used without which the FRR, as measured from a simulation experiment on the same data, was higher (3.2%). A slightly lower FRR (1.2%) and slightly higher FAR (0.6%) were obtained, again in a simulation experiment on the same data, when allowing four trials for verification per session.

Near zero percent errors were obtained by Plamondon et al. [116] with a multi-level signature verification system, which is based on three types of signature representation: global parameters, velocity signals and trajectory image information. The verification process is carried out in three levels: the first level calculates the weighted distance between the simple dynamic parameters of the test and reference signatures, the second level computes *intrinsic local correlation* between corresponding portions of the test and the reference signature velocity signals, and the third level calculates the distances between the test and the reference signature trajectory images by *local Elastic Matching*. The three verification steps were combined into a global classification function, while personalised thresholds were also used. Using a small database of genuine signatures and skilled forgeries, the results obtained were a FRR=0% and a FAR=0.5%.

It is generally considered that, although computationally expensive, function-based approaches to signature verification perform better than parameter-based ones [47;113]. On the other hand, the latter share the advantages of algorithmic simplicity, lower computational speed, and smaller storage requirements [87]. However, it would be almost impossible to assess the verification performance of these two types of approaches through an evaluation of the performance of the small number of different systems reported here, since the resulting performance of any system is very much dependent on the overall quality of implementation and database used. However, it

was shown that the combination of different types of representation has led to better performance than any single representation alone [23;116]. A comparison of the performance of the various dynamic signature verification systems discussed here is provided by Table 2.2.

Optional schemes

An on-line signature verification system has the advantage of being able to interact with the user. This means that optional parameters to resolve problematic entries or rejection decisions may be implemented for a commercial application environment promoting flexibility and practicality, while from a statistical aspect reducing the chance for error.

Fairhurst [49] proposed the introduction of an *enrolment validation module* in a practical automatic signature verification system, according to which individuals who failed to provide a satisfactory enrolment were allowed a retry having up to 3 attempts. The results obtained with a database of more than 8,000 signature samples generated during public trials with the KAPPA^[1] system showed that the use of the *enrolment validation module* brought a reduction in the FRR from 6.9%, with a fixed enrolment of 10 samples per enrollee, to below 1% with the enrolment validation retry facility. The evaluation of the quality of enrolment samples has also been investigated by Vielhauer et al. [144] with the help of a *Transitivity Check Matrix* computed on distance measures and a transitivity-based quality criterion, leading to significantly lower False-Enrolment-Rate (FER) and quality improvement during the verification process.

Additional options are included in the system proposed by Crane and Ostrem [43]. A simple enrolment criterion was used based on the total variance of the template –i.e. if the combined standard deviation was larger than some threshold the subject failed the enrolment criterion and was excluded. Furthermore, up to three tries per verification trial were allowed and thus, if the first signature failed the verification test then a second one was tested and if that also failed a third one was considered. Only if all three signatures for a particular verification trial failed was the subject rejected as an impostor. In addition the authors proposed an adaptive *template updating procedure*

Table 2.2: Dynamic signature verification systems.

Research Study	Signature Database	Feature Extraction	Classification Process	System Performance
Achemlal, Mourier, Lorette and Bonnefoy [8]	60 writers, 1 forger <u>training:</u> 10 genuine samples/writer (600 G) 10 corresponding forgeries (600 F)	40 parameters 10 personalised parameters selected	Linear Classifier threshold set: $T = 1.1 * S_v$ S_v : threshold above which all true signatures are accepted	11% FRR 8% FAR skilled forgeries (up to 3 verification attempts)
Lee, Berger and Aviczer [79]	10,000 signatures simple, statically skilled and timed forgeries	49 normalised features 15 personalised parameters selected	modified Majority Classifier	2.5% EER 0% FRR 7% FAR
Crane and Ostrem [43]	5,220 genuine signatures 58 writers 648 forgeries 12 trained forgers <u>training:</u> 10-12 genuine signatures/ writer	44 parameters 25 parameters selected (3-axis force- sensitive pen)	weighted Euclidean distance Threshold set to obtain the optimum trade-off between the 2 errors	1.5% EER with random forgeries 2.25% EER with skilled forgeries (up to 3 verification attempts)
Sato and Kogure [127]	110 genuine signatures 11 writers 330 skilled forgeries 3 forgers	- Shape - Motion - Writing Pressure (complex normalised functions)	Dynamic Programming Matching	1.8% FRR 0% FAR
Liu, Herbst and Anthony [84]	> 6,000 signatures 248 writers field test - over 6 months	- Pressure - Acceleration signals (2 accelerometers)	Regional Correlation averaging pressure correlation and acceleration correlation measures	17% FRR 0.4% FAR skilled forgeries 0.02% FAR random forgeries (Adaptation procedures used)
Plamondon, Yergeau and Brault [116]	<u>Forgery (IM) Test:</u> 8 writers: 3 genuine sign./writer 8 forgers -3 corresponding forgeries <u>True signatures (TR) Test:</u> 6 writers: 9 genuine sign./writer <u>training:</u> 3 genuine signatures/ writer	- global parameters - velocity signals - trajectory image information	- weighted distance - intrinsic local correlations - local elastic matching personalised thresholds	for multi-level verification system: 0% FRR (TR test) 0.5% FAR skilled forgeries (IM test)

according to which if the user successfully passed the verification trial – on the first attempt – his template was modified by adding the feature vector of the new signature with a weight of $1/8$ to the template vector. This *template updating procedure* was found to improve the verification performance by reducing the FRR. An assessment of the actual benefits provided by such procedures may be observed in the results of the large field test with the system proposed by Liu et al. [84]. The authors reported on the several hardware and program problems encountered during the trials and the respective actions taken in order to fix them. Furthermore, enrolment decisions, redesign and adaptation procedures were used to ensure higher functionality of the system, while the benefits from the use of four trials for verification were examined through a simulation experiment on the same data.

2.2 Visual Inspection of Signatures

Expert examiners

In the field of forensic science a number of specialised techniques exist and may be used to assist the visual inspection of signatures. According to Campbell [32], these include the stereomicroscopic examination of the writing on the paper, while chemical tests on the paper and the writing trace may be also performed, and different types of light and filters can be used. Special cameras may assist the naked eye and devices performing electrostatic detection can recover impressive details assisting further the work of expert examiners.

From the engineering point of view, little interest has been shown in examining the human capabilities in inspecting the authenticity of handwritten signatures. From this perspective, the corresponding performance of an individual relies solely on his/her discretion and/or expertise depending on whether the person is an expert or a non-expert. While the relative data available to document experts is a confidential issue, in the Handwritten Signature Verification (HSV) field, information on this matter is available through only a limited number of research studies. In this context, Ramesh and Murty [118] provided their signature samples to two teams of document examiners for verification. The results obtained from the human experts displayed a 100% success in correctly identifying simple forgeries, and a 75% success in

identifying skilled forgeries, while with respect to the genuine signatures the experts identified correctly 82% of the samples. Within a different framework, Brault and Plamondon [27] comparatively assessed the opinion of the imitators employed in their study and that of an expert document examiner in ranking eight signatures on their apparent imitation difficulty, and found the opinions almost inverted.

Non-experts

With respect to the performance of non-experts in inspecting the authenticity of handwritten signatures, again very few studies exist in the field of HSV. As reported in Chapter 1, Randolph and Krishnan [119] have cited previous studies stating that the majority of forgeries accepted by bank tellers are totally unskilled, and that under test conditions, tellers accept 10-50% of forgeries as valid signatures and accept nearly 100% of forgeries in their actual jobs. Fairhurst and Kaplani [51] in an experiment with human subjects, constituted mostly of university students and staff members, obtained an average of 73.8% correct classification of some 100 signature samples containing 5 different types of signatures. In particular an average FRR of 44.7% and an average FAR of only 7.7% were obtained, suggesting that humans are relatively good in spotting forgeries. A distinction should be made here, highlighting the fact that the former study represents a field test, while the latter, and more recent one, a laboratory test.

2.3 Conclusions

A review of research studies in the field of automatic signature verification has been presented both from a static and a dynamic perspective. The several studies reported were examined in relation to the content of the section within which they were described. A brief overview of the verification performance of several static and dynamic systems revealed the obvious superiority of on-line systems over off-line systems. It was explained how direct comparisons between different research studies is not possible, as different databases, methods and implementation strategies are used. Moreover, different types of forgeries may be even considered, with some studies using random or simple forgeries, while some others use highly skilled ones. Another issue is the number of reference samples used for the construction of the

reference template. While some studies have reported on as few as 2-3 samples, a great number of studies have considered as many as 10-15 reference samples for training. In addition, several optional schemes described tend to overestimate the verification performance of a system. It is therefore obvious that any comparative assessment of the various systems would be misleading. Furthermore, it is stressed that generally a large database is essential for accurate results, as well as a fair number of participants in the several tests. With respect to field tests it was shown that very few studies have expanded out of the boundaries of laboratory tests.

Concluding, the relatively poor performance of static signature verification systems, especially with respect to skilled forgeries, highlights the need for improvement of the static approach in the field. With respect to on-line systems showing relatively high performance, a number of optional schemes may be considered in order to achieve greater flexibility, aiding their introduction to the competitive market. Finally, the need for a deeper investigation into the human inspection of signatures is emphasised, as human checking procedures remain in common use.

The study reported in this thesis is particularly involved with a static approach offering improvements in the static verification of signatures while providing an in-depth investigation into related human inspection tasks. The next chapter describes the acquisition process for the signature data used in the study reported in this thesis, and the construction of a signature database for use in the design and testing of an automatic signature verification system and several human perception experiments to be reported in later chapters.

Chapter 3

Data Collection

This chapter discusses the means and methods used for the construction of a handwritten signature database for its use in the several experiments and applications to be reported in later chapters. An experimental procedure with human subjects was the source of the large number of genuine signatures and forgeries collected. The acquisition of signatures was dynamic as the signing process was performed on a digitising tablet. Further issues involving the format of the data and the digitised signature images are reported, while the procedures used are illustrated in detail.

3.1 Introduction

Signature data may be obtained either through off-line or on-line methods. Off-line methods involve the use of a scanner or a CCD camera as input device and therefore only the 2D signature image is captured. On the other hand, on-line methods can capture dynamic signature information in addition to the static image representation. The most common means of on-line signature acquisition is the graphics tablet device or digitiser, although the use of a camera-based acquisition system, incorporating a visual tracker of the pen-tip position in the writing surface, has been proposed in the

literature [95]. The use of digitising tablets or instrumented pens (strain-gauges, piezoelectric pens, accelerometers, etc.) to capture dynamic handwritten signature information goes back as early as the late 1970s, according to data input devices reviewed by Lorette and Plamondon [87]. Meeks and Kuklinski [93] further report on the type of errors that are likely to occur when digitisers are used for data collection. These are classified as *spatial errors*, related to the position coordinates, *temporal errors*, which are due to irregularities in the sampling period or non-simultaneous sampling, and *intrinsic errors*, which are caused by change in the pen tilt and inaccurate firmware processing. Digitiser technology has since developed further providing higher accuracy and reliability, and user-friendly interfaces, while introducing the measurement of parameters in addition to the standard x- and y-coordinates, with higher sensitivity in the pen-tip pressure captured, and other characteristics such as pen altitude, azimuth, etc. A great number of parameters can be further extracted from the ones directly captured by the tablet, such as the pen-tip velocity, acceleration, and so on.

3.2 Data Acquisition

The acquisition of signature data, in this study, was dynamic and was achieved by means of a digitising tablet accompanied by a specialised stylus. The hardware and software employed are described in the following sections.

3.2.1 Digitiser

The digitiser used was a WACOM ArtPad II tablet (Model: KT-0405-R) operating with a WACOM UltraPen Ink (Model: UP-401) with either black or blue ink (Figure 3.1). The tablet specifications as reported in the user manual include the following.

- Active area (W x D): 128.0 x 96.0 mm
- Resolution: 2540 lpi

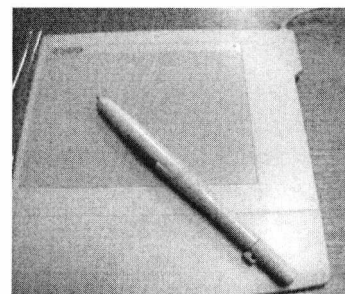


Figure 3.1: The WACOM ArtPad II tablet with Ultrapen.

- Accuracy (overall with pen): $\pm 0.5\text{mm}$, maximum
- Pressure levels: 256 levels
- Maximum reading height: 5mm or less
- Maximum report rate: 205 points per second
- Origin position: upper left

According to WACOM on-line publications [7], the communication between the tablet and the stylus is achieved through radio waves (Figure 3.2). The tablet transmits an electro-magnetic signal to the pen, which in turn modifies it and sends it back to the tablet for position and pressure analysis. A grid of wires below the tablet’s screen alternates approximately every 20 microseconds to accommodate the transmission and reception of the information. The tablet sends the information to the computer via its serial port.

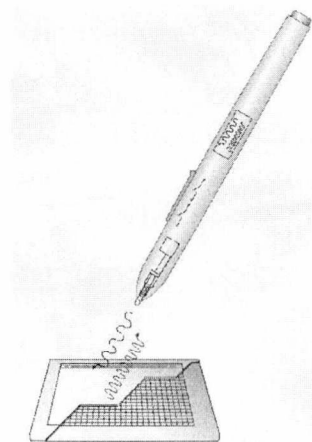


Figure 3.2: Transmission of information between pen and tablet [7].

3.2.2 Data Format

A generic capture program, developed in the Department of Electronics at the University of Kent by Cobbah [35], was the interface used for the acquisition and storage of the information transmitted to the computer’s serial port by the digitiser. The generated data included the following parameters: time stamp, x- coordinate, y- coordinate, and pressure. A sample signature file is illustrated in Figure 3.3, where the above parameters are displayed in columns in the exact sequence in which they are reported. The time stamp recorded specifies the system time at which an event occurs –such as a change in any of the parameter values

Time	X	Y	Pressure
10:25_04 - Notepad			
73627054	5430	4643	0
73627059	5536	4614	0
73627064	5567	4604	0
73627069	5597	4594	0
73627074	5597	4594	1
73627078	5703	4512	1
73627083	5753	4413	8
73627088	5778	4309	18
73627093	5782	4182	30
73627098	5753	4048	41
73627103	5686	3926	52
73627108	5567	3840	52
73627113	5410	3819	48
73627117	5220	3867	45
73627122	5015	4011	55
73627127	4825	4260	71
73627132	4660	4604	80
73627137	4548	5013	69
73627142	4505	5415	41
73627147	4532	5737	17
73627152	4625	5945	11
73627156	4767	6022	17
73627161	4931	5989	30
73627166	5104	5875	45
73627171	5277	5670	63
73627176	5447	5372	56
73627181	5590	5027	22
73627186	5693	4676	5
73627191	5746	4385	7
73627195	5746	4202	21
73627200	5730	4152	40
73627205	5637	4164	58
73627210	5538	4302	67
73627215	5429	4538	78
73627220	5319	4870	86

Figure 3.3: Sample signature file.

captured. This may explain a small drop in the time interval observed from 5msec to 4msec, which occurs every 8th recording. It should be noted that the maximum sampling frequency of the tablet, as specified, is 205Hz – i.e. a maximum sampling interval of 4.878msec. Large time gaps simply denote no change in the parameter values between the time stamps. Furthermore, the position co-ordinates captured by the tablet are reported in $1000 \times \text{cm}$, and the pressure ranges from 0 to 256 levels. Note that the origin position is the upper left corner of the tablet and hence, the y-coordinate will need to be inverted in a Cartesian-axis based implementation.

The signature image may be reconstructed, in a digitised form, from the x- and y-coordinates captured by the tablet in conjunction with sequential information provided by the time stamp. Figure 3.4(a) shows the pen down points of a signature, as a result of the position information captured by the tablet, and Figure 3.4(b) shows the continuous trace of the signature signal.

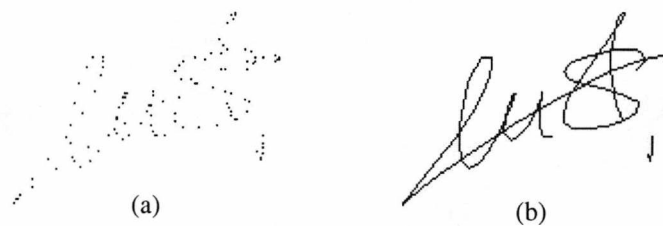


Figure 3.4: (a) Pen-down points and (b) reconstructed signature line of a sample.

3.2.3 Pen-Up Information

Along with the pen-down points extracted, the digitiser also captures position points of the pen when it is very close to the tablet but yet not touching its surface. According to the tablet specifications earlier reported, the tablet captures pen movements up to 5mm above its surface. These points where the pen is in the air moving within the recording area are denoted as pen-up points or pen-ups. The pen-up points are characterised by zero pressure, and thus can be easily located and algorithmically excluded from the reconstruction of the digitised signature image. Pen-up recordings may be observed in Figure 3.3 for zero pressure. An example of the pen-up points captured in a genuine signature sample is displayed in Figure 3.5(a),

whereas in Figure 3.5(b) both the pen-up (red line) and pen-down (blue line) contour trace is shown.



Figure 3.5: (a) Pen-up points and (b) pen-up and pen-down information of a sample.

Signatures containing only one segment usually display very few pen-ups – only at the beginning and end parts – in comparison to signatures made up of many segments, since in that case many pen-lifts are required. Poor forgeries usually display more pen-ups than their genuine counterparts, which could be attributed to hesitation observed on the part of the forger, causing more interruptions in the signature line, retracing or indecisive movements just before signing. On the other hand, increased hesitation may involve higher pen-lifts that are not captured by the tablet. Nevertheless, extensive pen-up recordings are not very common, with respect to the data collected in this study.

The nature of the pen-ups also differs between genuine signatures and poor forgeries, with the former being generally smooth, while the latter usually exhibit tremor. The pen-up movement (red line) recorded in two genuine samples of a 5-segment signature is displayed in Figure 3.6. Although the first sample (Figure 3.6 (a)) seems to begin with a unique pen-up pattern, this is not apparent in the second genuine sample (Figure 3.6 (b)). This shows that pen-up movement is not necessarily very consistent between the genuine samples of a signer. Two imitations of the previous signature, performed by different forgers, exhibit respectively too little pen-up movement (Figure 3.6 (c)) and too much (Figure 3.6 (d)), showing a different imitation approach employed by the two forgers. This suggests that pen-up information, because it is invisible to a potential forger, can make an imitation distinct from its original. On the other hand, according to studies reviewed by Plamondon and Lorette [113], pen-up movements “*are much more variable, less precise and less repeatable*”.

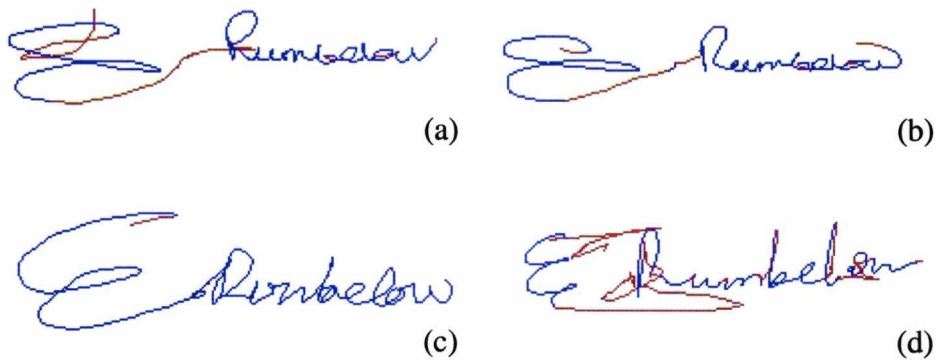


Figure 3.6: Pen-up and pen-down information for (a) and (b) genuine samples, and (c) and (d) forgeries.

3.3 Signature Acquisition

The acquisition of signatures was achieved through a data collection experiment with human subjects. Both genuine signatures and forgeries were collected, the latter being a result of an imitation test in which subjects participated. The experimental set up and a detailed description of the experimental procedure follows.

3.3.1 Experimental Set Up

For the acquisition of signatures the digitiser earlier reported was employed. In order for the signing process to be as natural as possible it was deemed necessary for the signatures to be performed on paper, which would be placed over the tablet, while making use of the UltraPen with Ink tip rather than the polyacetal tip that would allow ‘invisible’ drawings on the bare surface of the tablet.

According to Plamondon and Maarse [115], “*like any highly skilled motor process, fast handwriting is considered a ballistic phenomenon, that is, a motion controlled without instantaneous position feedback*”, while “*...a slower writing process is probably a matter of position and visual feedback*”. Since, signing is generally considered a highly skilled motor process, and hence, might not necessarily require

instantaneous position feedback, the imitation task, on the other hand, being generally a slower writing process seems to highly depend on position feedback. Therefore, for experimental purposes a special layer on the tablet's screen was implemented (see Figure 3.7), fixed on the left side of the tablet having a rectangular open area (3.2cm x 8.0cm) in the middle for the insertion of signatures. This ensured that the inserted paper was properly supported, and that the subjects were provided with a user-friendly device hiding away the sophisticated details and providing a more natural signing platform. In addition, the signing area was restricted to the box provided, while closely simulating restrictions applied in the signing space available in most implementations encountered in common point-of-sales applications.

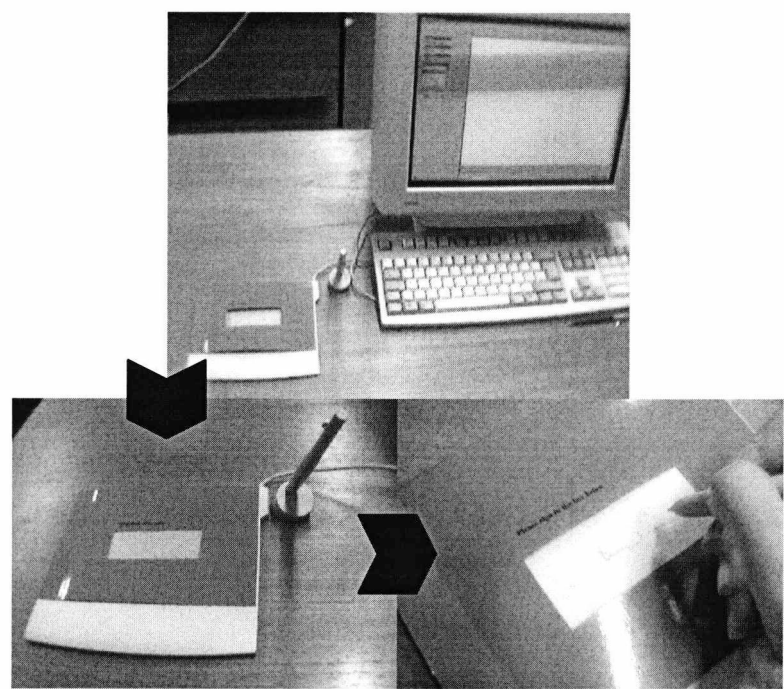


Figure 3.7: Digitiser experimental set up and signing process.

3.3.2 Experimental Procedure

Four volunteers donated a number of genuine samples of their signature for experimental purposes, while it was explained to them that their signature would be subjected to imitation. These subjects are denoted as target signers throughout this thesis, and their signatures as target signatures. One of the volunteers, besides his original signature, also donated his signed initials (Figure 3.8(e)) which he sometimes

used as a signature form, and which is a practice not uncommon among signers. The five target signatures, which were seen to represent a variety of styles, were initially collected signed on a piece of paper, which was then laminated. The scanned images of the five target signatures are displayed in Figure 3.8. The five target signatures were then presented to a group of 20 different subjects (designated “forgers”) who were given the task of imitating (“forging”) the target signatures. The “forgers”, who were inexperienced with respect to the task of imitating signatures, were presented with each of the target signatures in the following sequence Target 1, Target 2, Target 3, Target 4 and Target 5. For each target, they were required to submit five forgeries signed on the graphics tablet, after having about a minute of practice on a separate piece of paper with the laminated picture of the signature in view. Note that the forgers were allowed to keep the respective target signature in view during the imitation process. The subjects were also allowed to delete and resubmit samples in which they made a mistake. This also ensured that mistakes made due to lack of familiarisation with the digitising tablet would be narrowed down to minimum. The test was carried out in two sessions, in order to avoid the effects of tiring or lack of concentration on the part of the forgers. The twenty subjects were also requested to submit 5 samples of their own signature, but this was not obligatory. As a result, 19 subjects submitted 5 genuine samples each, while only one subject declined. Furthermore, the target signers were asked to submit 30 genuine samples of their signature, which were collected in three equal sessions. In all cases the subjects were allowed to delete and resubmit samples that were wrongly entered.

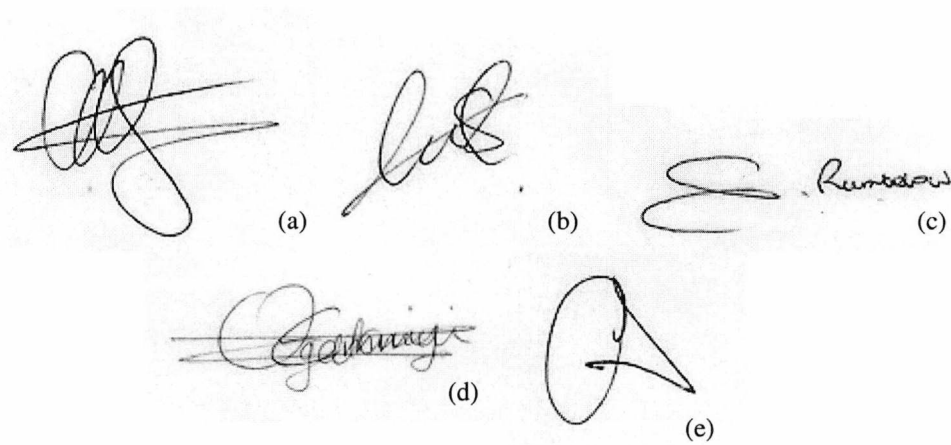


Figure 3.8: Scanned images of (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 signature.

In total, 244 genuine signatures were collected, from which 149 were genuine samples of the target signers and 95 genuine samples of the forgers' own signature. Note that one sample from Target 5 was irreversibly corrupted, due to problems encountered with the storage of the data linked to the capture program, and therefore was excluded from the database. In addition, a total of 500 forgeries were collected, which may be regarded as *skilled*¹ forgeries, as the original target signature was provided, initial practice was given (although restricted to a short period of time) and deletion of mistakes and resubmission was also allowed. Furthermore, the 95 genuine samples submitted by the forgers could later act as *random*¹ forgeries in an automatic signature verification system. Some examples of the forgeries submitted for each of the five target signatures are illustrated in Figure 3.9. Note that the images are directly extracted from the pen-down information acquired by the tablet. It is apparent that the forgeries range from very clumsy imitations to those closely approximating the original samples.

3.4 Data Preprocessing

The signature data collected have the advantage of being dynamically acquired, and thus both static and dynamic information about the signatures is available. The reconstructed digitised signature images from the position information captured by the tablet, as earlier described, will be used in the experiments reported in later chapters wherever the static signature images are required. This removes the need for long scanning sessions and extensive pre-processing techniques, such as noise reduction, thinning, etc., that the raw scanned images would require. It should be noted that the reconstruction of the signature images from the sampled coordinates does not apply any limitations to the findings reported in later chapters regarding static approaches, since this is a widely acceptable static approach simulation, whereby the digitised signature points are processed in a scanning pixel-by-pixel fashion of the image plane as would be the case with a binary scanned signature image [114].

¹ For a definition on the different types of forgeries see Chapter 2.

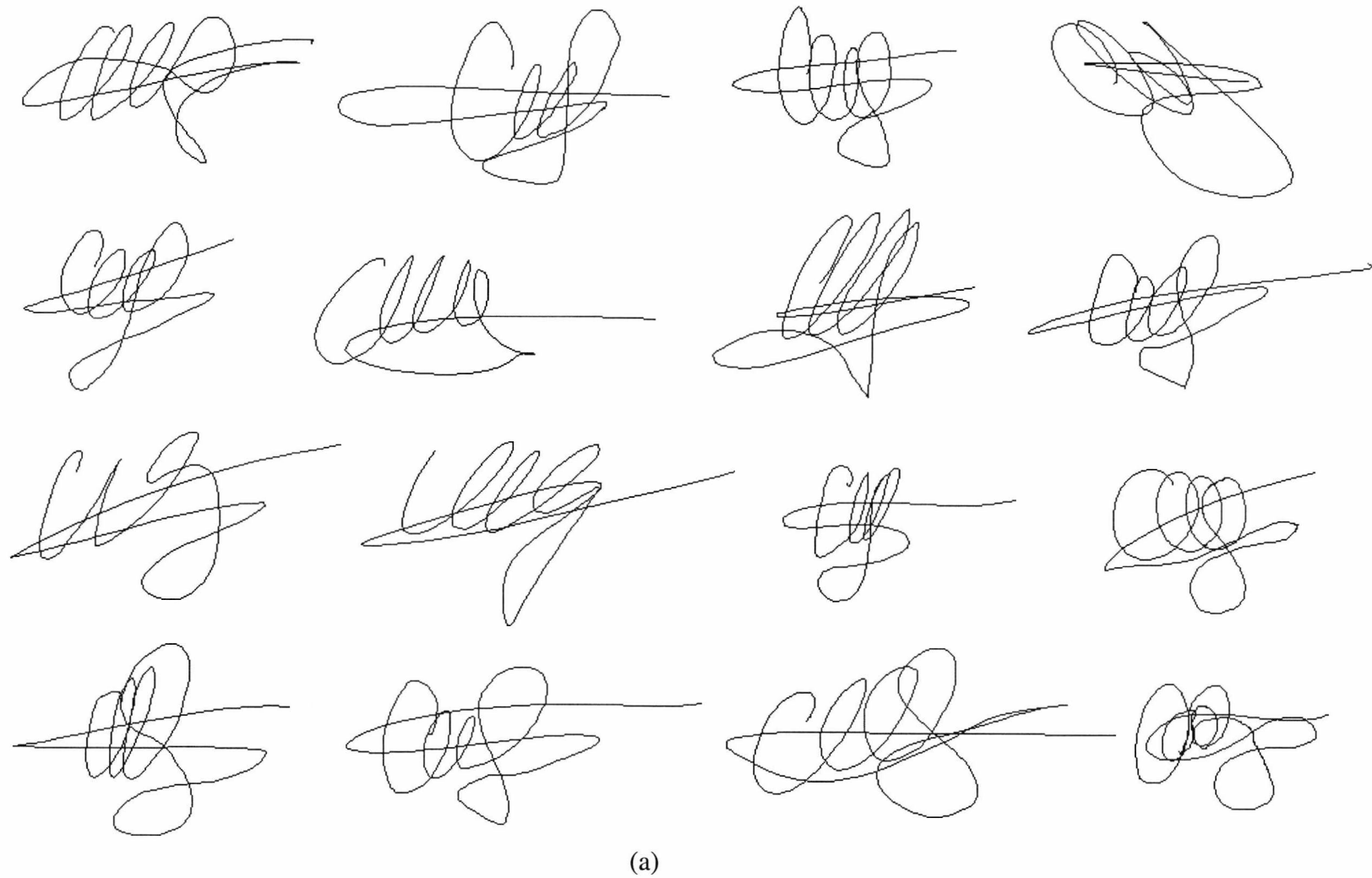


Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5.

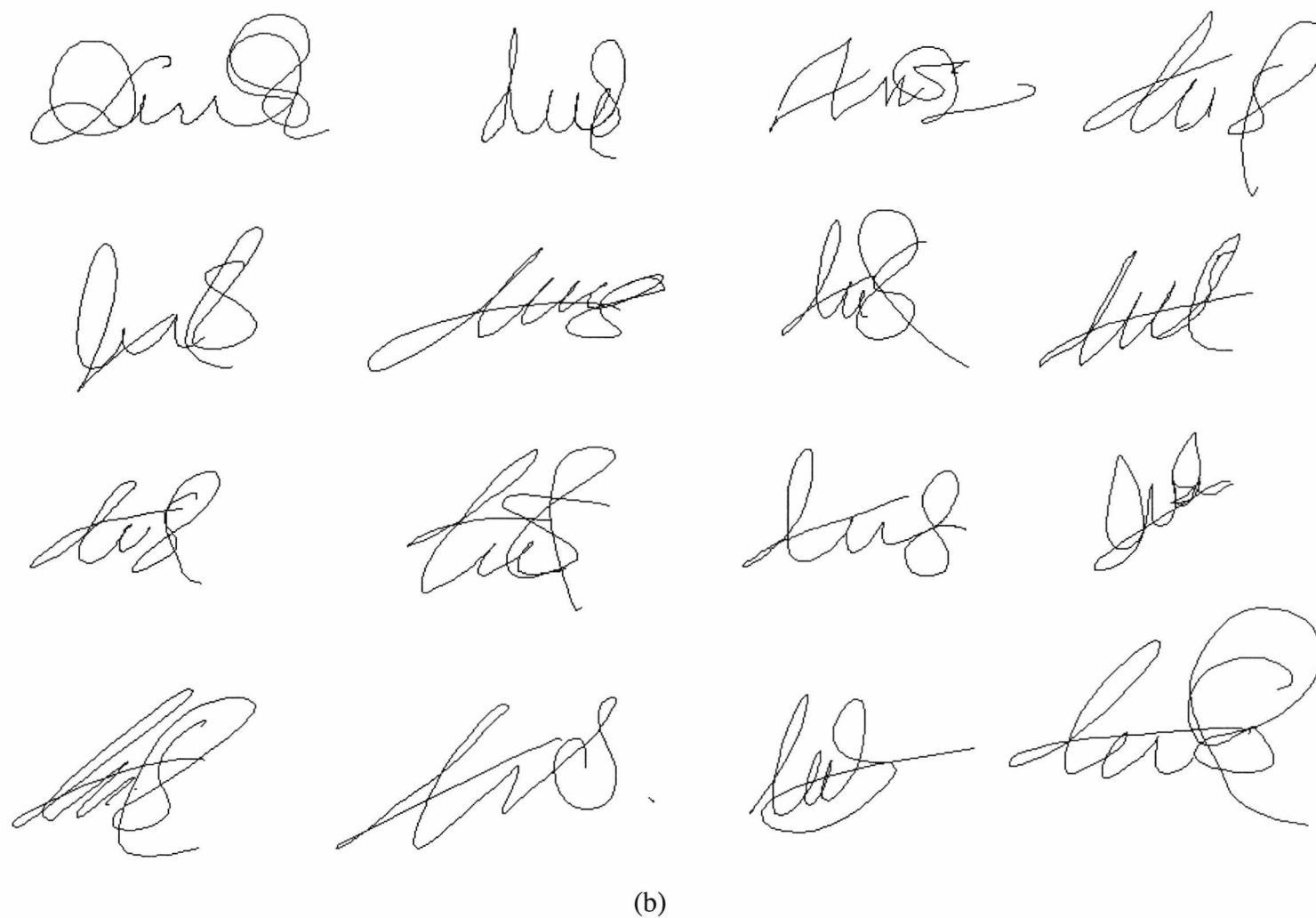


Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).

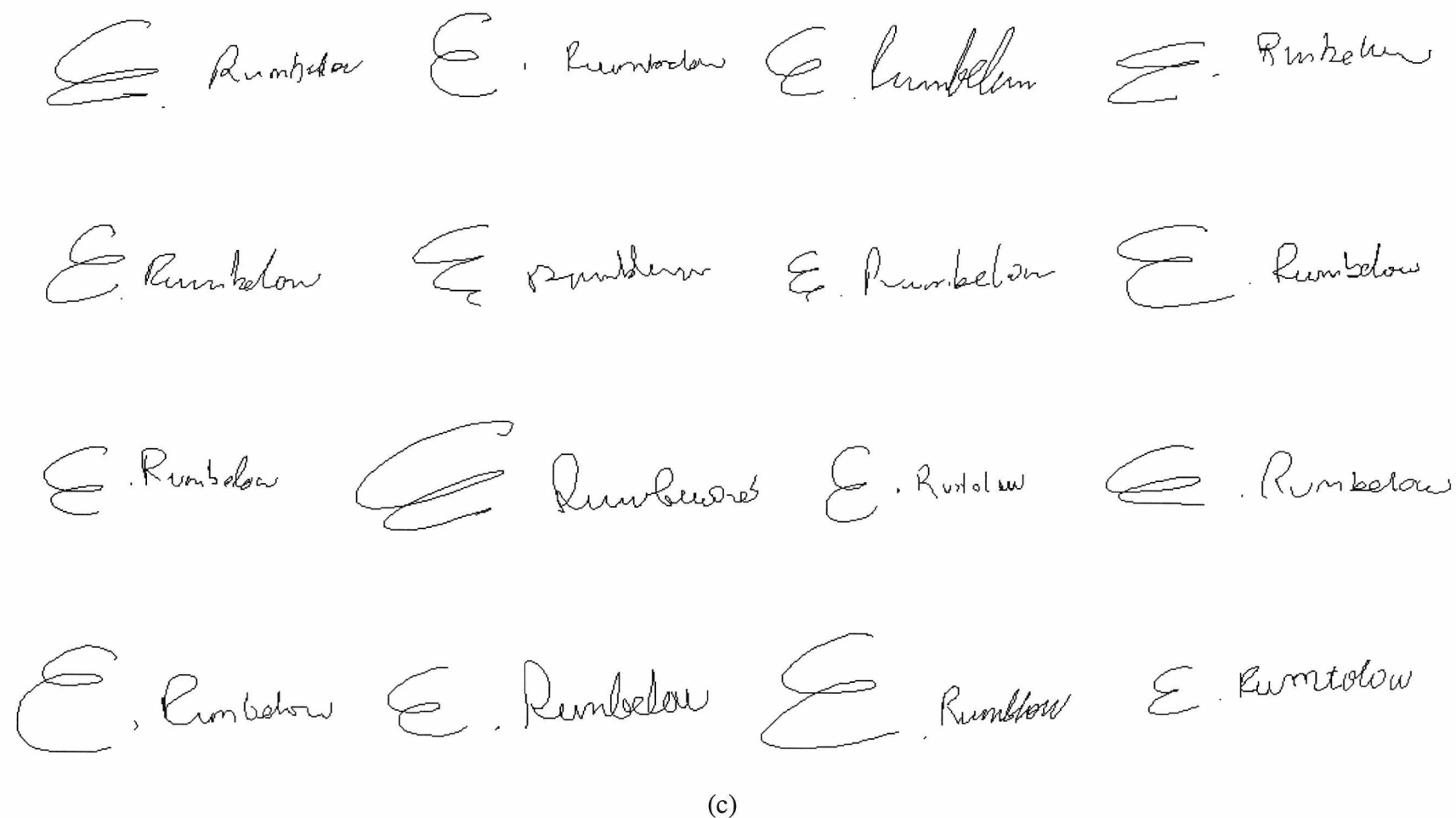


Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).

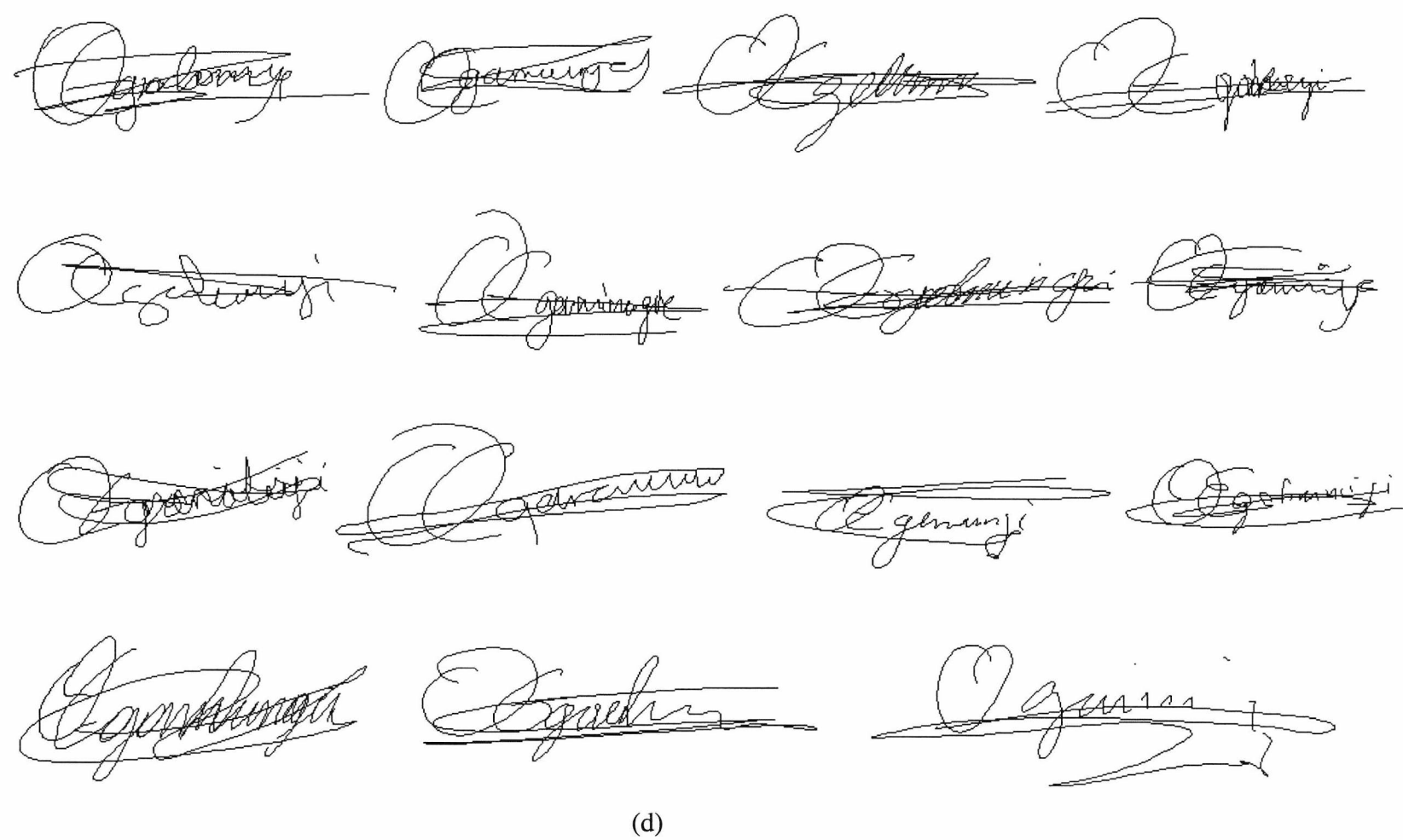


Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).

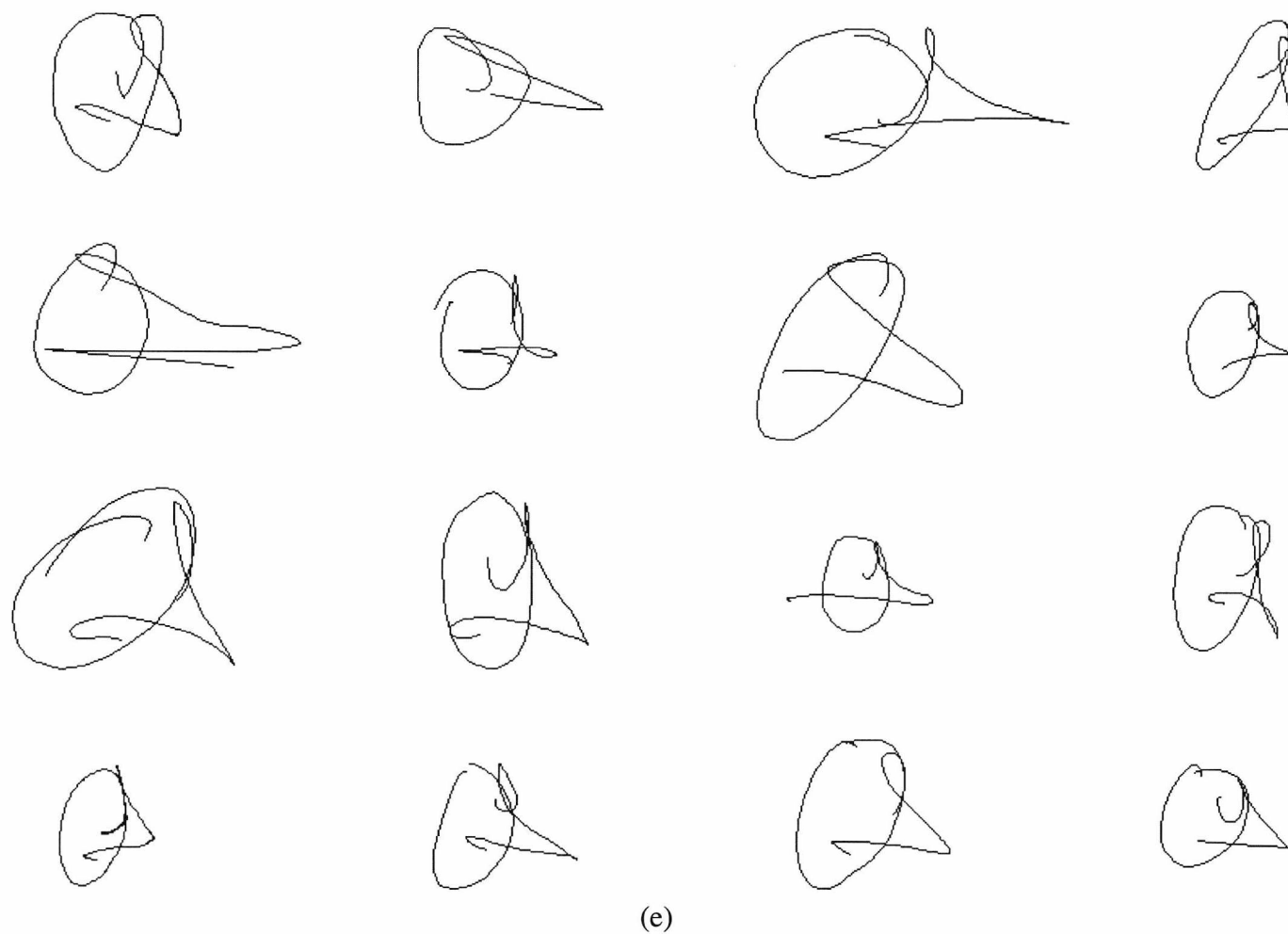


Figure 3.9: Various forgeries for (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5 (continued).

The pre-processing required here was limited to pen-up detection and extraction of the pen-down points alone, for the reconstruction of the signature images. This was based on the value of the parameter of pressure. Non-zero pressure denoted pen-down information, when the pen-tip was touching the tablet, whereas zero pressure meant that the pen was lifted from the tablet but yet close enough to be recorded. Note that the time parameter earlier described allowed the evaluation of the pen-up duration. The detection of the beginning and ending parts of the signature allowed the extraction of the pure pen-down signal. For signatures with pen-up periods in between, the detection of gaps was necessary and, hence, the segmentation of the signatures into their constituent components for further accurate exploitation. Furthermore, the implementation of the limited box-area provided for the signing process minimised the need for scale or rotation normalisation of the signatures used in later applications. Specifying an exact area for signing is expected to yield reference signature samples of similar scale, while reference to the horizontal level, as provided by the bottom side of the rectangular box, is expected to ensure that the natural baseline of the signatures is clearly visible.

3.5 Conclusions

A brief review of the available data acquisition methods for static and dynamic approaches was initially provided. The digitiser used for the dynamic acquisition of signatures was described, while the format of the generated data and the parameters captured by the tablet were also explained. An analysis of the pen-up information captured was provided, and differences observed between the pen-up patterns of genuine samples and forgeries were some of the issues further discussed. The experimental set up employed for the acquisition of signatures and the experimental procedure used were described, leading to the construction of a large signature database, containing a total of 244 genuine signatures and 500 forgeries, for its use in experiments and applications to be reported in later chapters. An illustration of some of the forgeries collected, corresponding to the five target signatures, was provided and further issues involving the preprocessing of signature data were finally discussed.

The following chapter is involved with the design and construction of an automatic signature verification system. Signature samples from the database described here will be used to train and test the verifier.

Chapter 4

Automatic Handwritten Signature Verification

In this chapter the machine verification of handwritten signatures will be assessed both through a static and a dynamic approach. For the purposes of this task a signature verifier was built based on parameter processing and the Euclidean classifier. The verification performance achieved with skilled forgeries, in terms of the equal error rate, resulted in a 21.2% EER for static and 6.6% EER for dynamic ASV, while the use of optimising techniques reduced these figures to 14.8% and a 3.3% respectively. Although it is acknowledged that better verification results have been reported in the literature (see Chapter 2), the verification rates obtained here are comparable to those reported in similar studies. Nevertheless, the various systems reported are based on different databases, different methods and different signatures – genuine and forgery samples – and, thus, are difficult to be adequately assessed and compared. The construction of the verifier will be described step by step through the system design, presenting the data sets, the enrolment stage, the feature selection, the classification and the threshold selection procedures used, while the feature extraction will be discussed with reference to the separate parts of the static and dynamic signature verification. Finally, the performance of the system will be assessed for each case and optimisation techniques will be extensively evaluated.

4.1 Introduction

The many methods reported in the literature of the automatic signature verification have been reviewed in Chapter 2. As a result, only some of the studies related to the construction of the verifier to be used in these experiments will be briefly discussed here, giving an indication of the performance that can be achieved. However, as mentioned earlier, a direct comparison of the performance of different systems reported in the literature would not be meaningful since different databases, types of forgeries, preprocessing techniques and implementation strategies, have been used in the various research studies. The performance of a verification system is usually judged in terms of the equal error rate (EER), which, as discussed in Chapter 2, is the point at which the FRR and FAR curves intersect. It is also common to report the FRR and FAR values. However, because these are threshold dependent, this is not especially helpful for the comparison of systems. On the other hand, the error tradeoff curve is a suitable way of presenting the relation between the Type II/ Type I error rates.

For the static verification of signatures, Qi and Hunt [117] studied the performance of two feature representations, global geometric features and local grid features, separately as well as combined through a multi-scale verification function, with a database of 450 signatures. The separate performance of the subsystem using 7 global geometric features and the Euclidean distance measure resulted in an FRR of 8.8% and an FAR of 15.7% when using skilled forgeries. On the other hand, the threshold was set to minimise the total error rate, which was recognised to generally lead to an underestimation of the error rates, but because of the comparative nature of the study this threshold selection approach was considered acceptable. Furthermore, as many as 15 genuine samples per singer were used for training the system, while only 5 separate samples were used for testing. In another comparative study, Ramesh and Murty [118] evaluated the performance of different classifiers and different types of features, with a database of 650 signatures. It may be observed that the performance of the subsystem, using 8 global geometric features (calculated as in the previous study [117]), with the Confidence Interval classifier and 95% confidence results in about 4% of error with respect to genuine signatures, in about 58% of error with

respect to simple forgeries and about 77% of error with respect to skilled forgeries. Also in this study, 15 reference samples from the 20 submitted per writer were used for system training, while the remaining 5 for system testing.

Moment invariants are popular 2-D shape descriptors that provide a scalar result based on the global shape of an image and can be invariant under certain transformations such as translation, scaling and rotation. They have been extensively used in the field of pattern recognition as shape measures for image retrieval and object recognition [94;104]. In the field of handwritten signature verification projection moments (one-dimensional moments of the signature projection functions) have been widely used [20;118] as well as the moments of inertia (2nd order moments) [98]. Allgrove [10] used complex moments and complex moment descriptors, which are considered more enhanced invariant shape descriptors compared to ordinary moments, with the Euclidean distance classifier for the handwritten signature verification. A minimum EER of 22.07% was achieved for a reference model size of 26 samples for the complex moments, while a minimum EER of 25.13% was attained when using the complex moment descriptors and a reference set of 14 samples. It may be observed that, in both cases, when using a small reference model size of 5 samples the EER was about 34%.

For the dynamic verification of signatures, Achemlal et al. [8] extracted a set of 40 dynamic parameters including maximum and average values of the x- and y- velocity components, duration features, etc. and selected 10 parameters using a single Jackknife method, based on a criterion which allows a large discrimination between true signatures and forgeries. The selection of the best parameters was specific for each user and was based on 10 genuine samples and 10 skilled forgeries. The FRR and FAR obtained in the verification process were 11% and 8% respectively. Lee et al. [79] proposed a set of 42 features which they advanced to a set of 49 normalised features to use with different feature selection approaches, depending on the availability of forgery data for training, while based on distance measures between samples. The performance of a Majority classifier and the best individualised 15 features selected from the 49 normalised feature set was demonstrated, resulting in an FRR less than 1% and an FAR of 20%, while a proposed modified version of the Majority classifier yielded a 2.5% EER, and an asymptotic performance of 7% FAR at

zero FRR, with a forgery database consisting of simple, statically skilled and timed forgeries.

4.2 Automatic Signature Verification

For the automatic verification of handwritten signatures, both a static and a dynamic approach were considered. The implementation of the two approaches will be analysed in detail after a first description of the general system design as developed in this chapter.

4.2.1 Signature Database

The signature database used for the purposes of this study was earlier described in Chapter 3. The separation of the signatures in a training set and a test set for their implementation in the signature verifier was as follows. The 10 first genuine samples out of the 30 samples submitted by each target signer served for the construction of the Reference Set. The remaining 20 genuine samples served for testing purposes and, together with the corresponding forgeries, formed the Test Set. Three Reference Sets were considered, a Small Reference Set containing the first three reference samples submitted by each signer, a Medium Reference Set, which was the one mainly used in this implementation, containing the first 5 reference samples, and a Large Reference Set containing all 10 genuine signatures of the target signers. In addition, a Limited² Reference Set was also formed including only one sample per writer. This is a special case that serves in a way which will be discussed later. In all cases the first 10 genuine samples were retained for training purposes, irrespective of which Reference Set was used. Moreover, as the task of forgery collection in a real world application is not feasible, for the training phase of this system only genuine signatures were used.

The Test Set contained 99 genuine signatures and 500 skilled forgeries. Note that having a greater number of forgeries than genuine signature samples suggests that the

² The Limited Reference Set contained the 4th sample in the sequence of 10 produced by the target signers for comparison reasons with experiments developed in later chapters.

total error – measured as the total number of erroneous classifications out of all the signatures tested – will be heavily weighted by the FAR. A Zero-Effort (ZE) test was also performed to test the system against random forgeries. The ZE Set contained a total of 694 signatures, including the genuine and forgery samples contained in the Test Set – where for every target signature tested only samples corresponding to the other targets were used – while a further 95 genuine signatures belonging to the forgers themselves were included.

4.2.2 System Design

As mentioned in Chapter 3, the preprocessing applied included the detection of pen-ups and the segmentation of signatures into their constituent parts. For the static representation of signatures the reconstructed digitised signature images were used also as described earlier.

The verification system developed is illustrated in the schematic of Figure 4.1. The several stages involved include an enrolment stage, where the original signer submits a specific number of reference samples for the training of the system, a feature extraction and feature selection stage, where a number of parameters are extracted to represent the signatures and an optimum set is selected to provide the feature set best representing the signature database.

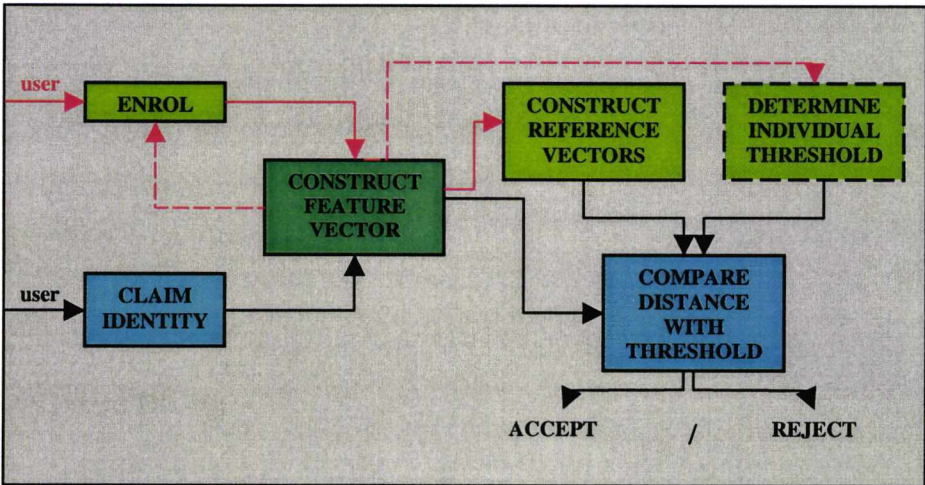


Figure 4.1: Schematic of handwritten signature verifier.

In the testing phase, the user enters a PIN code in order to get access to the system and his/her signature is evaluated in order to verify his/her claimed identity. The respective parameters are extracted and a classifier is used to obtain a measure of the distance of the user's signature from the reference set. If the resulting distance is smaller than or equal to a predefined threshold the signature is accepted as 'genuine' otherwise rejected as attempted 'forgery'. The processing stages are now further described in more detail.

Enrolment

At this stage the signer submits the requested number of reference samples, which are used to construct the reference vector, containing the mean values of the parameters measured across the constituent reference samples, and the standard deviation vector, containing the parameters' standard deviation values for the Reference Set, as illustrated below. The two vectors are stored in the system for further use in the verification stage.

$$R_w = \begin{Bmatrix} r_{w1} \\ r_{w2} \\ \vdots \\ r_{wn} \end{Bmatrix} \Rightarrow r_{wi} = \begin{Bmatrix} f_{wi1} \\ f_{wi2} \\ \vdots \\ f_{wik} \end{Bmatrix} \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} R_{Mw} = \begin{Bmatrix} m_{w1} \\ m_{w2} \\ \vdots \\ m_{wk} \end{Bmatrix} \\ R_{Sw} = \begin{Bmatrix} s_{w1} \\ s_{w2} \\ \vdots \\ s_{wk} \end{Bmatrix} \end{matrix}$$

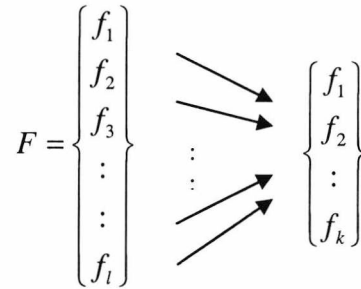
where R_w is the set of n reference signatures r_{wi} of writer w , each represented by a feature vector containing k parameters f_{wij} . R_{Mw} and R_{Sw} are respectively the w th writer's mean and standard deviation vectors containing respectively the mean value m_{wj} for each feature j calculated according to equation (4.1), and the corresponding standard deviation value s_{wj} measured according to (4.2).

$$m_{wj} = \frac{1}{n} \cdot \sum_{i=1}^n f_{wij} \quad (4.1) \quad \text{and} \quad s_{wj} = \sqrt{\frac{\sum_{i=1}^n (f_{wij} - m_{wj})^2}{n-1}} \quad (4.2)$$

For features that are described by an integer rather than a floating point, such as the number of components or the initial direction of a signature, a calculation of the mode (the most frequently occurring value) is used instead of the mean value. For features that are described by larger integers and are more likely to vary significantly, such as the number of x- or y- velocity zero-crossings, the mode value was not found very appropriate and, hence, the mean value rounded to the nearest integer was used.

Feature selection

The problem of feature selection of the k best features from an initial feature vector of l features extracted (with $l > k$) for the representation of signatures, was handled with a simple procedure described below. Although this process could be easily applied for the extraction of individual sets of best features for each target signature, it is only applied here for the extraction of a common feature set with the best overall features describing all target signatures.



The feature selection procedure involves the calculation of a *Feature Reliability Factor (FR)*, given by equation (4.3), for each feature j as described by its mean and standard deviation values stored in the writers' reference vectors. The FR factor is related to the *fractional uncertainty* or *precision* of a measurement in error analysis [139] for measuring the quality of an estimation³.

$$FR_j = \frac{sd_j}{|mean_j|} \quad (4.3)$$

³ Note that the fractional uncertainty as related to the mean and standard deviation values is best described as the ratio of the standard error of the mean over the absolute mean [139]. However, because this study involves feature extraction in different signatures the standard deviation expressing the uncertainty in any one measurement seems to be more suitable than the uncertainty in measuring the mean, as given by the standard error of the mean.

Thus, features with large FR values are eliminated leaving only features with small FR factors to best represent the reference sets. The factor FR expresses the intrinsic variability of a target signature with respect to the particular feature and, hence, the best feature set selected contains the features that are the least variable in the reference set. However, the factor FR does not guarantee the best system performance, since it provides no information about how well a particular feature separates the two classes: genuine signatures and forgeries. On the other hand, this evaluation is not possible for a system implementation where the training set does not include any representatives from the forgery class. Hence, based on the FR factor the features selected should at least perform well with respect to the attained FRR.

Another issue is the number of features selected to form the feature vector. This number usually depends on the application developer and the storage space available as described by the requirements of various applications. The optimum number of features is both user and application dependent. Usually, too few or too many features lead to poorer performance, while there is a specific number of features for which the system reaches an optimum result. Feature sets of 7-25 parameters have been commonly encountered in the literature of ASV (see Chapter 2).

Classification

The Euclidean distance classifier being the most popular method in the literature of signature verification, also known to offer good verification performance, was selected for the classification process. The Euclidean distance metric defined in equation (4.4) was used to measure the distance between the feature vector of a test signature T and the mean reference vector of the corresponding genuine signature R_{Mw} weighted by the respective standard deviation reference values. The distance measure is denoted as $D(T, R_{Mw})$, the f_{tj} is the j th feature of the test signature t and m_{wj} the reference mean of the j th feature for writer w , whose signature the test sample is claimed to be, while s_{wj} is respectively the reference standard deviation of writer's w j th feature.

$$D(T, R_{Mw}) = \sqrt{\sum_{j=1}^k \left(\frac{f_{tj} - m_{wj}}{s_{wj}} \right)^2} \quad (4.4)$$

Threshold selection

If the resulting distance measure for a test signature is smaller than or equal to a pre-defined threshold θ the test signature is accepted as “genuine”, otherwise it is rejected as a “forgery”.

- if $D(T, R_{Mw}) \leq \theta \Rightarrow$ “genuine” else
- if $D(T, R_{Mw}) > \theta \Rightarrow$ “forgery”

Considering that the training set contains only signatures that belong to the ‘genuine’ class, the selection of an adequate threshold should only depend on the reference samples. Thus, a frequently encountered threshold selection criterion in the literature of ASV, by which minimisation of the equal error rate is achieved and, hence, better performance is guaranteed, cannot be used here. Finally, both a common threshold for all the target signers and an individualised threshold specialised for each target signer were implemented and the corresponding merits will be assessed later.

Common threshold

Assuming a normal distribution for each of the features of the genuine signatures of a writer, a common threshold θ is determined according to Nagel and Rosenfeld [97]. For the random variable $z = (f_{ij} - m_{wj}) / s_{wj}$ with a standard normal distribution, a given value may be selected using the statistical tables of the normal curve areas [139], corresponding to the probability that a feature value will fall within z standard deviations on either side of the mean value. For example, a value of $z=1.96$ will yield a probability of 95% that the feature value belongs to the normal distribution with mean m_{wj} and standard deviation s_{wj} , while a $z=2.575$ to a probability of 99% respectively. Therefore, a threshold θ may be selected to correspond to the desired boundaries as defined by the selected value of z . Thus the distance must be at most:

$$\sqrt{\sum_{j=1}^k z^2} = \sqrt{kz^2} = \sqrt{k}z$$

for a signature to be accepted as “genuine”. Hence, for a selected value of z the threshold θ is given by $\sqrt{k}z$, where k is the number of features. A threshold value equal to $\theta = 2.575\sqrt{k}$ is chosen to minimize the number of false rejections.

Individual threshold

Depending on the inherent variability that the reference samples of each target signature exhibit, individual thresholds were derived. The intra-class distance of each of the reference samples from the centre of the cluster had to be measured with the same distance measure employed in the verification stage for the threshold values to have perfect correspondence in the comparison process. However, because each of the reference samples took part in the construction of the mean and standard deviation vectors, its distance from the mean would not correspond to the true value. Thus, for each reference sample r_{wi} new mean and standard deviation vectors were computed from the remaining reference samples – thus excluding the reference sample in question – and its distance from the new reference mean R'_{Mw} was computed. Hence, each reference sample for each target signature was associated with a distance value that reflected its variability from the rest of the target's reference set. Two options for the individual threshold selection occur: either the mean of these distances as related to each of the reference samples, or the maximum distance. In general the mean of the distances seems to yield better results. However, for some target signatures the maximum distance seems to work considerably better. This will be discussed in Section 4.3.3. Thus, the individual threshold ϑ_w for writer w is given by equation (4.5) where α is a constant empirically selected, and initially assigned a value of 1.5.

$$\vartheta_w = \alpha \cdot \frac{\sum_{i=1}^n D(r_{wi}, R'_{Mw})}{n} \quad (4.5)$$

4.3 Static Verification

The static verification of signatures involves the extraction of features from the digitised signature images, by examining all the pixels in the image frame and extracting the relevant information from the foreground (black pixels) image. An image array holds the bitwise information contained in the image with 1 corresponding to the foreground and 0 to the background pixels (white pixels). The calculations hereafter performed on the signature images make use of this image array.

4.3.1 Features

For the static approach involved in this study the use of global geometric feature representation of the signatures was deemed most adequate as it may be seen to better reflect human inspection protocols involved in a first encounter with the signature image. A few simple global features are first extracted, so as to later provide with a rough indication about possible performance outcome when using global features as a comparative basis with performances achieved when using more advanced shape descriptors, such as the moment invariants.

Global features

The following global features, commonly encountered in the literature of ASV, were extracted to represent the signature images:

- Total number of foreground pixels in the signature image
- Number of components
- Maximum vertical projection
- Maximum horizontal projection
- Point of maximum vertical projection – first x point
- Point of maximum horizontal projection – first y point
- Width-to-height ratio (W/H)

The vertical and horizontal projections⁴ are the projections of a signature onto the x-axis and y-axis respectively and are calculated according to Figure 4.2. Each projection i on to the x-axis contains the number of foreground pixels included in that column, and each projection j onto the y-axis contains the number of foreground pixels included in that row. The column or row which

```
// calculate vertical projection
for (i=0; i<Image_Width; i++){
    pixels_in_column=0;
    for (j = 0; j<Image_Height; j++){
        if (image_array[i][j] == 1){
            pixels_in_column++;
        }
    }
    x_projection[i]= pixels_in_column;
}

// calculate horizontal projection
for (j=0; j< Image_Height; j++){
    pixels_in_row=0;
    for (i = 0; i< Image_Width; i++){
        if (image_array[i][j] == 1){
            pixels_in_row++;
        }
    }
    y_projection[j]= pixels_in_row;
}

//calculate max x projection
max_x_projection=0;
for (i=0; i<Image_Width; i++){
    if (x_projection[i] > max_x_projection){
        max_x_projection= x_projection[i];
        point_max_x= i;}
}

//calculate max y projection
max_y_projection=0;
for (j=0; j< Image_Height; j++){
    if (y_projection[j]> max_y_projection){
        max_y_projection= y_projection[j];
        point_max_y=j;}
}
```

Figure 4.2: C++ Code for the calculation of the vertical and horizontal projections.

⁴ An extensive analysis of the vertical and horizontal projections of a signature is given in Chapter 5.

corresponds to the maximum vertical or horizontal projection respectively, is denoted the point of maximum vertical/horizontal projection. The first x or y point is the first point encountered in the respective projection for which the projection > 0. The number of components of a signature is calculated from its vertical projection based on the gaps (number of gaps + 1) detected within the first and last x point.

These global features are simple features that describe the shape of a signature and are likely to be successfully forged in a good forgery. More challenging shape descriptors are the moment invariants which describe advanced geometric properties of a shape, and therefore could provide a better shield against forgery penetration but, also, could reflect a larger intra-class variation.

Moments

Moment invariants are popular geometric shape descriptors that can be invariant to translation, scale and rotation change of a 2-D image. The moment of order (p+q) for a continuous function f(x,y) is defined as:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (4.6)$$

The central moments of order (p+q) are translation invariant as the origin (\bar{x}, \bar{y}) is set at the centre of gravity of the image, and are defined as:

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (4.7)$$

where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

For a digital image the integrals in formula (4.7) are replaced by summations giving (4.8). In a discrete binary image f(x,y)=0 for a background (white) pixel and f(x,y)=1 for a foreground (black) pixel.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (4.8)$$

The central moments of order $p + q \leq 3$ can be directly derived from the ordinary moments [56]:

$$\mu_{00} = m_{00}, \quad \mu_{10} = 0, \quad \mu_{01} = 0$$

$$\begin{aligned}
\mu_{11} &= m_{11} - \bar{y}m_{10} \\
\mu_{20} &= m_{20} - \bar{x}m_{10} \\
\mu_{02} &= m_{02} - \bar{y}m_{01} \\
\mu_{30} &= m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2 m_{10} \\
\mu_{03} &= m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2 m_{01} \\
\mu_{21} &= m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2 m_{01} \\
\mu_{12} &= m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2 m_{10}
\end{aligned}$$

The normalised central moments, which are translation and scale invariant, are given by (4.9).

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (4.9)$$

$$\text{with } \gamma = \frac{p+q}{2} + 1 \quad \text{and } p+q = 2, 3, \dots$$

The seven Hu moment invariants derived from the second and third order normalised central moments, which are invariant to translation, rotation and scale are defined as:

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
\phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
&\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
\end{aligned}$$

Moments of higher orders can be derived but, due to the higher computational complexity involved and the difficulty in their direct association with a specific shape quality, they have not been extensively used in the literature. In this study the

normalised central moments of order 4 were also used, directly derived from the central moments of order 4, which are given by:

$$\mu_{40} = m_{40} - 4\bar{x}m_{30} + 6\bar{x}^2 m_{20} - 3\bar{x}^4 m_{00}$$

$$\mu_{31} = m_{31} - 3\bar{x}m_{21} + 3\bar{x}^2 m_{11} - \bar{y}m_{30} + 3\bar{x}\bar{y}m_{20} - 3\bar{x}^2 \bar{y}m_{10}$$

$$\mu_{22} = m_{22} - 2\bar{y}m_{21} + \bar{y}^2 m_{20} - 2\bar{x}m_{12} + 4\bar{x}\bar{y}m_{11} + \bar{x}^2 m_{02} - 3\bar{x}^2 \bar{y}m_{01}$$

$$\mu_{13} = m_{13} - 3\bar{y}m_{12} + 3\bar{y}^2 m_{11} - \bar{x}m_{03} + 3\bar{x}\bar{y}m_{02} - 3\bar{y}^2 \bar{x}m_{01}$$

$$\mu_{04} = m_{04} - 4\bar{y}m_{03} + 6\bar{y}^2 m_{02} - 3\bar{y}^4 m_{00}$$

In order to select the set of moments that would be most adequate for the current verifier, the factor FR earlier defined and given by equation (4.3) was calculated for every moment in the reference set of each target signature. Since the features had to be at least translation invariant in order to sufficiently compare the signature images, the three sets of seven moment invariants of order up to 3: the central moments, the normalised central moments and the Hu moments, were evaluated. Tables 4.1, 4.2 and 4.3 show the respective FR values measured. The smaller the FR values the higher the reliability of the particular moment in the sense that it better represents the signature image, as its reference samples show small variability with respect to that moment. The average values for the columns and the rows are also displayed, providing an estimate of the average reliability of the whole set of moments for the individual target signatures, but also the average reliability of each particular moment for all target signatures. The last row rightmost cell represents the total average, giving only a rough estimation of the overall reliability of the set of moments for the whole training set.

A comparison of the FR values in Tables 4.1, 4.2 and 4.3 shows that the set of normalised central moments overall behaves better than the set of central moments or the set of Hu moments. The average reliability of the set of normalised central moments for each of the target signatures is superior to the respective reliability displayed by the other two sets. However, at the single moment level the average reliability of a few particular moments from the other two sets outperforms the respective FR values of the set of normalised central moments. Nevertheless, overall the set of normalised central moments seems to provide higher reliability than the

other two sets, with the central moments exhibiting higher stability than the Hu moments with respect to the training data set used in this study.

Table 4.1: FR factor for the seven central moments based on targets’ reference set.

Feature	FR					Average
	Target 1	Target 2	Target 3	Target 4	Target 5	
μ_{20}	0.323	0.547	0.197	0.254	0.351	0.334
μ_{02}	0.305	0.302	0.220	0.558	0.271	0.331
μ_{11}	5.552	0.412	0.604	1.218	0.615	1.680
μ_{30}	0.788	1.322	0.527	0.458	2.064	1.032
μ_{12}	0.434	0.670	0.223	0.882	1.124	0.667
μ_{21}	0.924	0.871	0.485	0.693	0.447	0.684
μ_{03}	1.372	0.616	0.414	0.760	1.526	0.938
Average	1.385	0.677	0.381	0.689	0.914	0.809

Table 4.2: FR factor for the seven normalised central moments based on targets’ reference set.

Feature	FR					Average
	Target 1	Target 2	Target 3	Target 4	Target 5	
η_{20}	0.148	0.345	0.191	0.046	0.237	0.193
η_{02}	0.135	0.123	0.197	0.338	0.151	0.189
η_{11}	3.750	0.220	0.614	1.037	0.608	1.246
η_{30}	0.845	1.335	0.542	0.570	2.191	1.097
η_{12}	0.626	0.511	0.198	0.608	1.037	0.596
η_{21}	0.830	0.710	0.491	0.562	0.276	0.574
η_{03}	1.857	0.627	0.412	0.636	1.538	1.014
Average	1.170	0.553	0.378	0.542	0.863	0.701

Table 4.3: FR factor for the seven Hu moments based on targets’ reference set.

Feature	FR					Average
	Target 1	Target 2	Target 3	Target 4	Target 5	
ϕ_1	0.133	0.261	0.187	0.058	0.143	0.156
ϕ_2	0.356	0.550	0.423	0.089	0.809	0.445
ϕ_3	0.897	1.066	0.534	0.505	0.493	0.699
ϕ_4	0.993	1.097	0.703	1.052	1.089	0.987
ϕ_5	1.361	1.753	1.004	1.889	1.385	1.478
ϕ_6	1.508	1.261	0.757	1.855	1.346	1.345
ϕ_7	9.279	1.842	0.970	1.200	1.042	2.867
Average	2.075	1.119	0.654	0.950	0.901	1.140

Since the number of features also plays an important role in the performance of a system, as earlier discussed, and since the normalised central moments of order up to 3 proved of high reliability, the normalised central moments of order 4 were also derived and their FR values measured are displayed in Table 4.4. From the average FR values of the single moments of order 4 it is obvious that the reliability of these higher order moments is also good compared to the lower order ones. In addition the average reliability shown for the set of twelve normalised central moments is better for three out of the five target signatures when compared to the smaller set of seven normalised central moments. Since the overall reliability offered by the set of 12 normalised central moments of order up to 4 is better than the smaller set of 7 normalised central moments up to the order of 3, the larger set will be considered for the construction of the feature vector for static signature image representation.

Table 4.4: FR factor for the twelve normalised central moments based on targets' reference set.

Feature	FR					Average
	Target 1	Target 2	Target 3	Target 4	Target 5	
η_{20}	0.148	0.345	0.191	0.046	0.237	0.193
η_{02}	0.135	0.123	0.197	0.338	0.151	0.189
η_{11}	3.750	0.220	0.614	1.037	0.608	1.246
η_{30}	0.845	1.335	0.542	0.570	2.191	1.097
η_{12}	0.626	0.511	0.198	0.608	1.037	0.596
η_{21}	0.830	0.710	0.491	0.562	0.276	0.574
η_{03}	1.857	0.627	0.412	0.636	1.538	1.014
η_{40}	0.259	0.607	0.347	0.127	0.468	0.362
η_{31}	0.603	0.492	0.867	0.885	1.896	0.949
η_{22}	0.342	0.398	0.347	0.480	0.337	0.381
η_{13}	1.333	0.334	0.358	3.000	0.608	1.127
η_{04}	0.222	0.297	0.324	0.674	0.241	0.352
Average	0.913	0.500	0.407	0.747	0.799	0.673

4.3.2 System Performance

Skilled forgeries

Using the set of twelve normalised central moments up to the order of 4 and the common threshold $\theta = 2.575\sqrt{k}$ with $k= 12$ (i.e. $\theta = 8.92$), yielded an overall correct

verification of signatures’ authenticity equal to 51.42% with an FRR=4.04% and an FAR=57.4%. As expected from earlier discussion the given threshold resulted in a low FRR. However, the FAR is very high contributing to a large total error of 48.58%. Since the Test Set consists of 500 forgeries and only 99 genuine samples, the FAR plays a far more important role in the overall error attained than the FRR does. Thus, a tighter threshold leading to smaller FAR can result to a significantly higher overall performance. Hence, the total error is a weak factor for the description of the performance of a system since it is highly dependent on the number of genuine and forgery samples used. Here, since minimisation of the FRR is the design consideration adopted, a large FAR is the cost with a consequently large total error. Figure 4.3 shows the FRR and FAR curves against a range of threshold values.

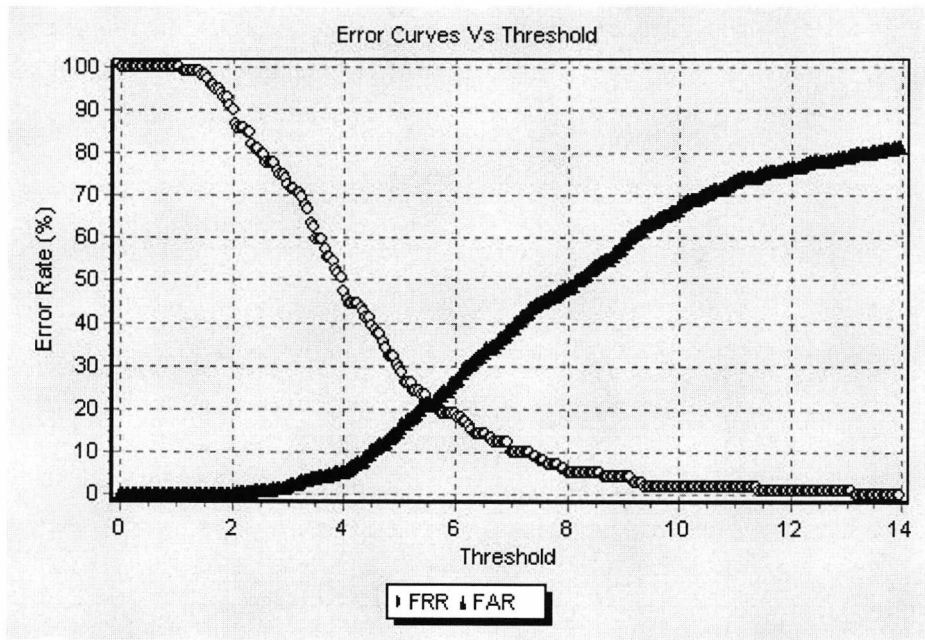


Figure 4.3: Static verification –Error curves vs threshold.

A more accurate way of describing the performance of a system is the EER, the point at which the two error curves intersect. It may be observed that an EER=21.2% occurs for a threshold of $\theta=5.5$, which is a good system performance comparable to the corresponding static verification performance using moment invariants reported in the literature [10;118]. In addition it may be observed from the curves that a slightly smaller value for the common threshold could yield a small improvement in the system performance in terms of the FAR, while sustaining the same FRR. It is also apparent that as the FRR decreases with higher threshold values the FAR increases,

and vice versa. Thus, the selection of threshold controls the attained performance in terms of the two error rates. The criterion for the determination of the common threshold in the implementation of this system, as earlier described, was the minimisation of the FRR.

Another way of analysing the performance of a verification system is through the error tradeoff curve which is the Type II versus the Type I error rate (Figure 4.4). A value for the FRR gives the respective resulting value of the FAR and vice versa. The error tradeoff curve is a suitable method of directly comparing the verification performance of various methods.

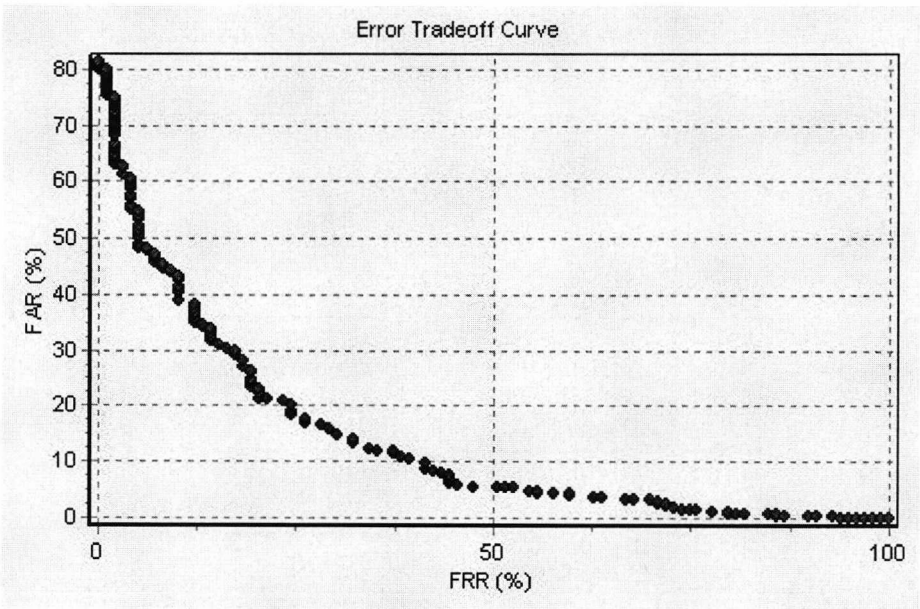


Figure 4.4: Static verification –Error tradeoff curve.

Figures B.1, B.2 and B.3 –Appendix B show the respective performance graphs for the sets of seven central moments, seven normalised central moments and seven Hu moments respectively, while Table 4.5 presents the attained EERs for these sets. It is apparent that the performance of the set of twelve normalised central moments is better than the results achieved using the other moment sets, having the lowest EER of 21.2%. A more direct comparison of the error tradeoff curves between the four different sets is provided by Figure 4.5. It is apparent that the two sets of normalised central moments have the best performance, with the larger set showing better rates in the most important regions of low FRR ($FRR \leq 7\%$) and low FAR ($FAR \leq 7\%$). Although the set of twelve normalised central moments was chosen with respect to the

attained FR values, exhibiting higher stability for the respective reference set, the system’s performance gave evidence that this set of moment invariants also worked best with respect to the set of forgeries.

Finally, the system’s performance for the set of seven global features is also provided (Figure B.4 –Appendix B) for a rough comparative assessment of the effectiveness of moment invariants as opposed to simple global features. The error tradeoff curve for the set of seven global features (Figure 4.5) is comparable to that of the seven central moments, although the latter is generally better. With a resulting EER= 30.9%, the set of seven global features performed better than the set of seven Hu moments but worse than the rest of the sets.

Table 4.5: Static verification –Equal Error Rates for the different feature sets.

Feature Set	EER (%)
12 Norm. Centr. Moments	21.2
7 Norm. Centr. Moments	22.0
7 Central Moments	28.9
7 Hu Moments	37.0
7 Global Features	30.9

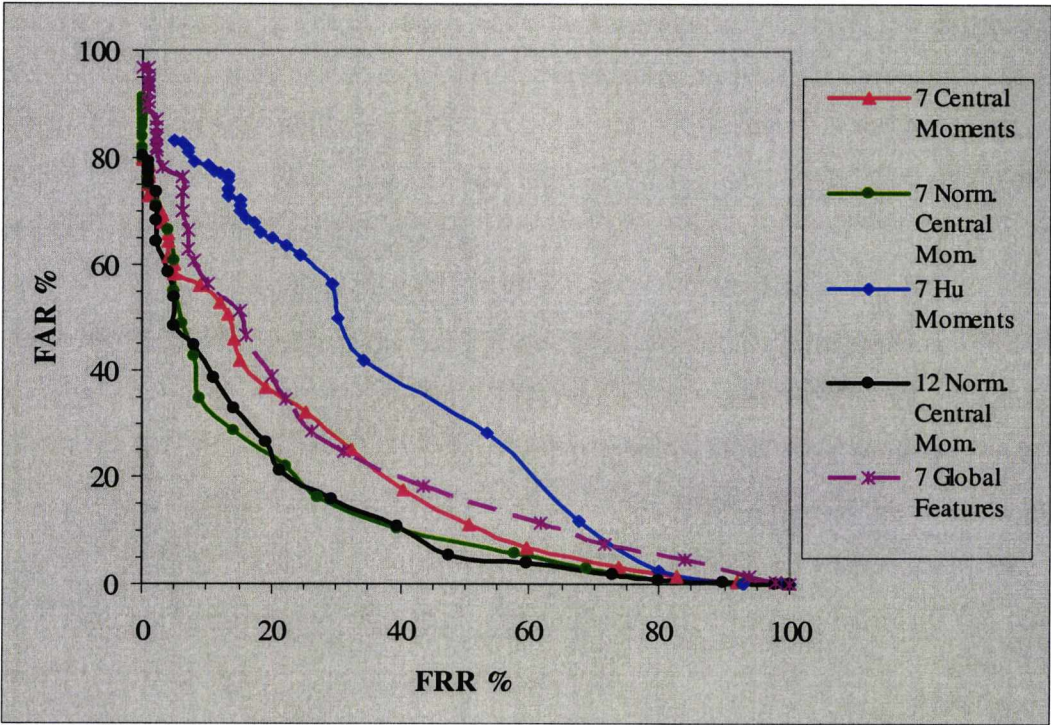


Figure 4.5: Static verification –Error tradeoff curves for the different feature sets.

Random forgeries

The performance of the system was also tested against random forgeries using the ZE Set described earlier. The resulting FAR for random forgeries, for the set of twelve normalised central moments and the common threshold, was 17.6%. Thus, significantly lower than the FAR obtained with skilled forgeries. The resulting FAR for a range of threshold values is displayed in Figure 4.6.

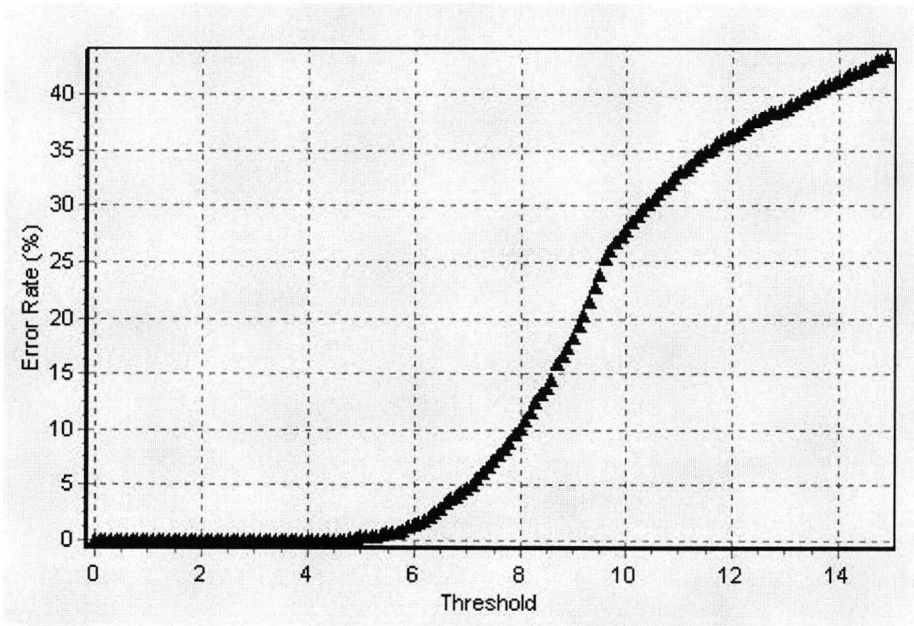


Figure 4.6: Static verification –FAR vs threshold for random forgeries.

4.3.3 Optimisation

Individual threshold

The performance of the system for skilled forgeries, examined separately for the target signatures, in terms of the attained error rates for the common threshold, are displayed in Table 4.6 together with the corresponding equal error rates. In terms of the resulting equal error rates Target 2 exhibits the best performance, with Targets 3 and 5 following with relatively low EERs. Although Targets 2 and 5 result in a zero FRR for the common threshold their corresponding FAR is very high.

Table 4.6: Static Verification –Error rates for the target signatures.

	Target 1	Target 2	Target 3	Target 4	Target 5
FRR %	5.00	0.00	0.00	15.00	0.00
FAR %	53.00	73.00	48.00	25.00	88.00
EER %	20.0	15.0	18.6	20.5	18.9

A better analysis of the individual targets’ performance is provided by Figures 4.7-4.11 corresponding to Targets 1-5 respectively. The common threshold is marked with a red dashed line. It is apparent that for all the targets the same FRR could be achieved with a smaller threshold resulting in a smaller FAR, thus improving the system’s performance. Since each target signature performs differently in terms of the FRR and FAR for different threshold values, an optimisation of the system’s performance could be achieved if individual thresholds were assigned to each target signature. The following performance graphs are displayed as an indication of the need for individual threshold selection and a demonstration of the system’s performance with respect to different signatures. Note that the individual threshold selection cannot be based on these graphs, but only on the training set as earlier described.

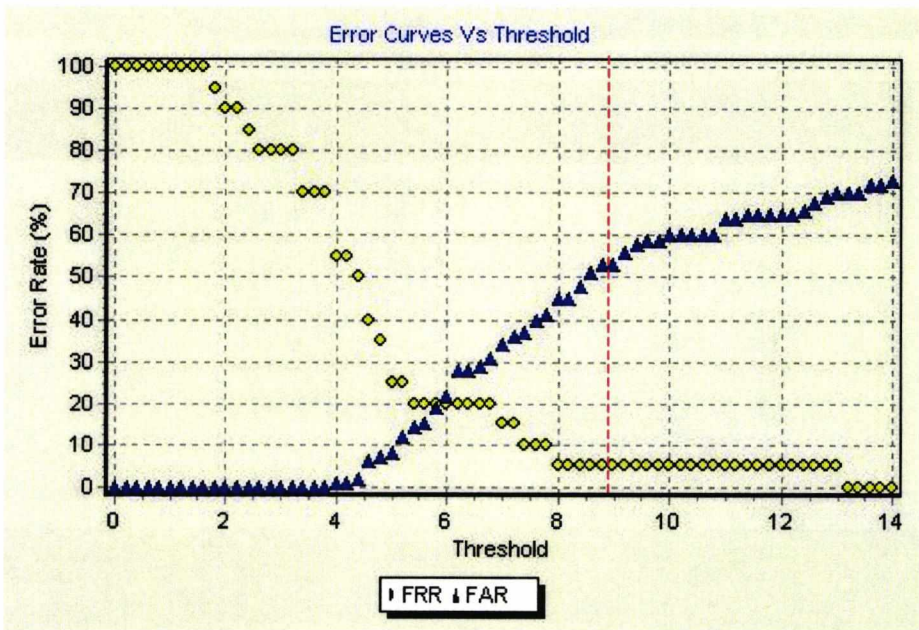


Figure 4.7: Static verification –Error curves vs threshold for Target 1.

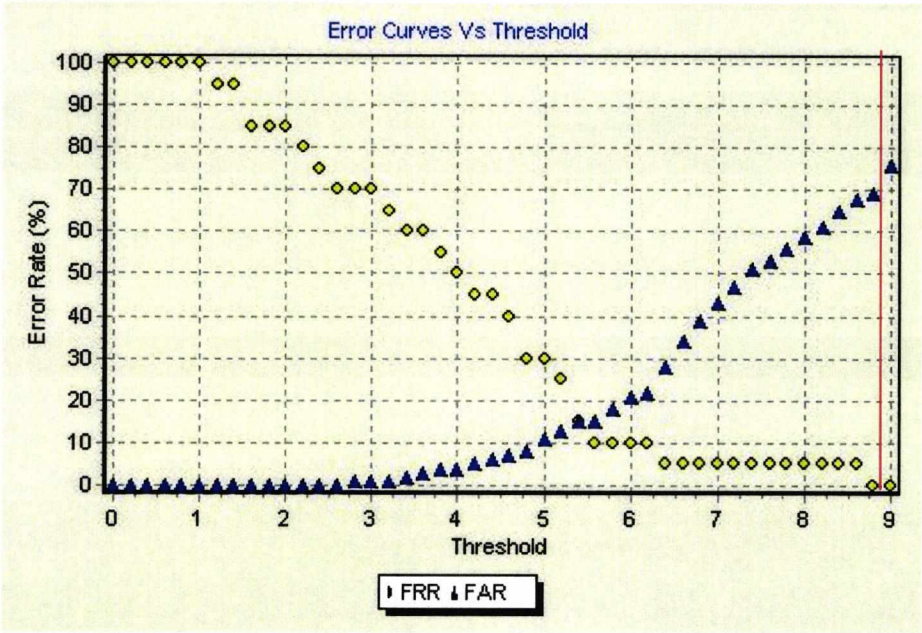


Figure 4.8: Static verification –Error curves vs threshold for Target 2.

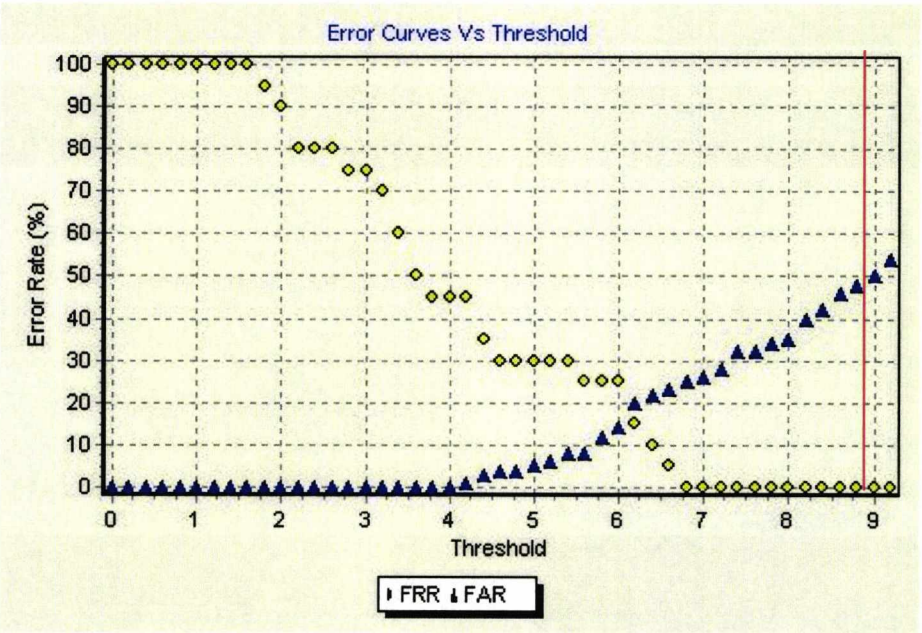


Figure 4.9: Static verification –Error curves vs threshold for Target 3.

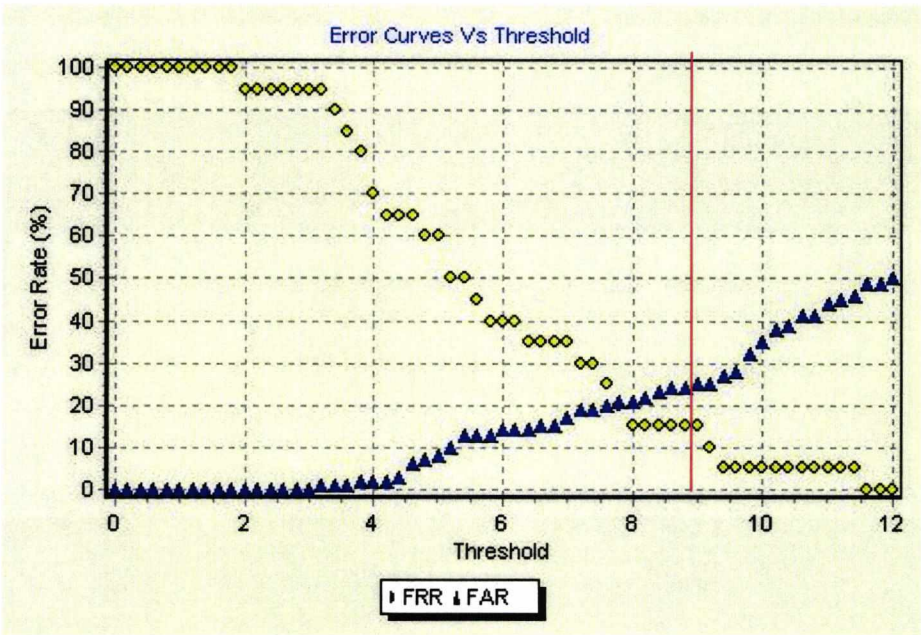


Figure 4.10: Static verification –Error curves vs threshold for Target 4.

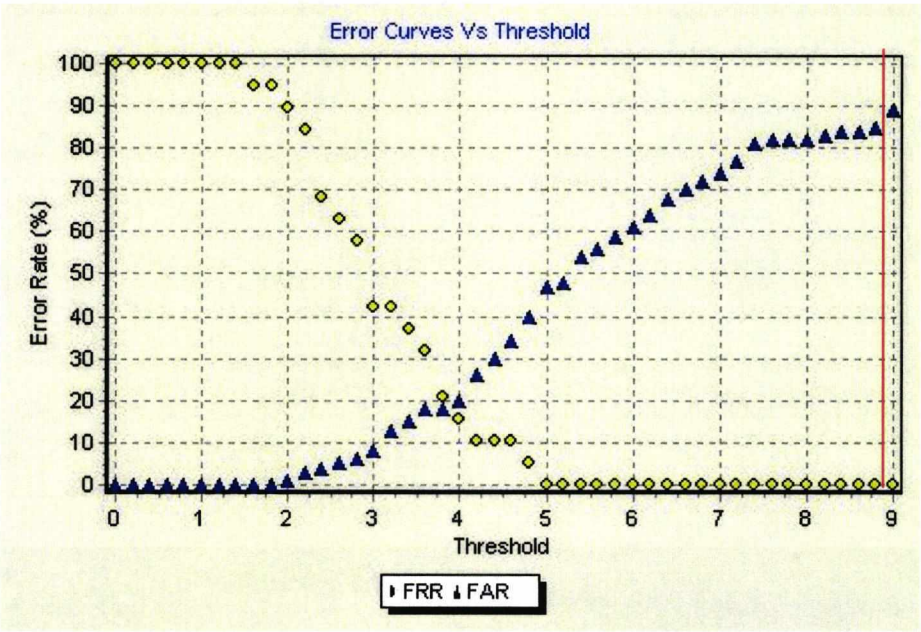


Figure 4.11: Static verification –Error curves vs threshold for Target 5.

The individual threshold values for the target signatures are determined according to equation (4.5) as earlier described. The distance measures of each of the reference samples of a target signature r_{wi} from the mean value, determined by the remaining reference samples in the set, are displayed in Table 4.7. The mean value of the

distances is also displayed, as well as the maximum distance in each reference set, which appears in bold. According to (4.5) the individual thresholds are given by: $\vartheta_w = \alpha \cdot \text{mean}(\text{dist})$, with α empirically set to 1.5, in order to reduce the FAR, while permitting enough space for the intrinsic variability of the targets to sustain a low FRR. It is evident that for Targets 1, 2 and 5 the maximum distance seems to give a better boundary of the intrinsic variability of the respective targets rather than the mean distance measure. However, the maximum distance seems unsuitable for Target 3 and even more so for Target 4, whose 2nd reference sample displays an absurdly large distance from the rest of the reference samples, and hence, would result in a very high FAR if the maximum distance was to be implemented for the determination of its individual threshold. Although a different option would be to exclude the obvious outliers from the mean or maximum calculation, with the limited number of samples available this could lead to an underestimation of the intrinsic variability of the respective targets and consequently lead to a higher FRR. Finally, the mean of the distances was preferred, as a suitable measure of the intra-class variability of the targets, for its use in the individual threshold determination procedure.

The resulting performance sustained the FRR at the same level (4.04%) and reduced the FAR to 55.4%, leading to an improvement of 3.5%. Smaller values of a would result to smaller FAR values, but at the same time compromising with larger FRR. The error rates corresponding to each of the targets are displayed in Table 4.8. Although an improvement was observed for the system's performance in terms of the FAR, this does not qualify for the improvement caused by the use of individualised thresholds. This may be examined from the error rates of Figure 4.3 for different threshold values. Indeed, a slightly lower common threshold for the same FRR returns a smaller FAR. By adjusting the z value to 2.5, corresponding to a probability of 98.76%, the performance of the system results in an FRR=4.04 and an FAR=55%. Thus, almost the same result is achieved with the implementation of individual thresholds.

On the other hand, an observation of the individual error curves for the five target signatures and a comparison with the dashed line of the common threshold, makes it clear that individual thresholds, if properly selected, can result in a significant

improvement – the best example is given by the graph of Target 5. This leads to the conclusion that the mean distance measure used in the individual threshold selection does not reflect the actual intrinsic variability of the targets, or equally that the five first samples are not good representatives of the respective genuine class. This may be supported by the distances generated for Targets 3 and 4, which are those which show the poorest performance results when using individual thresholds (compare Tables 4.6 and 4.8). It is proposed that an improvement in the threshold selection scheme could be made by assigning individual α values. This will be examined in Chapter 8, giving enhanced solutions for the improvement of the individual threshold selection scheme.

Table 4.7: Static verification –Intra-class distance measures for the targets’ reference set.

Target	rwi	$D(r_{wi}, R'_{Mw})$	mean
1	1	5.882	4.713
	2	5.541	
	3	4.801	
	4	2.902	
	5	4.438	
2	1	8.135	4.653
	2	5.602	
	3	1.669	
	4	5.823	
	5	2.036	
3	1	18.654	6.927
	2	2.457	
	3	1.911	
	4	2.279	
	5	9.332	
4	1	3.670	9.523
	2	31.321	
	3	3.675	
	4	3.064	
	5	5.901	
5	1	4.297	4.860
	2	6.196	
	3	4.193	
	4	5.846	
	5	3.767	

Table 4.8: Static verification –Error rates for the target signatures using individual thresholds.

	Target 1	Target 2	Target 3	Target 4	Target 5	Total
FRR %	15.00	5.00	0.00	0.00	0.00	4.04
FAR %	34.00	43.00	66.00	55.00	79.00	55.40

Reference Set

The size of the Reference Set plays an important role in the system's performance. A Reference Set consisting of a large number of reference samples per signer generally leads to better results than a smaller set, since the intrinsic variability of the reference signatures is better described with more samples. On the other hand, for many applications the contribution of a large number of samples for the enrolment process would mean nuisance to potential customers, thus making the use of smaller number of reference samples more practical. Indeed, the practical difficulty of ensuring large reference sets is a practical problem encountered in many biometric systems. Besides the Reference Set of 5 samples per signer – denoted as Medium Reference Set – used up until now, a larger set of 10 samples – Large Reference Set – and a smaller set of 3 samples – Small Reference Set – were also tested and the resulting error rates for the common threshold are displayed in Table 4.9. In addition, the extreme case where only one genuine sample is available for reference – Limited Reference Set – is also displayed. In order to provide with a rough estimation on the performance of the system under the extreme conditions a standard deviation value of 1 was assumed for all the features, so as to satisfy the requirements, for a standard deviation value, of the common threshold selection procedure. The distance measure in the extreme case takes the form of (4.10), where m_{wj} is here the j th feature value of the only genuine sample available for the particular writer w .

$$D(T, R_{Mw}) = \sqrt{\sum_{j=1}^k (f_{tj} - m_{wj})^2} \quad (4.10)$$

It is obvious that the Limited Set performed poorly in terms of the resulting FRR for the common threshold (Table 4.9). This is to be expected, since the threshold was measured based on an assumed standard deviation of 1 – which does not correspond to the actual variance that the features of the reference samples exhibit for larger sets. Although this assumption is made in order to give a rough indication of the performance of the system in the extreme case, a direct comparison of its performance with the other reference sets should be avoided. Further care should be taken for a different threshold selection procedure in this extreme case.

A comparison of the error rates attained for the common threshold shows that the larger the set the smaller the FRR and the larger the resulting FAR. A reduction in the

FRR was anticipated, as with a larger set the mean reference vector is better described, thus leading to a reduction in the FRR. In addition, it is apparent that an increase in the number of reference samples leads to a decrease in the EER. Figure 4.12 shows that the decrease in the EER is greater from 3 to 5 reference samples, while a large reduction is also observed when using up to 10 reference samples for training. Thus, according to this data, the use of 5 reference samples for training is a fair choice especially for applications where the acquisition of more samples is not very practical.

Table 4.9: Static verification –Error rates for different reference sets.

	Reference Set			
	Limited	Small	Medium	Large
FRR %	20.20	15.15	4.04	2.02
FAR %	47.20	49.40	57.40	67.80
EER %	37.4	31.2	21.2	14.8

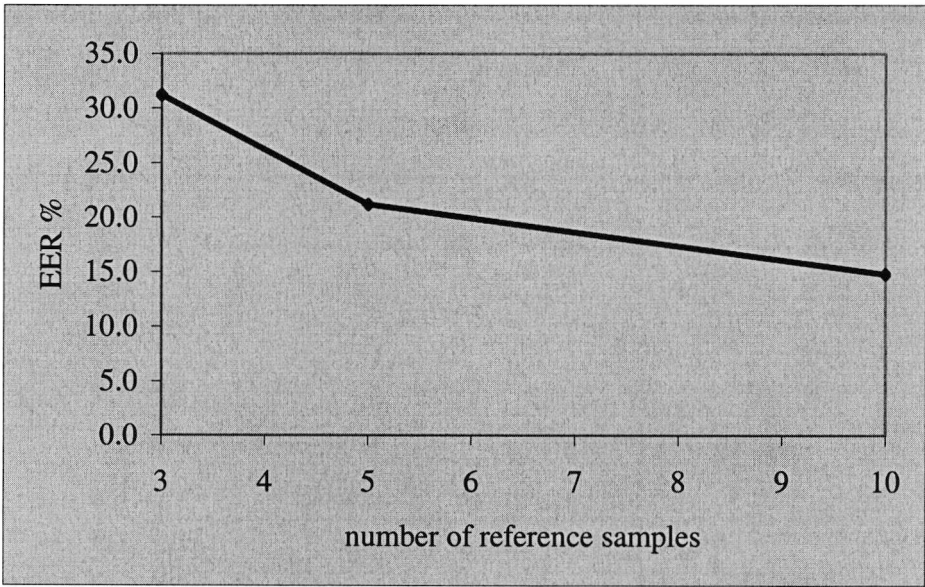


Figure 4.12: Static verification –EER vs number of reference samples.

A more analytical description of the system’s performance for the alternative reference sets is provided by Figures B.5-B.7 –Appendix B. A comparison of the system’s performance for the different reference sets is given by the error tradeoff curves of Figure 4.13. It is obvious that the Large Set performs better than the Medium Set, and that considerably better than the Small Reference Set. The relatively

poor performance of the Limited Set is shown by its error tradeoff curve, which is also provided as an indication.

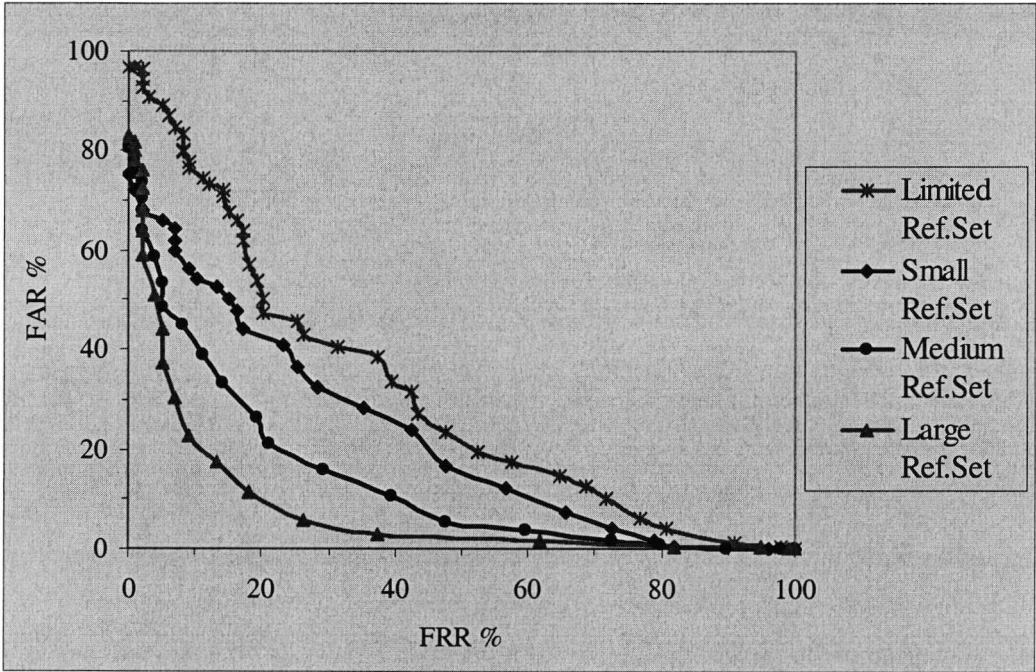


Figure 4.13: Static verification –Error tradeoff curves for different reference sets.

Furthermore, a comparative performance of the system for random forgeries and different reference sets is given by Figure 4.14 (individual graphs are provided in Figures B.9, B.10 –Appendix B). Note that, since the distance measure is different for the Limited Set, a direct comparison of the error rates with the other sets based on the same threshold values must be avoided. However, its individual performance graph is included (Figure B.8 –Appendix B). The resulting FAR for random forgeries and the common threshold was 17.6% for the Medium Set, 20.87% for the Large Set and 18.96% for the Small Set. As an indication only, with respect to the earlier assumption made for the extreme case and for the common threshold, the resulting error rate for the Limited Set was 33.8%. It is observed (Figure 4.14) that for threshold values greater than 9.5 the larger the reference set the higher the FAR, which is in agreement with the change in the corresponding error for skilled forgeries earlier analysed. However, for smaller threshold values the curves corresponding to the Small Set and the Large Set are almost superimposed, while the curve of the Medium Set shows smaller error rates than the other two.

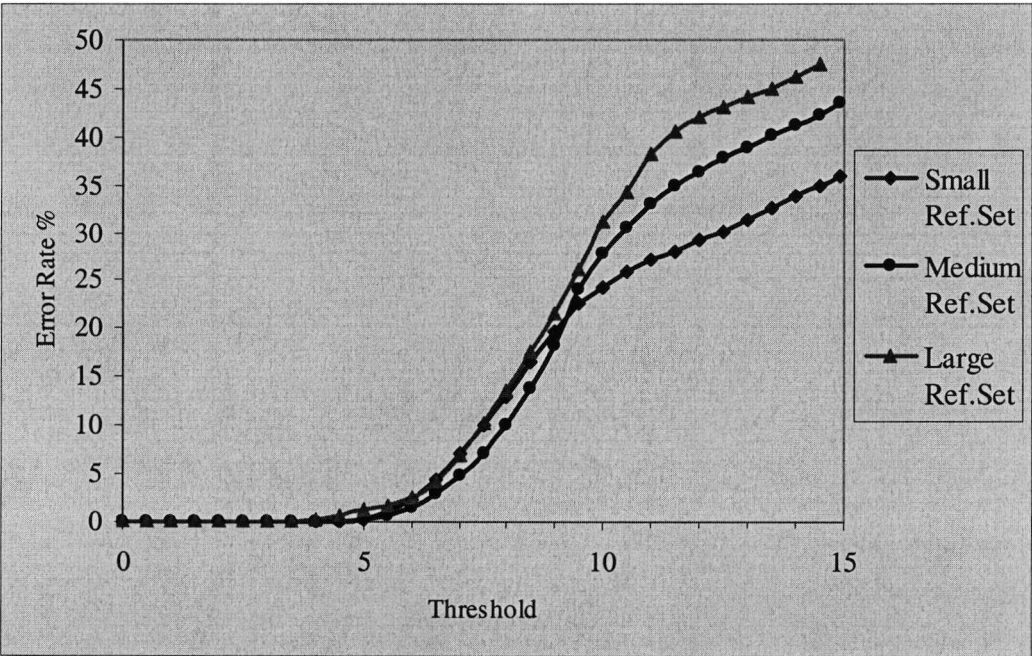


Figure 4.14: Static verification –FAR curves for random forgeries and different reference sets.

Thus, it was shown that an increase in the number of reference samples requested during the enrolment process can bring improvements in terms of the attained FRR, although this may result in a higher FAR. On the other hand, in situations where the acquisition of a large number of reference samples is not very feasible, a medium set of 5 reference samples is a more practical solution which proved to give a comparatively good performance also with respect to both skilled and random forgeries.

4.4. Dynamic Verification

For the dynamic verification of the handwritten signatures the on-line data captured by the tablet, which include the position coordinates, the timestamp and the pressure levels, were used. As extensively analysed in Chapter 3, the preprocessing applied involved the elimination of the pen-up signal before the start and after the end of the signature trace, and the detection of gaps – long pen-up specimens – between the beginning and end of the signature for its segmentation into its constituent components.

4.4.1 Features

The features that can be extracted from the on-line data can be both static and dynamic. As it is generally considered very difficult for a forger to imitate both the shape of a signature and its dynamics, and since it is more common to encounter an approximation of one of the two, the dynamic verifier design was based on both static and dynamic features. The static features calculated here were based solely on the recorded coordinates and not on the reconstructed digital signature images as implemented for the static verification of signatures in the previous section.

The static features measured include the number of components that constitute the signature, which is measured based on the zero-pressure values; the width/height ratio calculated from the maximum and minimum coordinate values; the total shift ratio calculated from the last and first coordinate values; the pure-width/width and pure-height/height ratios, where the pure-width and pure-height correspond to the total width and height of the signature excluding gaps; the total length of the signature represented by the sum of the distances between points excluding gaps; the average path-tangent angle, which is formed by the line connecting two consecutive points and the x-axis; and, finally, the initial direction of the signature, which is the direction of the line connecting the first two points of the signature and measured by coding the directions {E, NE, N, NW, W, SW, S, SE} from 0 to 7 moving counter-clockwise.

The dynamic features extracted include the temporal features: total signing time measured from the moment the pen first touched the tablet to the last pen-down point; and the total pen-down time, which is the total time that the pen was in contact with the tablet, thus excluding the duration of gaps. The initial time t_0 represents the moment of the first pen-down point of the signature. Time dependent features are features based on the time parameter such as the velocity and acceleration signals.

The x and y velocity components, u_x and u_y , can be generated from the time-derivation of the x and y position coordinates, while the α_x and α_y acceleration components can be extracted from the derivatives of the respective velocity components over time. Alternatively, using polynomial modelling on the digital data,

the first and second time derivatives of the position signals $x(t)$ and $y(t)$, will generate the x and y velocity and acceleration components respectively. Fitting a first-order polynomial through the data, the derivative may be computed according to the following equations [146] for forward (4.11), backward (4.12) and central differentiation (4.13). Note, that there is no second derivative for the first-order polynomial.

$$x'_K \approx x_{K+1} - x_K \quad (4.11)$$

$$x'_K \approx x_K - x_{K-1} \quad (4.12)$$

$$x'_K \approx \frac{x_{K+1} - x_{K-1}}{2} \quad (4.13)$$

Using polynomials to model the digital data has the advantage of representing the discrete digital data with a continuous signal, while making it easy to extract the derivatives. Higher order polynomials make use of more neighbours at the sides of the points for the fit, and thus generally offer a better description for the data. According to [146], the two-point differentiators produce non-recursive digital filters and so do higher-order polynomial differentiators. It should be mentioned that in the literature it is usually encountered that the data are either initially smoothed or processed by a combination of a differentiator with a smoothing filter as in [62;116], using a form of cubic polynomials. The first and second derivatives of a cubic (third order) polynomial are given by the following equations [146].

$$x'_K \approx \frac{-x_{K+2} + 8x_{K+1} - 8x_{K-1} + x_{K-2}}{12} \quad (4.14)$$

$$x''_K \approx \frac{x_{K+2} - x_{K+1} - x_{K-1} + x_{K-2}}{3} \quad (4.15)$$

The first derivative of the third order polynomial (4.14) was used on the position data for the extraction of the x and y velocity components, while the second derivative (4.15) was used for the generation of the acceleration components. Forward and backward equations for the cubic polynomial derivatives were used at the beginning and end of the signature respectively, as well as when gaps occurred for the starting points of the new component and the last points of the previous component. It should be noted that sometimes undesirable high peaks occur around the gap areas in the velocity and especially the acceleration signal for signatures with many components. An alternative approach would be to disregard the two points before and after the gaps, but that would also lead to a significant waste of data especially in the case were

signatures consist of many components. The writing speed is given by equation (4.16) and, similarly, the magnitude of the total acceleration by (4.17).

$$u = \sqrt{u_x^2 + u_y^2} \tag{4.16}$$

$$a = \sqrt{a_x^2 + a_y^2} \tag{4.17}$$

An example of the x and y velocity components and the resulting total velocity signal is provided by Figure 4.15, and the respective acceleration signals by Figure 4.16.

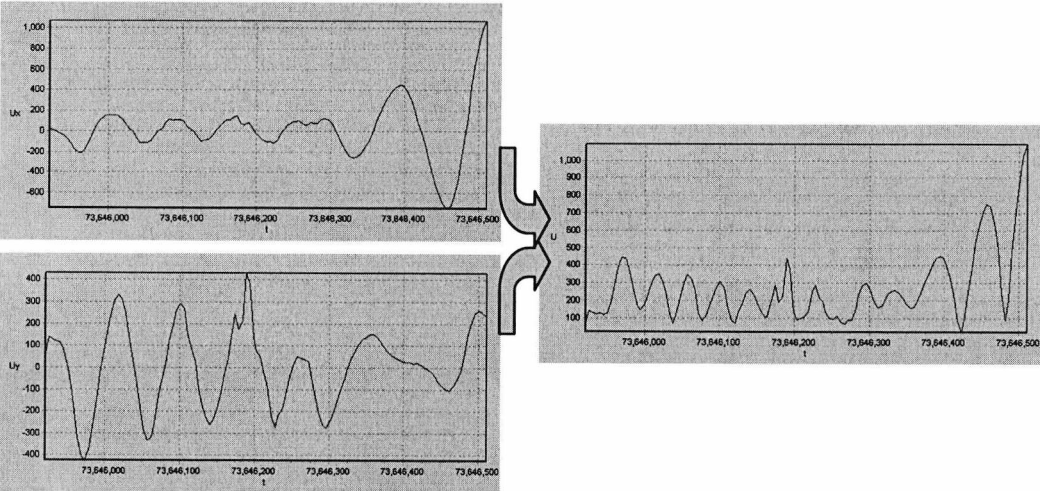


Figure 4.15: Velocity components as a function of time $u_x(t)$, $u_y(t)$, and total velocity signal $u(t)$.

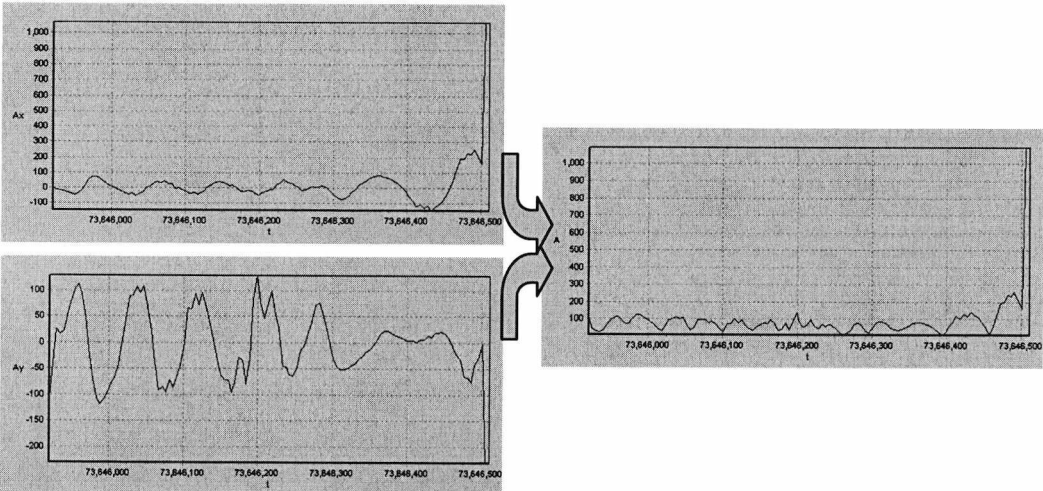


Figure 4.16: Acceleration components as a function of time $\alpha_x(t)$ and $\alpha_y(t)$, and the total acceleration $\alpha(t)$.

Note that the u_y was inverted for a Cartesian-axis representation, as the origin of the tablet used for data collection is positioned upper left. The time parameter is

represented by the timestamp and together with the velocity and the acceleration signals, needs to be normalised for a translation into standard units.

A number of measurements are made based on the velocity and acceleration signals, such as the maximum, minimum and average values, as well as the time instants when certain values occurred. It should be mentioned that the velocity signal is regarded as particularly important in the literature of signature verification [78]. Finally, the parameter of pressure was also measured and represented by its average and maximum value. As a result, a total of 40 features were extracted, most of which were encountered in the literature as being particularly effective in characterising signatures [79;8;99]. The complete feature list is summarised in Table 4.10.

The quality of these features was evaluated with respect to the value of factor FR, as with the static features in the previous section. The FR values for the 40 features with respect to the reference set for each of the target signatures are displayed in Table 4.11. The average FR value for all the targets represents the average reliability of the feature, where the smaller the resulting value the smaller the intrinsic variability of the targets with respect to that feature. Thus, the best 20 features, having the smallest average FR values, were selected from the set of 40 features. An alternative selection method would be to sort the features for each of the targets according to the FR values generated and then select the best common set. The best features selected are marked with a tick (✓) and their average FR value is underlined (Table 4.11). Therefore, the constructed feature vector includes the best 20 features selected for the five targets altogether, as summarised in Table 4.12.

It is interesting to note here, that almost all of the static features initially extracted were selected in the best 20 feature set, as their low average FR values demonstrated their effectiveness over a number of dynamic features. This suggests that some properties of the static signature image are more consistent than some kinematic attributes of the signing process. It is also noteworthy, that some of the lowest FR values are dominated by static features, highlighting their importance in the on-line verification of signatures. The feature vector constructed consists of 7 static and 13 dynamic features.

Table 4.10: Dynamic verification –Extracted features.

Feature	Description
0	Total signing time
1	Total pen-down time
2	Number of components
3	Width/height ratio
4	Total shift ratio
5	Signature path length
6	Initial direction of the signature
7	Number of u_x zero-crossings
8	Number of u_y zero-crossings
9	Number of a_x zero-crossings
10	Number of a_y zero-crossings
11	Average speed
12	Maximum speed
13	Moment of maximum speed - initial time
14	Average acceleration
15	Maximum acceleration
16	Moment of maximum acceleration - initial time
17	Average pressure
18	Maximum pressure
19	Pure-width/width ratio
20	Pure-height/height ratio
21	Average path-tangent angle
22	Minimum u_x
23	Maximum u_x - average u_x
24	Maximum u_y - average u_y
25	Maximum u_x - minimum u_x
26	Maximum u_y - minimum u_y
27	Maximum u_x - maximum u_y
28	Moment of maximum u_x - initial time
29	Moment of minimum u_x - initial time
30	Moment of maximum u_y - initial time
31	Moment of minimum u_y - initial time
32	Duration of $u_x > 0$
33	Duration of $u_x < 0$
34	Duration of $u_y > 0$
35	Duration of $u_y < 0$
36	Average $u_x > 0$
37	Average $u_x < 0$
38	Average $u_y > 0$
39	Average $u_y < 0$

Table 4.11: FR factor for the 40 features and the target signatures.

Feature	FR					Average
	Target 1	Target 2	Target 3	Target 4	Target 5	
0	0.082	0.465	0.117	0.339	0.088	0.218
√ 1	0.099	0.047	0.033	0.056	0.088	<u>0.065</u>
√ 2	0.500	0.000	0.000	0.042	0.000	<u>0.108</u>
√ 3	0.042	0.162	0.111	0.127	0.106	<u>0.110</u>
4	2.118	5.400	6.302	1.679	0.553	3.210
√ 5	0.127	0.151	0.028	0.104	0.077	<u>0.097</u>
√ 6	0.286	0.000	0.000	0.283	0.319	<u>0.178</u>
√ 7	0.134	0.093	0.131	0.082	0.323	<u>0.153</u>
√ 8	0.132	0.077	0.087	0.074	0.293	<u>0.133</u>
√ 9	0.212	0.083	0.092	0.109	0.158	<u>0.131</u>
√ 10	0.173	0.083	0.095	0.092	0.136	<u>0.116</u>
√ 11	0.133	0.124	0.033	0.104	0.129	<u>0.105</u>
12	0.353	0.263	0.231	0.422	0.052	0.264
13	0.074	0.789	0.646	0.361	0.165	0.407
√ 14	0.189	0.124	0.177	0.081	0.134	<u>0.141</u>
15	0.618	0.185	1.195	0.581	0.123	0.540
16	0.166	1.301	0.488	0.694	0.572	0.644
17	0.212	0.106	0.245	0.713	0.172	0.290
18	0.141	0.062	0.122	0.525	0.047	0.179
√ 19	0.018	0.048	0.046	0.085	0.000	<u>0.039</u>
√ 20	0.142	0.081	0.036	0.135	0.000	<u>0.079</u>
√ 21	0.043	0.037	0.017	0.040	0.021	<u>0.032</u>
22	0.054	0.307	0.090	0.877	0.139	0.293
23	0.401	0.136	0.296	0.083	0.196	0.222
24	0.112	0.034	0.467	0.242	0.093	0.190
25	0.198	0.248	0.172	0.398	0.051	0.213
√ 26	0.151	0.031	0.254	0.137	0.050	<u>0.125</u>
27	0.704	0.320	0.476	0.174	0.360	0.407
28	0.066	0.770	0.610	0.369	0.261	0.415
√ 29	0.074	0.125	0.097	0.503	0.090	<u>0.178</u>
30	0.307	0.343	0.803	0.404	0.114	0.394
31	1.323	0.569	0.113	0.414	0.158	0.515
32	0.124	0.885	0.113	0.332	0.155	0.322
33	0.070	0.090	0.107	0.753	0.190	0.242
34	0.141	1.074	0.168	0.562	0.165	0.422
35	0.052	1.020	0.023	0.127	0.211	0.287
√ 36	0.164	0.211	0.050	0.096	0.146	<u>0.133</u>
√ 37	0.157	0.207	0.114	0.181	0.226	<u>0.177</u>
√ 38	0.136	0.079	0.056	0.183	0.075	<u>0.106</u>
√ 39	0.093	0.101	0.059	0.156	0.281	<u>0.138</u>

Table 4.12: Dynamic verification –Selected features.

Feature	Description
0	Total pen-down time
1	Number of components
2	Width/height ratio
3	Signature path length
4	Initial direction of signature
5	Number of u_x zero-crossings
6	Number of u_y zero-crossings
7	Number of a_x zero-crossings
8	Number of a_y zero-crossings
9	Average speed
10	Average acceleration
11	Pure-width/width ratio
12	Pure-height/height ratio
13	Average path-tangent angle
14	Maximum u_y - minimum u_y
15	Moment of minimum u_x - initial time
16	Average $u_x > 0$
17	Average $u_x < 0$
18	Average $u_y > 0$
19	Average $u_y < 0$

4.4.2 System Performance

Skilled forgeries

The performance of the dynamic verification system with the feature vector of 20 best features, the Medium Reference Set and a common threshold of $\theta = 2.575\sqrt{k}$ with $k=20$ (i.e. $\theta=11.52$), resulted in 95.83% correct classification, a 14.14% FRR and a 2.2% FAR. Although the common threshold by definition should minimise the FRR, the resulting rate is relatively high, while a very small FAR was achieved contributing to a small total error of 4.17%. Furthermore, an EER of 6.6% is attained for a threshold value of $\theta=13.35$. A clearer view of the system’s performance in terms of the attained error rates for a range of threshold values is provided by Figure 4.17, and the error tradeoff curve is displayed in Figure 4.18. It may be observed that a zero FRR maybe achieved in return for a 21.4% FAR, which conforms with the

performance requirements of POS applications (Type I error near zero with Type II error $\leq 25\%$ [79]).

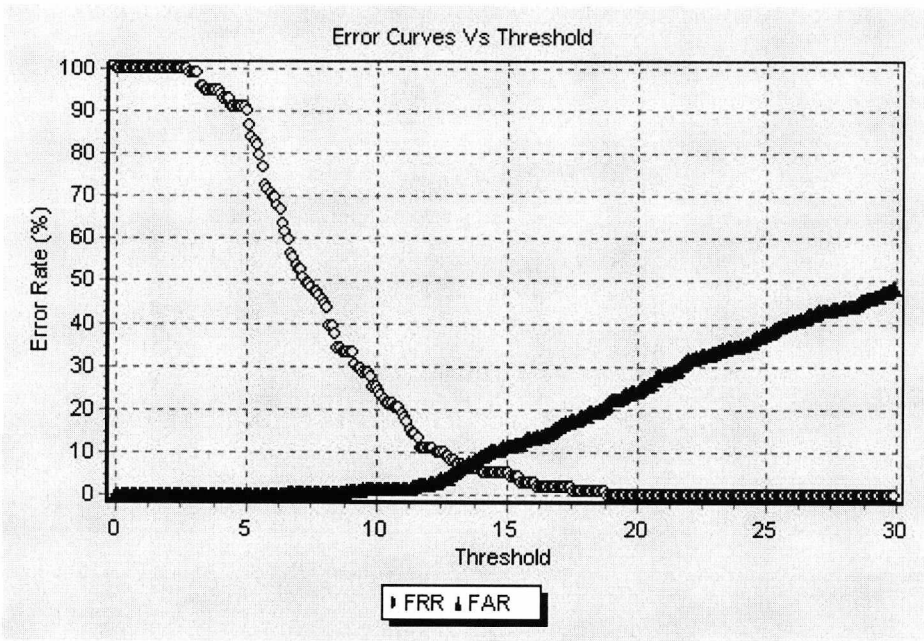


Figure 4.17: Dynamic verification –Error curves vs threshold.

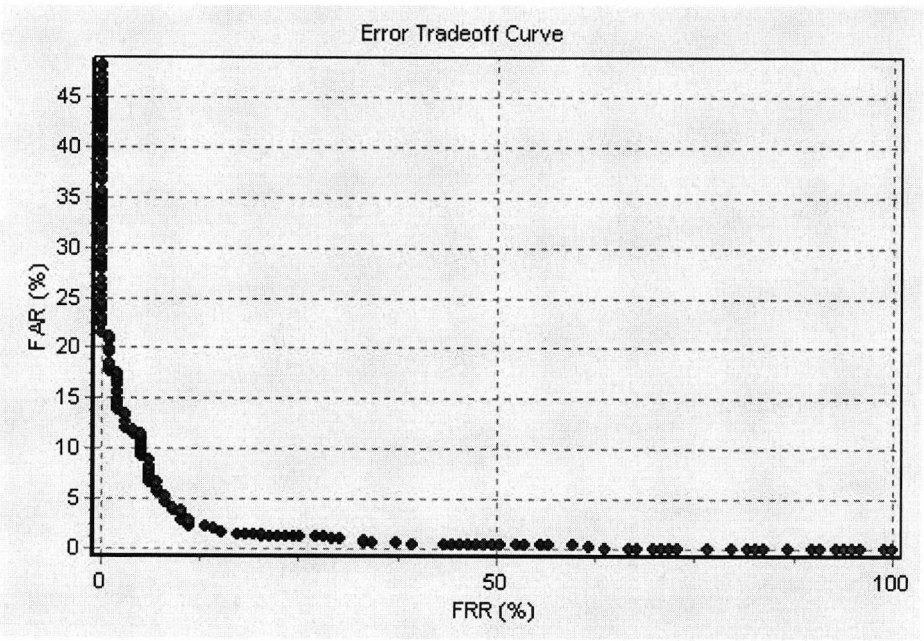


Figure 4.18: Dynamic verification –Error tradeoff curve.

Random forgeries

The performance of the system against random forgeries had a 99.83% success, thus, a small 0.17% FAR was attained, which was entirely attributed to false acceptances

with respect to Target 1 – i.e. for the rest of the target signatures no random forgeries were falsely accepted as genuine. The resulting FAR for a range of threshold values is displayed in Figure 4.19.

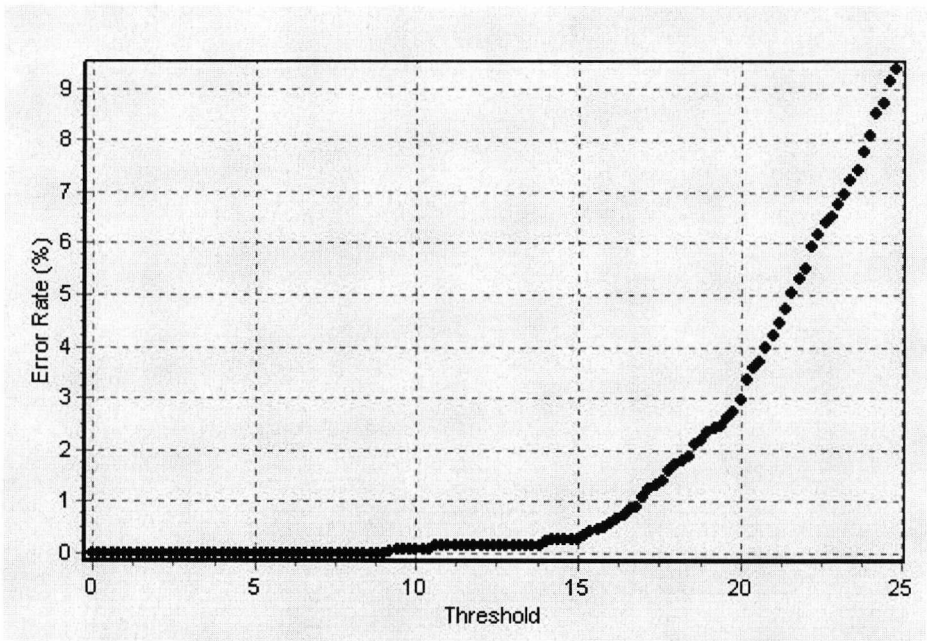


Figure 4.19: Dynamic verification –FAR vs threshold for random forgeries.

4.4.3 Optimisation

The resulting error rates for the common threshold and for each of the target signatures are displayed in Table 4.13 along with the equal error rates corresponding to each case.

Table 4.13: Dynamic verification –Error rates for the target signatures.

	Target 1	Target 2	Target 3	Target 4	Target 5
FRR %	5.00	40.00	15.00	5.00	5.26
FAR %	5.00	0.00	0.00	4.00	2.00
EER %	5.0	5.0	1.0	5.0	3.5

The analytical performance of the system for each of the target signatures, in terms of the resulting error rates for a range of threshold values, is displayed in Figures 4.20-4.24. The red dashed line shows the position of the common threshold. Apart from

Targets 2 and 3, the common threshold for the rest of the target signatures seems to lie close to the EER, thus resulting in low error rate values. Target 2 displays the highest FRR, while for a slightly larger threshold an improvement in the FRR for all targets could be made without any significant deterioration in the FAR.

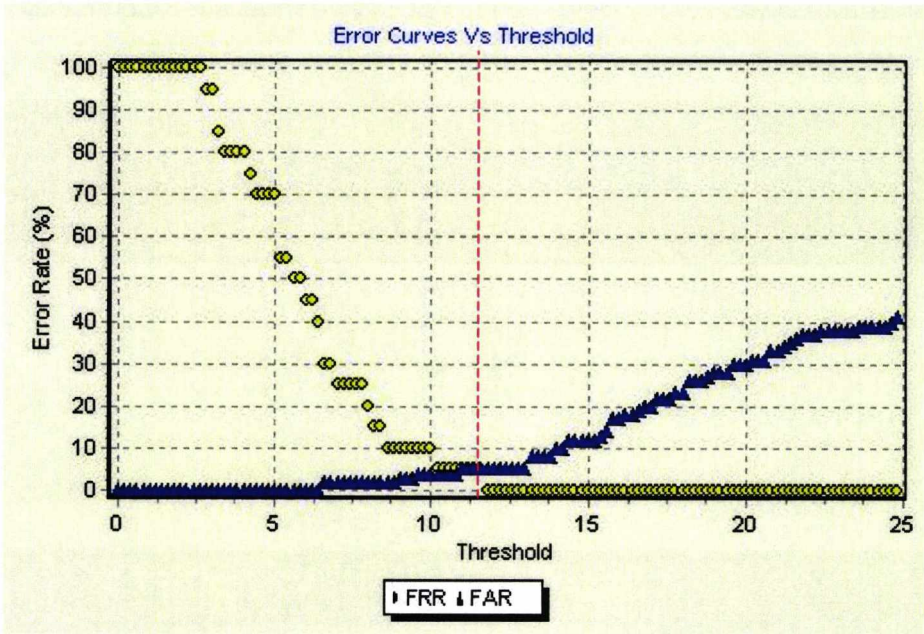


Figure 4.20: Dynamic verification –Error curves vs threshold for Target 1.

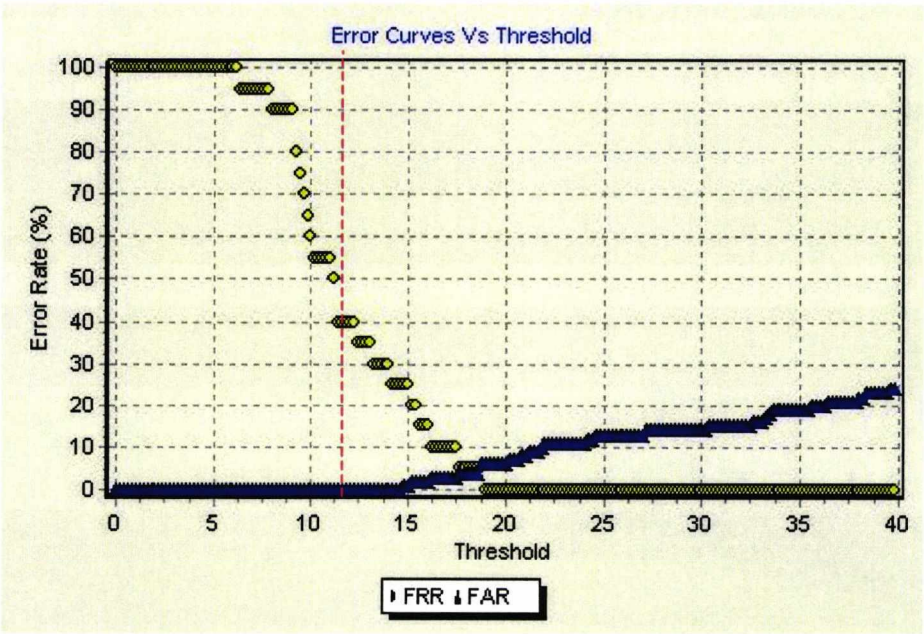


Figure 4.21: Dynamic verification –Error curves vs threshold for Target 2.



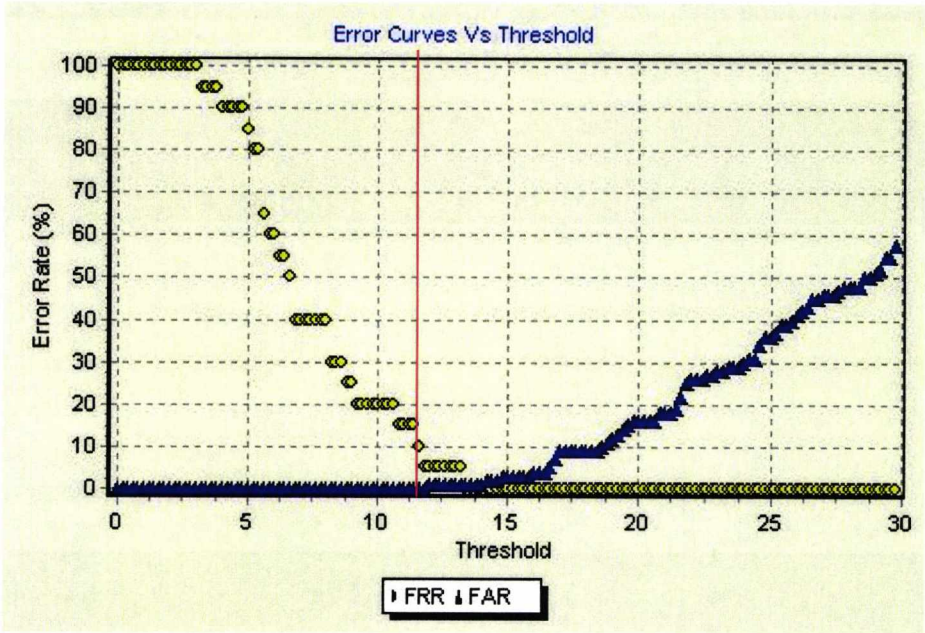


Figure 4.22: Dynamic verification –Error curves vs threshold for Target 3.

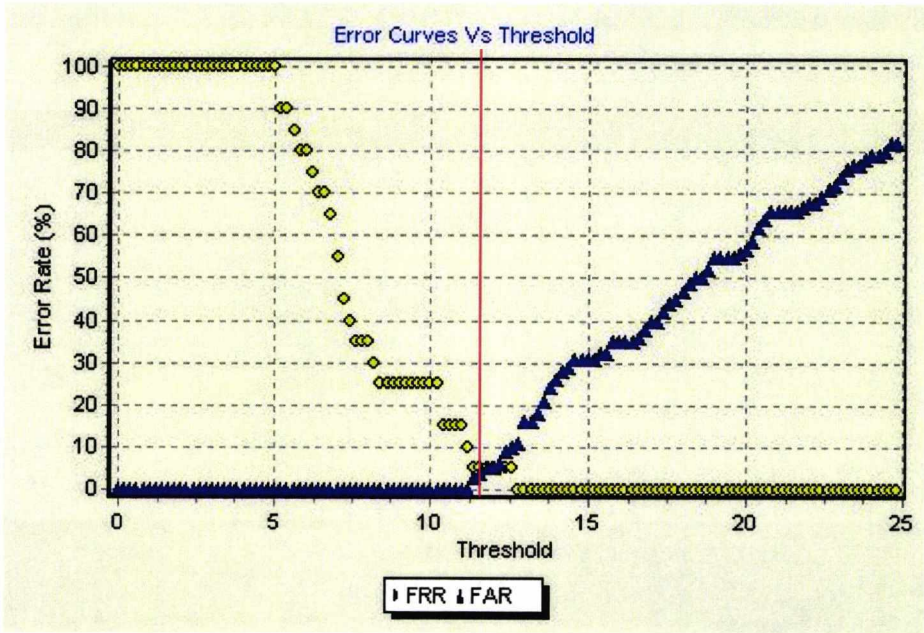


Figure 4.23: Dynamic verification –Error curves vs threshold for Target 4.

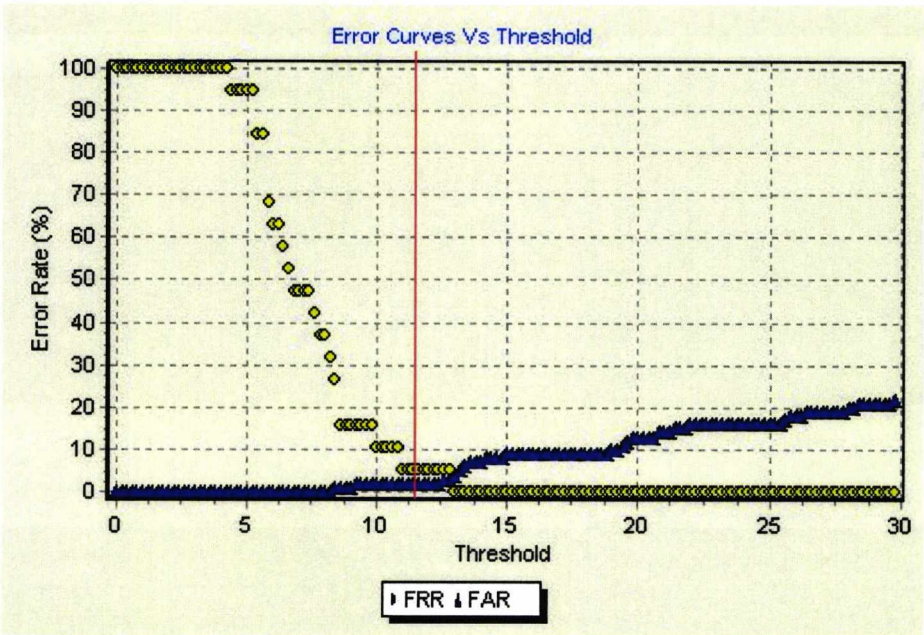


Figure 4.24: Dynamic verification –Error curves vs threshold for Target 5.

Similarly to the static verification approach, if individual thresholds are used according to (4.5) with $\alpha=1.5$, it is anticipated that this would lead to a larger FRR since the $1.5 \cdot \text{mean}(\text{dist})$ fails to even cover the maximum distance value in the five sample reference set (see Table 4.14). Indeed, for $\alpha=1.5$ a large FRR of 26.26% results with an FAR of 1.6%, while for a larger $\alpha=1.9$ the resulting error rates, of 9.09% FRR and 7.6% FAR, provide a lower performance compared to the actual performance of the system with a common threshold adjusted to a slightly larger value. In fact, according to Figure 4.17, for a threshold of $\theta=12.7$ the corresponding error rates are 9.09% FRR and 3.8% FAR. Thus, it is shown that in this situation the mean of the distances for individual threshold selection is not very appropriate. Moreover, in the static approach the threshold selection scheme (4.5) was used in order to tackle the large FAR by restricting the width of the threshold with individual values that described the intrinsic variability of the targets without causing a deterioration in the FRR. Here, the opposite situation has arisen, since the reduction of the FRR is of interest as the FAR is already very small. Therefore, a different approach to the mean of the distances should be followed.

The use of the maximum distance was also earlier proposed, which here seems to be more appropriate since it can relax the threshold values to accommodate higher intrinsic variability than the mean of the distances would. If the threshold selection procedure $\vartheta_w = \alpha \cdot \max(dist)$ is used with $\alpha=1.5$ this would relax the individual thresholds significantly, leading to a low FRR, with the cost of an increase in the already small FAR. Indeed, the results showed a 4.04% FRR and a 21% FAR, which is also a poorer performance than the respective 4.04% and 11.4% that the system results to for a common threshold equal to $\theta=15.2$ (see Figure 4.17). It is also apparent that for lower α values the FRR will be higher, while for larger α the FAR will increase. Therefore, this approach also seems inappropriate for this situation, where reducing the FRR with individual thresholds is of interest.

Consequently, individual threshold selection using the mean of the distances led to a considerable increase in the FRR, while the maximum of the distances, although it succeeded in reducing the FRR, it resulted in a considerably high FAR. Finally, both approaches performed much poorer than the respective performance of the system using common threshold.

It may be concluded that the 5 training samples are too few to adequately describe the intrinsic variability of the target signatures, leading to a relatively high FRR. In fact, this seems to be the case for Target 2, which is mainly responsible for the relatively high FRR attained (see Table 4.13). In order to explain this, an observation of the individual error curves for the five targets reveals that almost all targets perform similarly in the EER region, while the intersection of their two error curves occurs at very similar threshold values. This leads to the conclusion that individual threshold selection is not especially needed. However, Target 2 is the only target to perform very differently at the same threshold range, and thus, individual threshold selection proves necessary in this case. Moreover, individual threshold selection is considered beneficial for most applications, since there is always a high chance of finding a problematic signer for whom the system fails to perform in an optimum way.

Table 4.14: Dynamic verification –Intra-class distance measures for the targets’ reference set.

Target	rwi	$D(r_{wi}, R'_{Mw})$	mean
1	1	11.116	6.778
	2	3.012	
	3	11.689	
	4	4.384	
	5	3.689	
2	1	10.346	6.059
	2	7.193	
	3	4.867	
	4	4.200	
	5	3.691	
3	1	12.693	7.670
	2	3.978	
	3	3.893	
	4	10.374	
	5	7.413	
4	1	11.530	7.576
	2	13.533	
	3	3.829	
	4	2.719	
	5	6.269	
5	1	9.065	6.287
	2	6.090	
	3	3.910	
	4	5.680	
	5	6.692	

In a different approach, an optimisation of the resulting error rates for the system could be sought in a slightly more relaxed common threshold. According to the method described in the system design section the selected value of z corresponds to the probability that each of the features belongs to the normal distribution with mean and standard deviation values described in the respective reference vectors. The selected value of $z=2.575$ corresponds to a probability of 99%, i.e. the probability that a feature value will fall within z standard deviations on either side of the mean value. A higher probability is supported by higher z values. According to the normal error integral statistical tables [139], for a $z=3$ the probability is 99.73%, while a $z=4$ gives a probability of 99.994%. Hence, it is expected that higher z values will result in lower FRR values. With respect to the increased value of $z=3$, i.e. common threshold of $\theta = 3\sqrt{20} = 13.42$, the resulting error rates are 6.06% FRR and 6.6% FAR, thus very close to the EER. Furthermore, for $z=4$, i.e. $\theta=17.89$, the FRR is reduced to 1.01%, while the corresponding FAR is equal to 18.4%. Therefore, alternative solutions were given for the situation where a smaller FRR was of interest.

Reference Set

The performance of the system was also tested with the Small and Large Reference Set. A comparison of the attained error rates with respect to the three reference sets is provided by Table 4.15. It is apparent that the larger the set the smaller the FRR, which would be normally expected as a larger reference set provides a better description about the intrinsic variability of signatures. With respect to the FAR the Medium Set displays the lowest value, while all three sets have low FAR values. The EER is reduced with the increase in the number of reference samples, as is also shown in Figure 4.25. While there is a big reduction (40.5%) in the EER in moving from 3 reference samples to 5, there is only a 24.2% further reduction for an addition of another 5 samples (a total of 10). While there is obvious gain to be achieved by increasing the number of reference samples, the cost of the addition of two more samples in a reference set of three – in a hypothetical application design – may be compromised for the significant improvement offered by the bigger set.

Table 4.15: Dynamic verification –Error rates for different reference sets.

	Reference Set		
	Small	Medium	Large
FRR %	25.25	14.14	5.05
FAR %	4.60	2.20	4.40
EER %	11.1	6.6	5.0

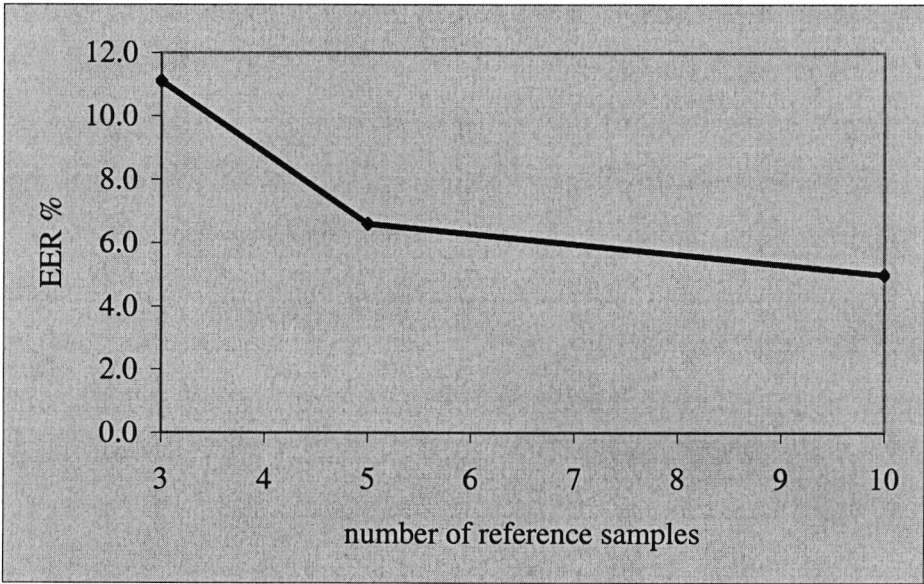


Figure 4.25: Dynamic verification –EER vs number of reference samples.

In the extreme case that only one reference sample is provided there are no standard deviation values to offer a normalisation in the distance measure, and hence, as earlier described, the distance measure is given by (4.10). On the other hand, in this dynamic verifier the parameters in the feature vector have very different measurements from one another with mean and standard deviation values of a different range, which leads, in this case, to the larger parameter values controlling the distance outcome. Consequently, in this extreme case the system does not perform to the optimum level. Moreover, the selected common threshold is inappropriate for the range that the distance values expand (see Figure B.11 –Appendix B). Thus, certain modifications would be needed in this case. For the aforementioned reasons, a comparison of the system’s performance for the Limited Set with the rest of the reference sets is not very appropriate. Nevertheless, its resulting error tradeoff curve is partially plotted (Figure 4.26), together with the respective curves of the other reference sets, as a rough indication of the system’s low performance when only one reference sample is available. The performance graphs for each of the reference sets are provided by Figures B.11-B.13 –Appendix B. An observation of the error tradeoff curves for the different sets shows that the larger the reference set the better the performance of the system. However, the Medium and Large Set seem to perform very similarly, although with a noticeable superiority of the Large Set in the important low error rate region.

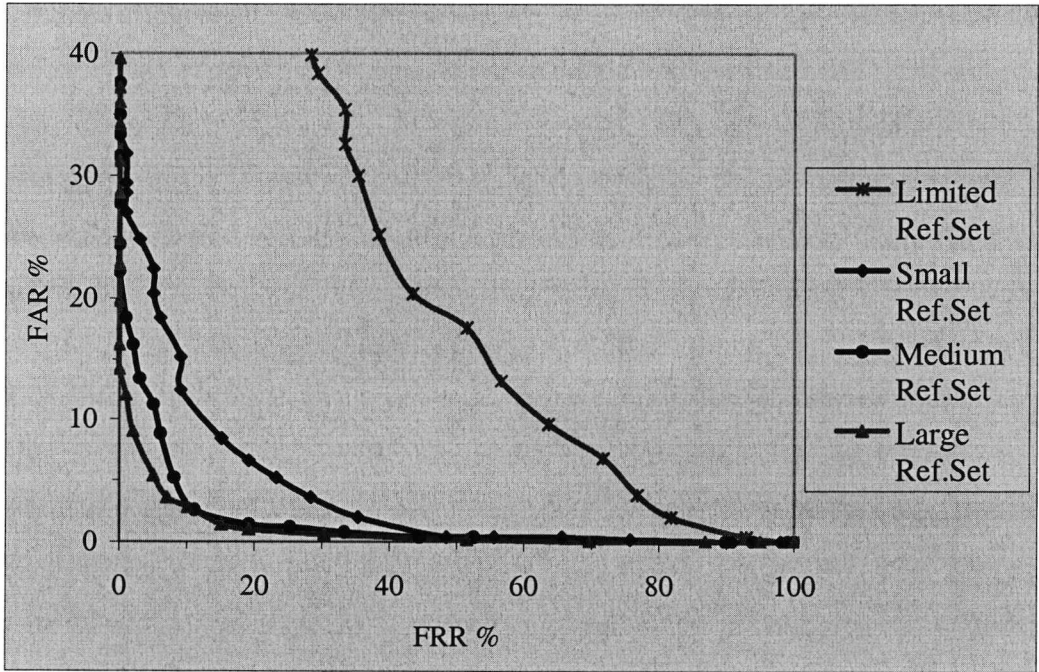


Figure 4.26: Dynamic verification –Error tradeoff curves for different reference sets.

The dynamic verification performance for random forgeries corresponding to the three reference sets resulted in only a 0.17% FAR for both the Medium and Large Set while a small FAR of 0.21% was also obtained using the Small Reference Set. The FAR curves for the three sets are displayed in Figure 4.27 for a range of threshold values. It is apparent that the Medium Set shows the best performance against random forgeries, while the Small set exhibits the weakest performance. The individual graphs for the different reference sets are provided by Figures B.14-B.16 –Appendix B.

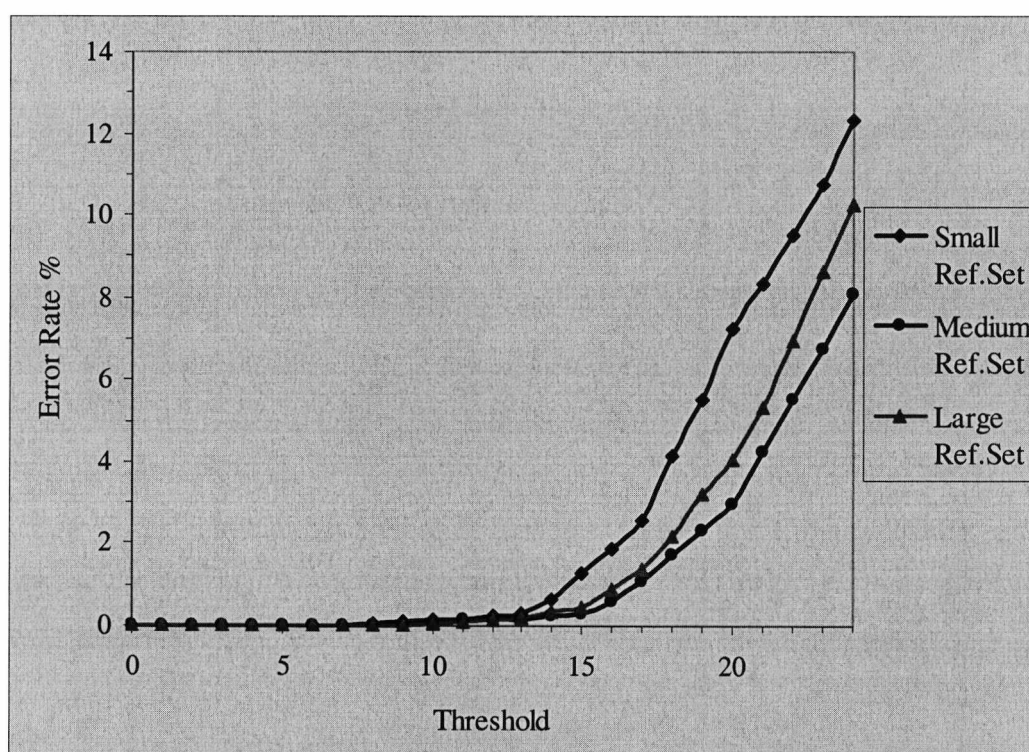


Figure 4.27: Dynamic verification –FAR vs threshold for random forgeries and different reference sets.

Number of parameters

The size of the feature vector used has an effect on the performance of the system. It is generally observed that too few parameters or too many parameters do not lead to the best performance. The performance of the system in terms of the attained EER for different sizes of the feature vector is presented in Figure 4.28. The various sizes of features sets tested were selected according to the feature selection procedure based on the FR factor earlier described. For a feature vector of 10 parameters an improved performance was measured with an EER of 3.7%, and furthermore for the common

threshold (with $k=10$) a 10.1% FRR and a 1.4% FAR were obtained. Finally, an optimum performance can be achieved for a combination of the best feature set (of 10 parameters)⁵ and the Large Reference Set of 10 samples per writer. The resulting EER is 3.3%, while the error rates for the common threshold are 7.07% FRR and 3% FAR.

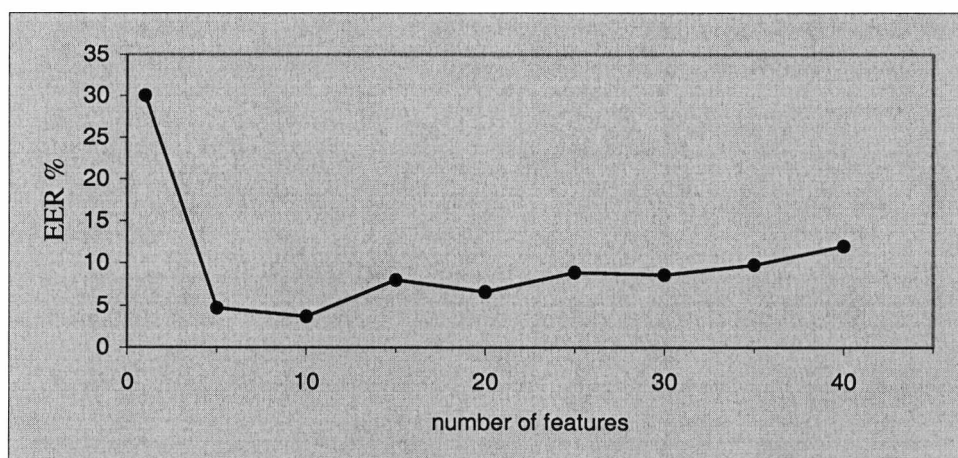


Figure 4.28: Dynamic verification –EER vs number of features.

4.5 Conclusions

The design and implementation of a handwritten signature verifier has been described, analysed and explained, while its performance in both an on-line and off-line approach was evaluated. Several optimisation methods were discussed and their merits were quantitatively assessed.

A serious effort was made for a realistic approach to the construction of the signature verifier approximating problems usually encountered in practical real world applications. Thus, the several methods used were not aiming at an optimal performance but rather at a more practical level. For these reasons, the training set only consisted of genuine samples, while the size of the reference set used was only 5 samples per signer. For the determination of the common threshold only the genuine signature class was considered, although a much better performance is guaranteed

⁵ Although, the EER for different sizes of feature sets may behave differently with respect to the Large Reference Set, thus the best feature set might be other than the 10 parameter one. Hence, further testing would be needed to determine the optimum result.

when forgery samples are also used, since it helps tuning the system to the optimum performance. The feature selection procedure used was also based on the reference set alone, while other methods proposed in the literature earlier described estimate the best feature set according to the minimum error rate achieved. In order for the latter to have the right effect, the forgeries used for training and those needed to test the system must be of a different type or collected under different conditions to avoid any biasing caused in the evaluation of the performance of the system.

With respect to the results reported in this chapter and the methods implemented a few important conclusions can be drawn. In the off-line verification three different sets of 7 moment invariants were computed, the central moments, the normalised central moments and the Hu moments, while the use of higher order moments was proposed and thus, an extended set of 12 normalised central moments up to the order of 4 was also computed. The selection of the best set was performed according to the value of the FR factor generated for each moment. A comparison of the performance of the system with the different sets verified the effectiveness of the FR factor. The set of higher order moments performed slightly better than just the lower order ones in the regions of low error rates. The normalised central moments showed a superior performance in comparison to the central moments or the Hu moments. In fact the Hu moments, being more advanced than the ordinary moments, have been considered more effective and are generally more popular in the literature of pattern recognition. However, for the verification of signatures, according to the results of this study, the Hu moments showed a very poor performance. The simple global feature set performed generally poorly compared to the moment invariants, however, its performance was comparable to the studies reported in Section 4.1 [118;117]. Note that these studies used as many as 15 genuine samples per signer for training, while the study in [117] also utilised a threshold set to minimise the total error rate which led to an underestimation of the error rates reported. It is acknowledged that lower error rates have been achieved in the literature of static signature verification [112] but for different methods, databases and approaches used. The 21.2% EER achieved for skilled forgeries with the 12 normalised central moments and a reference set of 5 samples is a good performance compared to the results reported in similar studies [10;118]. Moreover, it was shown that using a reference set of 10 samples leads to a much smaller EER of 14.8%.

In the dynamic verification 40 features were initially extracted including both static and dynamic parameters. The best set of 20 features was selected according to the FR factor generated for each feature. The performance of the system using a reference set of 5 samples yielded a 6.6% EER and for the common threshold a 14.14% FRR with a 2.2% FAR for skilled forgeries and a 0.17% FAR for random forgeries were attained. It was shown that using a larger reference set contributed to a significant improvement. Thus, a reference set of 10 samples resulted in a 5% EER, and for the common threshold in an FRR of 5.05% and an FAR of 4.4%. A test on different sizes of features sets showed the importance of this factor in the shaping of the error rates. It was shown that the best result was achieved for a feature set of 10 parameters, which led to an EER of as little as 3.7%, while for the common threshold a 10.1% FRR with a 1.4% FAR for skilled forgeries and a 0.17% FAR for random forgeries were achieved. A further improved EER of 3.3% was attained for both a reference set of 10 features and a feature set of 10 parameters. Overall, the performance of the system was comparable to other similar studies reported in the literature [8;79;43].

Finally, the individual threshold selection procedure was not found effective neither with respect to the static nor the dynamic approach. In both cases, a small adjustment in the common threshold provided a solution for either a small improvement or a different balance in the error rates. An enhanced scheme for the individual threshold selection method will be proposed in Chapter 8.

In conclusion, a comparison between the static and dynamic signature verification performance profiles easily shows the superiority of the dynamic approach with the advantages of higher feature efficiency, lower error rates and significantly lower computation time. On the other hand, depending on the application and the available data, the appropriate method must be used, and where only static information is available the performance characteristics reported here are to be expected.

In the following chapters, experiments with human subjects will be carried out in order to obtain important information about characteristics of signatures that lead systems to under-perform and make predictions about the vulnerability of problematic signatures in terms of the two types of errors, in an effort to overcome this weakness

in the construction of more robust automatic signature verification systems. The next chapter is focused on the study of one of these signature attributions: the degree of a signature's complexity.

Chapter 5

Handwritten Signature Complexity

Two of the most important handwritten signature attributions which characterise individual signing styles and which are known to affect the performance of both human and machine verification processes, are the complexity of signatures and their intra-class variability. This chapter is involved with an investigation of the former quality. The complexity of signatures is examined from the point of view of static signature image analysis, regarding the signature image as a whole. The notion of complexity is initially approached through human perceptual judgments on the visual complexity of signatures obtained from an experiment with human subjects. A method is further proposed which quantitatively estimates the complexity of a signature image reflecting the experimental findings on human perception integrated with an automatic approach.

5.1 Introduction

Signing styles among signatures generally vary and may be seen to reflect a different degree of inherent complexity. Despite the large volume of research in the field of ASV, very few studies have been concerned with the complexity of handwritten signatures.

Brault and Plamondon [25;27] proposed an algorithm to quantitatively measure the difficulty that a potential forger would experience in an attempt to reproduce a signature dynamically. Their model is based on the concept that the procedure of imitating a signature is made up of a sequence of overlapping subtasks, each subtask including a perception, preparation and execution step. The proposed imitation difficulty coefficient is mainly a function of the number of singularities contained in the signature and the rate of geometric modifications per unit of time [26]. In addition to the automatic approach, the imitation difficulty of the signatures is observed from the point of view of the forgers that produced the imitations and of an expert document examiner. The complexity of a signature is expressed in terms of the imitation difficulty of the signature and is mainly determined dynamically. However, a static component is included in the perception stage, where the difficulty in perceiving a spatial target in a signature is expressed as a function of the average angle formed by the target sides. Furthermore, in considering the concept of the imitation difficulty of a signature, Kim et al. [75] proposed a measure for the difficulty entailed in forging a feature value of a signature, called the degree of difficulty for forgery (DDF function). The value of the DDF function for each feature is mainly determined by the distance between the distributions of true signatures and of general handwriting, defined by their mean and standard deviation values, for that feature. The DDF function is used to define personalized weights in a signature's feature set.

A different approach in defining the complexity of handwritten signatures is proposed by Kaplani and Fairhurst [72], using solely static signature information. An algorithmic solution, based on the projections of the signature image, is proposed which reflects human perceptual criteria. The complexity of signatures has also been discussed by a number of other authors. Nalwa [98] reports that the simplicity of a signature is characterized by the domination of a few low-curvature strokes. Furthermore, Baltzakis and Papamarkos [21] suggest that the number of *closed loops* is an indication of the degree of complexity involved in the signature line.

It should be noted that the general concept of complexity has been repeatedly used in the past in the field of psychology, to describe the objective complexity of shapes,

mainly polygons [31;77;154]. Some of the physical measures of the shapes described in these studies, include: the number of turns on the contour, line length, arc length, size of angles, area, compactness, jaggedness, and the use of statistical moments.

5.2 Perceived Signature Complexity

The notion of complexity and perceived complexity of form was introduced in the field of psychology as early as the 1950s. Several studies, reviewed by Kolers [77] and Zusne [154], reveal valuable information about the perceived complexity of shapes, mostly polygons. A brief description will follow, first because the ‘handwritten signature’ may be included in the general class of ‘shapes’ and secondly because the findings relate to human perception of shape complexity.

Zusne [[154] ch.6] indicates the existence of a difference between the perceived complexity and the actual complexity of a shape. Moreover, Seiler and Zusne [133] have demonstrated in their experimental findings an increasing underestimation of the complexity of random polygons with increasing objective complexity and decreasing exposure time. Attneave [18] demonstrates that the main element that determines the perceived complexity of shapes is the number of *independent turns*⁶ in the contour, with the symmetry, angular variability and compactness (perimeter-squared/area)⁷, being secondary factors. These physical measures were also confirmed by Arnoult [16] to be good predictors of complexity judgments. Similar parameters have been reported in [31;55;77;136]. Furthermore, Attneave [18] showed that “*whether the shapes were angular, curved, or mixed made no significant difference in judged complexity*”. It is also expressed, as reviewed by Zusne [154], that *familiarity*⁸ and *association value*⁹ affect judged complexity, so that more-familiar shapes having also high association value are judged less complex. Concerning the familiarity judgments

⁶ Independent sides or contour turns are counted as the total number of turns in an asymmetric shape and about half the total turns in a symmetric shape.

⁷ In fact, P^2/A was found to intercorrelate with the angular variability, thus minimally contributing to the prediction of judged complexity and was therefore not used as a predictor variable.

⁸ As related to previous exposure to the shapes.

⁹ Measured from the associative response given by subjects through verbal associations to random shapes. Experiments on familiarity and meaningfulness (related to associational process) are reported in [16].

on shapes, Arnoult [16] states that *“there are certain invariances in the forms of the real world, and nonsense forms will be judged as familiar to the extent that these familiar physical properties appear in them”*.

More specifically, in the field of handwritten signature verification, rankings on the imitation difficulty of signatures were obtained in [27] from the imitators first, and an expert document examiner second, and were both compared against an objective measurement of signature dissimilarity, based on Dynamic Time Warping, proposed by the authors under the hypothesis that *“the more difficult a signature is to imitate, the more the imitations should differ from the original”*. The document examiner results were shown to be almost opposite to the other methods and were interpreted as follows: *“the examiner usually considers a regular, harmonic signature more difficult to imitate because even a small deviation could be easily detected. On the other hand, from an imitator point of view, it is more difficult to accurately reproduce a noisy signal than a smooth one”*. In a different approach, the relation between perceived signature complexity and human performance, in terms of the achieved error rates in verifying handwritten signatures, is evaluated in [51;52], where information is also inferred about the susceptibility of a signature to forgery in practical scenarios where human checking occurs.

In the following section an experiment with human subjects will attempt to sketch an outlook on how people perceive the complexity of handwritten signatures.

5.2.1 Experimental Procedure

A group of thirty-six subjects, mostly comprising university students of different nationalities, participated in the experiment. The subjects were asked to view a sample for each of the five target signatures (Figure 3.1) and provide an estimation of its perceived complexity, on a scale of 1 to 10. The subjects were given no specific information about the criteria needed for the required judgment, nor any type of definition of the term ‘complexity’. The genuine signature sample selected to represent each of the target signatures was the 4th in the sequence of 10 produced by the signers. This ensured a basic familiarization of the signers with the digitiser, while

maintaining their interest before the repetition of the signing process became tiresome. An illustration of the experimental interface is provided in Figure A.1 – Appendix A. In a subsequent questionnaire part of the subjects were asked to give details about the criteria and factors that influenced their decision and more importantly about the way they understood the concept of ‘complexity’.

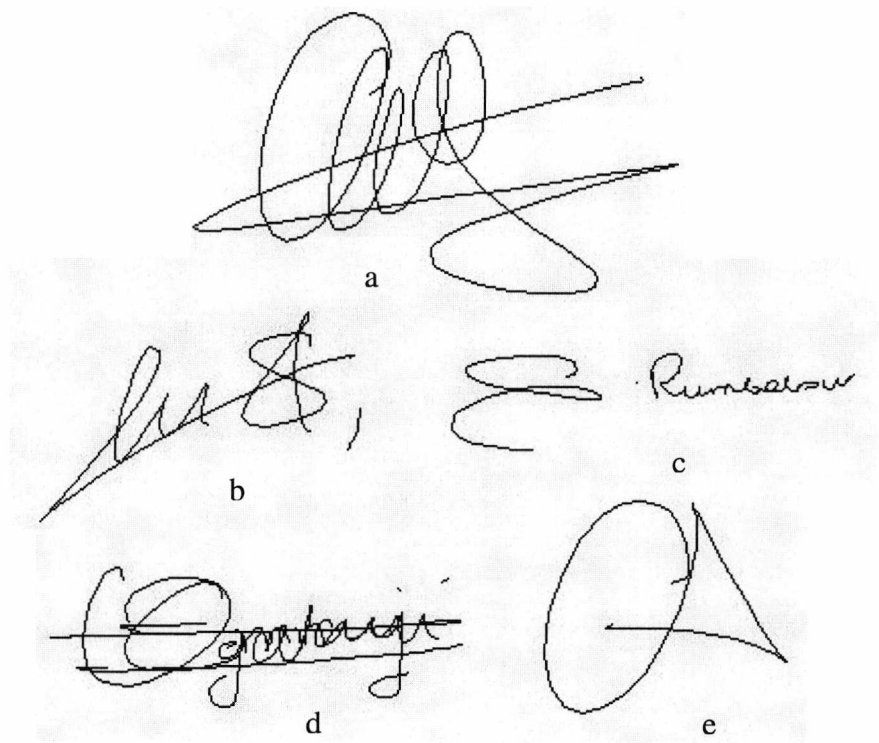


Figure 5.1: Signature images of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, and (e) Target 5.

5.2.2 Analysis of Results

Statistical measures of the perceived complexity judgements derived from the experimental data of the five target signatures are displayed in Table 5.1. A small disagreement is observed between the mean, median and mode values of the targets’ perceived complexity, which suggests that their frequency distributions exhibit a slight skewness¹⁰. In particular, those of Targets 2 and 3 appear to be positively

¹⁰ The degree of asymmetry of a distribution. Positively skewed data (rightward skewness), have the mean > median > mode, negatively skewed data (leftward skewness) have mean < median < mode. A symmetrical distribution (with the exception of the bimodal) has the mean = median = mode.

skewed. However, with respect to Targets 1 and 4, having almost negligible differences between the three measurements of central tendency, a similar assumption based on a relative comparison of those differences would present inaccurate results. Target 5 represents a more unusual situation, which will be extensively discussed at a later point.

Table 5.1: Perceived complexity statistics for the target signatures.

Target Signature	Perceived Complexity			
	Mean	Median	Mode	St. Deviation
Target 1	5.8	6	6	1.6
Target 2	4.8	4.5	4	1.9
Target 3	4.1	4	3*	1.9
Target 4	8.2	8	8	1.3
Target 5	1.8	2	1	1.2

*Bimodal. Two modes exist (3 and 4). The smallest is displayed.

A visual observation is provided by Figure 5.2, which displays the frequency distributions of the perceived degree of complexity for each of the target signatures. With Targets 4 and 5 being at the extremes of the scale, with high and low complexity estimates, respectively, Targets 1, 2 and 3 are assigned intermediate complexity values that are more difficult to precisely determine and rank. The effect of this can be more closely observed in Figure 5.2(a), (b) and (c). The frequency distributions of Targets 4 and 5 – Figure 5.2(d) and (e) – show evidence of a more unified and certain human response, as opposed to those of Targets 1, 2 and 3 that reveal a more confused decision-judgment. A support for this observation is provided by the range of the distributions or by the standard deviation values.

A further analysis of the frequency distributions and a comparison between them is provided by the box-plots of Figure 5.3. The upper and lower ends of a box indicate the 75th and 25th percentiles of the data. The line inside the box represents the median, the cross in the diamond corresponds to the mean, and its height to the confidence interval around the mean – with a confidence level of 95% – indicating the range within which the population mean is likely to lie. The dotted line connects the observations that fall within 1.5 IQRs (inter-quartile ranges). Possible outliers appear

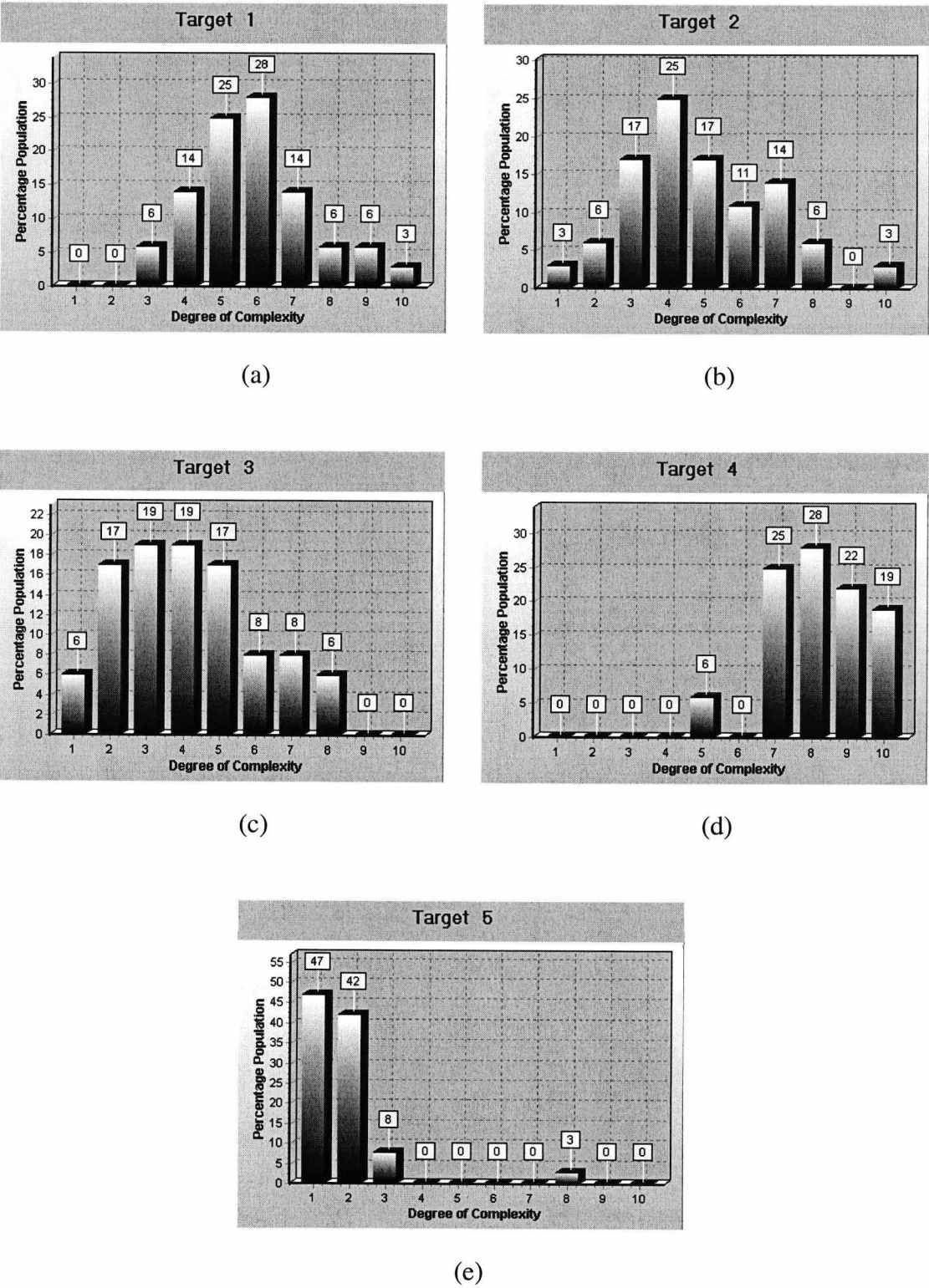


Figure 5.2: Frequency distributions of the perceived degree of complexity for target signatures: (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

as individual points marked as ‘+’ for near outliers (within 1.5 and 3.0 IQRs from the quartiles) and ‘o’ for far outliers (over 3.0 IQRs distance).

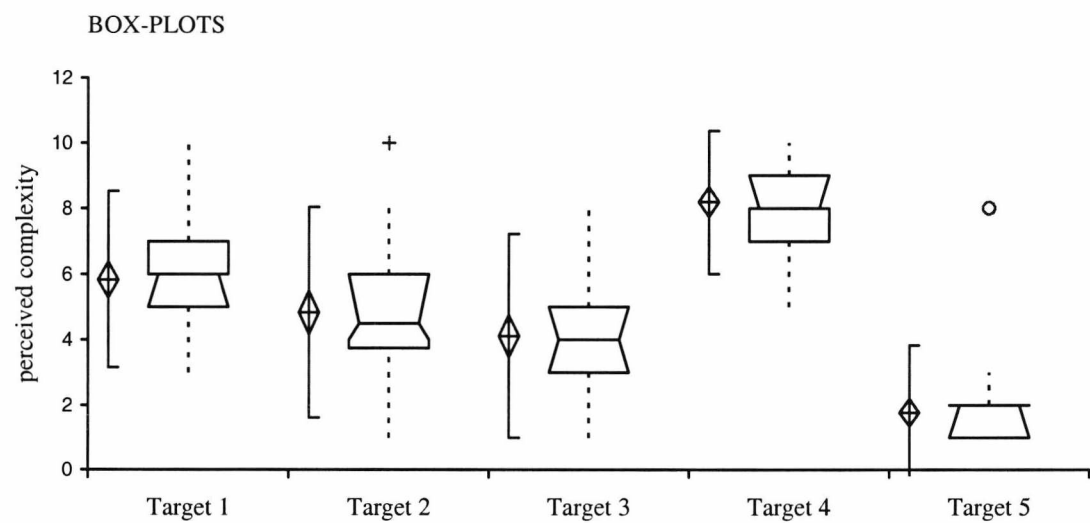


Figure 5.3: Box-plots of the perceived complexity judgments for the target signatures.

Elements that could reveal normality or possible skewness are the existence or not of outliers, the length of the upper tail compared to the lower tail of the box-plot and the position of the mean and median. Outliers above the upper range (e.g. Targets 2 and 5) or a mean value above the median line (e.g. Target 2) would be responsible for rightward skewness. A longer lower tail (e.g. Targets 4, 2) would reflect a slight leftward skewness. On the other hand, a median line evenly dividing the box and a mean value nearby, would exhibit normality. Hence, most of the target signature distributions exhibit near-normal behaviour.

The normality of the target signature distributions was, in fact, tested with the Shapiro-Wilk normality test¹¹. The null hypothesis of normality, i.e. that the data are random samples from a normal distribution, is rejected if the probability (p-value)

¹¹ According to [[14] ch.9] the Shapiro-Wilk test is overall the most powerful test of normality, with the Shapiro-Wilk W statistic given by $W = \frac{(\sum a_i X_{(i)})^2}{\sum (X - \bar{X})^2}$, where X are the sample values, \bar{X} the sample mean, $X_{(i)}$ the ordered sample values and a_i the optimal weights generated from the expected values of the n order statistics from the standard normal distribution and the corresponding covariance matrix. The W statistic may be viewed as the squared correlation coefficient between the ordered sample values and the corresponding normal scores, with values close to 1 indicating normality and smaller values indicating departure from normality. Here, the results of the Shapiro-Wilk normality test were generated with the help of a statistical software package.

associated with the test statistic is less than a chosen level of significance (for instance 0.05 for 95% confidence). Targets 1, 2 and 3 resulted in high W statistic scores with p-values greater than the 0.05 level of significance (p-values of 0.069, 0.23 and 0.099 respectively), thus, with 95% confidence the null hypothesis of normality is not rejected, concluding that the data come from a normal distribution. Target 4 resulted in a low W statistic (0.904) with a p-value (0.004) smaller than the 0.05 level of significance, thus resulting in a rejection of the null hypothesis of normality, but yet greater than the 0.001 level of significance, meaning that there is an extremely small probability that the null hypothesis is true. For Target 5 the resulting W statistic was very low (0.556) with an extremely small p-value (<0.0001), and hence, the null hypothesis of normality was rejected.

Outliers

As observed from Figure 5.3, Target 2 appears to have a possible near outlier and Target 5 a possible far outlier. Assuming a normal approximation for the distributions of the target signatures the suspect outliers were tested with Chauvenet's Criterion¹². The results revealed that both cases failed the criterion, with measurements for the expected number of samples as deviant as the suspect outlier smaller than the least required value of 0.5 (the resulting values of n were 0.22 for Target 2 and $7 \times 10^{-6} \approx 0$ for Target 5), and hence, the corresponding outliers could be considered for rejection. Thus, if it is assumed that the corresponding distribution of Target 5 could be approximated by the normal curve, the probability¹³ for the occurrence of the outlier would be as small as 2×10^{-7} .

The extreme case of Target 5, whose outlier appears to be highly improbable, will be examined and for this purpose a better representation is included in Figure 5.4. The outlier (subject 32) is spotted lying 5.2 standard deviations away from the mean. Although the hypothesis of normality was earlier rejected for the distribution of Target 5, the large deviation of the suspect measurement from the mean gives

¹² Chauvenet's criterion states that if the expected number of measurements at least as deviant as the suspect measurement is less than one-half ($n < 0.5$), then the suspect measurement should be rejected. According to [[139] ch.6], the calculation of n is carried out from equation: $n = N(1 - \text{Probability}(\text{within } ts))$, where N is the total number of samples and t the number of standard deviations s by which the suspect outlier lies away from the mean. The probability values are obtained from the normal distribution tables showing the percentage probability as a function of t .

¹³ Extended tables of the integral of the normal curve as a function of $z = |x - \mu| / \sigma$ exist in [24].

evidence that the sample in question is an outlier and could be rejected from the data set.

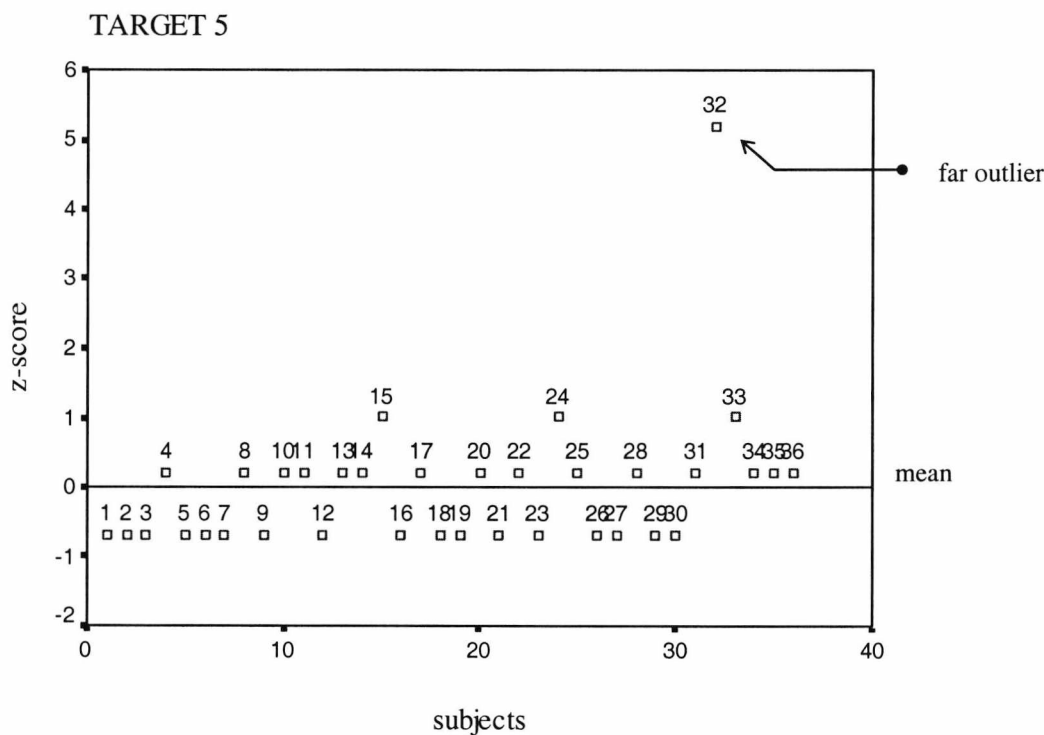


Figure 5.4: Perceived complexity z-scores ($z = (x - \bar{x}) / s$) of all subjects for Target 5.

The effect of the outlier on the normal approximation may be observed from the frequency distributions of perceived complexity for Target 5 including and excluding the outlier, Figure 5.5(a) and (b) respectively, having the normal curve superimposed.

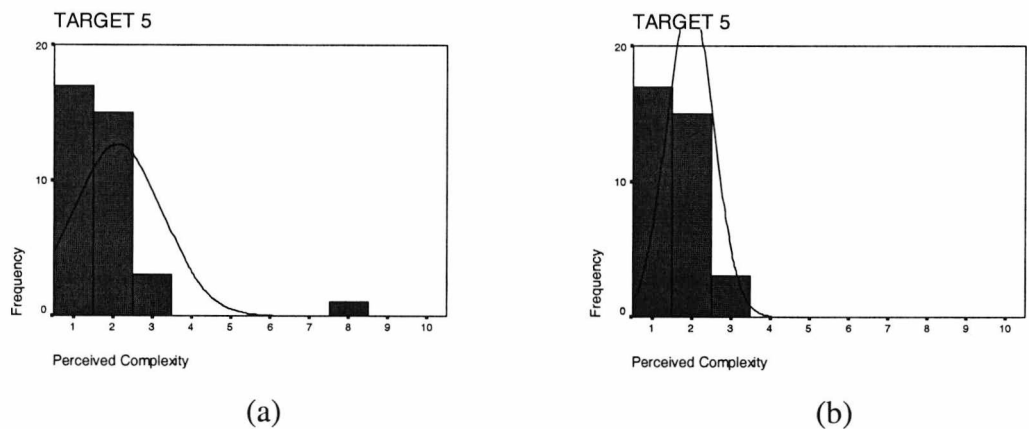


Figure 5.5: Perceived complexity frequency distribution for Target 5 (a) including and (b) excluding the outlier.

The measured statistics for the two cases are presented in Table 5.2. The absence of the outlier redirects the statistical values, thus achieving a significantly lower standard deviation and slightly sliding the mean to a lower value. On the other hand, the existence of the outlier creates a high degree of positive kurtosis¹⁴ and positive skewness to the frequency distribution, thus significantly deviating from the normal approximation¹⁵.

Table 5.2: Perceived complexity statistics for Target 5 including and excluding the outlier.

Statistics	Target 5	
	including	excluding
Sample Size	36	35
Mean	1.8	1.6
Median	2	2
Mode	1	1
St. Deviation	1.2	0.6
Skewness	3.7	0.6
Kurtosis	18.1	-0.5

Discarding data is a debatable issue. The rejection of data that are highly improbable could, in itself, be questionable. In the last analysis of Target 5’s distribution a sample was found extremely deviant from the mean and its occurrence could, therefore, be justified as a misunderstanding of the experimental task on the part of the subject, or as a misdirected perception of the task in question for the particular target signature, or in an extreme case as a reckless entry. On the other hand, it could also be an entry that reflects a part of the human population that perceives the specific signature shape in a more peculiar way, and this could be an interesting point in itself. Thus, for the reason that this experiment handles a sensitive and complex subject, that of human perception, the sample in question will not be rejected. However, its effect on the

¹⁴ Kurtosis indicates the relative peakedness or tail thickness of a probability distribution. Positive kurtosis corresponds to a distribution more peaked than normal and heavy tailed, and negative kurtosis to one flatter than normal and light tailed.

¹⁵ The normal distribution has zero degree of skewness and degree of kurtosis equal to 3. The measurements are made with kurtosis calculated as c-3, so that zero kurtosis would correspond to the normal distribution.

approximation of the distribution properties towards the estimation of the ‘true value’ was extensively analysed.

5.2.3 Perception Remarks

The results of the experimental study and an analysis of the subjects’ responses in the questionnaire indicate that a major criterion in determining the perceived complexity of a signature is the ‘readability’ or ‘intelligibility’ of the signature image. In other words, inspecting the signature involves an attempt to understand the signature by recognizing and identifying familiar letters or shapes. A second important factor in many cases was an attempt to reproduce or intuitively retrace the signature’s curve-line. In spite of the difficulty entailed in attempting to categorise the questionnaire responses quantitatively, it was evident that the two most important factors that determine the complexity of a given signature are its readability and its reproducibility.

Further elements that were regarded to affect the perceived complexity of signatures were also mentioned. These include the length of the signature, the curves/loops that overrun the signature’s body, the number of lines crossing over the signature image, and the apparent clarity of existing geometric shapes. These appeared to be secondary factors of relatively lower significance in defining the complexity of a signature. Moreover, they are indirectly embraced by either or both of the two main factors-processes, thus increasing or decreasing their apparent difficulty.

It should be noted here, that the last factor may be seen to relate to issues of familiarity and association, which showed to affect the judged complexity of shapes in the psychology studies earlier discussed, corresponding to geometric shapes that may have seemed to appear in some signatures (see, for instance, Target 5). In addition, the notion of complexity as perceived by the subjects, entails the imitation viewpoint discussed in the introduction, as one of the two main factors that were present in the determination of the degree of a signature’s complexity was its reproducibility.

5.3 Automatic Complexity Estimation

This Section aims in the determination of an automated algorithmic approach to quantitatively measure the complexity of a given signature. For this purpose, several approaches were considered and the range of possibilities will first be briefly described.

In an initial approach, the complexity of a signature was empirically found to correlate with some characteristics of the horizontal and vertical projections of the signature. These characteristics provide information about the simplicity of the projection spectra and include the following:

- Number of peaks
- Peak/valley ratio
- Width at half-max of the peak
- Spectrum density (number of black pixels/ width of projection)

A high peak/valley ratio and a small width indicate a well-defined peak. A projection spectrum of low density composed of a small number of well-defined peaks, characterises a simple spectrum and, hence, depicts a signature image of low complexity. For a complexity model to be considered, formed from the proposed factors, an accurate detection of the spectrum peaks is essential. The determination of the exact position and width of the peaks is troublesome, since there are peaks 'hidden' in the projection background which impose a difficulty for their proper detection.

A better approach of lower computational cost, that involves the evaluation of the projection as a whole, is based on an alternative method which involves the notion of Entropy, and this will be discussed in the following Section.

It should be noted that different approximations to that of the projection spectrums could also be adopted in defining the complexity of a given signature. A possible element that was found to highly correlate with the perceived degree of signature complexity is the number of *cross points*¹⁶ in a signature.

¹⁶ Points in the signature with at least three 8-neighbors [21].

5.3.1 Methodological Approach

The proposed algorithmic approach eventually adopted for the quantitative measure of the complexity of a given signature will follow after an introduction to the relevant background and a brief familiarisation with the notion of Entropy.

Relevant background

The notion of Entropy has been used extensively in the literature of pattern recognition, mainly as a variation measure for character image data, as proposed and reviewed, for example, in [73]. Côté [39] and Côté et al. [41;42] have proposed the use of Entropy for the extraction of the baseline of cursive script. The method involves computing histograms for different angles of projection and calculating the Entropy associated with each one; the histogram having the lowest Entropy is said to correspond to the writing direction.

It should be noted that the concept of Entropy has been widely applied in the field of image processing, to evaluate the information content of an image, as in [82]. This method involves computing the image histogram from the normalized frequencies of occurrence of the grey-levels in the image. The Entropy associated with the image histogram is independent of the size of the image or its pattern, and depends only on the frequency of each grey-level of the image. Some further studies have attempted to include the image spatial information in their definition of Entropy. Brink [29] uses a measure of local grey-level variation to weight the pixel illumination probabilities of the image, for use in a threshold selection process. Ghali et al. [54] have proposed the normalised rotational information of an image, based on Shannon's information theory. This, in short includes rotating an image through a number of possible angles on a pixel grid and recording the output pattern at each angle. A probability distribution is formed from the occurrence probability of each pattern. This technique was tested on rectangular, circular and triangular shapes and was also tried on a set of Helvetica characters. The normalised rotational information of an image was found to be proportional to the structural and geometrical image information, as it is shown to vary with the complexity of the image border irregularities.

Entropy

The concept of Entropy initially developed by Boltzmann – known as Thermodynamic Entropy – is a measure of the amount of ‘disorder’ in a system. Shannon [134] was the first to introduce Entropy in the field of information theory, as a measure of information, choice and uncertainty. Shannon’s Entropy H , which may be linked to Boltzmann’s Entropy S through the well-known constant k , has also been interpreted as a measure of the heterogeneity of a probability distribution [91].

For a discrete probability distribution $P = \{p_1, p_2, \dots, p_N\}$ Shannon’s Entropy is defined as:

$$H = -\sum_{i=1}^N p_i \log p_i \quad (5.1)$$

Note that, the logarithmic base is determined by the choice of a unit for the measured information. Thus, if the logarithmic base is 2 the Entropy is expressed in bits.

The two most important properties of H are:

- a) $H=0$ if and only if all the probabilities, except one, are equal to zero¹⁷ with the one remaining equal to 1 ($p_i=0 \ \forall i \neq j$ and $p_j=1$). This situation gives no information since the outcome is certain.
- b) H is maximum and equal to $\log N$, if all the probabilities are equal ($p_i = \frac{1}{N} \forall i$). The resulting distribution is uniform and reflects maximum uncertainty and information.

In fact, the more homogenous the probability distribution the higher is the uncertainty and the information and the greater the resulting Entropy.

Transferring a signature image into an x-y representation, based on the corresponding histograms in those axes, we may proceed with the following definitions. The vertical projection of a signature is the histogram computed from the projection of the signature image on to the x-axis and the horizontal projection is similarly computed from the projection of the signature image on to the y-axis.

¹⁷ In this case, for probability p_i equal to zero the expression is calculated as: $\lim_{p_i \rightarrow 0} p_i \log p_i = 0$.

The vertical/horizontal projections, of a signature image I , are computed by summing¹⁸ the pixels on each column/row and are respectively defined as:

$$P_v(j) = \sum_i I(i, j) \quad (5.2)$$

$$\text{and} \quad P_h(i) = \sum_j I(i, j) \quad (5.3)$$

where the binary image function $I(i, j)$ is 1 for black pixels (signature point) and 0 for white pixels (background), and where i refers to the rows and j to the columns.

The probability of occurrence of a single projection may be defined as the number of pixels on the particular projection – column or row – over the total number of pixels (N) in the signature [41]. The Entropy of the probability distribution $P=\{p_1, p_2, \dots, p_n\}$ with $p_j=P_v(j)/N$ or $p_i=P_h(i)/N$, depending to which histogram it corresponds, may be calculated from formula (5.1).

The vertical and horizontal projections of the five target signatures are displayed in Figure 5.6 together with their associated¹⁹ Entropy value. It is apparent that the more homogenous a histogram is, the higher is the Entropy associated with it. On the other hand, the more defined the peaks of a histogram and the more uneven the distribution, the smaller the resulting Entropy. This is well justified, since the more homogenous a distribution, the higher the uncertainty about the outcome – for the reason that all outcomes are nearly equiprobable – and, hence, the greater the amount of information contained. Since Entropy is a measure of information and uncertainty, the higher the uncertainty the higher the resulting Entropy.

¹⁸ Another projection technique involves taking the maximum value of each column/row [88], but in the case of a signature image this would line up to the upper envelope characteristic [20] and the right signature profile [71], whereas the preferred summing technique provides us with information about the distribution of pixels within the signature image.

¹⁹ The Entropy is measured on the probability distributions and not on the projections of the signatures. However the vertical and horizontal projection histograms differ from their respective probability distributions only in terms of scale (by $1/N$), so the histogram image appears the same. For convenience the Entropy of the probability distributions of the projections will be declared as the Entropy *associated* with the respective histograms.

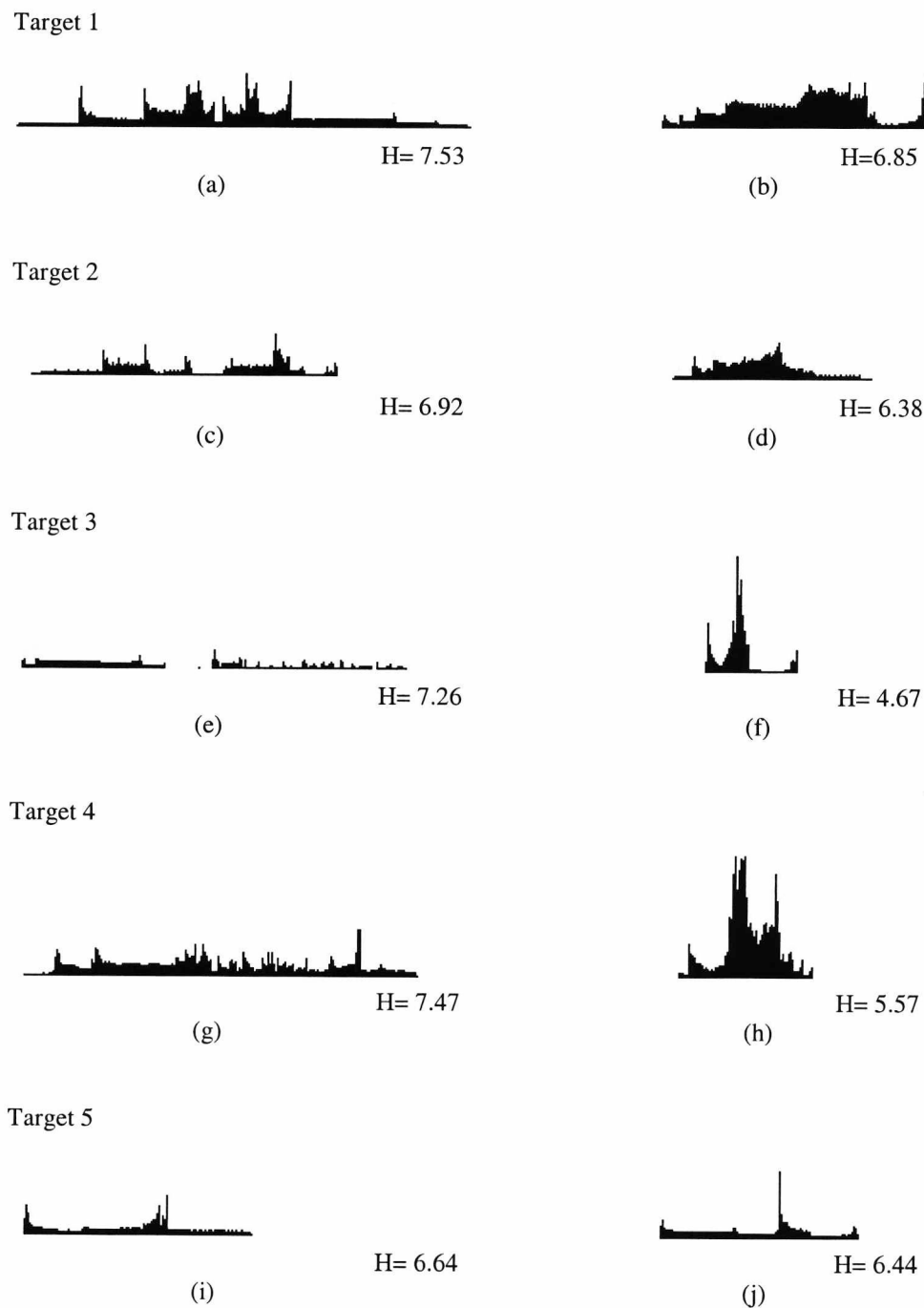


Figure 5.6: Vertical projections of Target: (a) 1, (c) 2, (e) 3, (g) 4, (i) 5, and horizontal projections of Target: (b) 1, (d) 2, (f) 3, (h) 4, (j) 5, along with their associated Entropy value.

Complexity

The complexity of a signature image may be reflected on its vertical and horizontal projections. Conversely, the form of the histogram spectrums may support conclusions about the complexity of the signature.

As mentioned earlier, a histogram comprising a few well-defined peaks with high peak/valley ratio implies separability among the signature elements; and, hence, denotes a signature of high readability and therefore high reproducibility. According to the experimental results on human perception previously described, these are qualities of low signature complexity. In addition, as previously mentioned, the former spectrum properties correspond to low Entropy and poor information content. It can also be argued that the amount of information involved depicts, in a way, the degree of complexity entailed. Finally, a positive relation may be assumed between the complexity of a signature image and its associated Entropy.

By definition Entropy is an additive quantity, so that the Entropy of a system will be the sum of the Entropy in each subsystem. Assuming that the two probability distributions are independent events, x and y , the joint probability will be $p(i, j) = p(i) \cdot p(j)$ and, consequently, the joint Entropy will be $H(x, y) = H(x) + H(y)$.

Therefore, the complexity of a signature may be measured as a result of the additive effect of the Entropy corresponding to each projection. However, if the complexity was defined as a direct function of the Entropy sum, the complexity estimates for the following cases would present undesirable results.

- a) Similar probability distributions but of different densities.
- b) Distributions of high homogeneity but of low density
- c) Similar and rather homogenous probability distributions, of the same density but of different length.

The first case reflects situations where similar probability distributions would result in similar Entropy values – as in Figure 5.6(e) and (g) or as a poor approximation in (b) and (d) or perhaps (f) and (h) – but the corresponding complexity should be different since the background densities are very dissimilar. The corresponding complexity

should be greater for a distribution of higher density. In addition, for the second case, spectrums as in (i) and (j) appear to be very simple with, however, high Entropy values. This is due to the high degree of homogeneity in their probability distributions. Indeed, Target 5, as verified by the experimental results of human perception (see Section 5.2.2), exhibits very low complexity. Thus, the density of the histograms should be taken into account in the complexity measure. Furthermore, the last case relates to signatures of similar and rather homogenous probability distributions with similar densities but spanning over different lengths. Thus, this reflects a situation where one is spread out, whereas the other is more condensed. Since the difference in density would not affect the Entropy value, the longer one would correspond to a higher Entropy estimate. This is because by definition, for equiprobable events, H is a monotonic increasing function of the number of events (with more possible events there is more uncertainty) [134]. On the other hand, a signature that is spread out over a wider length should score less on the complexity scale than one which has more information squeezed in a smaller distance, but yet higher than one with lower density. Consequently, the complexity measure should be normalized over the length.

Considering the above, the complexity of a signature image may be defined as:

$$\text{Complexity} = O_{lx} \cdot H_x + O_{ly} \cdot H_y \quad (5.4)$$

where H_x and H_y are the Entropy values associated respectively with the vertical and horizontal projections and O_{lx} , O_{ly} are the respective normalisation operands, expressing the density per unit of length, as defined below.

$$O_{lx} = \frac{N}{lx} \quad (5.5)$$

$$\text{and} \quad O_{ly} = \frac{N}{ly} \quad (5.6)$$

where lx is the length of the vertical projection – width of signature, and ly the length of the horizontal projection – height of signature, with blank spaces removed.

The complexity measure proposed, described by equation (5.4), solves the aforementioned problems. It should also be mentioned that the complexity measure,

by definition, is dependent on signature rotation, since a rotation of the signature image results in different histograms and thus different Entropy values. In this study the signature images are assumed with their original baseline (see Chapter 3). In addition, people’s judgements were also made on the original images. Thus, a later comparison between the two would be better justified. On the other hand, if the proposed complexity measure is to be used as part of a signature verification procedure, rotation normalization of the signature might be judged essential at a preprocessing stage. However, this could affect the quality of the signature image and, hence, its projections.

5.3.2 Analysis of Results

The proposed measure of complexity was applied to the same signature samples used in the human perception experiment. The complexity results generated algorithmically were shown to produce the same complexity ranking among the targets as the one determined by the human judgements. Table 5.3 displays the results along with the common generated ranking, with 1 corresponding to a simple signature and 5 to a complex one.

Table 5.3: Human and computer estimates of target signature complexity.

Target Signature	Complexity Estimates		
	humans	computer	ranking
Target 1	5.8	123.3	4
Target 2	4.8	66.7	3
Target 3	4.1	65.9	2
Target 4	8.2	149.2	5
Target 5	1.8	53.4	1

Regression analysis

A correlation test between the human and computer estimates of signature complexity resulted in a correlation coefficient, the *Pearson product moment coefficient of correlation*^[92], *r* equal to 0.914, which proves there to be a strong linear relationship between the two variables. A linear regression modeling of the data is displayed in Figure 5.7 together with the resulting values of the regression statistics.

The linear relation between human and computer signature complexity estimation is best described by equation: $y = 0.051x + 0.2648$. The independent variable x represents the computer estimation of signature complexity and the dependent variable y the human perception of signature complexity. Thus, the human perception of signature complexity may be predicted from the automatic measurements after they are applied in the fitted regression line equation.

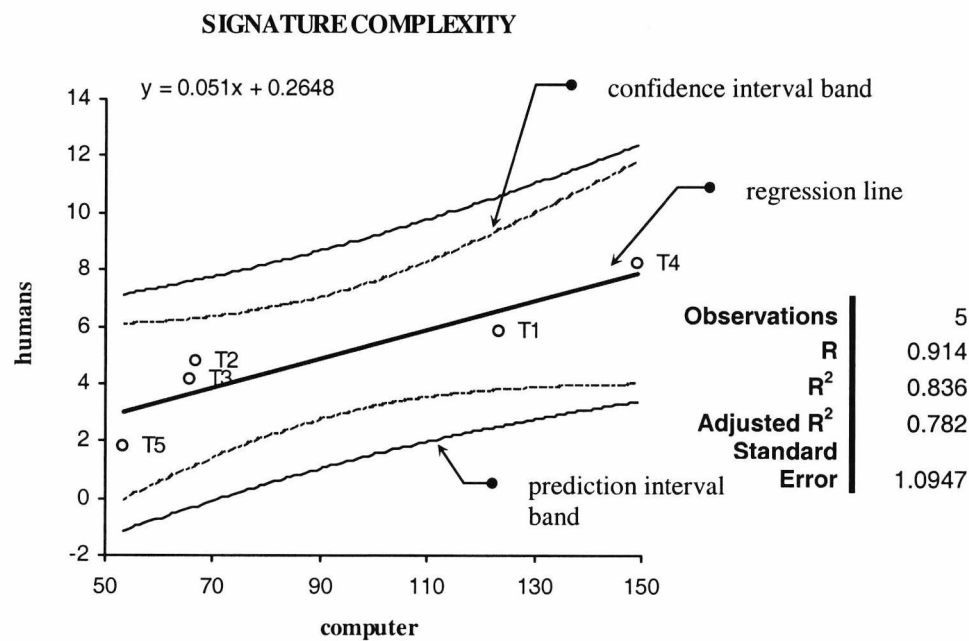


Figure 5.7: Linear regression between human and computer estimation of signature complexity.

The slope of the fitted regression line is equal to 0.051 and the intercept (point at which the line intercepts the y-axis) is 0.2648. The confidence interval band shows the range within which the population regression line is likely to lie with confidence level 95% and the prediction interval band gives the range within which future observations are likely to lie with the same confidence level.

The R^2 statistic, or coefficient of determination, is the proportion ($100R^2\%$) of the variation in the dependent variable y explained by the regression model. The closer the R^2 value is to 1.0 the better the fit of the regression line. An R^2 value of 0.836 shows the regression is a good fit. The adjusted R^2 takes into account the size of the

sample, usually for reasons of comparison with other models. A small number of observations will decrease the R^2 . The Standard Error of the Estimate is another measure of how well the model fits the data. Thus, a small deviation of the observations from the fitted regression line indicates that the line is a good fit.

The standardized residual plot of Figure 5.8(a) displays the standardized deviation of the observed y values (human perception) from the predicted line. The residuals have to be normally distributed for the regression assumptions to be met. Figure 5.8(b) shows that the residuals follow the normal curve; however, the results are poor due to the very small number of observations.

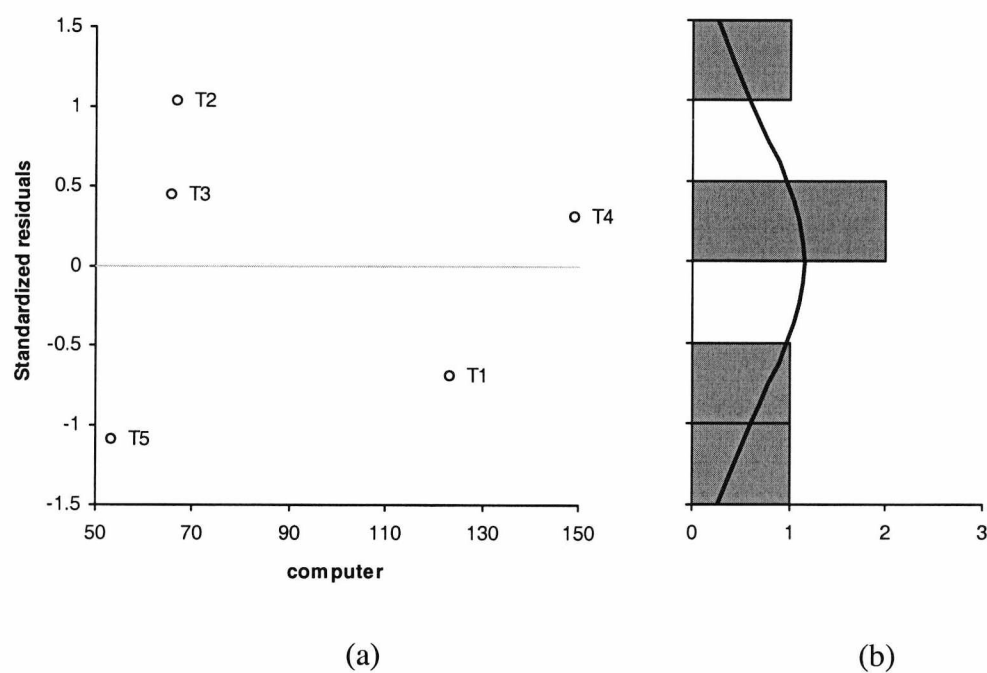


Figure 5.8: (a) Standardized residual plot, and (b) its frequency distribution.

The results of the linear regression have showed that the fitted regression line is a good approximation of the relationship of the two variables, allowing the prediction of signature complexity estimates as perceived by humans, with a good accuracy from the automated computer model.

Zusne [[154] p.277] states that judging complexity is a “psychophysical” function and that “physiological arousal and preference statements are dependent variables that are affected by past experience, set, meaning, and emotional significance, hence

cannot be expected to vary in a rectilinear fashion with measured complexity.” On the other hand, the linear regression is an acceptable approximation of the monotonic relation between the two variables – subjective and objective complexity.

5.4 Conclusions

The existent background relevant to signature complexity, as well as perceived complexity of objects-shapes was first briefly reviewed. The perceived complexity of handwritten signatures was approached through an experiment with human subjects and its results were extensively analysed and discussed, with the aid of the frequency distributions, the statistical values and the comparative box-plots. The normality of the distributions was evaluated visually (from the graphs) theoretically (from the box-plots) and pragmatically (from the results of objective normality tests). Suspect outliers in the distributions, were spotted and tested, and the appropriate decisions were justified. General perception remarks include the two major criteria that concentrate the human perception of signature complexity, the ‘readability’ or ‘intelligibility’ of a signature image and its ‘reproducibility’. Several approaches to automatic signature complexity estimation were suggested and discussed. A method was proposed to automatically measure the complexity of a signature, based on the notion of Entropy associated with the signature’s image projection histograms. The automatic signature complexity estimates produced proved to highly correlate with the human perceptual results. Finally, a linear regression between the human and the computer estimates of signature complexity arrived in an optimum regression equation that could predict future human responses on signature complexity from the automatically generated complexity values with a good accuracy.

Concluding, this chapter offered a genuine insight into human perception of signature complexity and proposed an algorithm for the quantitative measure of the complexity of a signature, from a static point of view, reflecting human-like perception. This is an important finding as prior research studies in this area [25;27;75] engaged in a definition and evaluation of the degree of a signature’s complexity make use of only dynamic signature information and, thus, reflect the difficulty of a signature to be imitated dynamically, while this study relies on only static information. The static

approach of the signature image as a ‘whole’, implemented here, engages a universal perception, as processing highly dependant on certain signature types is absent from this approach. Moreover, the participants in the experiment were of various nationalities, while the target signatures exhibit a range of different styles. The necessity of the static approach is also supported by the wider range of applications where a static signature image is encountered, as opposed to those where dynamic signature data is readily available.

Finally, many applications could benefit from the incorporation of human knowledge, which in many aspects is found to be superior against a judgement-free machine-based system. Investigation into human perceptual capabilities could offer some insight into areas where automatic signature verification systems fail. On the other hand, the development of automatic algorithms that reflect human-like perception could prove beneficial for both computer-based processing but also in enhancing human inspection procedures. Further research into this will be discussed in later chapters.

The next chapter investigates another important signature characteristic – the intrinsic variability of signatures – in a subjective and an objective approach, i.e. as perceived by humans and estimated by machines.

Chapter 6

Handwritten Signature Variability

This chapter is involved with the investigation of the intrinsic variability of handwritten signatures. The intra-class similarity of signatures will be analysed mainly in terms of a comparative relative consistency among the target signers, both in a subjective and an objective approach, decided by the variability among reference samples based entirely on the signatures' shape. A human perception experiment will be the source of the subjective signature variability results. For the objective measurement several methods will be considered and the one closest to the human perceptual judgements will be further evaluated, while its discrimination power will be comparatively assessed. Furthermore, issues related to signature normalisation and smearing application will also be discussed.

6.1 Introduction

A person's signature displays some inherent variability primarily as a result of the complex nature of the signing process, but also depending on the writing conditions, the physical and psychological state of the signer, the element of time, etc. The degree of signature variability typically encountered if standard conditions apply is a personal

characteristic and differs between signers. However, extreme conditions may introduce variability which is not a characteristic of the signer.

The calculation of the intrinsic variability of a person's signature is widely associated with the use of an Elastic Matching procedure. This is a function-based approach, which usually requires knowledge about the dynamics of the signature, and may be carried out by means of Dynamic Programming Matching techniques, between pairs of genuine signature samples. Among the great number of studies in this area, a few significant ones are briefly described here.

The intra-class dissimilarity between the reference signatures of a signer is calculated, in [25], with the help of a dissimilarity index computed after an Elastic Matching procedure between pairs of samples. This approach involves the calculation of the minimum accumulated distance resulting after optimally matching the points of one reference signature with the other. Dimauro et al. [45] and Congedo et al. [36] propose a measure of the local stability in on-line signatures. The local stability index is obtained from the frequency of *direct matching points* identified through an Elastic Matching procedure between a signature and each of the other reference signatures. In an off-line approach, after segmenting the signature into *arbitrarily-shaped primitives*, with the help of direction information, Sabourin et al. [124] proposed a static similarity measure using dynamic programming for the matching of a reference primitive set and a test primitive set, using *local interpretations of test primitives* and *relational similarity measures* between pairs of primitives [123]. Other studies that use dynamic programming techniques with offline data to measure the dissimilarity among reference signatures include [63;153], which are based on signature projections.

Finally, the most common approach encountered is a parameter-based one. The dissimilarity between reference samples, in a parametric approach, may be calculated as the distance between their parameter values from the mean values of the reference set (see, for instance, [11;12;97]). Moreover, the variability of a person's signature is reflected, in a way, by the mean and the standard deviation values of the feature vectors in a reference set.

6.2 Perceived Signature Variability

In the field of psychology, studies involved with the perceived similarity of shapes are reviewed by Zusne [[154] ch.6], most of which are concerned with the dimensionality of the psychological space involved in the judging process of shape similarity. However, due to the varying experimental procedures used by different investigators these studies “...yield little by way of any firm conclusions”. Nevertheless, several of these studies are worthy of a brief mention here.

In his study, Attneave [17] rejected the hypothesis that the psychological space involved in judging the similarity of shapes is Euclidean and supported a different hypothesis known as the “city-block” metric. In other words, the total psychological difference of two shapes, in a two-dimensional space, is not given by the Pythagorean theorem, but seems to approximate the sum of the differences on the individual dimensions. The shapes Attneave used in his experiments were parallelograms varying in size and angularity, squares varying in area and reflectance, and triangles varying in area and angularity. Concerning the relationship between the similarity judgments on the shapes and their physical variables, the author concluded that “*the judged difference between stimuli was approximately proportional to the difference between the logarithms of their physical values*”.

On the other hand, Stenson [137] found the use of a Euclidean distance model more appropriate in his study. Furthermore, the author suggested the use of 6 psychological dimensions and concluded that four main physical variables – complexity, curvature, curvature dispersion and straight-length dispersion – were important predictors of the psychological similarity of shapes.

In the field of signature verification specifically, the perceived variability between signature samples of a signer is an unvisited area upon which this work aims to cast some light. The following human perception experiment was carried out for the purpose of acquiring some knowledge about the perceived intra-class dissimilarity of genuine signatures.

6.2.1 Experimental Procedure

Fifty subjects, mainly university students, of different nationalities, participated in this human perception experiment. The subjects were presented with five genuine samples²⁰ for each of the five target signatures, displayed in Figure 6.1, and were asked to rank the target signatures according to their consistency level – with 1 corresponding to the most consistent target signature up to 5, which corresponds to the most inconsistent one. It was explained that a highly consistent target signature would be expected to have a high visual similarity among its 5 genuine samples and, conversely, a very inconsistent target should present a high dissimilarity between its five samples.

To assist the participants, directions were given for the search of the most consistent target signature first, then for the most inconsistent and lastly for a ranking of the remaining sample groups. It was assumed that requesting a perceived variability value for the target signatures would be too complex a task for the participants to carry out, and would most probably result in a significant degree of confusion. On the other hand, requesting only a relative ranking of the target signatures was considered to be a much simpler process. The duration of the test was about 3-4 minutes, varying slightly among individuals.

6.2.2 Analysis of Results

The perceived relative variability of each of the five target signatures is illustrated by the frequency graphs of Figure 6.2, where the frequency of occurrence of each relative variability ranking position, with respect to the other targets, is displayed.

²⁰ These are the first 5 samples produced by the target signers, and are displayed (Figure 6.1) in sequence from the first one (leftmost) to the fifth one (rightmost sample). Note that Target 1's third sample has a small interruption in its binary image. This is due to very light pressure applied by the signer at that point, which was recorded by the digitiser as zero-pressure.

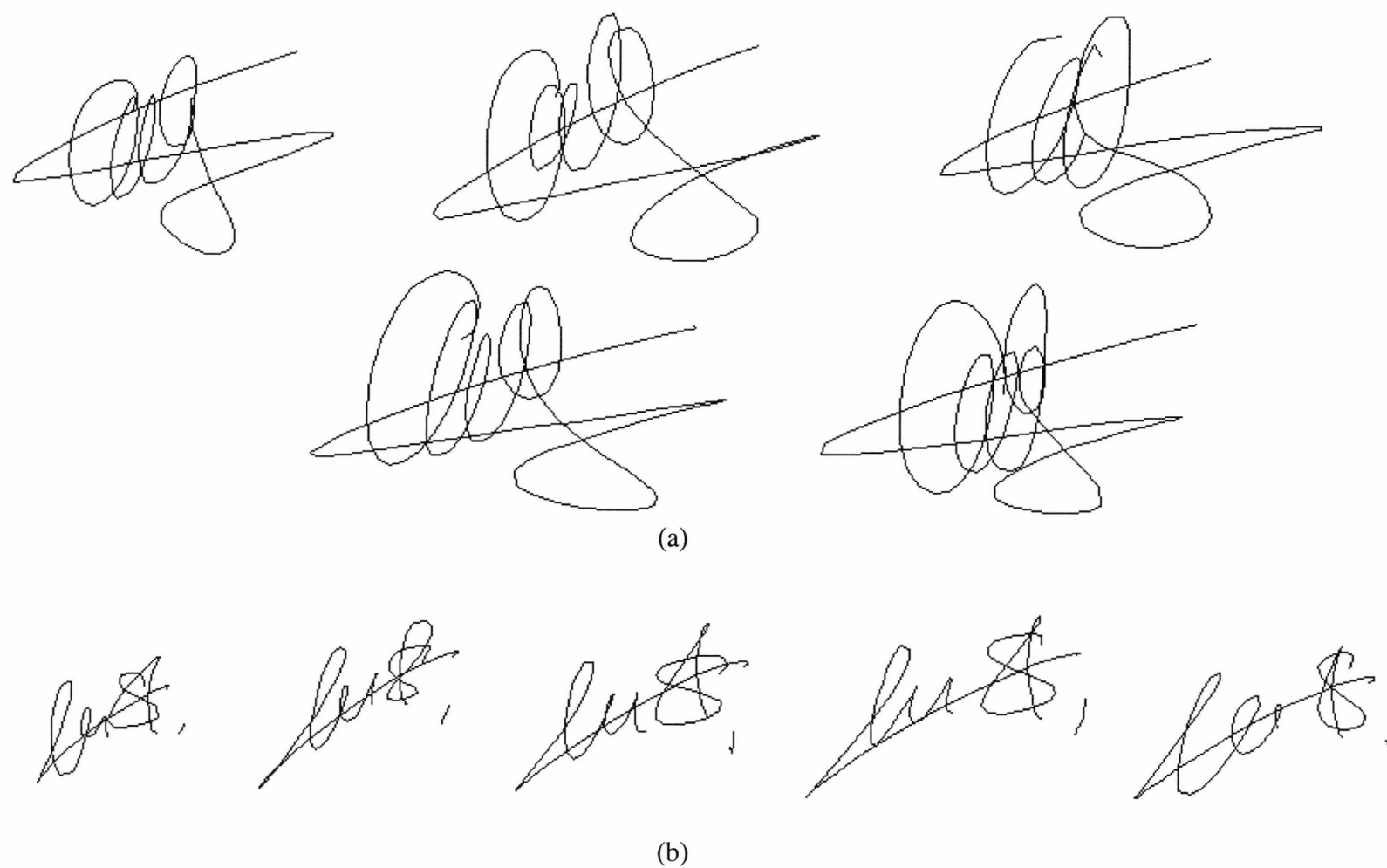
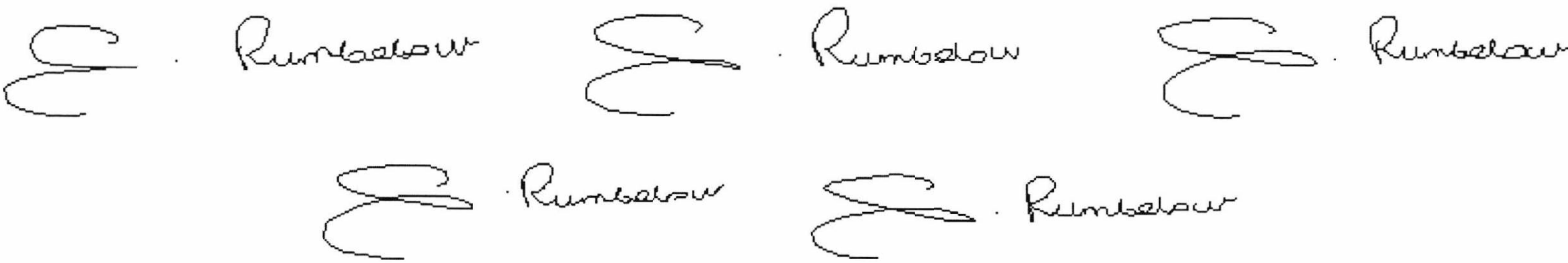
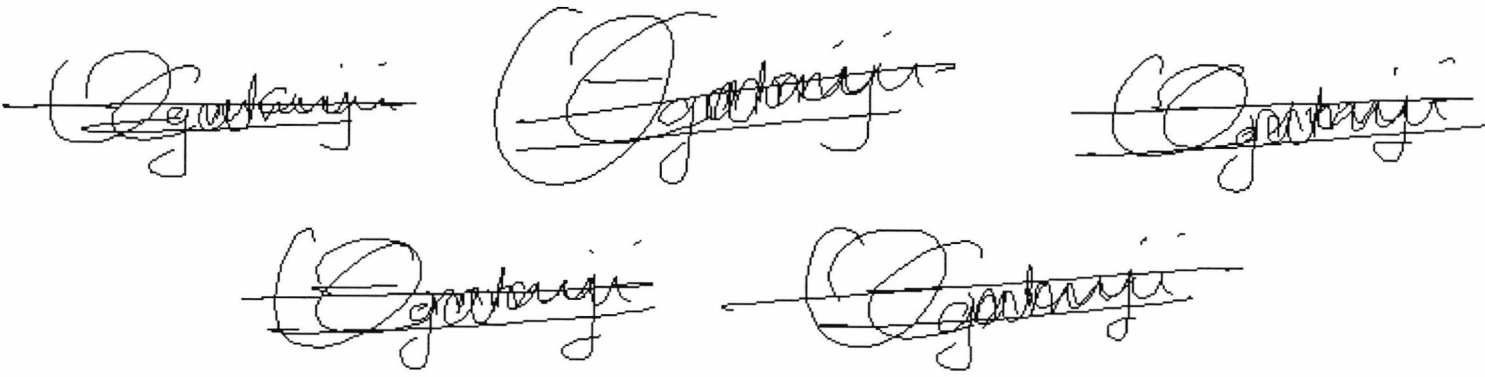


Figure 6.1: Genuine samples of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, (e) Target 5.



(c)



(d)



(e)

Figure 6.1: Genuine samples of the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4, (e) Target 5 (continued).

The frequency histograms reveal some important concepts. Targets 3 and 4, Figure 6.2(c) and (d), are generally judged as very consistent signatures, resulting in a relative variability ranking position of 1 and 2 respectively – a judgement embraced by a very large percentage of the participants (72% for Target 3 and 56% for Target 4). On the other hand, Targets 1, 2 and 5 are perceived to possess much higher relative variability ranking positions. However, these were more difficult to precisely estimate, and hence reveal a greater confusion and disagreement among the subjects. More specifically, Target 1, Figure 6.2(a), is assigned the 4th position in the variability ranking by 42% of the subjects, while Targets 2 and 5, Figure 6.2(b) and (e), reflect a much higher uncertainty, thus posing a difficulty in precisely estimating the relative ranking position of these targets. Hence, the task of ranking the five target signatures, according to their intrinsic variability, becomes more troublesome.

It is very important to note here that the nature of the variability ranking data, being measured on an ordinal scale rather than a continuous numeric scale, does not allow their evaluation with parametric statistics. Moreover, the application of normality tests would be inappropriate as the required assumption of treating observations of continuous variables is not met here. It is generally agreed that non-parametric statistics should be used with ordinal data. Non-parametric methods do not rely on the estimation of parameters (i.e. mean, standard deviation, etc.), which describe the distribution of the data. Moreover, they are known as distribution-free methods. However, the frequency distributions of the relative variability ranking data are presented here in order to obtain a much better view of the perceived variability ranking judgements, assisting their extensive evaluation.

Using the mean as a measure of central tendency and the standard deviation as a measure of variability in the case of ordinal data would be mathematically incorrect, since the intervals between the data points are not necessarily the same. The equivalent non-parametric measures are the median and the IQR (inter-quartile range), respectively. The non-parametric statistics of the variability ranking judgements for each of the five target signatures and for a sample size (N) of 50 subjects, are displayed in Table 6.1.

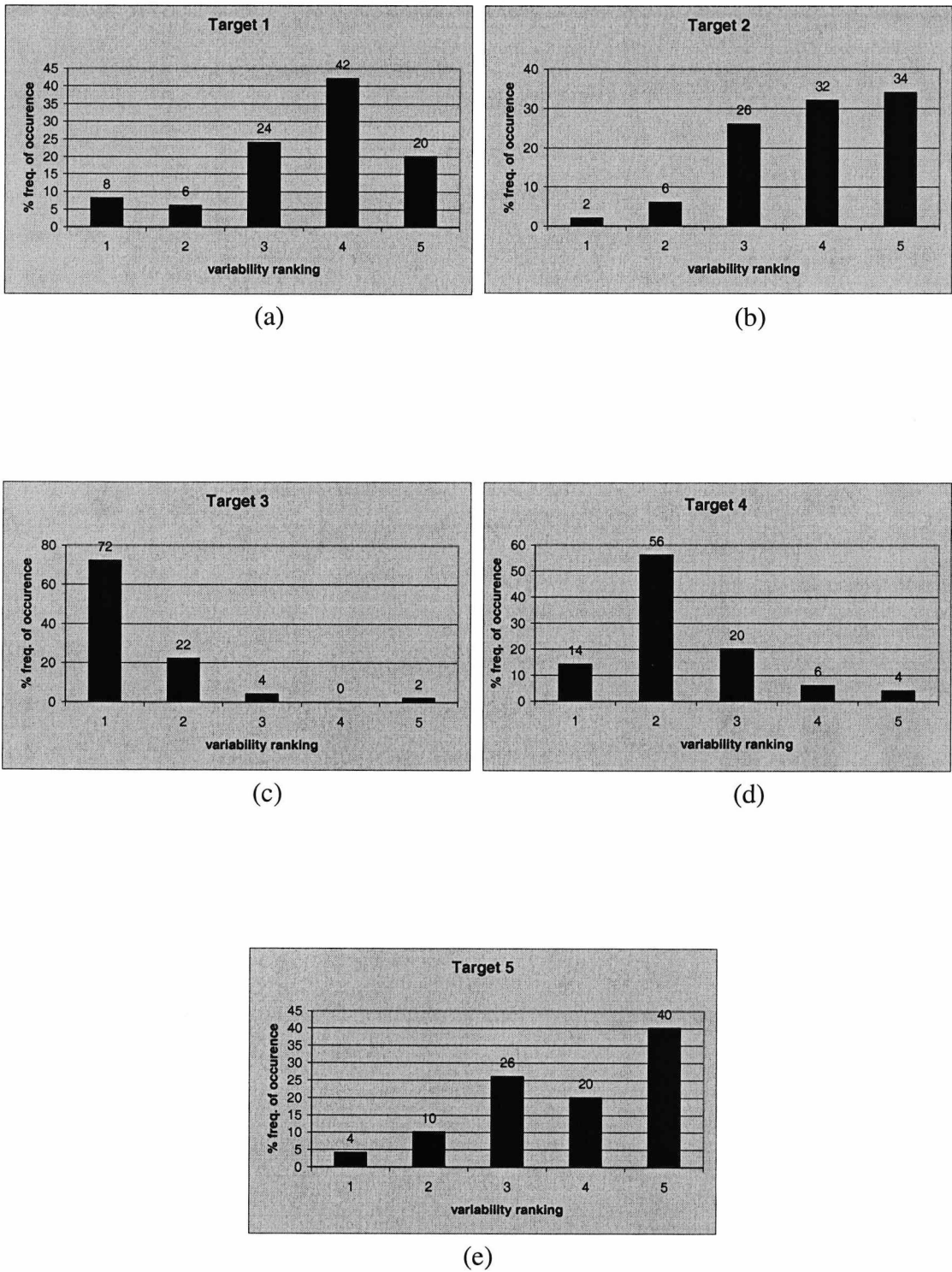



Figure 6.2: Frequency graphs of the perceived relative variability ranking position for the target signatures: (a) Target 1, (b) Target 2, (c) Target 3, (d) Target 4 and (e) Target 5.

Table 6.1: Relative variability ranking statistics for the target signatures.

Statistics	Target 1	Target 2	Target 3	Target 4	Target 5
N	50	50	50	50	50
Median	4.0	4.0	1.0	2.0	4.0
Mode	4	5	1	2	5
IQR	1.0	2.0	1.0	1.0	2.0

Indeed, Targets 1, 2 and 5 resulted in equal median values, enhancing the difficulty encountered in precisely estimating the exact ranking position for these targets. On the other hand, the mode statistic provides a better solution, expressing the most frequently encountered ranking judgement. According to their mode values, Targets 2 and 5 both fight for the 5th ranking position. However, the 5th ranking judgement has a greater frequency of occurrence for Target 5 (40%) than Target 2 (34%). Hence, ordering the ranking positions from 1 to 5, the following variability ranking order of the five target signatures may be assumed (Table 6.2).

Table 6.2: Variability ranking order (A) of the target signatures.

Ranking Order	Target Signature	Variability
1	Target 3	<i>low</i>
2	Target 4	
3	Target 1	
4	Target 2	
5	Target 5	<i>high</i>

Nevertheless, considering the unclear frequency distributions of the rather inconsistent Targets 2 and 5, as well as their wide-range spread – also denoted by the high IQR values, may without a doubt question the formerly suggested variability ranking. Having the two first ranking positions (assumed by Targets 3 and 4) almost unanimously defined, with the middle and upper variability ranking positions being far more vague, it may be concluded that the ranking of consistent signatures, rather than inconsistent ones, seems to be a much easier task for humans, who show a high degree of confusion and disagreement when the variability of signatures increases.

In an attempt to analyse the results a little further, the box-plots of Figure 6.3 are displayed, providing a visual comparison between the perceived variability rankings of the five target signatures. The asymmetry of the distributions is emphasised by the form of the box-plots. The median in Target 1 falls upon the upper quartile, and in Targets 3 and 4 upon their lower quartiles, while all three targets have an IQR as small as 1.0. For Targets 2, 3 and 4 the median, with a 95% confidence, is precisely 4.0, 1.0 and 2.0 respectively, whereas for Target 1 with the same confidence level the interval of the median is between 3.0 and 4.0, and for Target 5 between 3.0 and 5.0 – its entire IQR. Note that the representation of the parametric statistics is not included here, for reasons previously analysed, but only non-parametric statistic results are displayed in the box-plots.

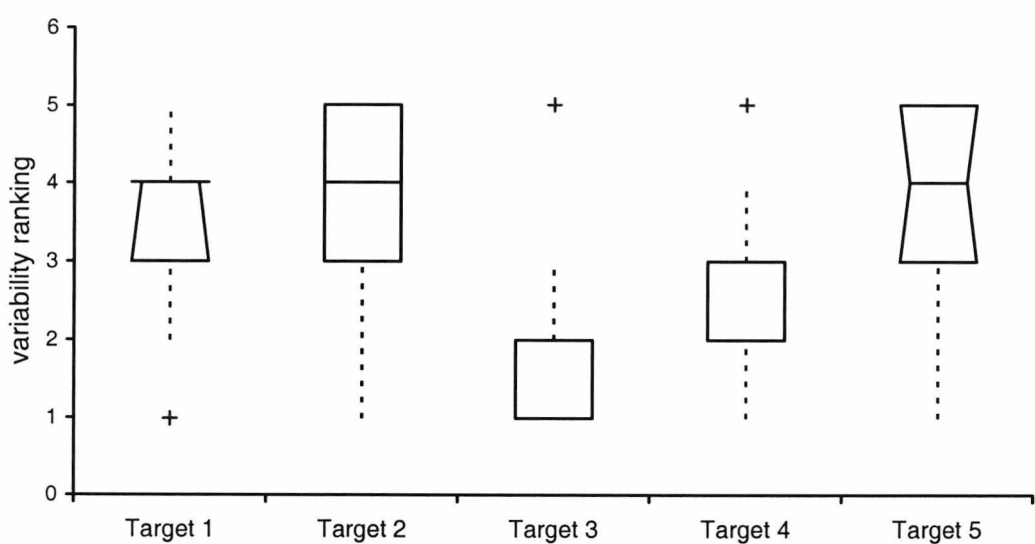



Figure 6.3: Variability ranking box-plots for the target signatures.

A closer look at the box-plots reveals that Targets 1, 3 and 4 appear to have possible outliers. The existence of outliers usually has little effect on the median or IQR values. The new statistics of the relative variability rankings were calculated excluding the outliers of Targets 1, 3 and 4. Indeed, discarding the outliers had no effect on the median and IQR values, although, for Target 1 the median with a 95% confidence is now precisely 4.0, and not in the interval between 3.0 and 4.0 as it previously was. Finally, the outliers were not discarded, due to the nature of the data, the small extent of the ordinal scale, and the unpredictability of the human factor.

It should be mentioned that the overall ranking order most frequently encountered in the subjects’ responses, from the most consistent to the most variable target, was actually the one earlier suggested (A) – resulting from the evaluation of the individual most frequent ranking positions. A second most popular ranking order was found to be the following (Table 6.3).

Table 6.3: Variability ranking order (B) of the target signatures.

Ranking Order	Target Signature	Variability
1	Target 3	<i>low</i>
2	Target 4	
3	Target 5	
4	Target 1	
5	Target 2	<i>high</i>

Although, these two rankings (designated (A), (B) respectively), may be possible outcomes, with ranking (A) being slightly more probable than ranking (B), the perceived variability ranking results, however, are very unclear at the level of inconsistent signatures, for an overall variability ranking to be proposed with confidence. Nevertheless, it may be concluded that the intrinsic variability of consistent signatures, in general, is more easily perceived by humans, than the variability of highly inconsistent ones.

Furthermore, the following signature pairs: Targets 3 and 4, and Targets 2 and 5 display, on the one hand, close variability judgements, but on the other hand, resulted in considerably different complexity judgements, as analysed in Chapter 5. This suggests that the perceived complexity of a signature does not necessarily influence its judged variability. However, if the complexity of the target signatures did not affect the majority of the judgments, a small number of highly deviant estimations, including the outliers, may – perhaps – be accounted for by the targets’ degree of complexity.

6.3 Objective Signature Variability

In this study, for the measurement of the inherent variability between genuine signature samples of a signer, a heuristic approach making use of the signatures' shape was found most appropriate.

A survey of a number of shape analysis techniques used for shape representation, description and matching, exists in [85]. These are divided into boundary and global approaches, using scalar transform or space-domain techniques. The global scalar transform techniques include the use of moments, morphological methods and Shape Matrices. The latter shape descriptor is information preserving and has been tested on objects, geometrical figures and alphabetical characters. Moreover, similarity measures between two shape matrices have been proposed in [57;140].


The use of Shape Matrices for the verification of handwritten signatures was examined by Sabourin [121]. An improved similarity measure between two Shape Matrices was proposed, more suitable for line images as it cancels the bias that the large amount of background pixels produce. The effect of some preprocessing algorithms, that emphasize local signature characteristics and human perceptual features, was evaluated. These included loop filling, smearing, smearing with loop filling, and morphological closing with loop filling, and it was shown that they had little effect on the performance of the verification system with the new similarity measure, although smearing produced the lowest error rate.

6.3.1 Methods Employed

Four different groups of methods were considered in this study for the estimation of the dissimilarity between genuine signatures. The dissimilarity between pairs of genuine samples was measured using the Elastic Matching procedure described in [25], Entropy-based variation measures for character image data as proposed and surveyed in [73], grid-related variations measured between the signature images, and Shape Matrices similarity measures as proposed in [57;121;140].

The Average Entropy, the Distribution Entropy and the Average Entropy Difference were assessed in measuring the degree of variation in a datum comprising the five genuine samples for each of the five target signatures. On the basis of the work reported in [73], the preferred Entropy-based variation measure was the Average Entropy Difference, since it satisfies all of the following criteria: *boundedness*, *independency*, *monotonicity* and *constancy*. Working with the complicated images of handwritten signatures, rather than simple characters, poses a difficulty in perfectly aligning them within the datum, and this may considerably affect the performance of the method. Although it was anticipated that this variation measure would be independent of the shape complexity (according to the *monotonicity* criterion), the resulting variability ranking (C) of the five target signatures (Table 6.4) suggests a dependence on their degree of complexity, at least with respect to the complexity measurement proposed in Chapter 5. Indeed the resulting variability values showed a correlation (*Pearson correlation coefficient* $r=0.71$) with the objective complexity of the target signatures and a rather strong correlation ($r=0.84$) with their subjective complexity.

Table 6.4: Variability ranking order (C) of the target signatures.

Ranking Order	Target Signature	Variability
1	Target 5	low
2	Target 2	
3	Target 1	
4	Target 3	
5	Target 4	high

It is interesting to note that the same variability ranking order (C) of the target signatures was produced with the Elastic Matching procedure, performed as in [25]. This is not surprising as it is generally expected for the Elastic Matching method to result in a higher degree of dissimilarity for signatures of higher complexity, since a more complex signature would need more iterations to be made in order to match one of its samples with another.

Furthermore, a grid-related approach was considered, where variations were calculated for pairs of genuine signature images divided into segments of 2x2 or 3x3

pixel areas, based on the local differences in the number of foreground pixels contained in the corresponding pixel boxes. The accumulated distance was normalised by the size of the segments used and the number of pixel boxes, having number of foreground pixels >0 , contained in the reference image. The variability rankings obtained for different grid sizes were different, which delivered a lack of reliability in this approach.

Finally, similarity measures between pairs of Shape Matrices were analysed using the measures proposed by Sabourin [121], Taza and Suen [140], and Goshtasby [57]. The weighted similarity measure proposed by Sabourin, which is an improved version of the similarity measures of [140] and [57] – for application on line images – was preferred for reasons that will be later explained. Contrary to the previous variability methods discussed, the variability ranking of the five target signatures obtained with the Shape Matrices was close to the perceived variability judgements, and hence, this method was selected for measuring the similarity between genuine samples. In addition, the results also suggest that the dissimilarity measure does not necessarily depend on the complexity of signatures, and this is an approach that this study would like to consider.

Shape Matrices - SMs

The $m \times n$ Shape Matrix M of a shape is acquired from a polar quantization of the digitised shape (Figure 6.4), as described in [57;140]. According to the definitions in these studies, with O being the shape's centre of gravity and OA its maximum radius r , the radius OA is divided into $(n-1)$ equal parts; and the concentric circles with centre O and radii $r/(n-1)$, $2r/(n-1)$, \dots , $(n-1)r/(n-1)$, are further

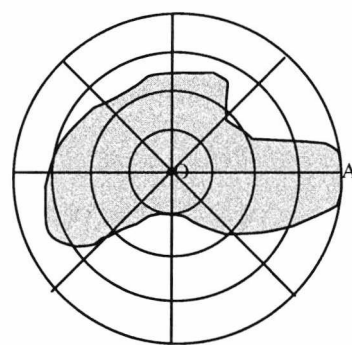


Figure 6.4: Polar quantization of shape.

divided, moving counter-clockwise, into m equal arcs of $\theta=360/m$ degrees. The shape matrix is constructed by sampling the binary value of the shape at the points where the circles (matrix columns) intersect the radial lines (matrix rows). The value of the shape matrix $M[i, j]=1$ if the sampling point of polar coordinates $(\rho, \theta) = (ir/(n-1), j(360/m))$ is a foreground pixel (black pixel), otherwise $M[i, j]=0$ corresponding to a

background pixel (white pixel). The values of m and n may be chosen within the limits $m \leq 2\pi r$ and $n \leq r$, otherwise over-sampling occurs.

Taza and Suen [140] have proposed a weighted similarity measure between two shape matrices taking into consideration the redundancy that is embedded in the sampling of the inner circles as opposed to the outer ones. The redundancy of a circle is said to be inversely proportional to the radius of the circle and is defined as:

$$R_i = \frac{m}{\text{circumference_of_circle_i}} \quad (6.1)$$

and the weight for column $[i]$ is defined as:
$$W_i = \frac{1}{R_i} \quad (6.2)$$

with $W_0 = \frac{n-1}{m}$, for column $[i=0]$.

The similarity measure between two shape matrices as proposed by Sabourin [121] for the handwritten signatures, is defined as:

$$\text{Similarity} = \frac{wcp}{wcp + wnc} \quad (6.3)$$

with $0 \leq \text{Similarity} \leq 1$, and where wcp is the sum of weighted number of corresponding black sampling points in each column between the two SMs, wnc is the sum of weighted number of non-corresponding sampling points in each column between the two SMs. Note that the weights for each column are measured according to equation (6.2).

Implementation

According to the signature database of this study, the dimensions and the quality of the digitised images the following sampling values were chosen as most appropriate. A maximum radius fixed to $r=200$ to ensure that the outmost circle would include all of the signature images, considering maximum signature image dimensions of 150×270 pixels. The number of sample points on r was set to $n=101$ to obtain a sampling unit of $r/(n-1)=2$ pixels, and the number of sample points on the maximum radius circle was taken as $m=628$, half the number of its total perimeter points ($2\pi r$).

Note that a selection of different values would change the amount of sampling imposed, although, a small change would produce similar results.

Similarity measures between pairs of reference signatures were obtained from equation (6.3). The length of the radius r was kept to the fixed value mentioned above, thus scale invariance – which would be obtained if r was adjusted to the maximum radius of each shape – was not considered. Besides, Sabourin [121] reported better results without scale invariance. On the other hand, the choice of a fixed r value can have other implications such as some signatures being more over-sampled than others, if a considerable size difference between the signatures exists. However, the over-sampling effect is removed by having a weighted similarity measure. In addition, implications due to the effect of the background in the previous case are also cancelled out by using the improved similarity measure (6.3) proposed by Sabourin.

Translation invariance was obtained by arranging O to the centre of gravity of the signature, using the following equations:

$$\begin{aligned} \text{Vertical centre of gravity:} \quad VCG &= \frac{\sum_i i \cdot P_h(i)}{\sum_i P_h(i)} \quad (6.4) \end{aligned}$$

$$\begin{aligned} \text{Horizontal centre of gravity:} \quad HCG &= \frac{\sum_j j \cdot P_v(j)}{\sum_j P_v(j)} \quad (6.5) \end{aligned}$$

where $P_h(j)$ and $P_v(i)$ are the horizontal and vertical projections²¹ of the signature image.

Rotation invariance can be obtained by aligning the OA to the baseline of the signature defined by an angle of φ degrees measured from the horizontal axis.

$$\begin{aligned} \text{Baseline angle:} \quad \varphi &= \tan^{-1} \frac{(VCG_{right} - VCG_{left})}{(HCG_{right} - HCG_{left})} \quad (6.6) \end{aligned}$$

where VCG_{right} and VCG_{left} are the vertical centres of gravity for the right part and the left part of the signature image respectively. Similarly, HCG_{right} and HCG_{left} are the

²¹ For details on the horizontal and vertical signature projections, see Section 5.3.1.

horizontal centres of gravity for the right and the left part of the signature image respectively. These are calculated from equations (6.4) and (6.5) applied in turns to both halves, right and left part, of the signature image.

Methods based on the maximum horizontal projection for the estimation of the baseline angle of signatures, such as those in [14;40;117], would be best applied to handwriting-like signatures in the form of a handwritten name (American signing style), since the maximum concentration of pixels, in this case, occurs along the baseline. However, when dealing with signatures that entail shape-like characteristics in the form of curves and loops – mostly encountered in European signatures – the baseline might not be described by points of maximum projection, and hence these methods would in most cases fail to detect a correct baseline angle. For this type of signatures, equation (6.6) may give better results. An example of correct baseline detection in this case is illustrated by Figure 6.5(b), where the baseline is denoted by the dotted line. However, signatures that contain strokes well above or below the main body of the signature might again fail a correct baseline detection. Baseline detection method (6.6) would also give good results for handwritten signatures written at an angle (see for example Figure 6.5(a)), since it is not affected by lines running through the body of the signature, something that the maximum horizontal projection methods are highly dependent on. The five target signatures used in this study include most of these different signature styles and equation (6.6) has given better results, although, still failing to correctly detect the baseline in some samples of Targets 1 and 5 (see, for instance, Figure 6.5(c)).

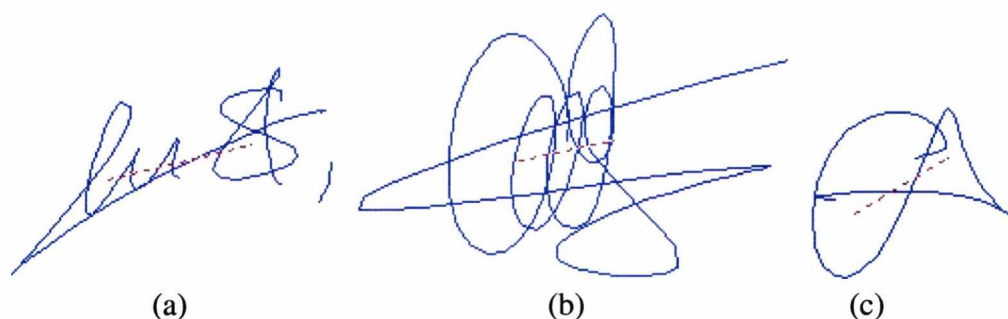


Figure 6.5: Baseline detection for (a) Target 2 (Sample 4), (b) Target 1 (Sample 5), and (c) Target 5 (Sample 2).

Furthermore, Table 6.5 shows the resulting baseline angle measurements, expressed in degrees, for the five samples of each of the five target signatures. The measurements that appear in bold correspond to baseline angles erroneously identified, whereas the rest of the measurements appear to be, at least, very nearly correct. A justification of this may be provided by a simple visual comparison against the sample images of Figure 6.1.

Table 6.5: Baseline angle measurements (in degrees), for the five samples of the target signatures.

Targets	Samples				
	1	2	3	4	5
T1	3.0	-2.4	-7.2	1.7	12.0
T2	18.9	32.5	16.6	14.2	21.5
T3	4.4	4.8	4.8	3.1	2.3
T4	1.5	1.2	-1.3	-2.4	1.2
T5	3.6	31.3	32.6	21.8	15.9

The objective variability rankings were obtained according to the following process. For each of the target signatures, each of the five reference samples was compared, in turn, with the four remaining reference samples of that target signature, and a similarity value was produced for each pair. The similarity between each reference sample and the rest of the reference set was calculated as the average of the pair-similarity values. The reference sample that produced the highest average similarity value for a target signature was chosen to represent the degree of intra-class similarity of the target's reference set. The overall variability ranking of the target signatures was formed by sorting the intra-class similarity values from high to low, thus producing the converse order that would appear in a variability rank.

The discrimination power of a method was assessed by calculating the similarity between the representative sample of a target signature with random samples from the other targets. A threshold²² equal to the smallest similarity value occurring within the pairs of the representative sample compared with the rest of the reference set, was selected to decide whether the resulting inter-class similarity was smaller than the

²² The selection of a more enhanced threshold is beyond the scope of this Chapter. Such a simple selection procedure of the representative reference signature and the threshold value is very common in the literature, as found in [12;63;74].

intra-class of the target (see for example Table 6.6 and 6.7). Thus, the higher the distance between the two the higher the discrimination power of the method would be.

Among the four Shape Matrix similarity measures compared, the improved weighted similarity measure proposed by Sabourin (6.3) for the handwritten signatures, showed indeed the best discrimination among signature samples of different targets. Although, the weighted similarity measure of Taza and Suen [140] for planar shapes, and the improved similarity measure of Sabourin without weights [121], produced the same variability ranking for the target signatures with that of (6.3), their discrimination power, however, was poorer (see Table C.2a, b and C.3a, b –Appendix C). The similarity measure for planar shapes initially proposed by Goshtasby [57], even though it gave high similarity values (close to 1) for the reference signatures, showed a discrimination power, however, which was very poor (see Table C.1a, b –Appendix C). Note that the range of the similarity values obtained through the similarity method [140] is $[-\infty, 1]$, whereas for the rest of the methods is $[0, 1]$.

6.3.2 Analysis of Results

The resulting similarity values based on equation (6.3) for the five target signatures are displayed in Table 6.6. The target signature is displayed on the left top corner as T1, T2, T3, T4 and T5 corresponding respectively to Targets 1, 2, 3, 4 and 5. The target signature reference samples are denoted as RS. The representative sample is highlighted and the threshold value appears in bold, selected as previously described. The representative sample of each target signature is then compared against random samples (TS)²³ from the rest of the targets. An 'X' is placed where the resulting inter-class similarity is greater than or equal to the threshold. An example of the inter-class similarity values obtained from comparisons with a few test samples is provided by Table 6.7. These serve as an indication of the range of the similarity values and the discrimination power of the method. However, a more extensive evaluation of the verification performance of this method is beyond the scope of this chapter.

²³ The same random samples were used to test the discrimination power of the four similarity methods.

Table 6.6: Intra-class similarity of target signatures.

T1	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.02	0.06	0.05	0.04	0.04
RS2	0.02	1	0.04	0.04	0.04	0.04
RS3	0.06	0.04	1	0.04	0.03	0.04
RS4	0.05	0.04	0.04	1	0.03	0.04
RS5	0.04	0.04	0.03	0.03	1	0.04

T2	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.12	0.05	0.03	0.02	0.05
RS2	0.12	1	0.06	0.04	0.03	0.06
RS3	0.05	0.06	1	0.06	0.05	0.05
RS4	0.03	0.04	0.06	1	0.05	0.05
RS5	0.02	0.03	0.05	0.05	1	0.04

T3	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.04	0.05	0.05	0.06	0.05
RS2	0.04	1	0.1	0.09	0.11	0.09
RS3	0.05	0.1	1	0.11	0.08	0.09
RS4	0.05	0.09	0.11	1	0.14	0.1
RS5	0.06	0.11	0.08	0.14	1	0.1

T4	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.11	0.11	0.13	0.1	0.11
RS2	0.11	1	0.07	0.1	0.09	0.09
RS3	0.11	0.07	1	0.1	0.11	0.1
RS4	0.13	0.1	0.1	1	0.13	0.12
RS5	0.1	0.09	0.11	0.13	1	0.11


T5	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.02	0.03	0.02	0.03
RS2	0.03	1	0.01	0.02	0.01	0.02
RS3	0.02	0.01	1	0.04	0.06	0.03
RS4	0.03	0.02	0.04	1	0.01	0.03
RS5	0.02	0.01	0.06	0.01	1	0.03

Table 6.7: Inter-class similarity of target signatures.

Target	RS	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8
T1	RS3	0.04	0.03	0.05	0.02	0.04	0.02	0.04	0.02
		X	X	X		X		X	
T2	RS2	0.05	0.05	0.02	0.02	0.03	0.05	0.02	0.02
		X	X			X	X		
T3	RS5	0.02	0.02	0.05	0.03	0.02	0.01	0.03	0.02
T4	RS4	0.05	0.04	0.04	0.03	0.04	0.06	0.03	0.04
T5	RS1	0.01	0.02	0.01	0.01	0.03	0.03	0.01	0.01
			X			X	X		

The inter-class similarity values show that Targets 3 and 4 had no inter-personal samples surpass their assigned threshold, which may be partly explained by their high intra-class similarity. The intra-class similarity of each target signature, as previously described, is taken as the average similarity between the representative sample and the rest of the reference samples for the particular target. Hence, the resulting objective variability ranking for the five target signatures, from low variability (or high similarity) up to high variability (or low similarity) is formed (Table 6.8).

Table 6.8: Variability ranking order (D) of the target signatures.

Ranking Order	Target Signature	Variability
1	Target 4	<i>low</i>
2	Target 3	
3	Target 2	
4	Target 1	
5	Target 5	<i>high</i>

Although translation invariance is essential for the correct comparison between the Shape Matrices of two polar sampled signature images, so that only corresponding radial lines and circles are compared, rotation invariance was not found essential for reasons explained in Chapter 3. However, the effect of rotation invariance was tested according to the description given in the implementation stage. The application of orientation adjustment produced a small effect in the variability ranking, as Targets 1 and 2 resulted in the same intra-class similarity value, and a slight change in the intra-class similarity values of the targets (see Table C.4a –Appendix C). Furthermore, the inter-class similarity results (see Table C.4b –Appendix C) showed an improvement in the discrimination power of the method, concerning Targets 1 and 2, but a negative discrimination performance regarding Target 5, as almost all the test samples passed its intra-class similarity threshold. The improvement observed for Targets 1 and 2 was because of the high inherent variability of their baseline orientation (see Table 6.5). On the other hand, the deterioration in the discrimination performance regarding Target 5 was due to the limitations of the baseline detection algorithm reflected in the shape-like signature samples, as it is based on the difference of the centres of gravity in the left and right parts of the signature image, rather than the main body of the signature. As reported earlier, Target 5 had the most false baseline detections in its

samples (see Table 6.5). Mainly, for this reason and for further reasons explained in Chapter 3, assuming the original orientations of the signature images was, finally, preferred.

As previously discussed, scale invariance was not considered in the implementation of the similarity measure. This is mainly because of the assumption that humans may consider size difference between signature images as a reason for variation, thus reflecting this in their variability judgments. Nevertheless, according to [121], better verification performance was achieved without scale invariance, thus when the radius r was assigned to a fixed value, rather than when adjusted to the maximum radius of each signature shape. Finally, applying orientation or scale adjustments may contradict with the way humans perceive and judge the variability of signatures.

It may be observed (Table 6.6) that all similarity values are very small, in fact closer to 0 than anything near 1, something that shows dissimilarity rather than similarity. The reason for these small values is the nature of the signature image, made up of very fine lines, rather than printed characters or compact shapes, for which this method is more appropriate, despite Sabourin's enhanced measurement for line images. A way to increase these similarity values would be to intensify the signature shape. Smearing²⁴ is a way to strengthen or wipe out the signature's fine details, depending on the degree of smearing applied, adding some more volume to the image. Sabourin [121] experimented using smearing with the similarity measure, with a threshold set to 30 pixels, and reported a better system performance with lower error rates.

The selected threshold for smearing is highly dependent on the size of the images in the database. Consequently, for a larger signature image a higher degree of smearing (higher threshold) is necessary to create the same effect that a smaller degree of smearing would require to suit a signature of a smaller size.

²⁴ Horizontal smearing is performed as follows. For every horizontal line of the signature image, the smearing algorithm checks if the number of white pixels found between two consecutive black pixels, of the same line, is smaller or equal to a threshold value, then the white pixels are turned into black, otherwise they are left untouched.

In this study, different degrees of smearing, with thresholds of 5, 10, 15, 20, 25 and 30 pixels, are applied to the signature images, and the resulting similarity values are displayed in Table 6.9. The similarity values obtained without the application of any smearing are also reported.

Table 6.9: Intra-class similarity values for different degrees of smearing.

	No	Smearing					
	Smearing	5	10	15	20	25	30
Target 1	0.04	0.07	0.19	0.29	0.43	0.49	0.56
Target 2	0.06	0.13	0.22	0.37	0.42	0.46	0.47
Target 3	0.1	0.19	0.35	0.42	0.44	0.45	0.39
Target 4	0.12	0.24	0.41	0.4	0.44	0.47	0.52
Target 5	0.03	0.05	0.08	0.17	0.27	0.26	0.27

It is obvious that the application of smearing, even with a threshold value as small as 5 pixels, significantly raised the similarity values, at first almost equally among the five targets. For a smearing of up to 10-pixel threshold the initial variability ranking (D), where no smearing was applied, is still preserved. However, even in this stage of a 10-pixel threshold the similarity values have already started to increase with a different rate of change among the targets (see Figure 6.6). Any further increase of the smearing threshold affects the variability ranking of the five targets, and for a threshold greater than 20 pixels the variability ranking is very differently rearranged. Moreover, the representative reference samples of the targets also change.

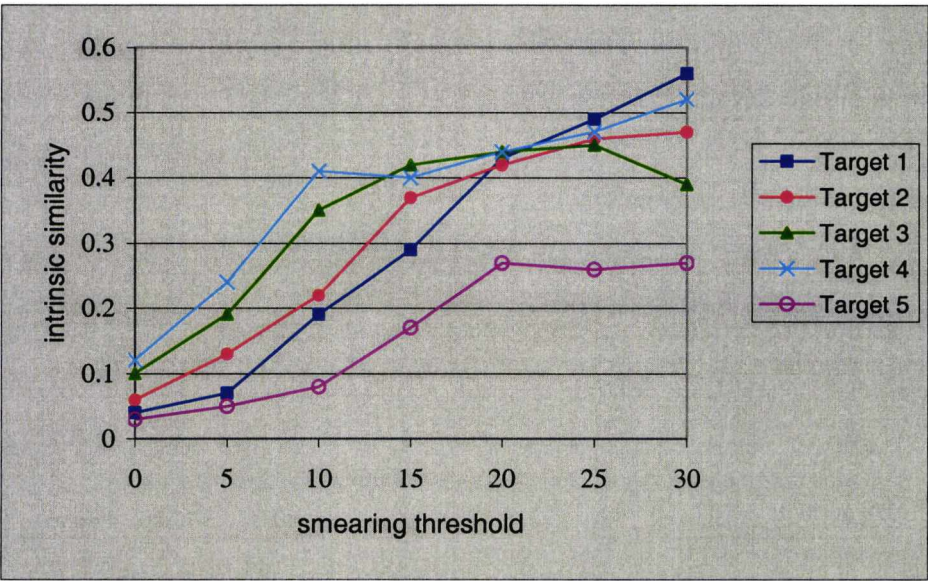


Figure 6.6: Effect of smearing on the intrinsic similarity values.

In fact, this odd change in the targets intra-class similarity increase rates, is due to the great difference in the individual signing styles. The five target signatures are indeed very different, ranging from a handwriting-like signature (Target 3), to shape-like signatures (Targets 1 and 5) made up of several loops. A high degree of smearing seems to favour some signatures over some others. For instance, Target 1 (Figure 6.6) seems to be favoured by the further increase of the smearing threshold, but that is not the case for Target 3. This can be better explained by reference to Figure 6.7. Because Target 1 is made up of big and wide closed loops the effect of smearing does not substantially alter appearance, until the smearing threshold is large enough to accommodate the interior loop distances. On the other hand, Target 3, because it consists of small and narrow closed loops, is immediately changed by the effect of smearing, not leaving any space for further filling when larger degrees of smearing are applied.

Therefore, the closeness and the wideness of the loops, together with the size of the signature, are major factors that determine the degree to which the applied smearing will have an effect. Hence, care should be taken when large amount of smearing is applied as the effect on different styles of signatures is shown to be very different.

A small degree of smearing may be generally found beneficial. In this experiment, smearing of a 5-pixel threshold seems to be advantageous to all the five target signatures, increasing the intrinsic similarity values almost equally among them. In addition, it enhances some of the fine details of the signature line – usually angles, crossing points and very narrow loops – perhaps points of concentration of the human eye. It is important to note that such a small degree of smearing does not alter the signature image nor the variability ranking order. A threshold of 10-pixel might at first appear beneficial, but the effect on the signature image (see Figure 6.7) is significant, and would perhaps change human perception regarding the respective signatures. Finally, the strength of the intrinsic similarity values is not the main concern of this study, where the relative intra-class variability of the target signatures is assessed and thus the produced ranking is of major importance.

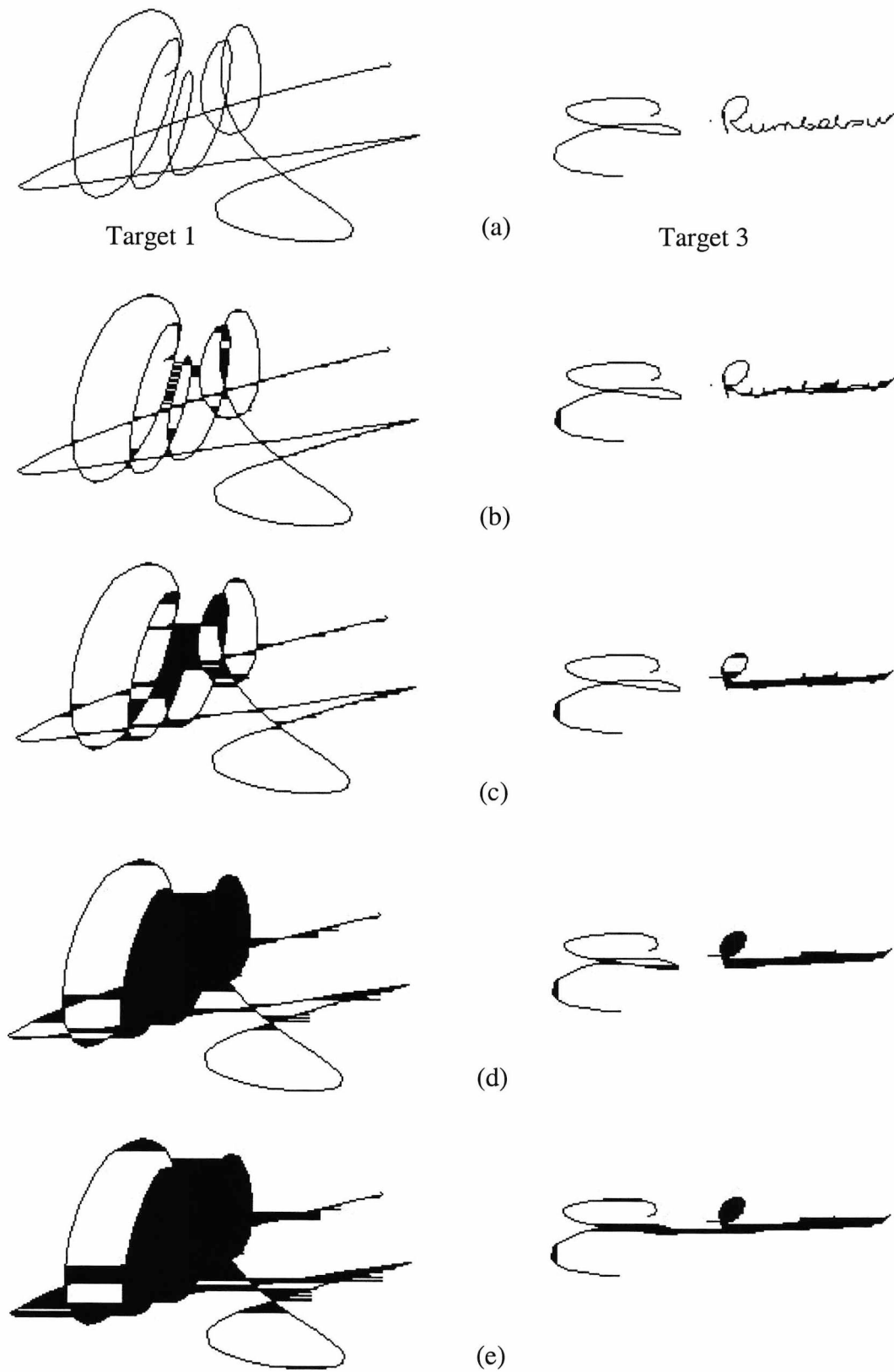


Figure 6.7: Effect of smearing on original signatures (a), for thresholds of (b) 5, (c) 10, (d) 20 and (e) 30 pixels.

Correlations

Assuming the variability ranking (A) from the perceived relative target variability results, a correlation with the ranking produced from the SMs' similarity measures, resulted in the following (Table 6.10). The *Spearman rank correlation coefficient* r_s ²⁵, suitable for correlations between ordinal data, proved there to be a high correlation ($r_s=0.8$) between the subjective and objective relative rankings produced. Application of small degrees of smearing gave the same correlation value. However, for smearing thresholds of 15 and 20 pixels, the correlation between subjective and objective ranking results is significant ($r_s=0.9$), showing that such a degree of smearing brings the objective results closer to human perception. On the other hand, much higher smearing threshold values produce the opposite effect ($r_s=0.3$).

Furthermore, correlations between the objective similarity results, including different degrees of smearing, and the subjective and objective complexity results of Chapter 5, are presented (Table 6.11). The *Pearson correlation coefficient*²⁶ between the objective intra-class similarity and the objective complexity of signatures showed a poor correlation ($r=0.451$) between the two variables, while an also low correlation statistic ($r=0.656$) was obtained between the objective intra-class similarity and the perceived complexity. This is a positive result, since, as mentioned earlier, this study was seeking a similarity measure that would not directly depend on the signature's degree of complexity. A smearing of a 15-pixel threshold gives even lower correlation values between a signature's intra-class similarity and its objective ($r=0.324$) or subjective ($r=0.644$) complexity. Higher degrees of smearing prove there to be a correlation between a signature's intrinsic similarity and degree of complexity, which can be explained since more complex signature images would perhaps assist the smearing to take place, which in turn would result in higher similarity values. Finally, the significant correlation results between the objective similarity measure without any smearing against smearing of 5- and 10-pixel thresholds ($r=0.994$ and $r=0.978$ respectively) justifies the application of a low degree of smearing on the signature data.

²⁵ The non-parametric alternative to Pearson correlation for the use with ordinal variables [135].

²⁶ Here the use of the Pearson correlation statistic is possible since the data it is applied on can be considered as continuous (discrete measurements can still be treated as continuous data for the sake of analysis).

Table 6.10: Non-parametric correlation between subjective and objective similarity rank orders.

Spearman correlation coefficient r_s	Objective Similarity	Obj. Similarity (sm. 5p)	Obj. Similarity (sm. 10p)	Obj. Similarity (sm. 15p)	Obj. Similarity (sm. 20p)	Obj. Similarity (sm. 25p)	Obj. Similarity (sm. 30p)
Subjective Similarity	0.800	0.800	0.800	0.900*	0.900*	0.300	0.300

*. Correlation is significant at the .05 level (2-tailed).

Table 6.11: Pearson correlations between subjective/objective complexity and objective similarity measures.

	Perceived Complexity	Objective Complexity	Objective Similarity	Obj. Similarity (sm. 5p)	Obj. Similarity (sm. 10p)	Obj. Similarity (sm. 15p)	Obj. Similarity (sm. 20p)	Obj. Similarity (sm. 25p)	Obj. Similarity (sm. 30p)
Perceived Complexity	1	.913*	.656	.676	.749	.644	.775	.794	.871
Objective Complexity	.913*	1	.451	.443	.543	.324	.545	.598	.798
Objective Similarity	.656	.451	1	.994**	.978**	.846	.644	.501	.303
Obj. Similarity (sm. 5p)	.676	.443	.994**	1	.972**	.866	.659	.524	.330
Obj. Similarity (sm. 10p)	.749	.543	.978**	.972**	1	.905*	.783	.664	.473
Obj. Similarity (sm. 15p)	.644	.324	.846	.866	.905*	1	.896*	.796	.524
Obj. Similarity (sm. 20p)	.775	.545	.644	.659	.783	.896*	1	.978**	.816
Obj. Similarity (sm. 25p)	.794	.598	.501	.524	.664	.796	.978**	1	.906*
Obj. Similarity (sm. 30p)	.871	.798	.303	.330	.473	.524	.816	.906*	1

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

6.4 Conclusions

This chapter has evaluated the intra-class variability of signatures both in terms of human perception and automatic objective approaches. A novel and in-depth investigation into human perceptual judgements with respect to the relative consistency of signatures was carried out. The subjective signature variability results showed that the intrinsic variability of consistent signatures was more easily perceived and judged by humans, rather than the relative variability of inconsistent signatures, which resulted in greater confusion. The experimental results also suggested that the perceived complexity of a signature does not necessarily influence its judged intra-class variability. Given the nature of the experimental data, being measured on an ordinal scale rather than a continuous numeric scale, the evaluation of the subjective variability ranking data demanded the use of non-parametric methods.

Several objective methods, encountered in the literature of ASV or proposed here, for the measure of the intrinsic variability of genuine signatures were tested. However, the method based on Shape Matrices (SM) was chosen as the one closest to the human perception experimental results, and one which also demonstrated good discrimination power. This is in agreement with the good performance results reported in a relevant signature verification study earlier reviewed [121]. A more suitable signature baseline detection method was proposed to cover a range of signature styles, compared to methods such as in [14;40] which are more suitable for handwriting-like signatures. However, limitations were observed, while the implications of orientation adjustment were further assessed. The application of different degrees of smearing, for boosting the low similarity values, was efficiently analysed and compared, with small degrees of smearing being more beneficial than the selection of high smearing thresholds such as that (of 30 pixels) implemented in [121]. Correlation tests showed that the objective (SM) similarity measure is highly related to the subjective similarity results, but the quite unclear perception results may question any further assumptions. Secondly, a low correlation value between the objective variability measurements and the subjective and objective complexity results of Chapter 5, showed that the objective similarity measure analysed does not necessarily depend on the complexity of the signature image, an assumption also supported by the human perception experiment.

Finally, the next chapter evaluates the knowledge gained about the perceived complexity and intrinsic variability of signatures in conjunction with human performance in verifying handwritten signatures, leading to scenarios of predicting the likelihood of a signature's susceptibility to forgery in human checking procedures.

Chapter 7

Human Visual Inspection of Handwritten Signatures

In this chapter an analysis of the visual inspection of handwritten signatures will take place with data obtained from three experiments on human judgements. The human performance in correctly verifying the authenticity of signatures and identifying forgeries together with the resulting error rates will be extensively evaluated. Furthermore, the relation between the attained performance and the complexity of signatures as well as their intrinsic variability will be examined, and the likelihood of the signatures' susceptibility to forgery will be assessed. Finally, an analysis of human confidence in making verification decisions of this nature will allow the specification of possible system rejection scenarios that could assist human checking procedures carried out in everyday situations.

7.1 Introduction

Studies concerned with the perceptual tasks of form discrimination, recognition and identification are extensively reviewed by Zusne [154]. Some of the conclusions reached by the author, which are found to be important concepts of human perception directly related to this study, are briefly stated here:

- “*recognition errors decrease as exposure time of visual forms is increased*”
- “*...recognition is adversely affected by presenting a pattern for recognition in a retinal locus that is different from the one in which learning or familiarization take place*”
- “*rotating a figure affects its discriminability adversely*” and “*a shape rotated to a different position is less readily recognized than it is in the position in which it was originally learned*”
- “*...the less complex shapes have fewer degrees of freedom to vary, hence are not as discriminable as the more complex shapes. The very complex shapes, on the other hand, contain too much information to be processed in a limited time period, hence subjects commit errors or take longer to make the necessary discriminations*”

In order to clarify different methodologies and terminologies used, according to Hake [61] *discrimination tasks* require the subject to give a judgment on whether a stimulus pattern is different from another pattern or some other set of patterns, *identification tasks* include judgements on indicating which stimulus is present, *recognition tasks* involve subjects' judgements on whether a stimulus is familiar, and *judgmental tasks* require the subject to assign a value to a pattern based on a judgement, such as size, complexity, etc.

Eriksen and Hake [48] found a considerable improvement in the discrimination performance of subjects for series of stimuli that varied on two or three dimensions simultaneously compared with when they differed on any of the single dimensions alone. The dimensions employed in their experiment were size, hue and brightness. The tests included series of stimuli varying on each of these dimensions, as well as on each two dimensions and, finally, in all three dimensions together. Vanderplas and Garvin [143] found the complexity of random shapes (determined as the number of points) inversely related to correct recognition of their prototype and to correct rejection of its variations, while positively related to response time. In addition, it was suggested that “*failure of discrimination, resulting from increased complexity of the shapes, would be expected to result in more confusions of the shapes and their variations, thus affecting both the learning task and the recognition score*”.

Furthermore, White [145] in an experiment where subjects were asked to identify single forms within sets of five forms, concluded that the recognition of homogeneous sets of forms was more difficult than for heterogeneous sets, irrespective of whether the forms were simple or complex. On the other hand, simple and homogeneous sets were more easily recognized than complex and homogeneous, whereas for the heterogeneous sets there was no distinction of this kind.

In the field of document analysis and recognition several studies have been concerned with the exceptional ability of humans in reading and recognizing handwritten script, in an effort to build more robust automatic handwriting recognition systems. Some of these studies have attempted to measure the handwriting recognition performance of human readers, in order to identify an optimum recognition rate for automatic systems, whereas some others have carried out a more detailed study identifying perceptually important features involved in human handwriting reading and recognition. A brief description of these studies will follow, as the concept of human reading and recognition of cursive script may be seen as related to the human reading and verification of handwritten signatures.

Barrière and Plamondon [22] performed an experiment where human subjects were asked to identify letters in mixed-script handwritten words. The mean letter recognition rates reported were 86.6% for a group of 6 readers that had access to limited linguistic knowledge (as the words were written in a language other than their native) and 92.8% for 5 readers with access to an extended linguistic context (the text was written in their native language). In [103] the achieved average character recognition rate of 10 human readers was 96% and was comparable with the reported system performance. Linguistic knowledge was not applicable, since the handwritten samples were random letter sequences. Two separate tests with groups of 3 and 17 human readers, reported in [76], gave an average of 81.17% and 76.88% respectively, for case insensitive recognition of characters extracted from handwritten words. Note that linguistic context was also not available here. In addition, it was reported that the machine recognition accuracy was comparable with the average human recognition performance. Schomaker and Segers [131] reported a human word recognition rate of 87.9%, after exposure to the words of the lexicon. Furthermore, the authors discovered some findings, through a specialized experiment, about features in

handwriting that are important in the human reading process. More specifically, the first and last letters of the words were found to be very important for the recognition process, as well as vertical strokes, crossings, high curvature points, and curled endings of final strokes. Moreover, vowel characters were found less important than consonants for the word recognition process. Lorette [86] highlighted the importance of knowledge gained from human perception in order to design more adequate handwriting reading systems, and extensively analysed human perceptual properties of handwriting and reading. He suggested that the reading of handwriting consists of perception, recognition and interpretation processes. The proposed perceptually important elements include the trajectory of the ink trace, the visual shape of the handwritten image, the singularities and regularities, the fundamental down-strokes, the local relative positions, the relative sizes of primitives and letters, the discriminative signs, and the apparent fuzziness. On the other hand, it was suggested for recognition only the use of a small number of significant primitives, without considering the unstable parts of the handwriting.

Some of these findings may be extended for the perceptual processes used in signature recognition and verification by humans. In this case contextual information is not directly present, even though knowledge about possible letter combinations forming syllables, as well as familiarity with plausible surname instances, may assist the reading of certain types of handwritten signatures and, thus, the recognition-verification process, but that would not be the case for incomprehensible shape-like signature samples. Nevertheless, there is much to be gained from knowledge related to human perceptual processes regarding the reading and recognition of cursive script, while highlighting the limited available investigations concerning the human perception of handwritten signatures, their verification and identification of forgeries.

Studies in the signature verification literature, concerning human performance in verifying signatures or identifying forgeries are extremely limited. Fairhurst, Kaplani and Guest [52] have reported the performance of humans in verifying the authenticity of handwritten signatures in relation to their judged complexity. Randolph and Krishnan [119] report some of the elements that experts look for when spotting forgeries. These properties that frequently appear in forgeries are:

- *improper spelling*
- *shaky handwriting*
- *retracing and retouching*
- *vertical weaving*

The case of vertical weaving, as explained by the authors, results from a poor join between characters of the signature, since some forgers aim at precisely copying each letter in a signature individually. Furthermore, it is noted that “*inexperienced forgers often attempt to correct poor forgeries by retracing, retouching, thickening, or adding extra lines to them*”.

The above are issues that are highly related to the area of forensic science. However, little is known about the methods employed by forensic document analysts in examining the authenticity of handwriting, firstly, because of the confidential nature of the matter, secondly, as this is primarily based upon the knowledge and experience of the forensic expert [53], and thirdly, since relative documentation is not readily available. According to a forensic document examiner [15], the major signs that help detect forged signatures are:

- *written at speed which is markedly slower than the genuine signatures*
- *frequent change of the grasp of the pen or pencil*
- *blunt line endings and beginnings*
- *poor line quality with wavering and tremor of the line*
- *retracing and patching*
- *stops in places where writing should be free*

Further forgery properties, as stated by the author, include that “*any forgery will, of necessity, exhibit a considerable degree of similarity to the general run of genuine signatures in the more obvious features of letter design*”, and specifically “*some forgeries will resemble at least one genuine signature in almost every detail*”, whereas “*no two genuine signatures of any length are replicas of each other*”. Although the opinion of the expert document examiners generally differs from that of non-experts, for the simple reason that the first have undergone training and possess valuable experience, the aforementioned clues are important for the understanding of possible weaknesses that forged signatures exhibit.

7.2 Human Verification of Signatures

The performance of humans in verifying the authenticity of handwritten signatures was tested with the help of the following experiments. The experiments were designed so as to approximate possible signature checking scenarios that may be encountered in common practical situations such as in point-of-sales applications.

7.2.1 Experiment 1

The main purpose of this experiment was to test the performance of humans in visually inspecting the authenticity of handwritten signatures of different styles, reflecting a practical scenario of checking a signature against an available prototype reference sample.

Description

Thirty-six subjects²⁷, mostly university students and a fair number of university staff members, of various nationalities and different age groups, took part in this experiment. The subjects were presented with a range of signatures, both genuine and forged samples, belonging to the target signature groups – produced in the previous experiment described in Chapter 3 – and were asked to classify them as being either genuine or forged with respect to an original sample that they had simultaneously in view. For each target signature, 10 genuine reference samples and 10 forged samples of this target were displayed in a one-by-one fashion, while an original reference sample of the target was constantly in view. The original reference sample was the 4th in the sequence of 10 produced by the target signers, for reasons previously analysed (see Chapter 5). Similarly, the 10 forged samples were randomly drawn from the 4th forgery attempt out of the 5 permitted to the forgers, to ensure that the forgers' familiarisation with the digitiser had already taken place before the test sample of

²⁷ The same subjects that participated in the signature complexity experiment described in Chapter 5. A part of these subjects had participated in the forgery trials (see Chapter 3), however, with an interval of about a year between the two experiments it was assumed that this did not affect their performance in this experiment.

interest was generated, thus securing a fair forgery attempt. The 10 genuine samples were randomly drawn from the thirty reference samples previously produced by the target signers. Consequently the subjects were presented with 20 signature samples for each of the five target signatures, resulting in a total of 100 signature samples. The test signature samples were randomly pre-selected, as previously described, and shuffled so that the genuine and forged samples were displayed in a mixed manner. The subjects were reminded to take only a few seconds for each judgement, thus perhaps closely imitating a common practical scenario of checking a signature against the sample provided on the back of a credit card. The average duration of the experiment was between 25-30 minutes, although timings varied somewhat among individuals.

In addition to classifying a signature sample as forged or genuine, the subjects were also asked to assign a confidence value indicating the level of confidence they had in making each classification decision. The confidence level was measured on a scale of 1 to 10, with 1 corresponding to very little confidence and 10 to complete confidence. An illustration of the experimental interface is included in Figure A.2 –Appendix A. Furthermore, a subset of the participants was asked to complete a questionnaire after the end of the experimental session, stating the factors that they considered when making the classification decisions or, in other words, the signature characteristics that mostly influenced their decisions.

Results

The achieved human performance in visually inspecting handwritten signatures is presented in Table 7.1. The results of Experiment 1 showed that an average of 73.8% of the signatures was correctly classified as being either of the ‘genuine’ or the ‘forgery’ class. The remaining 26.2% was erroneously verified, of which 85.3% was due to genuine signatures being falsely rejected (Type I error) and 14.7% due to forgeries being falsely accepted (Type II error). This led to a False Rejection Rate (FRR) as high as 44.7% and a False Acceptance Rate (FAR) as small as 7.7%. The FRR is calculated as the percentage of genuine signatures being falsely rejected out of the total number of genuine signatures shown, and similarly the FAR is calculated as the percentage of forgeries being falsely accepted out of the total number of forgeries shown. It is evident that the subjects were strict enough not to let many forgeries

penetrate the system, which, however, caused many genuine signatures to be rejected as attempted forgeries. Consequently, it seems that humans are generally good in spotting forgeries, however, not very successful in recognising genuine samples. The latter will be further tested in the following experiments. Overall, the human performance in verifying the authenticity of handwritten signatures is promising, considering the unknown class of forgeries, and almost comparable with the performance in the easier task of handwriting recognition reported in the literature.

Table 7.1: Experiment 1 – Average human performance in verifying signatures.

Average Human Performance	
Correct Classification %	73.8
Total Error %	26.2
FRR %	44.7
FAR %	7.7

More analytically, the frequency distribution of the successful classification performance of the participants, with the normal curve superimposed, is displayed in Figure 7.1(a).

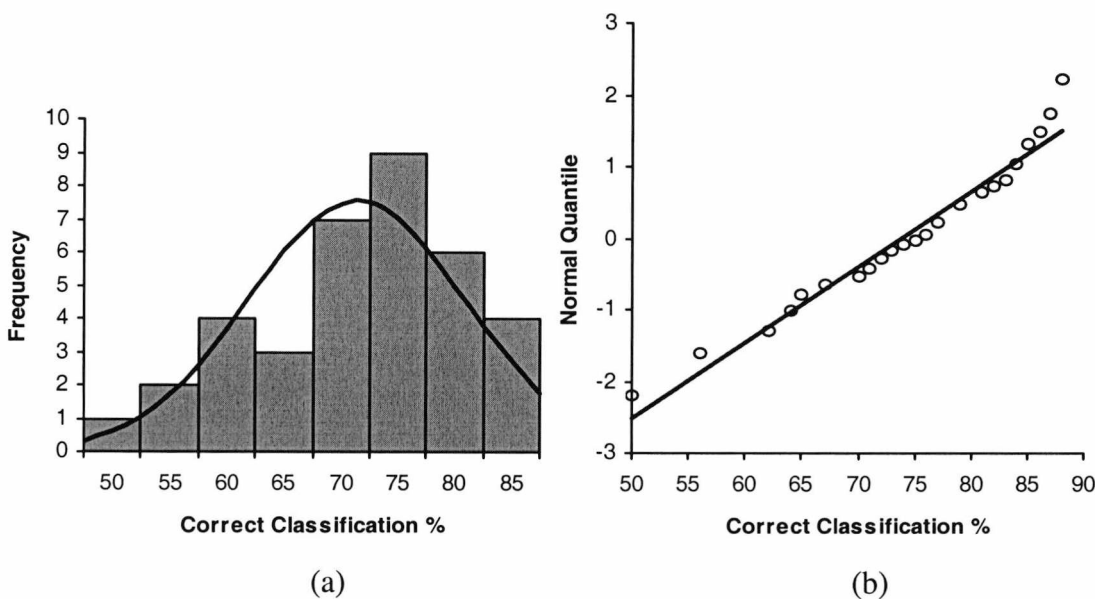


Figure 7.1: Experiment 1 – (a) frequency distribution and (b) normal probability plot of the participants' correct classification performance.

It is apparent that the distribution exhibits normal behaviour and this is also verified by the Shapiro-Wilk normality test, which resulted in a coefficient of $w=0.957$ with a

significant p -value=0.175. Further evidence of the normality of the distribution is provided by the normal probability plot²⁸ (Figure 7.1(b)), where the points of the plot lie very close to the straight line.

The statistical parameters of the distribution are further presented in Table 7.2. It is important to stress, here, that the interval of the mean, with 95% confidence, is between 70.63% and 77.03%. In other words, the expected average correct classification of the population would fall within the above interval in 95% of the cases. As a consequence of the total error being directly derived from the correct classification performance, its frequency distribution also behaves in a normal way, and the interval of its mean, with 95% confidence, can be derived by subtracting the correct classification highest and lowest interval values from 100%. Moreover, the standard deviation, Shapiro-Wilk coefficient, skewness and kurtosis values remain the same with the skewness, however, having its sign reversed, thus exhibiting skewness in the opposite direction.

Table 7.2: Experiment 1 – Statistics of the correct classification and total error.

Statistical Parameters	Correct Class. %	Total Error %
N	36	36
Mean	73.83	26.17
95% CI of Mean	70.63 to 77.03	22.97 to 29.37
Standard Deviation	9.46	9.46
Shapiro-Wilk coef.	0.957	0.957
Skewness	-0.613	0.613
Kurtosis	-0.135	-0.135

The frequency distributions of the FRR and FAR, with the normal curve superimposed, are displayed in Figure 7.2(a) and 7.3(a) respectively. The normality of the distributions was tested again with the Shapiro-Wilk normality test, and resulted in a p -value=0.046 for the FRR, which is greater than the 0.01 significance level, and since there is not sufficient evidence to reject the null hypothesis of normality it may be concluded that the observed FRR data come from a normal distribution. However,

²⁸ Also known as normal Q-Q plot (normal quantile-quantile plot), is the plot of the observed ordered data values against the associated quantiles of the normal distribution. The normal probability plots were generated with the use of a statistical software package. For observations that come from a normal distribution, the points should lie close to the straight line.

for the FAR the resulting p-value (<0.0001) was much smaller than the level of significance, thus rejecting the null hypothesis of normality. The non-normality of the FAR distribution can also be regarded from its normal probability plot (Figure 7.3(b)), where the observed values deviate significantly from the straight line. In addition, the position of the points with respect to the straight line reflects the heavy-tailedness of the distribution. Further evidence of the deviation from normality is provided by the high skewness and kurtosis values of the FAR (Table 7.3).

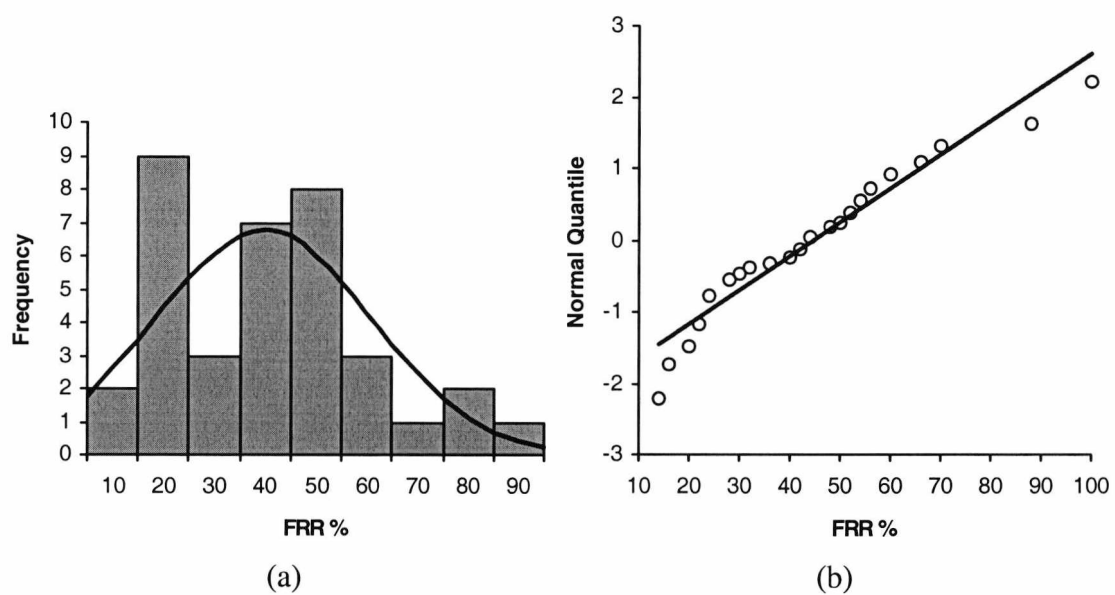


Figure 7.2: Experiment 1 – (a) frequency distribution and (b) normal probability plot of the FRR.

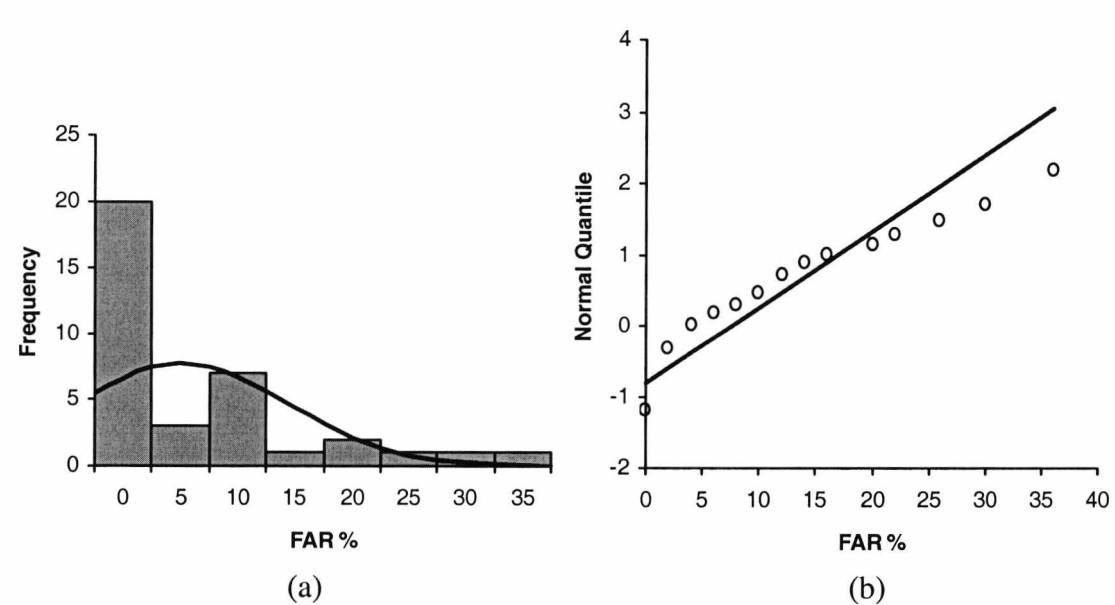


Figure 7.3: Experiment 1 – (a) frequency distribution, and (b) normal probability plot of the FAR.

The normality of the FRR distribution and the resulting interval values of its mean, allow the estimation that the population mean, with 95% confidence, lies between 37.51% and 51.83% (Table 7.3). On the other hand, the estimated interval of the FAR mean (4.52% to 10.82%) is an insufficient clue, since the FAR frequency distribution was proven to be non-normal. In this case, inference about its non-parametric equivalents²⁹ may be made. Thus, it may be reported that the median of FAR is equal to 4.00, the interval of the median, with 95% confidence, is between 2.00 and 10.00 and the IQR is 12.

Table 7.3: Experiment 1 – Statistics of the FRR and FAR.

Statistical Parameters	FRR%	FAR %
N	36	36
Mean	44.67	7.67
95% CI of Mean	37.51 to 51.83	4.52 to 10.82
Standard Deviation	21.16	9.31
Shapiro-Wilk coef.	0.939	0.803
Skewness	0.734	1.500
Kurtosis	0.278	1.780

Finally, the average correct classification and error rate values reported are fairly good estimators of the respective human performance in this experiment. However, a more extensive analysis has been given to acquire a better view of the relation between the reported results and the performance likelihood of the general population.

Questionnaire results

An analysis of the questionnaire responses revealed that the most dominant perceptual signature feature claimed to have played important role in the classification judgements of the participants was the shape of the individual characters within the signatures, or generally stated the signer’s writing style. Other important perceptual factors that seemed to have influenced the subjects’ decisions involved the angles of the signatures, the curves, and particularly distinctive features such as dots, double dots, underlines, etc. The holistic shape of the signatures was also claimed to have drawn attention, along with the signatures’ proportions, as well as its beginning and

²⁹ The non-parametric equivalent of the mean is the median, and of the standard deviation is the IQR (see Chapter 6).

ending parts. Other elements of lower perceptual significance, as mentioned by only a small number of participants, included the size of the signatures, apparent retracing or lack of continuous movement, the absence of smoothness in the signature line, as well as the individual characters' connectivity and direction within the signature.

In general, the responses given included features related to perception of the global shape of signatures, but most importantly to local signature characteristics that stand out in a first encounter with the signature image, which could be claimed to be points of fixation of the human eye. Furthermore, it is evident that the perceptual signature characteristics related to human verification of handwritten signatures, as uncovered in this study, have many common elements with those involved in the human reading and recognition of handwritten script, as suggested by the previously reviewed studies [86;131]. Although a number of primitives, such as the *flourishes*, *hooks at the beginnings*, or *long ends*, quoting the words of Lorette [86], may be regarded as needless 'noise' in human handwriting recognition processes, in human verification of handwritten signatures they may be viewed – as partly proven through this questionnaire – as perceptually important elements characterising individual signing styles.

In addition, it is evident that features primarily important to expert document examiners as earlier mentioned, such as evidence of retracing or disfluencies along the signature line, were mentioned only by a small number of participants, suggesting a fundamental difference in perception between forensic experts and non-experts. While the former concentrate on identifying signs of imitation, the latter seem to carry out processes of form perception, recognition and discrimination. On the other hand, it could be claimed that the identification of forgery signs is a next step to the fundamental processes of shape perception carried out with the immediate exposure to the signatures, and that the lack of prolonged contact with the signature images in the case of this experiment or any other POS application could partly explain the difference in judgement and performance that apparently exists between these two groups (i.e. experts and non-experts). Finally, knowing the type of processes carried out by humans when judging the authenticity of signatures could assist the proper development of training programs, suitable for the different needs of various types of applications, if human signature inspection is to be improved.

7.2.2 Experiment 2

The aim of this experiment was to test the performance of humans in verifying signatures with respect to some knowledge about the signature's intrinsic variability.

Description

The participants in this experiment were 28 subjects³⁰, mostly undergraduate and postgraduate university students of various nationalities, the majority of whom belonged to age groups ranging from 17-30 years. Similarly to Experiment 1, the participants were shown a series of signature samples, both forged and genuine, belonging to each target signature group and were asked to assign them to one of the two classes, genuine or forgery, giving also their confidence level in making each classification decision (based on the same scale). The difference was that in this experiment the subjects were provided with two original reference samples of the target signer simultaneously in view, based on which the comparison against the test sample would be made. As in Experiment 1, the number of test signatures the participants were presented with was the same, 10 genuine and 10 forged signatures per target signature group, randomly pre-selected from the acquisition trials³¹ and shuffled. The original reference target signatures displayed were the first two signature samples, in the sequence of 10, provided by the target signers. The subjects were advised to give a quick response for every test signature and were prompted to take a break, half way between the five target groups, if they felt tired. The time extension of the break taken varied among individuals from 10 minutes to a couple of hours, although a few of the subjects completed the second part of the test on a different day. In addition, the sequence of the target groups displayed was not fixed to any particular order, but the majority of the subjects followed the numbered order of the groups displayed, as was the case for Experiment 1. An illustration of the experimental layout is provided in Figure A.3 –Appendix A.

³⁰ A very small number of which had participated in Experiment 1. However, it is assumed that their previous participation did not affect their judgements, since Experiment 2 was carried out with about a year and a half interval from Experiment 1.

³¹ Similarly to Experiment 1, the forged samples were randomly selected from the 4th forgery attempt submitted by the forgers.

Results

The general performance of the subjects in this experiment, in terms of correct classification of signatures, total error attained, false rejection and false acceptance rates, is displayed in Table 7.4. An average of 81.8% of the signatures was correctly verified, and out of the remaining 18.2% which is the total error, 72.5% was due to Type I error and 27.5% due to Type II error. The rejection of genuine signatures is, also in this experiment, the major source of error occurring. In particular, the resulting FRR was 26.4% and the FAR was exactly 10%.

Table 7.4: Experiment 2 – Average human performance in verifying signatures.

Average Human Performance	
Correct Classification %	81.8
Total Error %	18.2
FRR %	26.4
FAR %	10.0

A further analysis of the subjects’ performance in correctly verifying the handwritten signatures is provided by Figure 7.4. The frequency distribution of Figure 7.4(a), with the normal curve superimposed, exhibits normality. This is emphasised by the normal probability plot of Figure 7.4(b), where the data points fall close to the straight line.

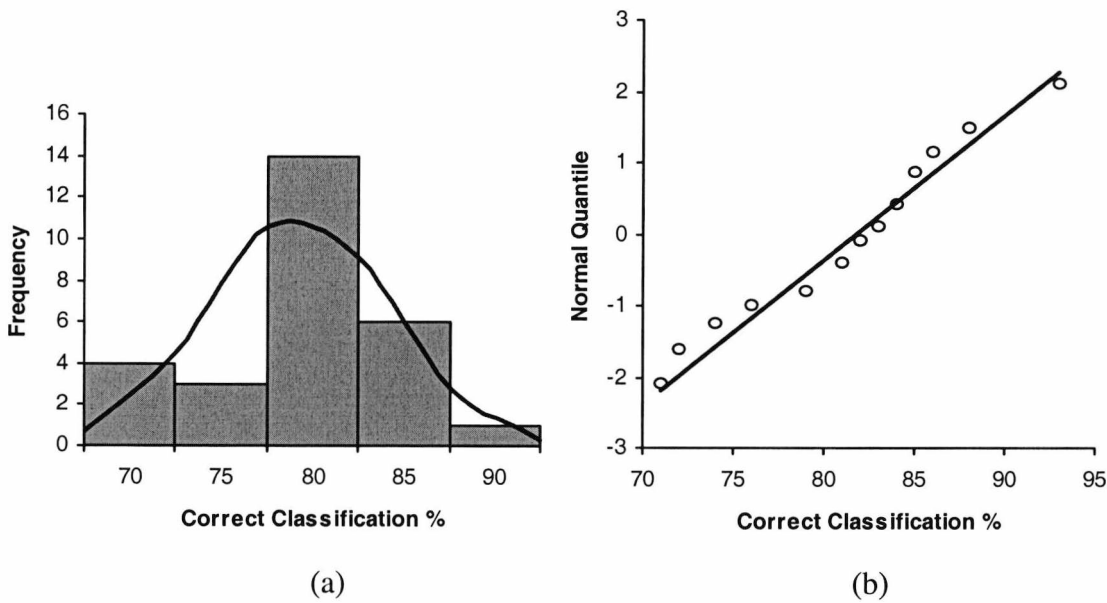


Figure 7.4: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the participants’ correct classification performance.

Indeed, the resulting coefficient of the Shapiro-Wilk normality test was $w=0.945$ (Table 7.5) with a very significant $p\text{-value}=0.152$, thus concluding that the observed values come from a normal distribution. For reasons described in the previous experimental analysis (Section 7.2.1) the distribution of the total error behaves in an analogous way, resulting to the same Shapiro-Wilk coefficient and exhibiting the same kurtosis and exactly opposite skewness effects.

The mean of the distribution for the correct classification performance (Table 7.5) is 81.79% and the standard deviation has a small value of 4.96%, which may also be graphically observed from the small range of its frequency distribution. Furthermore, the resulting interval of the mean for the correct classification performance of the population, with 95% confidence, is 79.86% to 83.71%, and from 16.29% to 20.14% for the population's mean total error.

Table 7.5: Experiment 2 – Statistics of the correct classification and total error.

Statistical Parameters	Correct Class. %	Total Error %
N	28	28
Mean	81.79	18.21
95% CI of Mean	79.86 to 83.71	16.29 to 20.14
Standard Deviation	4.96	4.96
Shapiro-Wilk coef.	0.945	0.945
Skewness	-0.409	0.409
Kurtosis	0.517	0.517

Analysing the FRR frequency distribution (Figure 7.5(a)), it appears to exhibit normality, which is also apparent from its normal probability plot (Figure 7.5(b)). The Shapiro-Wilk normality test resulted in a $w\text{-statistic}=0.954$ with a very significant $p\text{-value}=0.256$, thus concluding that the FRR data come from a normal distribution. The same may be concluded for the FAR frequency distribution (Figure 7.6(a)) but with smaller confidence, since the Shapiro-Wilk normality test resulted in a $w\text{-statistic}=0.911$ with a small $p\text{-value}=0.021$, but yet greater than the 0.01 level of significance. Thus the null hypothesis of normality was not rejected. A visual observation of this may be achieved through the FAR normal probability plot (Figure 7.6(b)). Further evidence of the near normality of the distributions is provided by the relatively small skewness and kurtosis values (Table 7.6).

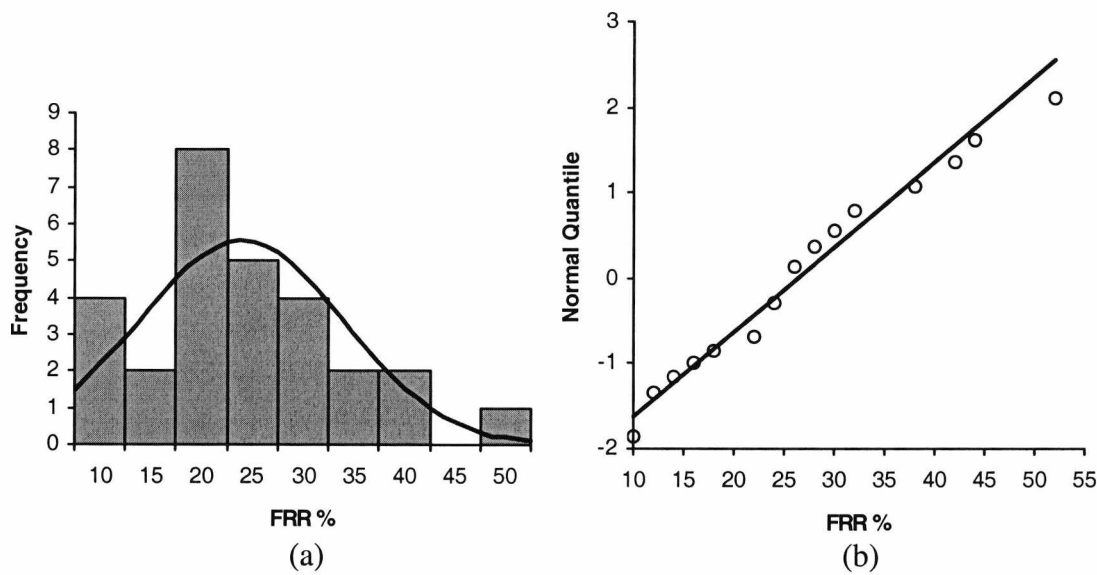


Figure 7.5: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the FRR.

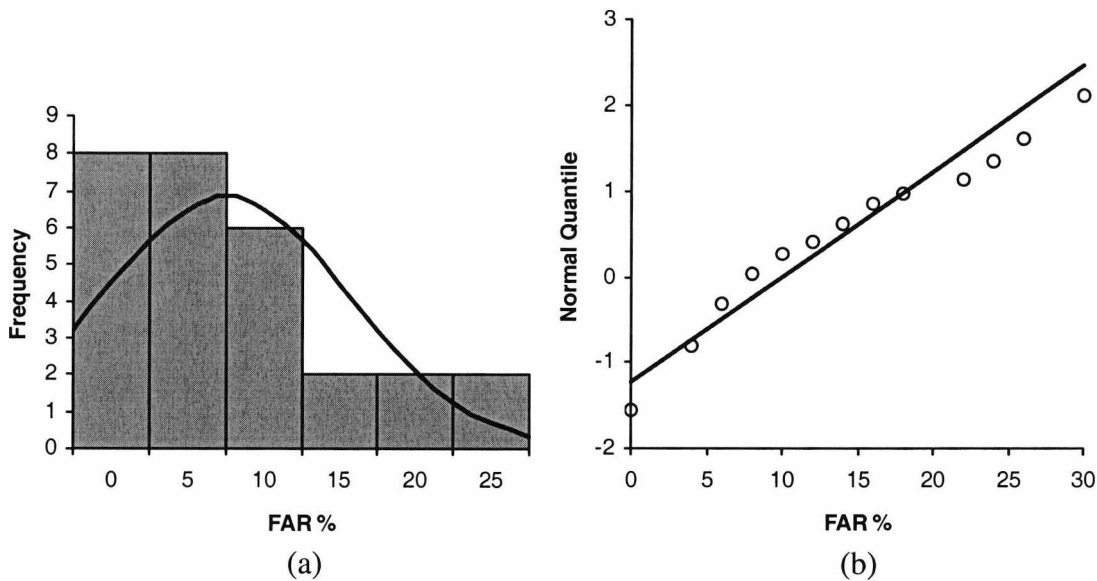


Figure 7.6: Experiment 2 – (a) frequency distribution and (b) normal probability plot of the FAR.

Table 7.6: Experiment 2 – Statistics of the FRR and FAR.

Statistical Parameters	FRR%	FAR %
N	28	28
Mean	26.43	10.00
95% CI of Mean	22.54 to 30.32	6.85 to 13.15
Standard Deviation	10.04	8.13
Shapiro-Wilk coef.	0.954	0.911
Skewness	0.539	0.913
Kurtosis	0.505	0.211

Finally, the mean FRR was 26.43% and its standard deviation as large as 10.04%, whereas for the FAR a mean value of 10% and a surprising standard deviation of 8.13% were respectively measured. On the other hand, the FRR population mean was estimated to lie between 22.54% and 30.32%, with 95% confidence, and for the FAR between 6.85% and 13.15%.

7.2.3 Experiment 3

In this experiment the aim was to test human performance in verifying handwritten signatures, when additional knowledge about the signature's intrinsic variability is available.

Description

Similarly to Experiment 2, twenty-eight subjects³², mostly university students of various nationalities – the majority of whom belonged to the age group of 21-30 years – participated in this experiment. Again, the participants were shown a series of genuine and forged signature samples, for each of the target groups, and were asked to classify them as being either of the 'genuine' or 'forgery' class. However, this time the subjects were provided with five reference samples of the original signer simultaneously in view. In addition, the subjects were also asked for their level of confidence in making the respective classification judgement. Again, the number of test signatures the participants were presented with was the same, 10 genuine and 10 forged signatures per target signature group, randomly pre-selected from the acquisition trials and shuffled. The original reference samples constantly on display were, in this experiment, the first five signature samples, in the sequence of 10 provided by the target signers, whereas the test samples were exactly the same as in Experiment 2. Note that the 5 genuine samples provided in the display were a disjoint set from the genuine signatures used for testing. The same procedure with Experiment 2 was also used regarding the element of time and the sequence of the target groups. An illustration of the experimental layout is provided in Figure A.4 –Appendix A.

³² A totally disjoint group of participants from that of Experiment 2. However, a small number of these had participated in Experiment 1. For reasons previously stated, this was assumed to have no influence on the judgements recorded.

Results

The average human performance in verifying handwritten signatures (Table 7.7), gave 84.1% correct classification, 15.9% of errors, of which 83% were accounted for by Type I error and 17% by Type II errors. The resulting false rejection rate was 26.5%, whereas the false acceptance rate was as small as 5.4%.

Table 7.7: Experiment 3 – Average human performance in verifying signatures.

Average Human Performance	
Correct Classification %	84.1
Total Error %	15.9
FRR %	26.5
FAR %	5.4

A similar analysis of results, as pursued in the previous experiments, is employed here. The frequency distribution of the subjects' correct classification performance, with the normal curve superimposed, is displayed (Figure 7.7(a)). The normal probability plot (Figure 7.7 (b)) gives evidence of near normal behaviour, with the observed data points lying on or very close to the straight line. The resulting Shapiro-Wilk normality coefficient was $w=0.960$ with a very significant $p\text{-value}=0.346$, and hence there is sufficient evidence that the data come from a normal distribution.

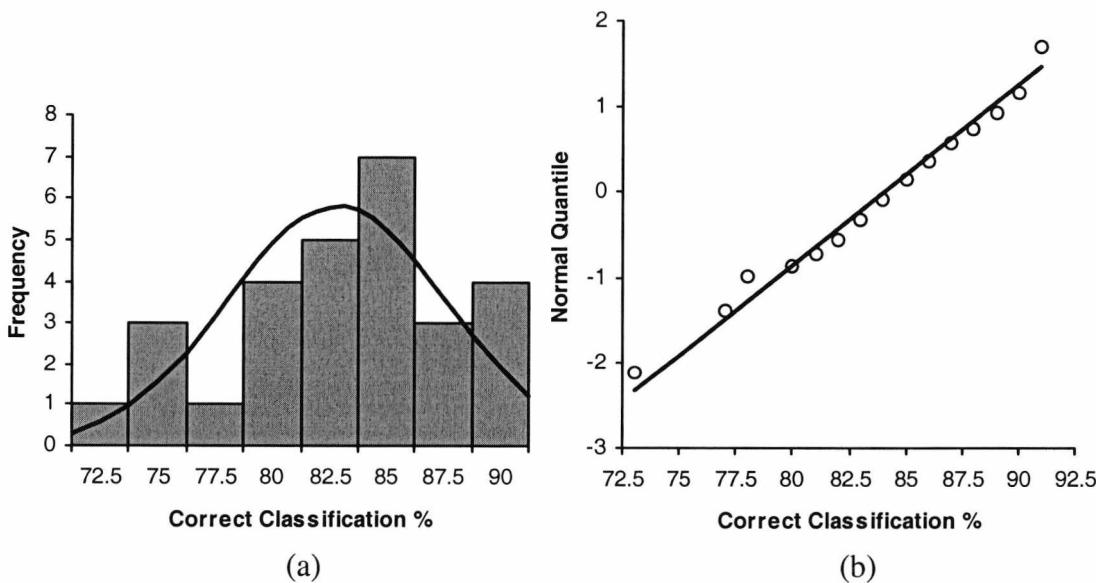


Figure 7.7: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the participants' correct classification performance.

Furthermore, the statistical values (Table 7.8) show a small standard deviation of 4.78%, and fairly small skewness and kurtosis values. With 95% confidence, the population mean was estimated to lie between 82.22% and 85.93% regarding the expected correct classification performance, whereas it was between 14.07% and 17.78% regarding the expected total error attained.

Table 7.8: Experiment 3 – Statistics of the correct classification and total error.

Statistical Parameters	Correct Class. %	Total Error %
N	28	28
Mean	84.07	15.93
95% CI of Mean	82.22 to 85.93	14.07 to 17.78
Standard Deviation	4.78	4.78
Shapiro-Wilk coef.	0.960	0.960
Skewness	-0.432	0.432
Kurtosis	-0.358	-0.358

The frequency distribution of the false rejection rate and its normal probability plot is displayed in Figure 7.8(a) and (b), respectively. With the resulting Shapiro-Wilk coefficient w-statistic=0.948 having a significant p-value=0.172 there is sufficient evidence that the FRR data come from a normal distribution.

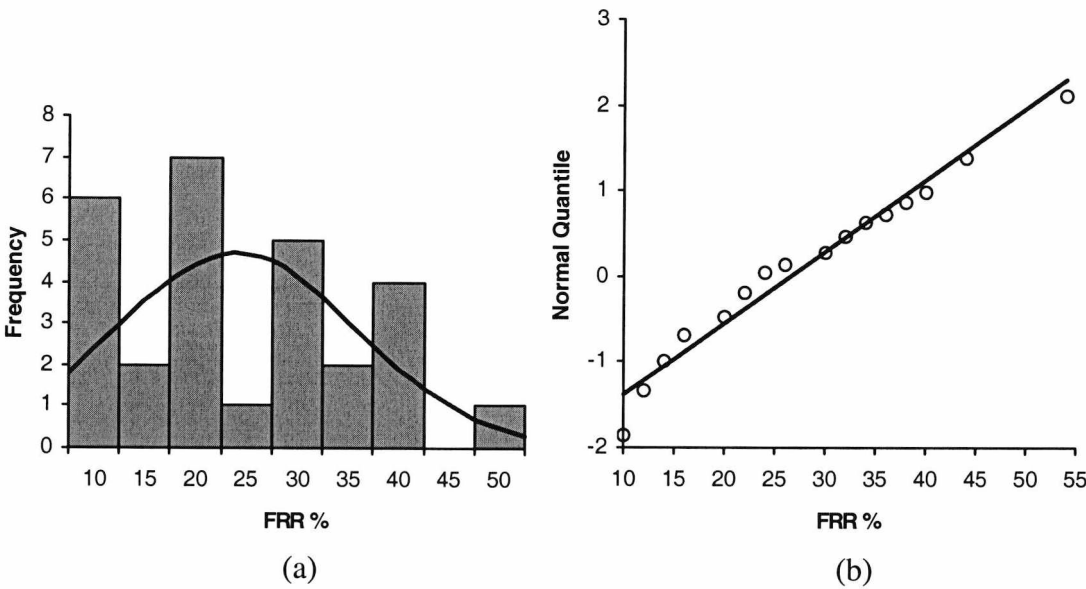


Figure 7.8: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the FRR.

Note that the high fluctuations regarding the frequencies of adjacent rates is a result of the selected intervals, and the distribution is approximating the superimposed normal curve. In addition, the normal probability plot shows that the data fall on or very close to the straight line, thus exhibiting normality.

Regarding the false acceptance rate, its frequency distribution (Figure 7.9(a)) also exhibits normal behaviour and the data points of the normal probability plot (Figure 7.9(b)) fall near the straight line. However, with a Shapiro-Wilk w -statistic=0.918 and a small p -value=0.031, but yet greater than the 0.01 level of significance, the null hypothesis of normality will not be rejected thus concluding that the data come from a normal distribution.

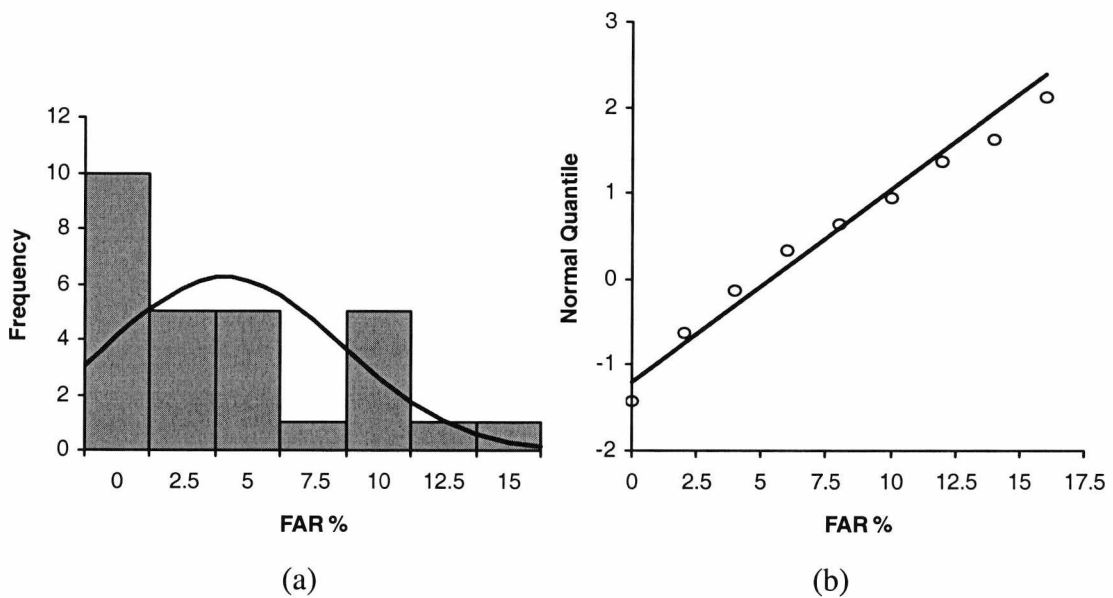


Figure 7.9: Experiment 3 – (a) frequency distribution and (b) normal probability plot of the FAR.

The statistical parameter values for the FRR and the FAR are displayed in Table 7.9. The mean of the FRR is 26.50% and the standard deviation as high as 11.92%, whereas the respective values for the FAR are 5.36% and 4.46%. The interval of the mean, with a 95% confidence is from 21.88% to 31.12% for the FRR and between 3.63% and 7.09% for the FAR, thus indicating the expected range of the population means.

Table 7.9: Experiment 3 – Statistics of the FRR and FAR.

Statistical Parameters	FRR%	FAR %
N	28	28
Mean	26.50	5.36
95% CI of Mean	21.88 to 31.12	3.63 to 7.09
Standard Deviation	11.92	4.46
Shapiro-Wilk coef.	0.948	0.918
Skewness	0.504	0.718
Kurtosis	-0.586	-0.196

7.2.4 Comparative Analysis

A comparative analysis between the three experiments is an evaluation of the human performance in visually inspecting the authenticity of handwritten signatures, in situations where no knowledge (Experiment 1) about the signature’s intrinsic variability exists and cases where little (Experiment 2) or greater knowledge (Experiment 3) is directly available. To provide an accurate comparison, an extensive statistical analysis will follow, making use of various statistical tests, in order to assess whether there is a significant difference between the population mean values of the correct classification performance, the FRR and the FAR among the three experiments or whether the obvious difference in the results reported could be purely attributed to chance. The statistical results will further assist to evaluate how the error rates change – in the population level – with respect to an increase in the number of reference samples provided for the human inspection of signatures.

A first look at the statistics of the three experiments (Table 7.10) reveals an improvement in the verification performance of the subjects from an average of 73.83% correct classification in Experiment 1 to an average of 81.79% in Experiment 2 and an 84.07% in Experiment 3. There is a substantial increase in the performance from Experiment 1 to Experiment 2, but only a small improvement with Experiment 3. In addition there is a reduction in the standard deviation from Experiment 1 to Experiment 2, whereas Experiments 2 and 3 show almost equal values. The interval of the mean, with 95% confidence, shows a distinct increase in the population mean correct classification performance from Experiment 1 to 2. However, the intervals of the mean for Experiments 2 and 3 seem to partially overlap, with the population mean

lower bound of Experiment 3 lying within the interval of Experiment 2. The median follows a similar increase as the mean, between the three experiments, and similar overlapping effects in its intervals for Experiments 2 and 3.

Table 7.10: Correct Classification (%) statistics for the three experiments.

Statistics	Experiment 1	Experiment 2	Experiment 3
N	36	28	28
Mean	73.83	81.79	84.07
Standard Deviation	9.46	4.96	4.78
95% CI of Mean	70.63 to 77.03	79.86 to 83.71	82.22 to 85.93
Median	75.50	82.50	84.50
IQR	14.75	4.25	6.00
95% CI of Median	71.00 to 79.00	81.00 to 84.00	82.00 to 87.00

A visual observation of all this is provided by the box-plots of Figure 7.10. The interval of the means is represented by the diamonds, and the overlapping effect can be viewed from the comparative extent of the diamond dimensions for Experiments 2 and 3. Possible close outliers are apparent in the case of Experiment 2. However, due to the delicacy of the experiment – dealing with subjective human judgements – and the difficulty of the task – requiring the verification of signatures' authenticity when the forgery class is unknown – the outliers will not be excluded, and further evaluation of their effect is outside the scope of this analysis.

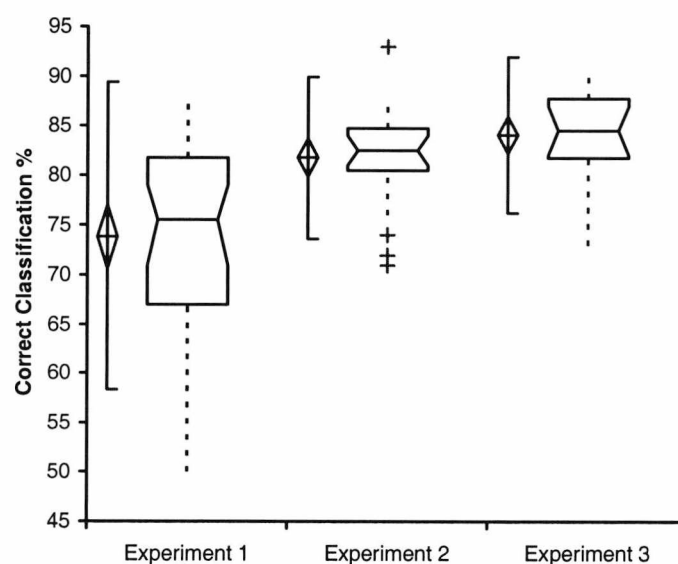


Figure 7.10: Box-plots of the correct classification performance in the three experiments.

In order to statistically assess whether there is a significant difference between the population mean of the three experiments or whether the differences occurred purely by chance, one-way Analysis of Variance (ANOVA)³³ could be performed. The required assumption that the three population probability distributions are normal is met, as previously analysed through the Shapiro-Wilk normality test results. A further requirement is that the three population variances are equal, although some departures are accommodated.

To test if this assumption is met a Homogeneity of Variance test³⁴ is performed, which resulted in a Levene statistic of 9.656 and a significance value of 0.000, thus rejecting the null hypothesis of equal population variances and accepting the alternative hypothesis that there are differences between two or more variances. Since the assumption of equal population variances is not met the one-way ANOVA is not directly appropriate for the comparison of the population means. Instead, multiple comparison tests³⁵ (one-way ANOVA Post Hoc Tests) that do not assume equal variances, such as the Tamhane's T2, can be used.

Table 7.11 displays the pairwise comparisons of the Experiment means. Comparisons between Experiment 1 and 2 as well as between Experiment 1 and 3 yielded a probability (that the population mean difference is zero) equal to 0.000, thus concluding that there is a highly significant difference between the population means of these Experiments. If the 95% confidence interval contains zero then the two population means do not differ. This is true for the confidence interval of the mean difference for Experiments 2 and 3, thus concluding, with a probability value equal to 0.234, that their population means do not differ.

³³ The one-way ANOVA is a statistical procedure that performs analysis of variance, for a single factor, testing the null hypothesis that two or more group means are equal in the population, by comparing the sample variance estimated from the group means to that estimated within the groups.

³⁴ Statistical procedure testing the equality of group variances, such as the Levene statistic which was calculated by means of a statistical software package. Note that this test is not dependent on the assumption of normality.

³⁵ Pairwise multiple comparisons can determine which means actually differ. All statistical tests have been performed with the help of statistical software packages.

Table 7.11: Multiple Comparisons of the correct classification means for the three experiments.

		Mean		95% Conf. Interval		
(I) Exp	(J) Exp	Difference (I-J)	Sig.	Lower Bound	Upper Bound	
Tamhane	1	2	-7.95*	0.000	-12.47	-3.44
		3	-10.24*	0.000	-14.71	-5.76
	2	1	7.95*	0.000	3.44	12.47
		3	-2.29	0.234	-5.50	0.92
	3	1	10.24*	0.000	5.76	14.71
		2	2.29	0.234	-0.92	5.5

*, The mean difference is significant at the 0.05 level.

A further comparison between the population means of Experiments 2 and 3 can be more robustly achieved if an evaluation of the data of these 2 groups is performed independently from that of Experiment 1. The Levene’s test for equality of variances resulted to a 0.019 statistic with a p-value of 0.891, greater than the 0.05 level of significance, thus concluding that the two experiments have equal population variances. A more robust test of the equality of the population variances, between only two groups, is the F-test³⁶. The resulting F-statistic of 1.076 is smaller than the critical value $F_{0.05}=1.905$ with a significant p-value=0.425, thus the null hypothesis cannot be rejected and hence it is concluded that the population variances for Experiments 2 and 3 are equal. Therefore, assuming equal variances a two-independent sample t-Test³⁷ for equality of the population means is performed (Table 7.12). The null hypothesis tested is that the difference of the means $(\mu_2-\mu_3) = 0$. The alternative hypothesis for one-tailed test is that the difference of the means $(\mu_2-\mu_3) < 0$, whereas for the two-tailed test is that $(\mu_2-\mu_3) \neq 0$. In a two-tailed test the null hypothesis of equality of the means cannot be rejected since the criterion of $|t| > t_{crit(2)0.05}$ is not met. However, the one-tailed test criterion of $t < -t_{crit(1)0.05}$ is met, and hence the null hypothesis is rejected and the alternative hypothesis $(\mu_2-\mu_3) < 0$ or

³⁶ A statistical test that performs a comparison between the variances of two data sets. The normality of the population probability distributions is a necessary requirement. The null hypothesis of equal variances is rejected if the F statistic is greater than the critical value (may be obtained from statistical F-tables [92]) for a specific confidence level α .

³⁷ The two-sample student’s t-test is performed for the statistical comparison of the equality of the population means of two groups. The sample sizes should be as equal as possible, and preferably smaller than 30, while the distributions should be normal. Both one-tailed and two-tailed tests can be performed.

$\mu_2 < \mu_3$ is accepted. In addition, since the p-value obtained (0.042) is smaller than the selected α value of 0.05, there is sufficient evidence to reject the null hypothesis.

Table 7.12: Two-independent sample t-Test assuming equal variances for the correct classification population means of Experiments 2 and 3.

	Experiment 2	Experiment 3
Mean	81.786	84.071
Variance	24.619	22.884
Observations	28	28
Pooled Variance	23.751	
Hypothesized Mean Diff.	0	
df	54	
t Stat	-1.755	
P(T<=t) one-tail	0.042	
t Critical one-tail	1.674	
P(T<=t) two-tail	0.085	
t Critical two-tail	2.005	

Therefore, after this extensive statistical analysis it is statistically proven that the population mean for the correct classification performance of Experiment 2 is smaller than that of Experiment 3. Finally, the population mean of the correct classification performance increases from Experiment 1 up to Experiment 3, with $\mu_1 < \mu_2 < \mu_3$.

As earlier described, the behaviour of the total error is analogous to the correct classification performance, since it is the remaining percentage from the 100%. Therefore, there is a reduction in the mean total error from 26.17% in Experiment 1, to 18.21% in Experiment 2, and 15.93% in Experiment 3. The interval of the mean, with 95% confidence, in Experiment 1 is from 22.97% to 29.37%, in Experiment 2 from 16.29% to 20.14%, and in Experiment 3 from 14.07% to 17.78%. Thus, again there is an overlapping between the intervals of the mean for Experiment 2 and 3, but now reflected in the opposite direction. Finally, while the population mean of the correct classification performance increases when moving from Experiment 1 up to 3 the opposite is expected for the population mean of the total error, thus $\mu_1 > \mu_2 > \mu_3$.

A comparison of the false-rejection-rate statistics between the three experiments (Table 7.13) reveals a reduction from 44.67% mean FRR in Experiment 1 to a

surprising 26.43% in Experiment 2, and a similar 26.50% in Experiment 3. Similar reduction rates are observed in the standard deviation values of the three experiments. Both the mean and the standard deviation values of the FRR in Experiments 2 and 3 are almost equal. Moreover, the intervals of their mean are almost superimposed. Thus, a difference between the FRR population means achieved in Experiments 2 and 3 cannot be inferred at this stage. A slightly different pattern is observed in the median and the IQR values of the FRR in the three experiments. A visual observation of all this is provided by the box-plots of Figure 7.11.

Table 7.13: FRR (%) statistics for the three experiments.

Statistics	Experiment 1	Experiment 2	Experiment 3
N	36	28	28
Mean	44.67	26.43	26.50
Standard Deviation	21.16	10.04	11.92
95% CI of Mean	37.51 to 51.83	22.54 to 30.32	21.88 to 31.12
Median	44.00	25.00	23.00
IQR	32.00	9.50	19.50
95% CI of Median	30.00 to 52.00	24.00 to 30.00	20.00 to 32.00

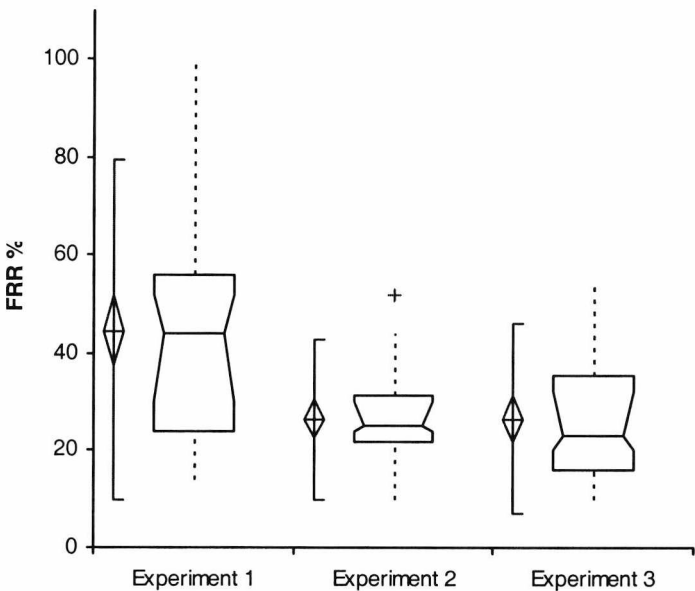


Figure 7.11: Box-plots of the FRR in the three experiments.

Similarly to the statistical analysis previously performed, the Homogeneity of Variance test is performed for the FRR of the three experiments to test the null hypothesis that the three population variances are equal. The resulting Levene statistic

is 8.279 with a significance of 0.001, which is smaller than the 0.05 level of significance, thus rejecting the null hypothesis and hence concluding that there are differences between two or more variances. Since the variances are not equal the Tamhane test of multiple comparisons is performed. The pairwise comparisons between Experiment 1 and 2, as well as between Experiment 1 and 3 (Table 7.14), proves to be a significant difference between their means, since the resulting probability, that the difference between the two means is zero, is p-value= 0.000. For the pairwise comparisons of Experiment 2 and 3, the 95% confidence interval contains zero, and hence the two means do not differ. Moreover, the resulting probability value, of zero difference between the two means, is 1.000!

Table 7.14: Multiple Comparisons of the FRR means for the three experiments.

		Mean		95% Conf. Interval		
(I) Exp	(J) Exp	Difference (I-J)	Sig.	Lower Bound	Upper Bound	
Tamhane	1	2	18.24*	0.000	8.36	28.11
		3	18.17*	0.000	7.87	28.46
	2	1	-18.24*	0.000	-28.11	-8.36
		3	-0.07	1.000	-7.34	7.19
	3	1	-18.17*	0.000	-28.46	-7.87
		2	0.07	1.000	-7.19	7.34

*. The mean difference is significant at the 0.05 level.

A more detailed analysis between the population statistics of Experiment 2 and 3 will follow. The F-test resulted in an F-value=1.409 smaller than the critical value $F_{0.05}=1.905$, with a significant p-value=0.189, hence the null hypothesis cannot be rejected, thus concluding that the two variances are equal. In order to compare the two population means, a two-independent sample t-Test assuming equal variances was performed (Table 7.15). All the assumptions are met, with both the FRR probability distributions for Experiments 2 and 3 approximating normality as earlier proved, with equal variances, and equal number of samples (28) randomly and independently selected from the populations. The resulting t-statistic (-0.024) does not qualify for the rejection criteria of either the one-tailed test – of either direction – or the two-tailed test, thus accepting the null hypothesis that the FRR population means of the two Experiments are equal.

Table 7.15: Two-independent sample t-Test assuming equal variances for the FRR population means of Experiments 2 and 3.

	Experiment 2	Experiment 3
Mean	26.429	26.500
Variance	100.847	142.111
Observations	28	28
Pooled Variance	121.479	
Hypothesized Mean Diff.	0	
df	54	
t Stat	-0.024	
P(T<=t) one-tail	0.490	
t Critical one-tail	1.297	
P(T<=t) two-tail	0.981	
t Critical two-tail	1.674	

With respect to the false-acceptance-rate, the FAR statistics (Table 7.16) show an increase of the mean FAR from 7.67% in Experiment 1 to 10.00% in Experiment 2, and a decrease to 5.36% in Experiment 3, with the standard deviation decreasing along the experiments. The interval of the mean, with 95% confidence, is 4.52% to 10.82% in Experiment 1, between 6.85% and 13.15% in Experiment 2, and 3.63% to 7.09% in Experiment 3, with the three intervals partially overlapping.

Table 7.16: FAR (%) statistics for the three experiments.

Statistics	Experiment 1	Experiment 2	Experiment 3
N	36	28	28
Mean	7.67	10.00	5.36
Standard Deviation	9.31	8.13	4.46
95% CI of Mean	4.52 to 10.82	6.85 to 13.15	3.63 to 7.09
Median	4.00	8.00	4.00
IQR	12.00	10.00	7.50
95% CI of Median	2.00 to 10.00	6.00 to 14.00	2.00 to 6.00

According to the Shapiro-Wilk normality test results earlier reported the FAR frequency distribution of Experiment 1 was proven to reject the null hypothesis of normality, whereas for Experiments 2 and 3, despite the relatively small p-values, the null hypothesis of normality was not rejected at the 0.01 level of significance. Therefore, a comparison of the non-parametric statistics for the three experiments

would be more accurate here. The FAR median values are equal to 4.00 in both Experiments 1 and 3, whereas a value twice as high (8.00) is obtained for Experiment 2. The IQR is reduced from 12.00 in Experiment 1, to 10.00 in Experiment 2 and furthermore to 7.50 in Experiment 3. The interval of the median, with 95% confidence, is shown to be overlapping between Experiment 1 and each of the other two. However, this is not the case between Experiments 2 and 3. A visual observation of the above is provided by the box-plots of Figure 7.12.

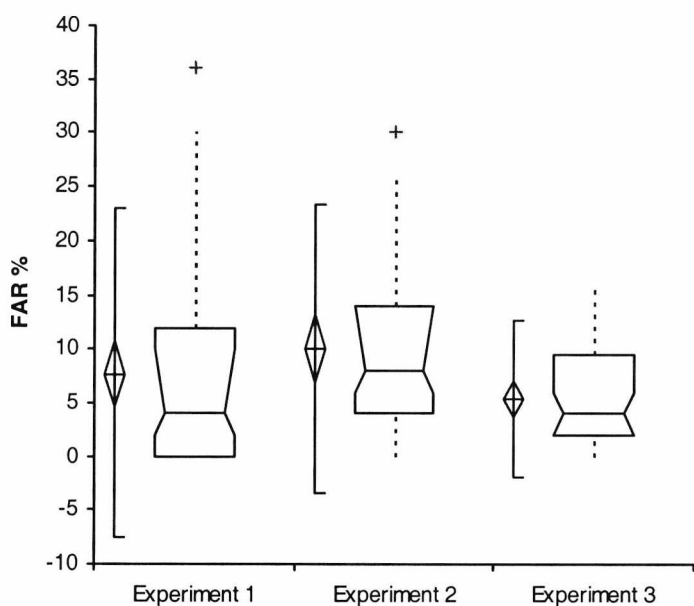


Figure 7.12: Box-plots of the FAR in the three experiments.

A non-parametric comparison between the three populations is obtained through a Kruskal-Wallis³⁸ one-way ANOVA test. A right-tail probability smaller than the 0.05 level of significance indicates that the data sets differ. The resulting p-value=0.084 (Table 7.17) was greater than 0.05, thus concluding that the three experimental sets of the FAR do not differ. However, the fact that a significant difference between the three experiments was not found does not mean that the samples are the same.

³⁸ This is the non-parametric equivalent to one-way ANOVA and tests whether two or more independent samples come from the same population. The distributions do not have to be normal and the variances do not need to be equal.

Table 7.17: Kruskal-Wallis One-way ANOVA for comparison of the three experiment FAR probability distributions.

Exp.	N	Mean Rank	Chi-Square	df	Asymp. Sig
1	36	43.26	4.949	2	0.084
2	28	55.71			
3	28	41.45			
Total	92				

Since the intervals of the mean and median for Experiments 2 and 3 are nearly distinct a two-independent sample test was performed for a further comparison between the two experiments. Since the normality of the probability distributions was accepted, although with low confidence, both the parametric two-independent sample t-test and its non-parametric equivalent Mann-Whitney U test will be performed. The t-test results assuming unequal variances – a decision based upon the low significant value (0.008) of Levene’s test for equality of variances – are presented in Table 7.18. Both in the one-tailed and two-tailed test the resulting t-statistic (2.650) is greater than the respective critical value, and thus the null hypothesis of equality of the means must be rejected and the alternative hypothesis –i.e. that $(\mu_2-\mu_3)>0$ or $\mu_2>\mu_3$ for the one-tailed test and $(\mu_2-\mu_3)\neq 0$ or $\mu_2\neq\mu_3$ for the two-tailed test – must be accepted. The small p-values indicate the very small probability of the differences occurring by chance.

Table 7.18: Two-independent sample t-Test assuming unequal variances for the FAR population means of Experiments 2 and 3.

	Experiment 2	Experiment 3
Mean	10.000	5.357
Variance	66.074	19.868
Observations	28	28
Hypothesized Mean Diff.	0	
df	42	
t Stat	2.650	
P(T<=t) one-tail	0.006	
t Critical one-tail	1.682	
P(T<=t) two-tail	0.011	
t Critical two-tail	2.018	

Furthermore, a similar conclusion is reached from the results of the Mann-Whitney U test (Table 7.19). The p-value obtained (0.025) is smaller than 0.05, which indicates

that the two independent samples, FAR data sets for Experiments 2 and 3, are from different populations.

Table 7.19: Mann-Whitney U test for comparison of the FAR populations of Experiment 2 and 3.

Exp.	N	Mean Rank	Sum of Ranks	Mann-Whitney U	Asymp.Sig.2-tail
2	28	33.34	933.50	256.50	0.025
3	28	23.66	662.50		
Total	56				

Finally, the comparison between the three experiments could be translated as a comparison of the human verification performance corresponding to different number of reference samples simultaneously in view. With Experiment 1 having only one original sample in view, Experiment 2 having two, and Experiment 3 having five reference samples displayed, Figure 7.13 presents the resulting error rates as a function of the number of reference samples provided. The mean values of the total error, the FRR and the FAR are plotted, together with their upper and lower interval bounds for a clearer view of the range within which the population mean, with 95% confidence, is likely to lie.

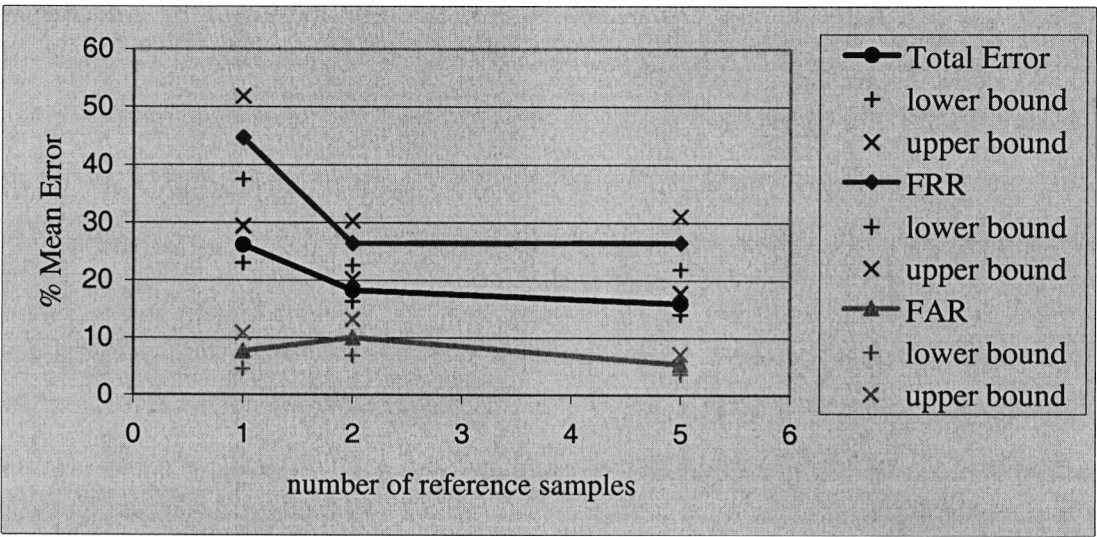


Figure 7.13: Mean error rate performance vs number of reference samples.

As a result the total error seems to decrease substantially (by almost 30%) when a second sample of the original signer is added in display during the verification

process, but only a smaller further reduction (by about 13%) is noticeable when the number of reference samples displayed increases to five. Similarly, the mean FRR displays a dramatic decrease (by about 41%) when one more reference sample is added to the original sample in view, but it remains about the same when three more reference samples – a total of five – are added. The mean FAR is increased, by about 30%, when a second reference sample is added to the one original in display. A further addition of reference samples seems to cause a reduction in the mean FAR, by about 46%, from having two samples to five reference samples in view.

Considering the population statistics, the mean total error is expected to decrease with an increase in the number of reference samples. The FRR is expected to decrease substantially with the addition of one more reference sample, but further reference samples cause no change in the attained rate. Lastly, for the FAR it is uncertain whether there will be an increase, a decrease, or no change with the addition of a second reference sample to the one initially displayed. However, the FAR is expected to decrease when more than two reference samples are displayed.

Therefore, the addition of one more reference sample to the original sample provided, has a considerable effect on the total error and the FRR, whereas a further addition of more reference samples seems to affect mostly the FAR. In particular, providing two original signature samples in view instead of one reduces substantially the mean FRR. This would be expected, since knowledge about the intrinsic variability of signatures becomes directly available. On the other hand, it seems that the mean FAR increases, and this can also be explained on the basis that a second reference sample would result in more tolerance regarding the classification of signatures as ‘genuine’, which in turn would result in more forgeries penetrating the ‘genuine’ class, thus more forgeries being falsely accepted. However, the increase of the mean FAR is not certain with respect to the population mean, and thus a generalisation here should be avoided. On the other hand, further increase in the number of reference samples – a total of five original signature samples in view – brings only a small decrease in the mean total error, along with a substantial decrease in the mean FAR. It seems that having more reference samples in view does not affect the amount of tolerance shown with respect to the ‘genuine’ class judgments, but makes the discrimination of forgeries much easier.

As a consequence, providing two original samples in view may be found beneficial for human inspection of signatures. However, the cost of the addition of more reference samples may be seen as unjustified. The reduction of the mean FRR with the addition of only one more reference sample to the one supplied is substantial, providing a 41% decrease. On the other hand, in the worst case the cost of having a small increase in the already small mean value of the FAR could be tolerated in return to the benefits of a significant reduction in the FRR. After all, humans seem to be relatively good at detecting forgeries but not so good at accepting genuine reference signature samples. It should be noted that the human performance in verifying signatures is generally expected to be substantially better in a test environment – as in the case of these experiments – rather than in a real environment. In practical situations, today, a vast amount of tolerance is shown with respect to the variability of signature samples, so as to avoid the nuisance that would be caused to the genuine signers by the recurring questioning of their samples. This has repercussions in the number of forgeries being accepted, which is increased since the judgement calls are very tolerant. It is proposed that the implementation of a two-reference sample scheme, as for instance on the back of a credit card, could allow more strict verification decisions to be made while providing reassurance that at least no extra nuisance will be caused to the genuine customers.

It is important to mention that small differences regarding the experimental procedures used in the three experiments could perhaps account for a small percentage of the attained performance results. This is because experiments with humans are very sensitive to the procedural layout, as psychological parameters can interfere with the judgmental processes. However, to an extent, psychological effects are an unavoidable phenomenon and they can often be seen in the response variations between individuals. Nevertheless, special care was taken, so as to retain the experimental procedure used as similar as possible between the three experiments. Finally, an extensive comparative analysis of the human verification performance in the three experiments was provided, evaluating the merits of supplying more reference signature samples for the verification of signatures by humans.

7.3 Complexity and Variability vs Verification Performance

An analysis of human performance regarding the verification of the individual target signatures can provide information about its relation to the signatures' degree of complexity and intrinsic variability. In the previous Section it was shown that sufficient knowledge about the intrinsic variability of signatures reduces the attained error rates. The question of how the degree of a signature's complexity or intrinsic variability affects its likelihood of being falsely rejected or of an imitation to pass as its original will be addressed in this Section.

Analysis

The individual error rates for each of the target signatures together with the results of the perceived complexity and perceived intra-class variability of the targets, as reported in previous Chapters, will be used in order to evaluate how these variables are related. The error rates reported in Table 7.20 are obtained from Experiment 1 and correspond to a situation where no knowledge about the intra-class variability of the targets signatures is available. The variations observed in the resulting target error rates can be accounted for by a number of factors: the different signature styles that the targets incorporate, thus relating to the degree of their complexity, the degree of intrinsic variability that the targets embody reflected through the genuine samples included in the test set, and the degree to which the forgery attempts were successful meaning how closely they approximated the originals. With respect to the latter, the skilfulness of the forgeries was unknown to the subjects and therefore it is expected to have been considered equally successful for all the targets. On the other hand, the degree of intrinsic variability of the target signatures plays an indirect role in forming the differences in the individual rates. Since the intrinsic variability of the targets is not known it is assumed to have been regarded by the subjects equally among the targets. However, the targets' intrinsic variability differs substantially, as reported in Chapter 6, and, while reflected in the test samples, it is expected to have shaped the errors accordingly. It should be mentioned that another parameter that may have affected the attained error rates by a small degree is the sequence in which the five targets were displayed and as a result of the considerable length of the experiment this might have partially influenced the subjects' judgement. However, this is considered to have had only a small impact on the judgements of the participants. Therefore, two

main factors may be considered responsible for the differences in the error rates: the degree of complexity and intrinsic variability of the target signatures as perceived by the subjects

Table 7.20: Experiment 1 – Average error rates for the target signatures.

	Total Error %	FRR %	FAR %
Target 1	26.67	46.11	7.22
Target 2	27.50	50.83	4.17
Target 3	21.67	26.39	16.94
Target 4	21.94	38.06	5.83
Target 5	33.06	61.94	4.17

In order to statistically assess whether there is a relationship between the complexity and intrinsic variability of the target signatures with their corresponding error rates, the Spearman rank correlation coefficient r_s was computed as the perceived similarity data were ordinal. The results (Table 7.21) reveal that the perceived intrinsic variability of the targets has a perfect positive correlation with the total error ($r_s=1.000$) and the FRR ($r_s=1.000$), but a negative correlation ($r_s=-0.872$) with the FAR. Furthermore, the rank correlation results obtained between the perceived complexity and the total error ($r_s=-0.400$), the FRR ($r_s=-0.400$) and the FAR ($r_s=0.205$) provide insufficient evidence to determine whether these variables are related. Thus, it is shown that the primary factor responsible for the different error rates obtained is the intra-class variability of the target signatures. Increasing perceived intra-class variability is associated with an increase in the total error, an increase in the FRR and a decrease in the FAR.

Table 7.21: Spearman’s rank correlation coefficient with error rates from Experiment 1.

<i>Spearman's r_s</i>	Perc. Complexity	Perc. Variability	Total Error %	FRR %	FAR %
Perc. Complexity	1.000				
Perc. Variability	-0.400	1.000			
Total Error %	-0.400	1.000**	1.000		
FRR %	-0.400	1.000**	1.000**	1.000	
FAR %	0.205	-0.872	-0.872	-0.872	1.000

** . Correlation is significant at the 0.01 level (2-tailed).

Apart from the significant correlation results, further evidence about the effect of intrinsic variability with respect to the error rates may be presented if a group of targets with similar complexity values is examined. The arrows displayed (Table 7.22) show the direction of change of the variables in question. Indeed, the pattern of change in the error rates with respect to increasing variability, for samples with similar complexity values, is the same as that previously reported. Therefore, increasing variability leads to an increase in the total error, an increase in the FRR and a decrease in the FAR, and thus more genuine signatures are rejected, on the grounds that they differ considerably from the original. However, as a result of the stricter judgements fewer forgeries are falsely accepted.

Table 7.22: Effect of variability in targets with similar complexity values.

		Variability Rank Pos.	Total Error %	FRR %	FAR %
<i>Group of similar complexities</i>	Target 2	4 ↑	27.50 ↑	50.83 ↑	4.17 ↓
	Target 1	3	26.67	46.11	7.22
	Target 3	1	21.67	26.39	16.94 ↓

In a similar way, in order to evaluate the role of complexity in the classification decisions and hence in the error rates obtained, it is essential that the intrinsic variability of the targets is kept the same. For this reason target signatures with a very similar degree of perceived intra-class variability were grouped together, thus forming the ‘stable’ set A scoring low in the perceived variability rank, and the ‘unstable’ set B of higher variability rank judgements. Although the targets in each set have similar perceived intra-class variability, the change in the error rates with increasing complexity is exactly the opposite for the two sets (Table 7.23). For the stable group A increasing complexity causes an increase in the total error, an increase in the FRR and a decrease in the FAR, whereas the opposite takes place for the unstable group B. Thus, with increasing complexity, a decrease is observed in the total error and the FRR, whereas the FAR is increased. It seems that increasing complexity, for the stable set, leads to stricter judgements with respect to the authenticity of signatures, resulting in more genuine signatures being rejected and fewer forgeries being falsely accepted. On the other hand, for the unstable set, the increasing complexity seems to have caused confusion to the subjects leading to more signatures being accepted, both

genuine samples and forgeries. However, the extremely small number of samples included in each group does not at this point validate these generalisations.

Table 7.23: Targets with similar perceived intra-class variability in Experiment 1.

Sets		Perceived Complexity	Total Error %	FRR %	FAR %
A <i>stable</i>	Target 4	8.2 ↑	21.94 ↑	38.06 ↑	5.83 ↓
	Target 3	4.1 ↑	21.67 ↑	26.39 ↑	16.94 ↓
B <i>unstable</i>	Target 1	5.8 ↑	26.67 ↓	46.11 ↓	7.22 ↑
	Target 2	4.8 ↑	27.50 ↓	50.83 ↓	4.17 ↑
	Target 5	1.8 ↑	33.06 ↓	61.94 ↓	4.17 ↑

In a different approach, if the error rates obtained from Experiment 3 are used, the available knowledge about the targets’ perceived intra-class variability should cancel out the actual errors caused by the inherent variability of the target signatures reflected through the test samples. Therefore, the effect of complexity on the error rates would now be obvious. Note that here, the five reference samples are considered enough for acquiring a general view about the targets’ actual intra-class variability. The individual error rates corresponding to each of the five target signatures, with intra-class variability knowledge available, are displayed in Table 7.24. A visual comparison between these error rates with those shown in Table 7.20 shows that a different pattern exists between the error rates and the individual targets in the two experiments.

Table 7.24: Experiment 3 – Average error rates for the target signatures.

	Total Error %	FRR %	FAR %
Target 1	17.32	30.71	3.9
Target 2	17.50	29.29	5.7
Target 3	10.89	13.57	8.2
Target 4	18.39	33.57	3.2
Target 5	15.54	25.36	5.7

The Spearman rank correlation results of the above error rates with the perceived complexity and intra-class variability of the targets (Table 7.25), demonstrate that a significant positive relationship exists between the perceived target complexity and the FRR ($r_s=0.900$), while with respect to the total error and the FAR a positive

(rs=0.800) and a negative (rs=-0.821) correlation coefficient were respectively obtained. On the other hand, no evidence of a correlation between the perceived variability and the total error (rs=0.100), the FRR (rs=0.000) or the FAR (rs=-0.051) were shown. Therefore, as discussed earlier, knowledge of the intra-class variability of the targets, available in Experiment 3, is shown to cancel out the effect of the actual inherent variability of the targets on the error rates.

Table 7.25: Spearman’s rank correlation coefficient with error rates from Experiment 3.

<i>Spearman's rs</i>	Perc. Complexity	Perc. Variability	Total Error %	FRR %	FAR %
Perc. Complexity	1.000				
Perc. Variability	-0.400	1.000			
Total Error %	0.800	0.100	1.000		
FRR %	0.900*	0.000	0.900*	1.000	
FAR %	-0.821	-0.051	-0.821	-0.975**	1.000

*, Correlation is significant at the 0.05 level (2-tailed).

**, Correlation is significant at the 0.01 level (2-tailed).

Note that in both rank correlation tables (Table 7.21 and 7.25) a significant positive relationship is apparent between the total error and the FRR, whereas a negative relationship is regarded with the FAR. In addition, a significant negative relationship is regarded between the FRR and the FAR. The first is justified as the largest part of the total error is accounted for by the FRR, and the second, as also discussed in Chapter 4, a change in the FRR generally leads to a counter-change in the FAR, and vice versa. Another point worth mentioning is the small negative correlation coefficient (rs=-0.400) obtained between the perceived complexity and the perceived intra-class variability, which provides insufficient evidence to whether these two variables are related. This suggests that possibly the perceived intra-class variability of the targets was not particularly influenced by their apparent complexity. However, due to the uncertainty of the result, it would be interesting to consider what a small opposite correlation between the two variables would mean. Consequently, a signature of high perceived complexity could be intuitively seen as slightly more stable than it actually is. Conversely, a rather stable set of genuine samples might show the signature to be slightly more complex than otherwise would be perceived.

Returning to the issue of complexity, it is shown that in Experiment 3 the major influence on the formation of the error rates was caused by the targets’ degree of complexity. The emerging pattern reveals that increasing complexity leads to an increase in the total error and the FRR, but to a reduction in the FAR. This is the actual pattern regarded for the stable set A (Table 7.23) in Experiment 1. Thus, despite the small number of samples in sets A and B, it seems that the actual patterns reported do not appear by chance. The new error rates for the two sets of similar variability ranking judgements (Table 7.26) validate the general complexity pattern reported for Experiment 3.

Table 7.26: Targets with similar perceived intra-class variability in Experiment 3.

Sets		Perceived Complexity	Total Error %	FRR %	FAR %
A stable	Target 4	8.2 ↑	18.39 ↑	33.57 ↑	3.2 ↓
	Target 3	4.1 ↑	10.89 ↑	13.57 ↑	8.2 ↓
B unstable	Target 1	5.8 ↑	17.32 ↑	30.71 ↑	3.9 ↓
	Target 2	4.8 ↑	17.50 ↑	29.29 ↑	5.7 ↓
	Target 5	1.8 ↑	15.54 ↑	25.36 ↑	5.7 ↓

Thus, knowledge about the instability of the targets in set B seems to have balanced back the change in the error rates reflecting the effect of complexity. As a consequence, different reductions or increments in the error rates are evident among the five targets, which is justified in order to accommodate the large shift in the direction of the error rates (from Table 7.23 to Table 7.26). A general reduction in the error rates from Experiment 1 to 3 would have been expected, as analysed in the previous Section. However, the different reductions with respect to the individual targets (Table 7.27) prove the effect of both the complexity and the intrinsic variability of the targets.

Information about the intra-class variability of the targets, which was supplied by the display of 5 reference samples in Experiment 3, seemed to be more beneficial for a simple and unstable target rather than for a complex and stable one. In fact the more complex and stable the target the smaller the percentage of the reductions and hence, the less effect had the knowledge about their intrinsic variability on the error rates. Furthermore, it is apparent from the results that it is the simple signatures that benefit

most from available knowledge about their variability, regardless of whether they are stable or unstable.

Table 7.27: Percentage of reductions affected both by the complexity and the intrinsic variability of the targets.

Description	Sample	Reductions %		
		T.E.	FRR	FAR
<i>complex + stable</i>	Target 4	16.2	11.8	45.1
<i>complex + unstable</i>	Target 1	35.1	33.4	46.0
<i>simple + stable</i>	Target 3	49.7	48.6	51.6
<i>simple + unstable</i>	Target 5	53.0	59.1	-36.7*

*.A negative value indicates an increase.

With the evident merits being reductions in the error rates for targets with combinations of different degrees of complexity and variability, the very unstable signatures with low-to-medium degrees of complexity (Targets 2 and 5), exhibit a significant increase in the FAR instead of a reduction. Nevertheless, it is seen that they generally benefit from available knowledge on their variability, as a result of their significantly large reduction in the FRR, compared to an increase of their relatively small FAR.

Discussion

A number of interesting points may be drawn from this extensive analysis regarding the inherent complexity and variability of signatures as perceived by humans and their relation to the error rates that are likely to occur during a visual verification of their authenticity. According to the experimental results of this study, complex signatures are more likely to lead to more rejections of their genuine samples during a human visual verification and to fewer forgeries being falsely accepted, as opposed to simple signatures. With respect to the intra-class variability of signatures, stable signatures are more likely to have fewer of their genuine samples being falsely rejected, and more forgeries being mistaken for genuine, in comparison to unstable signatures. This is an interesting result since one may have expected that a stable signature would not be the one to suffer from falsely accepted forgeries. However, it seems that a stable signature is more likely to be susceptible to a fraudulent sample being mistaken as genuine, rather than an unstable signature. This can be justified since in situations

where no knowledge about a signature's variability is available, as for instance in Experiment 1, the tolerance shown with respect to the variability of a signature by a human inspector, may be actually larger than the degree to which the stable signature actually varies, and this would lead to generally more samples being authenticated, including prospective forgeries.

Considering both the complexity and intra-class variability of signatures, as determined from the results of this study, a signature that is both complex and unstable would be more susceptible to a fraudulent sample being falsely accepted than a complex and stable, or a simple and unstable signature, would be. This may suggest that the combination of a complex signature with high intra-class variability could lead to more confusion for the subjects, thus resulting in more signatures being accepted, both genuine samples and forgeries, and hence this could be reflected in a low FRR and a high FAR. On the other hand, a signature that is both simple and stable resulted in a greater FAR than any other combination. However, this was not a result of confusion, but rather, as explained earlier, of higher instability intuitively expected.

The previous Section analysed how knowledge about the intrinsic variability of signatures, using as many as 5 reference samples in Experiment 3, results in smaller error rates. In this Section it was explained that by having as many as five reference samples in view for the purpose of visual verification of signatures almost cancels out the actual effect of the intrinsic variability of signatures on the verification performance, and therefore the main factor that affects the attained error rates is the perceived complexity of the signatures. The degree of relative reduction in the error rates of the target signatures, as observed when knowledge about their intra-class variability is available (Experiment 3) compared to when it is not (Experiment 1), reveals that simple signatures benefit more from the availability of this information. Signatures that are both simple and unstable seem to benefit the most in terms of the significant reduction in the total error and the FRR, however, exhibit also a substantial increase in the FAR. On the other hand, complex and stable signatures are the last to benefit from available information about their intrinsic variability. Finally, the simple and stable signature is the one posing the highest likelihood of susceptibility to forgery penetration, when no information about its stability is available. A way to

tackle this would be to demand additional reference samples to be implemented, and this would reduce the false acceptances by as much as 50%, according to the results of this study. However, more reference samples would not favour a significantly unstable signature, if reduction of the FAR is the priority. Being able to measure the perceived complexity and intra-class variability of signatures, issues covered in previous Chapters, could assist in the right implementations being made in order to manage the expected error rates, since the prediction about the likelihood of a false rejection or false acceptance error would now be possible.

7.4 Confidence in Verifying Signatures

According to the three experimental procedures described earlier, the subjects were asked to assign a confidence value along with every classification decision made (i.e. selecting the ‘genuine’ or ‘forgery’ class). The degree of confidence was measured on a scale of 1 to 10, with 1 indicating very little confidence and 10 absolute confidence in the decision taken. This additional task compensates for the lack of a third class marked as ‘unknown’ to which the subjects would resort if uncertain about the other two classes. To avoid the frequent clicking on the prospective ‘unknown’ class, the confidence level bar was employed instead, which ensured that one of the two main classes would be selected, while allowing an indication of the degree of uncertainty involved.

The confidence data, obtained from Experiment 1, resulted in an average confidence of 6.9 with a standard deviation of 1.3 (Table 7.28). The frequency distribution of the subjects’ average confidence (Figure 7.14) showed that a considerable number of 13 participants – out of the total $N=36$ – exhibited an average confidence of around 7 out of 10, whereas as only two had an average confidence as small as about 5, and another two with an average confidence almost reaching 10. Although, at the two extremes, it seems that the degree of confidence reflects a personality trait, in the general case the degree of confidence reported seems to depend on the actual difficulty encountered or the uncertainty involved in making the classification decisions in question.

Table 7.28: Degree of confidence in classification decisions of Experiment 1.

Statistical Parameters	Degree of Confidence
N	36
Mean	6.9
95% CI of Mean	6.5 to 7.3
Standard Deviation	1.3

Furthermore, the Shapiro-Wilk normality test resulted in a w-statistic=0.969 with a very significant p-value=0.398 much greater than the 0.05 level of significance, thus concluding that the data approximate the normal distribution. The 95% confidence interval of the mean – 6.5 to 7.3 – highlights the relatively high confidence of the population in making verification decisions of this type.

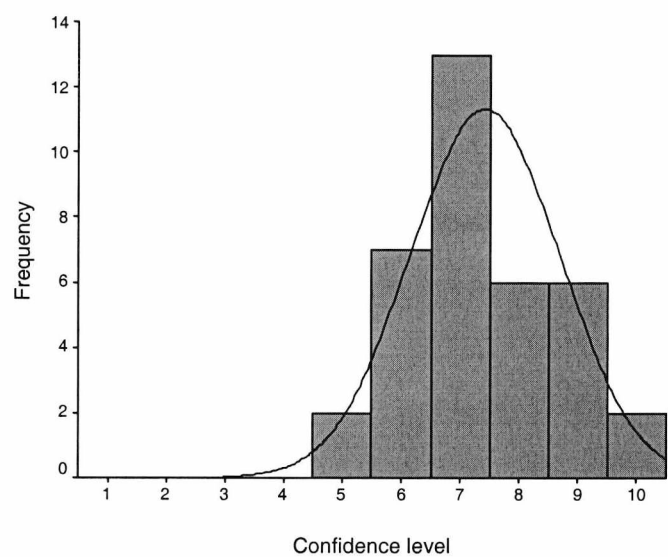


Figure 7.14: Degree of human confidence in verifying signatures.

With respect to the individual target signatures, the average degree of confidence has very similar values (Table 7.29), however, with Target 3 having the smallest. While not being able to make any inferences about significant differences between the means, it is suggested that the average degree of confidence does not seem to depend on the quality³⁹ of the targets, but rather reflecting the difficulty of the individual classification decision. The frequency distributions of the targets (Figure D.1 –

³⁹ Since also the mean confidence values for the targets corresponding to Experiment 2 and 3 (Table D.1 –Appendix D) are very similar. However, for Target 3 there seems to be a small increase in the confidence value compared to that of Experiment 1, which might be due to the high stability of the signature. However there is not enough evidence to support this, since a test for the difference of the means for the five paired sets is not available, and if all pairwise combinations are tested with a paired t-test the significance value will not reflect the overall differences.

Appendix D) resulted in a significant p-value in the Shapiro-Wilk normality test, and hence they exhibit normality.

Table 7.29: Degree of human confidence with respect to the target signatures.

Statistical Parameters	Degree of Confidence				
	Target 1	Target 2	Target 3	Target 4	Target 5
N	36	36	36	36	36
Mean	7.1	7.0	6.6	6.8	7.0
95% CI of Mean	6.6 to 7.5	6.5 to 7.4	6.2 to 7.1	6.4 to 7.3	6.5 to 7.5
Standard Deviation	1.3	1.4	1.4	1.5	1.4
Shapiro-Wilk statistic	0.973	0.982	0.973	0.977	0.948
p-value	0.516	0.811	0.517	0.639	0.091

Furthermore, the degree of confidence reported when deciding upon the authenticity of genuine samples rather than forgeries, shows a distinct difference between the two (Table 7.30). The Shapiro-Wilk normality test resulted again in a significant p-value, and hence both frequency distributions approximate normality (Figure 7.15).

Table 7.30: Degree of confidence for genuine and forged signatures.

Statistical Parameters	Degree of Confidence	
	Genuine	Forgeries
N	36	36
Mean	6.4	7.4
95% CI of Mean	6 to 6.8	7 to 7.9
Standard Deviation	1.2	1.5
Shapiro-Wilk statistic	0.980	0.971
p-value	0.753	0.442

A paired⁴⁰ t-test was performed to test the null hypothesis that the difference between the population means $(\mu_1 - \mu_2) = 0$. With a resulting t-statistic= -6.209, the rejection criteria were met both for a one-tailed test, with $t < -t_{(1)crit.} = -1.69$ and a two-tailed test with $|t| > t_{(2)crit.} = 2.03$. Thus, the null hypothesis is rejected and the alternative is accepted, and therefore $(\mu_1 - \mu_2) < 0$ and $(\mu_1 - \mu_2) \neq 0$ respectively. In both cases the resulting p-value is 0.000, and thus the difference between the means is very

⁴⁰ A paired test is performed when the two variables represent the same group or when there is a pairing in the observations. Here, the same subjects performed both judgements – i.e. classification of genuine signatures and of forgeries.

significant. Therefore, it is concluded that the average confidence when verifying genuine signatures is smaller than that for forgeries, suggesting that people are slightly more sceptical and more uncertain when judging genuine signatures rather than forged samples. This may be explained as forgeries generally appear more dissimilar from the original and, hence, they are perhaps easier to spot, whereas genuine samples, while closely approximating the original, cause more confusion as to whether they are forged or genuine considering that the intrinsic variability of the signatures is unknown and so is the ‘forgery’ class.

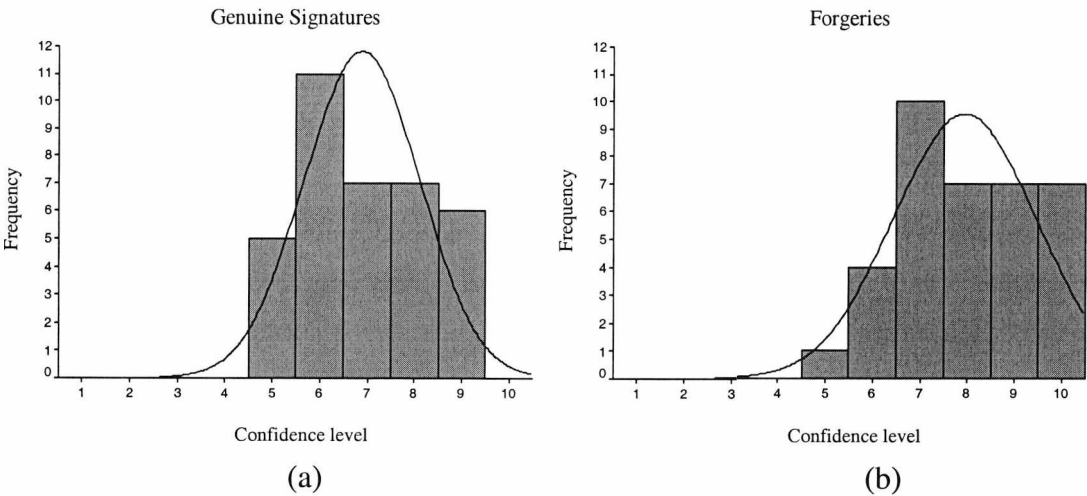


Figure 7.15: Degree of confidence in verifying (a) Genuine signatures, (b) Forgeries.

Confidence vs Error rates

The human performance in verifying signatures could be perhaps improved if the degree of confidence is employed as a rejection mechanism, where decisions below the lower confidence threshold are either discarded or passed to a more specialised verification method. According to [9] “...the objective of rejection is to detect problematic patterns by some means, and refrain from classifying them entirely or redirect them to a reject handler classifier”.

The percentage of rejections for confidence values greater than the specified lower limit is displayed in Figure 7.16. It is obvious that the rejections increase along with the increase of the lower confidence limit, in a non-linear fashion. Starting with a rejection of 0% for confidence values that vary within the whole range, a rejection of

17% is reached as soon as the confidence lower limit is set to 5, whereas an immense 79.5% rejection is resulted for classification decisions made with absolute confidence.

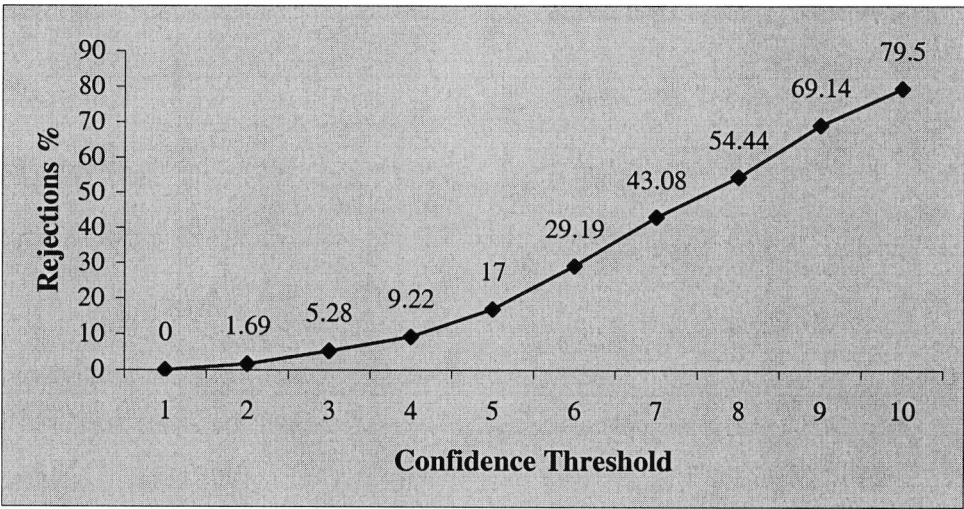


Figure 7.16: Rejection rate vs confidence threshold.

The schematic in Figure 7.17 represents an implementation of the degree of confidence in the verification process. If the confidence value c is greater or equal to the selected threshold θ a decision is taken about the authenticity of the test sample, otherwise the sample is passed to a rejection handler where a more specialized decision is made. The errors corresponding to these two verification stages are denoted as e_v and e_r respectively. The exact implementation of the rejection handler will be later discussed.

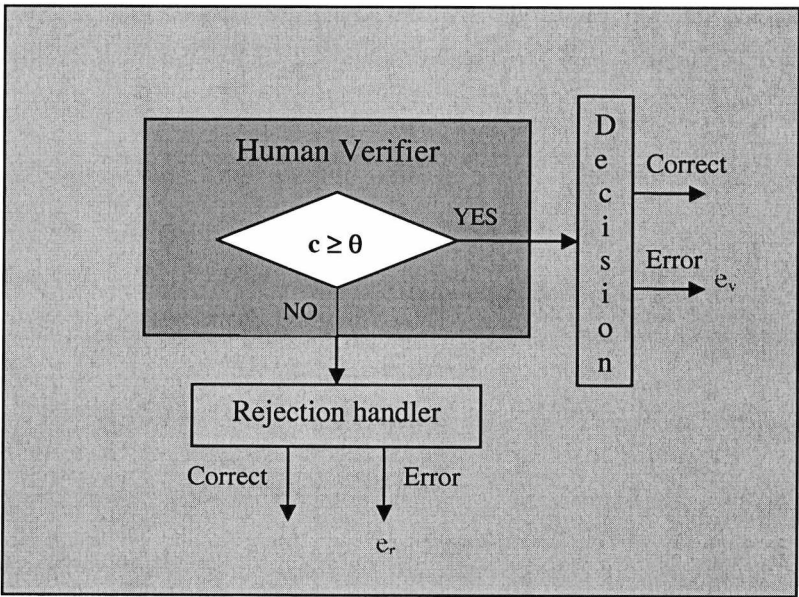


Figure 7.17: Implementation of the degree of confidence in the verification scheme.

With e_v being the first stage verification error as a result of the implementation of the confidence filter, e_v' could represent the relative verification error measured as the errors made with respect to the number of filtered signatures. Figure 7.18 shows the reduction of the two error rates plotted against the lower confidence threshold. The e_v' represents the actual reduction in the total error as a result of the rejection of lower confidence decisions.

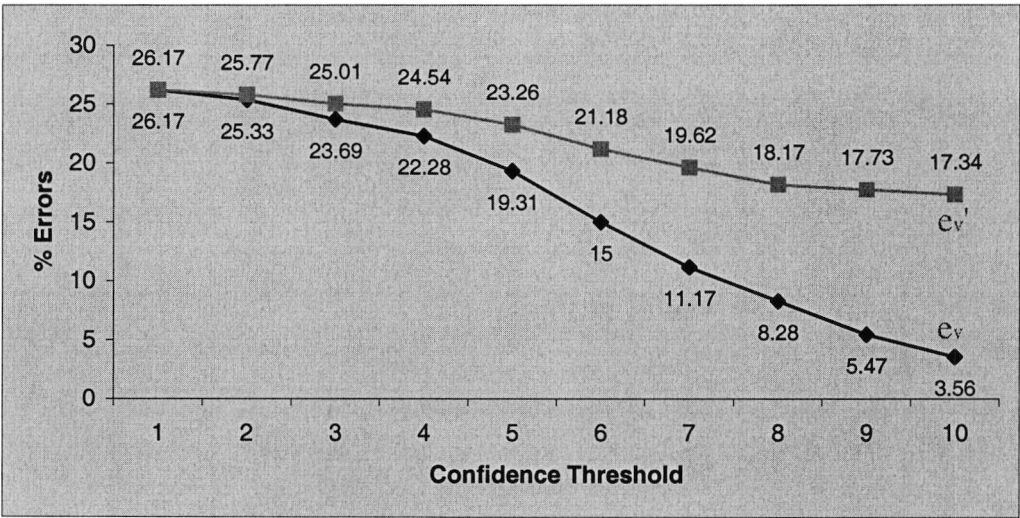


Figure 7.18: Verification error and relative error vs confidence threshold.

According to [9] the total error of the whole verification system will be:

$$e_{total} = e_v(r) + r \cdot e_r \tag{1}$$

where r is the rejection rate and $e_v(r)$ the first stage verification error corresponding to that rejection. It could be assumed that the reject handler utilizes different means for the verification of the problematic signature samples, as for example would be verification through the banking details of a person’s card presented at a cash point. In this case the acquired verification result would be 100% reliable, with an $e_r = 0$. Thus, the total error would be equal to the e_v , and the actual improvement of the systems’ error rate, due to the implementation of the confidence filter, would be considerable (see Figure 7.18) and could be adjusted with selection of the desired rejection rate. In order to gain the most out of the improved verification system, a high confidence threshold would give smaller total error. On the other hand, the nuisance that a high

rejection rate would cause to the customers and to the rejection handler – the bank in the previous example – would outweigh the actual benefits of the rejection.

A different rejection handler scenario that is less demanding and would cause a smaller degree of nuisance is, for example, the case where the teller demands from a client additional proof of their signature, perhaps by inspecting a second card displaying another sample of their signature. This would closely reflect Experiment 2 and its results could help evaluate the benefits of such an implementation. Hence, the rejection error would be equal to the mean error rate of Experiment 2, thus $e_r=18.21\%$. Therefore, the total error rate, given from equation (1), is displayed in Figure 7.19 against the confidence threshold. It is apparent that the total error is almost stable for very high threshold values. A confidence threshold of 5 or 6 would bring a good improvement to the system’s performance.

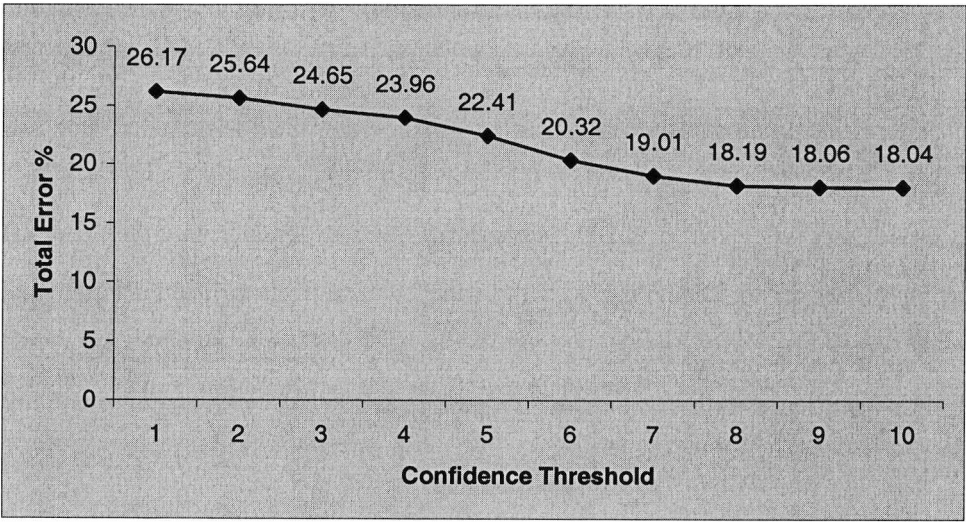


Figure 7.19: Total error (with $e_r=18.21\%$) vs confidence lower threshold.

Finally, a combination of both a very reliable rejection handler, such as the one initially described, and of a less reliable rejection handler, such as the second case reported, together with a desirable confidence threshold could lead to a substantial improvement in the verification of signatures by humans. A simple POS implementation would be a rejection mechanism being controlled by a 50% confidence with the use of the less reliable rejection handler, which requires the display of two reference signatures, whereas the use of the 100% reliable rejection

handler in situations where large transactions take place with respect to an initial confidence either set to a high value or controlled by the scale of the transaction.

In order for these suggestions to have greater value it is important that the FRR and FAR are similarly analysed. There are applications that show more interest in the FRR where the nuisance of the customers is their major concern, whereas others are more interested in the FAR as the reduction of fraud is of major importance.

The rejection rates for the genuine signatures (r_g) and for the forged samples (r_f) as a result of the implementation of the rejection mechanism controlled by the degree of confidence, are displayed in Figure 7.20. It is obvious that the rejection rate of the genuine signatures is increasing faster than that corresponding to the forgeries. This is not especially surprising since, as earlier described, people had generally higher confidence in classifying forgery samples than genuine signatures.

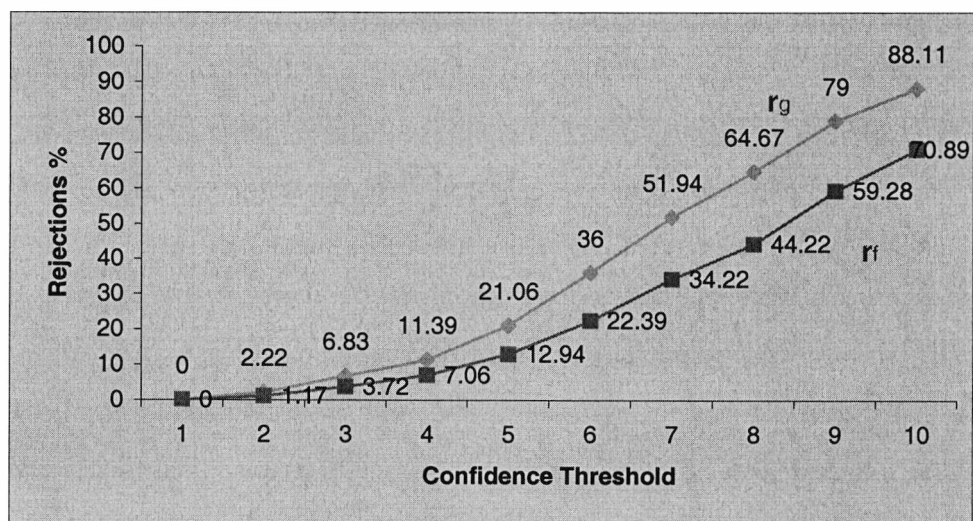


Figure 7.20: Rejection rates of genuine signatures (r_g) and forgeries (r_f) vs confidence threshold.

The resulting FRR of the first stage verification system (FRR_v) and its relative counterpart (FRR_v'), obtained from the genuine signatures falsely rejected with respect to the number of genuine signatures filtered by the confidence threshold, are displayed in Figure 7.21.

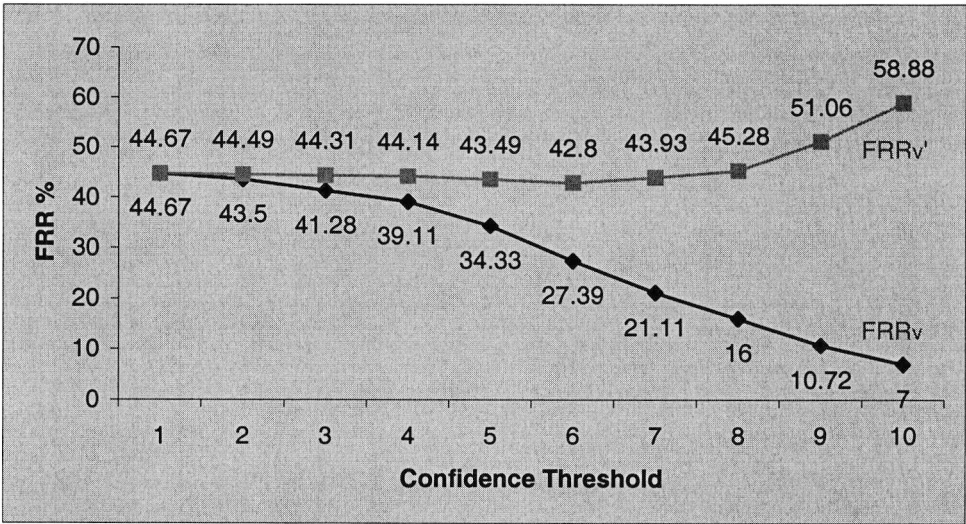


Figure 7.21: First stage FRR_v and the relative FRR_v' vs confidence threshold.

The FRR_v shows a substantial decrease with increasing confidence threshold, whereas the relative FRR_v' seems to remain almost stable with a very small decrease up till the confidence threshold of 6 and for higher confidence values it shows a considerable increase. This unexpected increase in the FRR_v' is because the rate at which the genuine signatures are rejected due to low confidence is greater than the corresponding reductions in the number of signatures falsely rejected as forgeries. Therefore, a substantial number of correct classifications are included in the decisions already rejected due to low confidence. Thus, an important point to note is that increased confidence does not show improvement in the FRR but rather the opposite.

Similarly to the total error analysis the total FRR would be equal to FRR_v in the case where a 100% reliable reject handler exists. However, with respect to the second scenario of a rejection handler closely simulating Experiment 2 the total FRR will be given by:

$$FRR_{total} = FRR_v + r_g \cdot FRR_r$$

(2)

where FRR_r is the false rejection rate of the rejection handler, and according to the resulting mean FRR of Experiment 2, it will be equal to 26.43. The resulting total FRR is displayed in Figure 7.22.

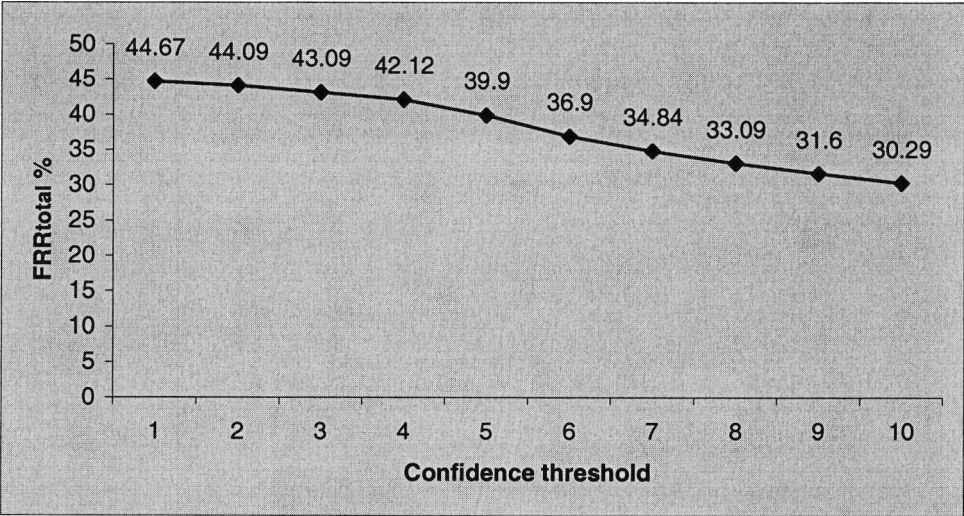


Figure 7.22: Total FRR of the system (with $FRR_t=26.43$) vs confidence threshold.

With respect to the FAR, the resulting first stage FAR_v and the relative FAR_v' , obtained from the forgeries falsely accepted as genuine out of the number of forgeries filtered out by the confidence threshold, are displayed in Figure 7.23.

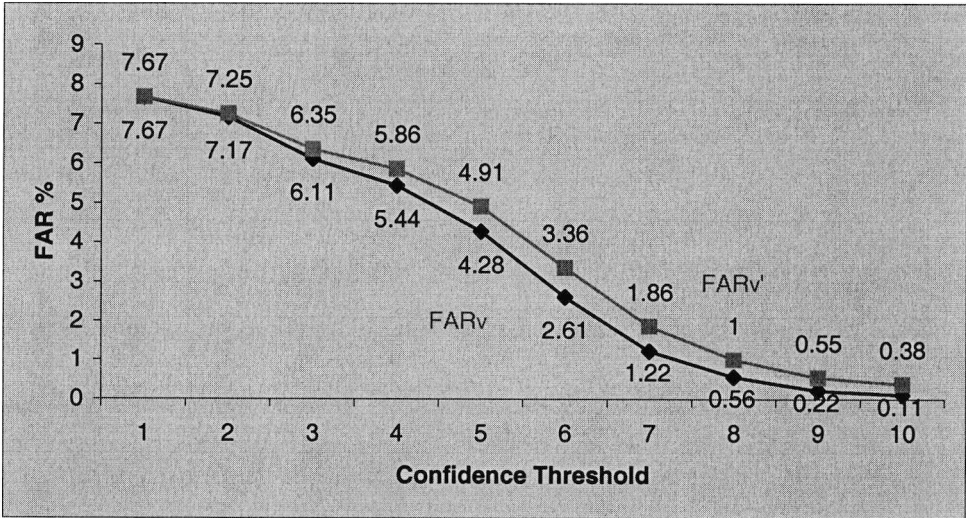


Figure 7.23: First stage FAR_v and the relative FAR_v' vs confidence threshold.

Both curves show a similar reduction with the FAR_v resulting in slightly smaller values for corresponding confidence threshold. Considering the reduction that the FAR_v' exhibits it is concluded that increasing confidence has actually given more accurate classification decisions regarding the forged signature samples. The total FAR if a 100% reliable rejection handler is assumed will be equal to FAR_v , exhibiting

much better rates for increasing confidence values. On the other hand if the second scenario is assumed with data obtained from Experiment 2, the total FAR will be given from the following equation.

$$FAR_{total} = FAR_v + r_f \cdot FAR_r$$

(3)

where $FAR_r = 10$, as given from the mean FAR of Experiment 2. The resulting FAR_{total} curve is displayed in Figure 7.24. It is apparent that for a confidence threshold up to the level of 7 the total FAR decreases, but for greater confidence values a substantial increase in the total FAR is observed. This is a result of the substantial increase in the rejection of forgeries for high confidence values, which multiplied by the considerable FAR_r gives a substantial additional weight to the calculation of the total FAR.

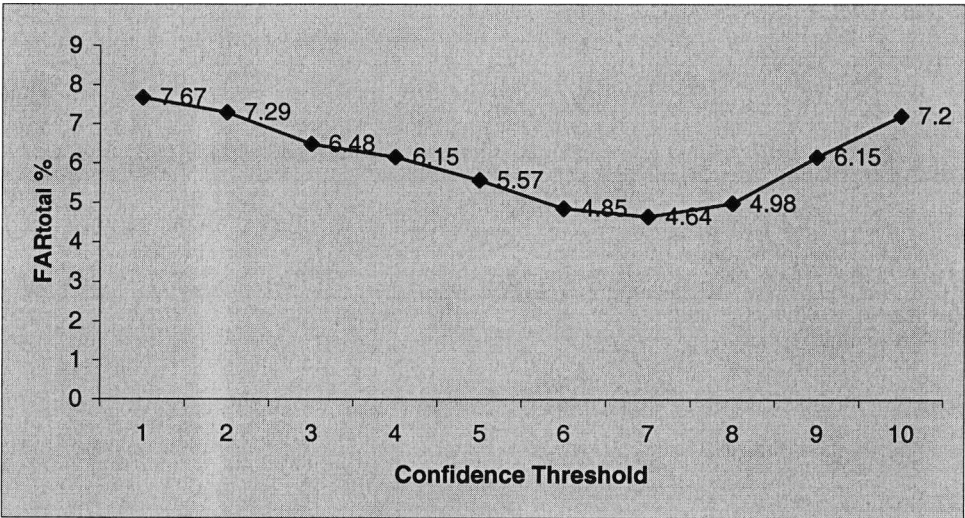


Figure 7.24: Total FAR of the system (with $FAR_r=10$) vs confidence threshold.

It is evident that an improvement can be achieved in the human performance of signature verification if rejection is implemented through the simplicity offered by the human confidence in decision making. It was shown how a “medium” confidence level of 50% in the classification decision taken could reduce the total error by 14.4%, the FRR by 10.7% and the FAR by 27.4%, with the drawback of a second signature sample required in 17% of the cases. Better error rates can be achieved together with slightly smaller rejections, if more reliable sources are included in the final decision. However, this has the major disadvantage of a greater nuisance caused for the same rejection rate. A combination of these two cases with different confidence thresholds

also provides a good solution. Furthermore, a second rejection handler ruled by a second confidence filter may also be implemented within the first rejection mechanism. Thus, for the case of a second reference signature required, another confidence filtering procedure would lead to smaller errors, although with the disadvantage of an additional rejection rate⁴¹ that would redirect the problematic samples to a more reliable second rejection handler. Finally, different combinations of rejection handlers and strength of confidence filtering may be implemented according to the nature and demand of the application in question, providing better error rates while tuning the amount of rejection to an optimum for a particular application level.

7.5 Conclusions

In this chapter, studies of the perceptual tasks of form discrimination and recognition, the human processes of reading and recognising handwritten script, the verification of signatures and the identification of forgeries were initially reviewed. The results obtained from three experiments testing the human performance in verifying handwritten signatures revealed that humans are relatively good in identifying forgeries but not so good in verifying genuine signature samples with an average of approximately 74% correct verification, about 45% FRR and around 8% FAR attained in Experiment 1. In order to tackle the immense FRR, Experiments 2 and 3, providing more reference samples in view, tested the new verification performance which resulted in a substantial reduction in the FRR and a general improvement in the correct classification. The responses obtained from a questionnaire of the perceptually important features for the verification process, gave evidence to common elements compared with those involved in human reading and recognition of handwritten script, while also distinguishing a different approach used in the verification of signatures than the one employed by document experts.

⁴¹ The rejection rates obtained for different confidence threshold values in Experiment 2 are provided by Figure D.2 –Appendix D. In addition the verification error, the FRR and FAR are plotted against the confidence threshold (Figure D.3). The respective graphs based on results of Experiment 3 are also included (Figures D.4-D.5).

An extensive statistical analysis further employed for the comparison of the three experiments, gave sufficient evidence of a reduction in the FRR with the addition of one more reference sample in view, but a further supply of reference samples left the FRR unaffected, while causing a reduction in the FAR. The implementation of a rejection mechanism based on the degree of confidence involved in the verification decisions taken by humans revealed a range of options by which improvement in the error rates may be achieved with the use of different rejection handlers or combinations to suit the needs of different applications. Finally, the effect of the degree of signatures' complexity and intrinsic variability on the verification performance was extensively evaluated, revealing many interesting points. Both increasing complexity and increasing intra-class variability, individually, lead to an increase in the total error, an increase in the FRR and a decrease in the FAR. Furthermore it was suggested that a combination of high complexity and high intra-class variability causes confusion to human subjects, resulting in a low FRR and a high FAR. It was further shown that complex signatures are less susceptible to false acceptance of fraudulent samples than simple ones, but, on the other hand, stable signers face higher risk in having forgeries mistaken as original signatures in situations reflecting common POS applications. Moreover, signatures that are both simple and stable are the ones most susceptible to a false acceptance of their imitation in common applications where no knowledge about their intra-class variability is available. In addition, it was shown that simple signatures seem to benefit more from the reduction in the FRR as a result of the increase in the number of reference samples employed in the verification process, with complex and stable signatures being the last to benefit from knowledge about their intrinsic variability.

In conclusion, significant insight was gained in the previously unvisited area of handwritten signature inspection as carried out by humans, revealing important findings as related to human performance in signature verification and factors that affect it considerably. Knowing the effect that factors, such as signature complexity, intra-class variability and number of reference samples, have on the human inspection performance, while being able to quantitatively measure the complexity and intrinsic variability of signatures, approximating human perceptual judgements, as reported in previous chapters, can give the opportunity for the prediction of a signature's susceptibility to forgery and lead to the implementation of different strategies being

most beneficial for individual signers. The study discussed in this chapter has concentrated on several implementation strategies and predictions, and hence sets the ground rules for future improvement in the human verification of signatures, while raising issues for the integration of the knowledge acquired from human judgements with automated signature verification processes.

The next and final chapter proposes a combined scheme for the integration of knowledge gained from the several human perception experiments carried out in the study of this thesis with the automatic signature verification system developed, drawing some final conclusions.

Chapter 8

Combined Scheme and Conclusions

A comparative analysis between human performance in inspecting the authenticity of handwritten signatures, as examined in Chapter 7, and the automatic signature verification system, as developed in Chapter 4, is initially carried out. Several combined schemes are proposed integrating knowledge and implementations developed in Chapters 4, 5, 6 and 7, offering the possibility of an improvement in the automatic verification of signatures. A subsequent section provides a summary of the work carried out in this thesis, in the context of contributions made in the field. A few general issues involving the static approach employed, limitations related to human experiments, the effect of normalisation procedures, and efficiency of the automatic algorithms implemented are discussed, while further developments are suggested and ideas are given for future work. Final conclusions are drawn setting out a clear direction for the future.

8.1 Human Inspection vs Machine Verification of Signatures

The performance of humans in inspecting the authenticity of handwritten signatures, as evaluated in the previous chapter, may be compared to the automatic verification of

signatures through the static approach developed in Chapter 4. Both cases were evaluated in terms of the Type I and Type II errors obtained for tests with different number of reference samples provided. The FRR and FAR obtained from the human inspection of signatures with 1, 2 and 5 genuine samples in view are displayed in Figure 8.1. Similarly, the attained error rates from the static verification of signatures with reference sets of 3, 5 and 10 samples are provided by Figure 8.2. Although the corresponding error rates –for the human inspection and the automatic verification– differ in their range, it may be observed that the pattern of change in the FRR with the increase in the number of reference samples provided is very similar for the two systems. On the other hand, the pattern of change of the FAR, for the two cases, displays an obvious difference. While for humans a decrease in the FAR is observed for further increase in the number of reference samples provided, the opposite takes place in the automatic approach.

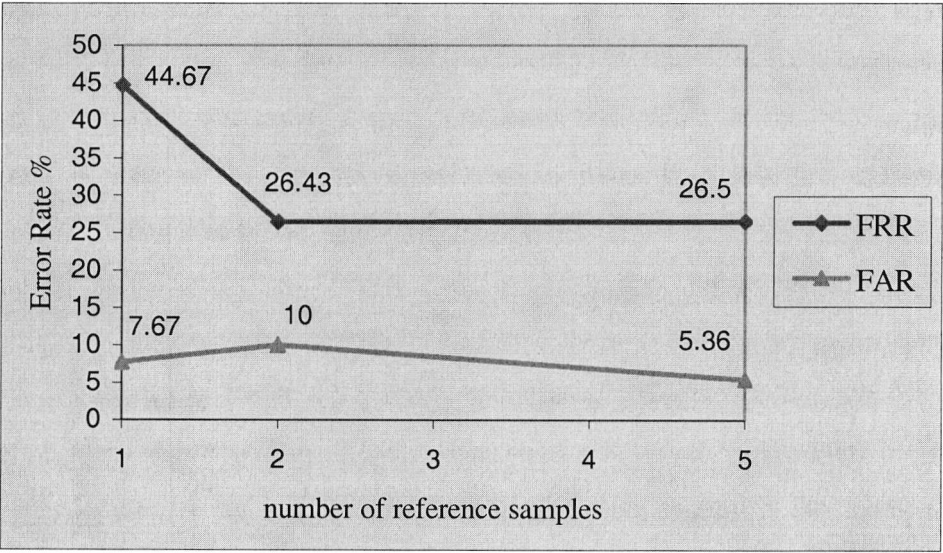


Figure 8.1: Human inspection of signatures –Error rates vs number of reference samples.

A closer observation reveals that the number of reference samples employed in the two cases is different. Although in the static verification of signatures a Limited Set containing only one reference sample was also tested, because of the different approximations enforced regarding the threshold and distance measures, a direct comparison of the error rates attained with those of other reference sets was avoided. Nevertheless, it may be considered that human and machine capacity differs and

hence, while one reference sample may be enough for humans to make an authentication decision, it is certainly not adequate for machine use (see Chapter 4), and conversely while many reference samples may guarantee a smaller FRR in an automatic approach they may cause confusion to humans wherever a fast decision is required.

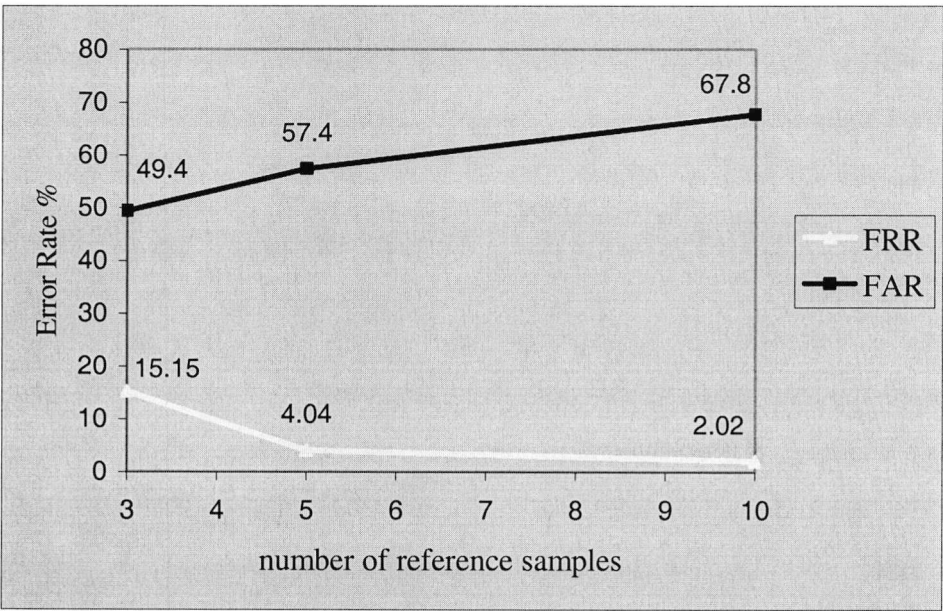


Figure 8.2: Static signature verification –Error rates vs number of reference samples.

The human performance results, in terms of the two error rates, may be compared to the respective performance of the static verifier for a different threshold value as may be realised through the error tradeoff curves corresponding to the different reference sets provided in Chapter 4. Thus, the human performance with only one reference sample in view is roughly comparable to the static verification system with as many as 5 reference samples enrolled, as may be discerned from Figure 4.4 for the respective error rates. Furthermore, the human performance with 5 genuine samples in view is roughly comparable to the system performance with as many as 10 reference samples enrolled, as may be observed from Figure B.7 –Appendix B. It is interesting to note that the performance of the static verifier for both the Limited and Small Reference Sets is much lower than the human performance with only one reference sample in view, as may be observed from Figures B.5 and B.6 –Appendix B for similar error rates.

A few general conclusions may be drawn concerning the human performance in inspecting signatures, as examined in Chapter 7, and the static verification system developed in Chapter 4. Initially, it may be said that the general performance of humans resulted in a relatively high FRR but rather low FAR, whereas the automatic system meeting the design criteria achieved a very low FRR with the cost of a very high FAR. The human performance in inspecting the authenticity of signatures was generally better than the corresponding performance of the static verification system, while achieving lower error rates with fewer reference samples available. It is noteworthy that the human performance with only one reference sample available outperformed considerably the respective static verification performance. In fact, the former is comparable to the machine performance with as many as 5 reference samples enrolled. Moreover, for the human performance there is evidence that a greater number of reference samples can lead to both a decrease in the FRR and a decrease in the FAR, while for the automatic verification it is observed that an increase in the number of reference samples leads to a decrease in the FRR at the cost of a substantial increase in the FAR. Another interesting point is that the human performance with as many as 5 reference samples in view comes close to the dynamic verification performance – based on the on-line verification system developed in Chapter 4 – when only 3 reference samples are used for training. Finally, with respect to the automatic verification performance using random forgeries, even the lowest error rates are incomparable to the corresponding human performance, which is known to produce a perfect result.

Finally, it may be concluded that while the static verification system requires many reference samples to perform comparatively to humans, there is still considerable room for improvement and this could be achieved with the integration of some insight gained from the advanced human inspection capabilities. In the next section the integration of human knowledge, as obtained in the last three chapters, with the static verification system developed in Chapter 4 will be proposed. The advanced human perceptual information will be used in the development of an enhanced individual threshold selection procedure offering a signature verification system approach at a personalized level.

8.2 Human – Machine Integration Scenarios

Two simple approaches may be used to integrate human knowledge, as obtained through the complexity and variability factors developed, into a static verification system: via individualised enrolment decisions or through the implementation of individualized thresholds.

Individualised thresholds

In Chapter 4 the selection of individual threshold values for the target signatures was proposed and the performance of the system in a static verification approach was evaluated. The individual threshold values were calculated according to equation (4.5) simply defined as $\vartheta_w = \alpha \cdot \text{mean}(\text{dist})$ where $\text{mean}(\text{dist})$ is the mean of the distances that each of the reference samples of a target signature exhibit with respect to the mean feature values as formed by the remaining samples in the reference set. The constant α was set to 1.5. However, the performance of the system did not show any improvement when compared to the use of a common threshold adjusted to a more suitable value. In this section, individual values for α will be used for the target signatures incorporating several findings developed in previous chapters. The individual selection of α will be based on the degree of complexity and intrinsic variability of the target signatures. Although it may be argued that the intrinsic variability of a target signature is reflected by the $\text{mean}(\text{dist})$ factor, the individual static signature verification results showed that this was not well described by the above factor. The method for the automatic calculation of the intrinsic signature variability analysed in Chapter 6, which reflects human judgements on the intra-class variability of the targets, is used as the variability factor incorporated in the estimation of α . The second factor implemented is the degree of the target's complexity, estimated with the method developed in Chapter 5. In Chapter 7, the relationship between signature complexity and intrinsic variability with the error rates attained in the human inspection of signatures was extensively assessed and discussed. It was concluded that increasing complexity leads to an increase in the total error, an increase in the FRR and a decrease in the FAR. The same effect on the error rates was observed with increasing intrinsic variability, but only when no knowledge about the intra-class variability of the targets was available. When five reference samples were

provided the effect of perceived variability was found negligible. As shown in the previous section the automatic verification system needs more reference samples to accomplish a performance comparable to the human. Thus, the effect of the intrinsic variability on the error rates attained when 5 reference samples are used with the static verification system may not be negligible. In fact, correlation results on the variability measures (as developed in Chapter 6) and the error rates attained with the static verification system when individual thresholds are used as previously described with common α showed indeed a minimal effect of the variability estimates on the FRR ($r=0.495$), while no effect on the FAR ($r=-0.04$). Note that the intra-class variability of the signatures is already inherent, by different means, in the distance measure and the *mean(dist)* factor of the individual threshold calculation. The corresponding correlation results for the complexity measures (as developed in Chapter 5) showed a significant positive linear relationship between the complexity estimates and the FRR ($r=0.93$) and a negative relationship with the FAR ($r=-0.82$). Thus, if the complexity factor is used to control the FAR, an increase in the FRR will be inevitable. However, the variability factor, having no effect on the FAR, could be used to some extent to control the increase in the FRR. The above reflect a general approach with regard to the total effect on the error rates, as at the individual level any single attempt to decrease any of the two error rates would result to an increase in the other.

The selection of individual α 's is based on the following reasoning. Increasing variability leads to an increase in the FRR and, hence, in order to decrease the FRR a variable signature must have a relaxed threshold in order to accommodate its large intrinsic variability. On the other hand, as increasing complexity leads to a decrease in the FAR, a simple signature exhibiting high FAR would require a stricter threshold in order to prevent the large false acceptance of fraudulent samples. Therefore, the variability and complexity factors may be adequately added for the estimation of α in order to achieve a minimisation in the error rates. In Chapters 5 and 6 the complexity and intrinsic variability of the target signatures was estimated in an automated static approach that closely reflected human judgements. Note that the intrinsic variability of the target signatures was calculated based on the first five reference samples, the same ones used in the static verification of Chapter 4 and the complexity based on the 4th reference sample also included in the reference set used in the automatic system.

The complexity and variability values reported were normalised according to (8.1) to fit into a scale of 0 to 1, in order to be used in the estimation of individual α 's. Thus:

$$Normalisedvalue = \frac{currentvalue - \min(value)}{\max(value) - \min(value)} \tag{8.1}$$

The value of α for a particular signer is estimated according to (8.2) with NV being the normalised variability value and NC the normalised complexity value automatically obtained.

$$\alpha = 1 + (NV + NC) \cdot 50/100 \tag{8.2}$$

Both the complexity and the intrinsic variability factors contribute an equal amount of 50% each in the calculation of α . In other words the combined factor is computed by a simple averaging of the complexity and variability factors. Thus, a very stable and simple signature will have both NV and NC factors small enough to force α to a small value just above 1. On the contrary a very variable and complex signature will have high NV and NC factors allowing a relaxation of the threshold in order to accommodate the high intra-class variability of the signature while ensuring a small FAR, as increasing complexity leads to a decrease in the FAR. The multipliers are chosen so that α will be at least equal to 1 and at most equal to 2, leading the individual threshold to an adequate multiplier for the mean of the distances. The resulting error rates according to this implementation are displayed in Table 8.1.

Table 8.1: Error rates for individual threshold selection Scheme 1 and different α 's.

		Target 1	Target 2	Target 3	Target 4	Target 5	Total
α	FRR %	5.00	5.00	0.00	0.00	0.00	2.02
	FAR %	48.00	31.00	35.00	55.00	79.00	49.60
α'	FRR %	5.00	5.00	0.00	0.00	0.00	2.02
	FAR %	48.00	43.00	41.00	53.00	82.00	53.40
α''	FRR %	5.00	10.00	0.00	0.00	0.00	3.03
	FAR %	48.00	21.00	33.00	57.00	72.00	46.20

It may be easily observed that the results reported in Chapter 4 for individual threshold and constant $\alpha=1.5$ (see Table 4.7), are improved with implementing signer dependent α values based on the complexity and intrinsic variability estimates. More specifically, the total FRR was decreased by 50% and the FAR was reduced by

10.5%. Furthermore, the actual improvement made with the use of the proposed scheme in relation to the use of a common threshold may be measured from the error tradeoff curve Figure 4.4 – which shows that for an FRR of 2.02% the system results in an FAR of 63% – and, therefore, a reduction by 21.3% in the value of the FAR is achieved.

Different weights may also be used to control the contribution of each of the two factors, thus controlling the resulting error rates. As, for instance, with $\alpha' = 1 + NV \cdot 60/100 + NC \cdot 40/100$ which aims at a reduction of the FRR, since the intrinsic variability factor is dominant, or inversely with $\alpha'' = 1 + NV \cdot 40/100 + NC \cdot 60/100$, which aims to a reduction in the FAR, since the complexity factor is now dominant. The corresponding effects on the resulting error rates may be observed in Table 8.1 for the two alternative α values having different weights. It is noteworthy, that the variability factor, as shown in the correlation results, is slightly weak and in the case of α' , although it is dominant, it fails to reduce the FRR any further. Furthermore, the selection of weights may depend on the application and the desirable error rates, since there are applications which require minimisation of the FRR while in others the minimisation of the FAR is of greater importance. A further investigation into the actual effect of the separate factors on the error rates would be needed for a more enhanced selection of weights.

A different scheme is the use of the maximum distance instead of the mean of the distances, as described in Chapter 4, but, as was explained previously, this was found unsuitable for some of the target signatures. The new scheme proposed here for the threshold determination is the use of the maximum distance for unstable signatures since this would assist in the minimisation of the FRR by widening the threshold, and the use of the mean of the distances for stable signatures in order to avoid a very relaxed threshold in situations where only a problematic sample is detected to marginally differ from the rest of the set. Therefore, according to the normalised variability score measured as previously described, if it is smaller than a threshold of $\theta_v=0.5$ the corresponding signature is judged as stable, otherwise as unstable. Note, that this approach overcomes the observed weakness in the variability factor. Hence, for the unstable signatures, the maximum distance is selected for the threshold

estimation i.e. $\vartheta_w = \beta \cdot \max(dist)$, while for the stable signatures the threshold is determined as previously from the mean of the distances i.e. $\vartheta_w = \beta \cdot \text{mean}(dist)$. Furthermore, the parameter β is now defined as $\beta = 1 + NC$ depending only on the normalised complexity score, since the variability score was already used. Table 8.2 displays the resulting error rates for this scheme with β determined by $(1 + NC)$, along with two other options: of β being determined as in (8.2) despite the earlier use of the variability score denoted here as β' , and β'' being assigned the constant value of 1.5 while using only the variability score for the threshold determination. The results show a considerably better performance when β is determined, as earlier proposed, by the factor $(1 + NC)$ for this new scheme where the variability score is used once for the selection of either the $\max(dist)$ or the $\text{mean}(dist)$ for the description of the individual threshold.

Table 8.2: Error rates for individual threshold selection Scheme 2 and different β 's.

		Target 1	Target 2	Target 3	Target 4	Target 5	Total
β	FRR %	5.00	0.00	0.00	0.00	0.00	1.01
	FAR %	60.00	73.00	33.00	64.00	64.00	58.80
β'	FRR %	5.00	0.00	0.00	0.00	0.00	1.01
	FAR %	60.00	94.00	35.00	55.00	92.00	67.20
β''	FRR %	5.00	0.00	0.00	0.00	0.00	1.01
	FAR %	53.00	98.00	66.00	55.00	92.00	72.80

As compared to the initial scheme where the threshold was determined by the mean of the distances i.e. $\vartheta_w = \alpha \cdot \text{mean}(dist)$ and α given by (8.2), the resulting error rates show a smaller FRR but a larger FAR. Thus, the new scheme using both the $\text{mean}(dist)$ and the $\max(dist)$ depending on the variability score resulted in a smaller FRR than the scheme using only the $\text{mean}(dist)$ and both the variability and complexity scores to control α . On the other hand, since the new scheme caused a wider relaxation of the threshold, the FAR increased. Nevertheless, the determination of β as $\beta = 1 + NC$ offered some control over the resulting FAR (see Table 8.2). Finally, this alternative threshold selection scheme gave a further reduction in the FRR with only a small increase in the FAR. Compared to the use of a common threshold, from Figure 4.4 it may be measured that a FRR of 1.01% is followed by a FAR of 75.4%. Therefore, with the alternative threshold selection scheme an actual

reduction in the FAR by 22% is achieved, which is slightly better than the initial scheme (8.2) proposed.

A different approach altogether for the individual threshold determination would be to solely use the complexity and variability factors with a common multiplier cm , thus removing the $mean(dist)$ factor completely. Thus, the individual threshold is defined as: $\vartheta_w = \alpha_w \cdot cm$, where cm is defined according to the common threshold method employed in Chapter 4, but with a different z value ($z=1.65$) –empirically chosen to correspond to a 90% probability, allowing the remaining percentage to be managed by the variability factor– and hence $cm = 1.65 \cdot \sqrt{N}$ where N is the number of features used. It is apparent that this scheme by definition involves less knowledge about the intra-class variability of the signatures. A correlation between the variability measures and the individual error rates for the target signatures under a common threshold reveal that a small negative relationship with the FRR ($r=-0.59$) and a positive relationship with the FAR ($r=0.84$). Thus, similarly to the human performance results, the effect of intrinsic variability on the error rates is greater when less knowledge about its degree is available. In order adequately to incorporate the opposite relationship between the variability factor and the error rates (i.e. negative relationship with the FRR and a positive with the FAR) in the combined factor, the averaging combination method (8.2) was employed with a similarity factor NS instead of the variability factor NV such that: $NS=1-NV$. The resulting error rates are displayed in Table 8.3. The average error rates obtained show a significant improvement with respect to the error rates attained for the common threshold (Table 4.5). In particular the total FRR is reduced by 25% and the total FAR reduced by 20.2%. Furthermore, the results are also comparable with the previous schemes proposed. Note, that a slightly larger value of z could offer a small reduction in the FRR.

Table 8.3: Error rates for individual threshold selection Scheme 3.

	Target 1	Target 2	Target 3	Target 4	Target 5	Total
FRR %	5.00	5.00	0.00	5.00	0.00	3.03
FAR %	45.00	39.00	41.00	46.00	58.00	45.80

In conclusion, three enhanced individual threshold selection schemes were proposed resulting in a significant improvement in the automatic system performance when compared to the use of a common threshold or a simple individual threshold selection technique. The ability to control the output error rates in the automated system was achieved – a fruitful result based on integration of the work carried out in all main chapters of this thesis through a number of human perception experiments and an automated signature verification approach.

Enrolment scheme

Although an individualised enrolment scheme was not efficiently examined, it is suggested that both an individualized threshold selection scheme based on the complexity factor is used and an enrolment decision procedure controlled by the variability factor. Thus, for instance a signer that is judged as ‘unstable’ by the variability factor estimated on 5 reference samples, could be prompted to submit additional samples in order to develop a more adequate template model to encapsulate the given intra-class variability, and thus reduce the FRR, while the complexity factor incorporated in the individualised threshold selection procedure would assist in a reduction of the FAR. However, a further investigation of the actual error rates attained is needed through an examination of the effect of additional samples in conjunction with individual threshold selection schemes proposed. According to the human performance results and perceptual judgments on complexity and intra-class variability it was found that, with respect to human inspection of signatures, it is the simple signatures that benefited the most from available knowledge on their intra-class variability with regard to reductions observed on the FRR. Thus, if an enrolment scheme is to be developed allowing the addition of more reference samples, a deeper investigation on the types of signatures that are more likely to be positively affected by the new scheme with respect to the automatic approach would be needed. Although many of the findings with regard to human perceptual observations work well with respect to the automatic approach, a generalization cannot be made here unless a complete investigation is carried out. This is supported by the opposite relationship regarded with respect to changes occurring in the two error rates, which is strongly evident in the automatic approach. Nevertheless, a possible scheme has been proposed here, offering some options for decision making in the construction of more robust static verification systems.

8.3 Summary of Thesis

A discussion of the need for biometric technology as a means for authenticating the identity of individuals in an attempt to tackle the increased fraudulent activities that are present nowadays was initially presented. The advantages of the handwritten signature over other biometrics for its use in applications such as a point-of-sales environment was discussed and the complex process of signing and forging was further analysed (Chapter 1). Several issues concerning the inherent variability of signatures, the different styles and variations encountered and the existence of forgeries, were approached with respect to potential difficulties presented to automatic verification systems. A review of common techniques, methods and systems developed in a large number of research studies in the field of automatic signature verification was given in Chapter 2. The several methods used were discussed and briefly analysed in the context of the major stages involved in building an automatic signature verification system, and evaluated separately through a static and a dynamic approach. Human performance in visually inspecting the authenticity of handwritten signatures was also examined, from the point of view of expert examiners and non-experts, although the relative studies in the field were extremely limited. The relatively poor performance of static signature verification systems observed, especially with respect to skilled forgeries, highlighted the need for improvement of the static approach in the field. With respect to on-line systems showing relatively high performance, it was found that a number of optional schemes could be considered in order to achieve greater flexibility, aiding their potential for introduction to the competitive market. Finally, the need for a deeper investigation into the human inspection of signatures was emphasised, as human checking procedures remain in common use.

The following chapters were engaged in the development of a novel approach both in terms of an automatic signature verification system implementation but also towards an investigation into human perceptual judgments involved in a number of visual processes related to the inspection of handwritten signatures, aiming at an integration of these two disciplines in an attempt to offer an improvement in the field both as

related to the automatic verification systems built but also with respect to the human inspection procedures used.

The acquisition process of signature data was a first essential step as the construction of a signature database was needed to be used in the design and testing of the automatic signature verification system and the several human perception experiments developed. The acquisition of signatures was carried out dynamically, by means of a graphics tablet device, while the source of the large number of genuine and forged signatures collected was an experimental test with human subjects. The acquisition device used, the format of the data, the experimental protocol, and the digitised signature images are some of the issues discussed in Chapter 3. An automatic signature verification system was developed in Chapter 4 and the several stages involved in the design of the system, enrolment, feature selection, classification and threshold selection procedures, were analysed. The feature extraction process was separately discussed in the corresponding static and dynamic approach. Furthermore, new approaches were proposed, evaluated and discussed. The performance of the system was evaluated for both static and dynamic verification and optimisation methods were extensively assessed. A comparison between the static and the dynamic signature verification results proved the superiority of the dynamic approach, with the advantages of higher feature efficiency, lower computation time, and significantly lower error rates. Finally, the verification results obtained in both the static and the dynamic approach were comparable to the error rates reported in the literature (see relevant systems discussed in Chapter 2). Note that the design considerations employed in the development of the automatic signature verification system reflect a realistic situation, where only few reference samples are available and no forgeries exist to train the system, while also a variety of different signature styles are likely to occur. Considering all the above, it could be said that the verification performance of the system developed is very promising, achieving low error rates compared to similar studies reported in the literature.

A number of human perception experiments were carried out in Chapters 5, 6 and 7, in order to obtain some insight about the two signature characteristics, the complexity and intrinsic variability, that are known to account for the large number of errors in relation to the automatic verification of signatures, and to investigate the possibility of

making predictions about the vulnerability of problematic signatures in terms of the two types of errors.

The complexity of signatures was examined from the point of view of static signature image analysis in Chapter 5. An experimental study with human subjects was carried out in order to assess the perceptual viewpoints and judgements of humans with respect to the degree of complexity inherent in signatures, while five signatures of different styles were employed. Furthermore, a novel method was developed to quantitatively estimate the degree of a signature's complexity reflecting human perceptual criteria. More specifically, the proposed signature complexity measure makes use of the notion of Entropy associated with the signature's image projection histograms, and the results extracted proved to highly correlate with the human perceptual judgements. Finally, a linear regression analysis between the human and the computer estimates of signature complexity arrived in an optimum regression equation that can predict future human responses on signature complexity from the automatically generated complexity values with a good accuracy.

The second important factor examined was the intrinsic variability of signatures. The intra-class similarity of signatures was analysed in Chapter 6, mainly in terms of a comparative relative consistency among the target signers, both in a subjective and an objective approach, and was entirely based on the signatures' shape. Subjective signature variability judgements were obtained through another experiment with human subjects. For the objective measurement several methods, proposed both in this study and in the wider literature, were assessed and the closest to the human perception results and of good discrimination power was employed. The method selected as most suitable was one formerly proposed in the literature, based on Shape Matrices (SM). The application of different degrees of smearing, for boosting the low similarity values, was efficiently assessed and compared, with small degrees of smearing being more beneficial than the selection of high smearing thresholds. The implications of orientation adjustment, based on a baseline detection method suggested to be more suitable for a range of signature styles, were further assessed. Correlation tests showed that the objective (SM) similarity measure was highly related to the subjective similarity results. A low correlation value between the objective variability measurements and the subjective and objective complexity results of

Chapter 5, showed that the objective similarity measure analysed does not depend on the complexity of the signature image. Therefore, the signature complexity and intrinsic variability were approached both through human judgements and also through automated measures that reflected human perception.

The performance of humans in inspecting the authenticity of handwritten signatures was examined through three experiments with human subjects, in Chapter 7. Each experiment was characterised by the availability of a number of genuine samples in view, where only one, two and as many as five original samples were provided corresponding respectively to the three experiments. A comparative analysis of the human performance in terms of the attained error rates in the three cases was performed and conclusions were drawn on the way the number of reference samples affects the subsequent error rates. Furthermore, the relationship between the error rates attained and the complexity and intrinsic variability of signatures was examined, which allowed predictions about the likelihood of a signature's susceptibility to forgery. Moreover, the identification of problematic signers was judged accordingly. Finally, an analysis of the human confidence in making verification decisions of this nature inspired a few rejection scenarios proposed, which were seen capable of assisting the human checking procedures carried out in typical practical scenarios.

A comparison between the static verification performance of the automatic system with that of the human visual inspection of signatures, in Chapter 8, gave evidence of the superiority of human performance being able to achieve lower error rates with fewer reference samples available. The use of knowledge gained from the advanced human perceptual capabilities was then suggested in order to improve the relatively poor performance of the static signature verification system. Therefore, two integration approaches were suggested: the use of individualised enrolment decisions and individualised thresholds. The latter was extensively evaluated through three enhanced individualised threshold selection techniques proposed, which proved to bring a significant improvement in the error rates obtained when compared to both options of using a common threshold and simple individualised thresholds. Finally, it was shown how methods reflecting human perceptual mechanisms may be used to assist automatic approaches into generating lower error rates.

8.4 Discussion

Some general comments and conclusions drawn are discussed here. In the literature review an evaluation of the static verification systems performance confirmed the need for improvement in the field. An automatic signature verification system was developed in this thesis with both a dynamic and a static approach. Unlike the dynamic signature verification, which proved to perform with a relatively high accuracy, the static signature verification showed very poor results. Although, based on the system design criteria, a very low FRR was achieved in the static approach, the FAR was significantly high. It should be noted here that in all situations, where reference to FAR was given, skilled forgeries were assumed unless stated otherwise. Hence, the need for improvement in the static verification of signatures was highlighted. For this reason, a number of human perception experiments were designed and performed and several findings were incorporated into automated processes trying to bring a different dimension into some ill-defined problems. The automated algorithms developed reflecting human perceptual judgements on the complexity and intrinsic variability of signatures employed a static approach of assessing the signature image as a 'whole', thus engaging a universal perception rather than respective processing which could be adequate for signatures of one culture but not of another. This was further supported by the fact that the participants in the experiments were of various nationalities.

Some attention should be drawn to a few limitations related to the human perception experiments. There are a number of factors that can affect to some degree the judgements made by human subjects. According to psychological studies [77], these include the duration of a test image being exposed to the subject, the culture of the subject (semantic attributes), the relative position of the images in an application form (visual syntax), the picture change rate, and a few others indicated here are: the total number of images shown (a large number could lead to possible weariness-fatigue of subjects), the allowance of pauses/breaks, and the sequence of signature images or classes shown. Although these are factors that can affect the subjective decisions made and to some degree are unavoidable, they are nevertheless considered to have had a minimal effect on the results obtained, as a relatively large number of

participants took part in the experiments, and special care was taken into the design of the experiments and the instructions given to avoid as much as possible influential control over the subjects. However, the fact that these experiments are best described as “laboratory” tests might have led to an overestimation of the human performance in inspecting the authenticity of handwritten signatures.

In the several automatic methods developed normalisation was generally not found beneficial. Moreover, it was not deemed necessary since a confined area was provided and inference of the horizontal level was given in the acquisition process of signature data (see Chapter 3), secondly as scale and rotation normalisation can cause distortion on the signature images and thirdly as information embedded in the shape and the orientation of a signature can be viewed as characteristic signature elements during visual inspection. The use of rotation normalisation incorporated in the variability method was not fully supported by the results obtained (see Chapter 6), and size normalisation was not considered based on assessment provided by an existing related study discussed. Although the use of normalised features in an automatic verification system avoids any possible distortion that would be caused by normalisation procedures involving resizing and rotating, experience showed that the use of normalised features in the dynamic signature verification system developed (see Chapter 4), other than related to translation or duration normalisation which are considered essential, such as the ones proposed in [79] were found less efficient than the respective non-normalised ones. This could be explained, for instance, by noting that the efficiency of a feature normalised by the respective maximum value would be degraded as one of the two factors could perform worse than the other, and hence, the situation where the best of the two factors is solely selected would be preferred.

A final comment involves the observation that predictions drawn from human perceptual judgements and methods reflecting human related processes proved to correspond well with respective automatic verification processes leading to considerable improvements. In particular, the complexity factor proved to largely affect the error rates in both the human inspection procedure but also in the automatic verification approach in the same way. With respect to the latter the high efficiency of the complexity measure proposed was demonstrated (Chapter 8). Although the variability measure was not found extremely effective with respect to the automatic

approach, the perceived variability judgements were highly correlated with the human performance results. In general, the human perceptual tests performed and findings drawn were found very beneficial for the static verification approach leading to a significant improvement in the performance and to a better understanding of error generation with respect to individual characteristic signature attributes.

8.5 Novelties and Contributions

A number of novel methods were developed for automatic signature processing and several original investigations were made with respect to human processing of signature information, substantially contributing to the field of handwritten signature verification, and hence are briefly discussed here.

- A method was developed to quantitatively estimate the degree of a signature's complexity based on the static signature image, reflecting human perceptual criteria.
- Human judgemental tasks on signature complexity and relative intra-class variability, based on the static signature images, were investigated.
- Human performance in inspecting the authenticity of handwritten signatures was investigated, especially with respect to the number of available reference samples provided, suggesting strategies for improvement.
- The effects of signature complexity and intra-class variability on the verification performance were extensively evaluated, allowing predictions about the likelihood of a signature's susceptibility to forgery during human signature checking and proposing methods for improvement.
- Human confidence in authentication decisions made during visual inspection of handwritten signature samples was extensively investigated proposing practical scenarios for reduction of errors.
- Several individualized threshold selection schemes were proposed for the integration of human perceptual knowledge on signature complexity and intra-class variability with a static verification system offering improvement in the

automatic approach, while raising issues for improvement in situations where human-computer intervention is required.

- Finally, among the smaller contributions are the use of normalised central moments, up to the order of 4, for the static verification of signatures and the use of a signature baseline detection method more appropriate for shape-like signatures.

An important contribution of this work is to the development of new strategies adapted to individual signers, leading to improvements in human checking applications, automatic signature verification systems and in situations where human-computer interaction is required. Another important area where this work could contribute is in the development of practical strategies and human training schemes for the reduction of signature verification errors during human visual inspection.

8.6 Suggestions for Future Work

Some suggestions may be proposed here for future work. Although the database used contained a large number of signature samples, it is suggested that some more signature types are collected to be used as target signatures, thus including corresponding genuine and forgery samples, for further testing with the complexity and variability algorithms and the static verification system developed. With respect to the acquisition of skilled forgeries it is considered that more time should be given to the imitators for practice, since that could lead to the generation of more dexterous forgeries. Furthermore, it is suggested that some testing is carried out with scanned images rather than signatures reconstructed from the position coordinates acquired dynamically, both with respect to experiments with humans but also with the automated processes proposed. This would require the application of more preprocessing methods (see Chapter 2), but would also lead to perhaps more accurate perceptual judgements, since information such as pen pressure would be apparent through an inspection of signatures written on paper, while, on the other hand, it could be measured automatically from the signature's grey levels.

With respect to the static signature verification system developed, an investigation into the extraction of more representative signature features – that are ideally both highly stable and discriminative – could be carried out aiming at a further improvement of the performance achieved. A suggestion is given for the examination of more simple moment invariants based on the signature image projections, while also envelope characteristics or more advanced global and local features could be assessed. The use of a combination of different signature representation types is also proposed.

Although the intra-class variability method adopted seemed to highly reflect human perceptual judgements and highly correlate with the human inspection errors obtained, it did not appear optimal for its use in the automatic approach – although it already brought an improvement in the performance. As this method was selected among several different intra-class variability measures, a small enhancement could be considered. The assessment of a different smearing process to be integrated in the variability measure is further proposed, involving both vertical and horizontal smearing combined with the logical AND function, a method that has been used before for document processing [71], since this may better reflect visual inspection processes carried out by humans as it is based on both vertical and horizontal directions.

Finally, it is suggested that a deeper investigation into different integration scenarios, involving the merging of knowledge obtained from human experiments with the automated verification processes, is carried out. An individualised enrolment decision scheme was already proposed and described. Perhaps a different approach to the *mean(dist)* and *max(dist)* factors could be also considered for the individualised threshold selection method, providing a better basis for the integration of the complexity and variability factors earlier proposed in this chapter.

8.7 Conclusions

The general approach followed in this thesis was the construction of a static and a dynamic signature verification system, concentrating on the former which showed a

significant weakness with respect to skilled forgeries. An improvement in the performance was sought in the highest ‘model’ that has always inspired the construction of machines: the human. Experimental tests with human subjects were carried out in order to define the two important signature attributions, complexity and intrinsic variability, that are considered to widely affect the performance of verification systems, and implement methods to quantitatively measure them. Further experimental tests on the human inspection of signatures’ authenticity were carried out, allowing a number of observations and predictions to be made with respect to several aspects that affect the two types of errors, including the number of reference samples provided, the degree of perceived complexity and intrinsic variability of signatures, and the amount of human confidence involved in signature checking procedures. The implementation of algorithms that can quantitatively measure a signature’s degree of complexity and intrinsic variability reflecting human perceptual judgements allowed predictions to be made about the susceptibility of a signature to forgery and false rejection. In this respect, these automated processes may be found useful to predict and consequently attempt to avoid errors occurring in situations where signature checking is carried out by humans. A few rejection scenarios were proposed based on human confidence levels during signature inspection, which showed an improvement in the error rates obtained and therefore, both prediction and reduction of errors is now possible.

However, the assessment of the human experimental studies described here was not limited to the improvements that they can offer to the human inspection of signatures, but were also evaluated in terms of improvement that they can bring to the automatic signature verification processes. In this respect, integration scenarios were proposed attempting to combine knowledge gained from the human inspection of signatures with the static verification approach developed. The signature complexity and intrinsic variability methods developed were integrated with the static signature verification process through the selection of individualised thresholds. Thus, a relaxation or restriction of the individual thresholds was applied based on an individualised factor determined by the normalised complexity and variability scores. The use of individualised enrolment decisions in conjunction with individualised thresholds was also suggested exploiting findings based on an increase in the number of available reference samples in relation to the complexity and variability factor.

In conclusion, inspired by the advanced human perceptual capabilities this study achieved improvements in two directions: human inspection and automatic verification of signatures. Several findings related to the human perceptual judgements evaluated in this study can be used to generate advanced training procedures to enhance human inspection strategies with respect to the authenticity of handwritten signatures. Furthermore, several of these findings can be used as an inspiration in the construction of more robust automatic signature verification systems with a significant improvement in performance. A final comment involves the realisation that predictions drawn from human perceptual judgements and methods reflecting human related processes proved the existence of a parallelism with respective performance outcomes depicted from static signature verification processes, which suggests that related human perceptual studies can offer considerable improvements in automatic processes, while promising a lot for the future. The following words of Lipkin [83] are used here in conclusion:

“Knowledge of what features of the object are useful to the human might suggest generally useful machine algorithms. But, obviously, it is unlikely that this will lessen the search for algorithms that 'work' whether or not they bear any resemblance to human perception. Nevertheless, understanding how the human 'sees' a picture is important in determining the utility of an algorithm no matter how it was derived.”

References

- [1] KAPPA signature verification public trials and public survey on biometrics. 1994. BTG plc, 101 Newington Causeway, London SE1 6BU. 1994.
- [2] Cyber-SIGN for Electronic Signatures. www.cybersign.com . 2000. Cyber SIGN Inc.
- [3] Automatic signature verification. www.app-davos.ch/ . 2001. App Informatik Davos.
- [4] Electronic signatures for the internet. www.cic.com . 2001. Communication Intelligence Corporation.
- [5] Signature capture, storage, verification and recognition. www.datavisionimage.com . 2001. DATAVISION Corporation.
- [6] Credit and debit card fraud reaches record high. www.apacs.org.uk [Press Release]. 5-11-2002. Association for Payment Clearing Services (APACS).
- [7] UltraPen - How it works. ftp://ftp.wacom-europe.com/pub/white_papers/tech_ultrapen_uk.pdf . 2002. WACOM Co. Ltd.
- [8] Achemlal, M., Mourier, M., Lorette, G., and Bonnefoy, J.-P., "Dynamic signature verification," in Grissonnanche, A. (ed.) *Security and protection in information systems* Elsevier Science Publishers B.V. (North-Holland), IFIP, 1989, pp. 381-389.
- [9] Aksela, M., Laaksonen, J., Oja, E., and Kangas, J. Rejection methods for an adaptive committee classifier. [Proc. 6th Int. Conf. Document Analysis and Recognition]. 2001. Seattle.
- [10] Allgrove, C, "A study of automatic signature verification techniques in the context of a practical document processing application." PhD Thesis, University of Kent, Canterbury, UK, 1999.
- [11] Allgrove, C. and Fairhurst, M. C. Enrolment model stability in static signature verification. [Proc. 7th Int. Workshop on Frontiers in Handwriting Recognition], 565-570. 2000. Amsterdam, International Unipen Foundation.
- [12] Ammar, M., Yoshida, Y., and Fukumura, T. A new effective approach for off-line verification of signatures by using pressure features. [Proc. 8th Int. Conf. on Pattern Recognition], 566-569. 1986. Paris.
- [13] Ammar, M., Yoshida, Y., and Fukumura, T., "Feature extraction and selection for simulated signature verification," in Plamondon, R., Suen, C. Y., and Simner, M. L. (eds.) *Computer Recognition and Human Production of Handwriting* World Scientific, 1989, pp. 61-76.
- [14] Ammar, M., Yoshida, Y., and Fukumura, T., "Structural description and classification of signature images," *Pattern Recognition*, vol. 23, no. 7, pp. 697-710, 1990.
- [15] Anderson, C. Document Examination. www.docexam.co.au/download.htm#Document . 2002.

- [16] Arnoult, M. D., "Prediction of perceptual responses from structural characteristics of the stimulus," *Percept.Mot.Skills*, vol. 11 pp. 261-268, 1960.
- [17] Attneave, F., "Dimensions of similarity," *American Journal of Psychology*, vol. 63 pp. 516-556, 1950.
- [18] Attneave, F., "Physical determinants of the judged complexity of shapes," *J.Exp.Psychol.*, vol. 53, no. 4, pp. 221-227, 1957.
- [19] Bae, Y. J. and Fairhurst, M. C. Parallelism in dynamic time warping for automatic signature verification. [Proc. 3rd Int. Conf. on Document Analysis and Recognition], 426-429. 1995. Montréal, Canada, IEEE Computer Society.
- [20] Bajaj, R. and Chaudhury, S., "Signature verification using multiple neural classifiers," *Pattern Recognition*, vol. 30, no. 1, pp. 1-7, 1997.
- [21] Baltzakis, H. and Papamarkos, N., "A new signature verification technique based on a two-stage neural network classifier," *Engineering Applications of Artificial Intelligence*, vol. 14, no. 1, pp. 95-103, 2001.
- [22] Barrière, C. and Plamondon, R., "Human identification of letters in mixed-script handwriting: an upper bound on recognition rates," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 28, no. 1, pp. 78-82, 1998.
- [23] Bauer, F. and Wirtz, B. Parameter reduction and personalized parameter selection for automatic signature verification. [Proc. 3rd Int. Conf. on Document Analysis and Recognition], 183-186. 1995. Montréal, Canada, IEEE Computer Society.
- [24] Bevington, P. R., *Data reduction and error analysis for the physical sciences* New York: McGraw-Hill Book Co., 1969.
- [25] Brault, J. J. and Plamondon, R. How to detect problematic signers for automatic signature verification. [Int. Carnahan Conf. on Security Technology: Electronic Crime Countermeasures, Zurich, Switzerland], 127-132. 1989.
- [26] Brault, J. J. and Plamondon, R. Can we measure the intrinsic difficulty of a signature to be imitated? [Proc. 5th Handwriting Conf.], 169-171. 1991. Tempe, Arizona.
- [27] Brault, J. J. and Plamondon, R., "A complexity measure of handwritten curves: Modeling of dynamic signature forgery," *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 2, pp. 400-413, 1993.
- [28] Brault, J. J. and Plamondon, R., "Segmenting handwritten signatures at their perceptually important points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 953-957, 1993.
- [29] Brink, A. D., "Using spatial information as an aid to maximum entropy image threshold selection," *Pattern Recognition Letters*, vol. 17 pp. 29-36, 1996.
- [30] Brittan, P. and Fairhurst, M. C., "An approach to handwritten signature verification using a high performance parallel architecture," in Impedovo, S. and Simon, J. C. (eds.) *From Pixels to Features III: Frontiers in Handwriting Recognition* Elsevier Science Publishers B.V., 1992, pp. 385-390.
- [31] Brown, D. R. and Owen, D. H., "The metrics of visual form: Methodological dyspepsia.," *Psychological Bulletin*, vol. 68 pp. 243-259, 1967.
- [32] Campbell, A., *Forensic Science: Evidence, Clues, and Investigation* Chelsea House Publishers, 2000, pp. 67-69.

- [33] Cardot, H., Revenu, M., Victorri, B., and Revillet, M.-J., "A static signature verification system based on a cooperating neural networks architecture," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 39-52.
- [34] Chollet, G. Speaker verification. [BMVA Symposium on Advancing Biometric Technologies]. Royal Statistical Society, London. 2002.
- [35] Cobbah, W. G. K., "The computer assessment of handwriting and drawing movement in Parkinsonism." PhD Thesis, University of Kent, Canterbury, UK, 2001.
- [36] Congedo, G., Dimauro, G., Forte, A. M., Impedovo, S., and Pirlo, G. Selecting reference signatures for on-line signature verification. [Proc. 8th Int. Conf. on Image Analysis and Processing], 521-526. 1995. San Remo, Italy, Springer.
- [37] Congedo, G., Dimauro, G., Impedovo, S., and Pirlo, G., "Off-line signature verification by fundamental components analysis," in Impedovo, S. (ed.) *Progress in Image Analysis and Processing III* Proc. 7th Int. Conf. on Image Analysis and Processing ed. Capitolo, Monopoli, Italy: World Scientific, 1994, pp. 331-337.
- [38] Cootes, T. F., Kang, H., Wheeler, G., Butcher, L., and Taylor, C. J. Face recognition using active appearance models. [BMVA Symposium on Advancing Biometric Technologies]. Royal Statistical Society, London. 2002.
- [39] Côté, M., "Utilisation d'un modèle d'accès lexical et de concepts perceptifs pour la reconnaissance d'images de mots cursifs." Ph.D. Thesis, Ecole nationale supérieure des télécommunications de Paris, France, 1997.
- [40] Côté, M., Cheriet, M., Lecolinet, E., and Suen, C. Y. Automatic reading of cursive scripts using human knowledge. I[Proc. 4th Int. Conf. on Document Analysis and Recognition], 107-111. 1997. Ulm, Germany, IEEE Computer Society.
- [41] Côté, M., Cheriet, M., Suen, C. Y., and Lecolinet, E. Détection des lignes de référence de mots cursifs à l'aide de l'entropie.
<http://www.cenparmi.concordia.ca/CENPARMI/publications.html> . 2000. CENPARMI.
- [42] Côté, M., Lecolinet, E., Cheriet, M., and Suen, C. Y., "Automatic Reading of Cursive Scripts Using a Reading Model and Perceptual Concepts," *International Journal on Document Analysis and Recognition*, vol. 1, no. 1, pp. 3-17, 1998.
- [43] Crane, H. D. and Ostrem, J. S., "Automatic signature verification using a three-axis force-sensitive pen," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, no. 3, pp. 329-337, 1983.
- [44] Daugman, J. Iris recognition - update on algorithms and trials. [BMVA Symposium on Advancing Biometric Technologies]. Royal Statistical Society, London. 2002.
- [45] Dimauro, G., Impedovo, S., Modugno, R., Pirlo, G., and Sarcinella, L. Analysis of stability in hand-written dynamic signatures. [8th Int. Workshop on Frontiers in Handwriting Recognition]. 2002. Ontario, Canada.
- [46] Dimauro, G., Impedovo, S., and Pirlo, G., "A stroke-oriented approach to signature verification," in Impedovo, S. and Simon, J. C. (eds.) *From Pixels to Features III: Frontiers in Handwriting Recognition* Elsevier, 1992, pp. 371-384.
- [47] Dimauro, G., Impedovo, S., and Pirlo, G., "Algorithms for automatic signature verification," in Bunke, H. and Wang, P. S. P. (eds.) *Handbook of Character Recognition and Document Image Analysis* World Scientific Publishing Company, 1997, pp. 605-621.
- [48] Eriksen, C. W. and Hake, H. W., "Multidimensional stimulus differences and accuracy of discrimination," *J. Exp. Psychol.*, vol. 50, no. 3, pp. 153-160, 1955.

- [49] Fairhurst, M. C., "Signature verification revisited: promoting practical exploitation of biometric technology," *Electronics and Communication Engineering Journal*, vol. 9 pp. 273-280, 1997.
- [50] Fairhurst, M. C. and Brittan, P., "An evaluation of parallel strategies for feature vector construction in automatic signature verification systems," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 21-38.
- [51] Fairhurst, M. C. and Kaplani, E. Strategies for exploiting signature verification based on complexity estimates. [BMVA Symposium on Advancing Biometric Technologies]. Royal Statistical Society, London. 2002.
- [52] Fairhurst, M. C., Kaplani, E., and Guest, R. M. Complexity measures in handwritten signature verification. [Proc. 1st Int. Conf. on Universal Access in Human-Computer Interaction], 305-309. 2001. New Orleans.
- [53] Franke, K. and Köppen, M., "A computer-based system to support forensic studies on handwritten documents," *Int. J. Doc. Anal. & Recogn.*, vol. 3 pp. 218-231, 2001.
- [54] Ghali, A., Daemi, M. F., and Mansour, M., "Image structural information assessment," *Pattern Recognition Letters*, vol. 19 pp. 447-453, 1998.
- [55] Goldstein, A. G. and Andrews, J., "Perceptual uprightness and complexity of random shapes," *American Journal of Psychology*, vol. 75 pp. 667-669, 1962.
- [56] Gonzalez, R. C. and Woods, R. E., *Digital Image Processing* New Jersey: Prentice-Hall Inc., 2002.
- [57] Goshtasby, A., "Description and discrimination of planar shapes using shape matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 6, pp. 738-743, 1985.
- [58] Gupta, G. and McCabe, A. A review of dynamic handwritten signature verification. http://cay.cs.jcu.edu.au/~alan/Work/HSV-Lit_rev.html . 1997.
- [59] Gupta, G. K. and Joyce, R. C. A Study of Shape in Dynamic Handwritten Signature Verification. 97/04. 1997. Townsville, Australia, Department of Computer Science, James Cook University.
- [60] Gurney, K., *An introduction to neural networks* UCL Press. (Quote obtained from: <http://www.shef.ac.uk/psychology/gurney/notes/index.html>), 1997.
- [61] Hake, H. W., "Form discrimination and the invariance of form," in Uhr, L. (ed.) *Pattern recognition: Theory, experiment, computer simulations, and dynamic models of form perception and discovery* New York: John Wiley & Sons, Inc., 1966, pp. 142-173.
- [62] Hastie, T., Kishon, E., Clark, M., and Fan, J. A model for signature verification. AT&T Bell Laboratories. 1992.
- [63] Herbst, B. and Coetzer, H. On an offline signature verification system. [Proc. 9th Annual South African Workshop on Pattern Recognition], 39-43. 1998.
- [64] Herbst, B. and Richards, D. On an automated signature verification system. [Proc. Int. Symposium on Industrial Electronics], 600-604. 1998. Pretoria, IEEE.
- [65] Huang, K. and Yan, H., "Off-line signature verification based on geometric feature extraction and neural network classification," *Pattern Recognition*, vol. 30, no. 1, pp. 9-17, 1997.

- [66] Hurley, D. J., Nixon, M. S., and Carter, J. N. Automatic ear recognition by force field transformations. [Colloquium on Visual Biometrics]. 2000. Savoy Place, London WC2R 0BL, the IEE.
- [67] Ismail, M. A. and Gad, S., "Off-line arabic signature recognition and verification," *Pattern Recognition*, vol. 33 pp. 1727-1740, 2000.
- [68] Jain, A. K., Griess, F. D., and Connell, S. D. On-line signature verification. <http://www.elsevier.com/locate/patcog> [Pattern Recognition, Article 1629]. 2002. Elsevier Science Ltd.
- [69] Jain, A. K., Hong, L., and Bolle, R., "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 302-314, 1997.
- [70] Jain, A. K., Ross, A., and Pankanti, S. A prototype hand geometry-based verification system. <http://citeseer.nj.nec.com/cs>. 1999. NEC Research Institute 2002.
- [71] Kaballieratou, E., "Σχεδίαση και ανάπτυξη συστήματος αυτόματης επεξεργασίας εγγράφου και αναγνώρισης χειρόγραφων χαρακτήρων συνεχόμενης γραφής, ανεξάρτητο συγγραφέα." (System design and development for automatic document processing and recognition of handwritten characters of continuous writing, independent of the writer). PhD Thesis, University of Patras, Greece, 2000.
- [72] Kaplani, E. and Fairhurst, M. C. A technique for estimating handwritten signature complexity based on human perceptual criteria. [Proc. PREP 2002 Conference]. 2002. Nottingham.
- [73] Kim, D. H., Kim, E. J., and Bang, S. Y., "A variation measure for handwritten character image data using entropy difference," *Pattern Recognition*, vol. 30, no. 1, pp. 19-29, 1997.
- [74] Kim, J., Yu, J. R., and Kim, S. H., "Learning of prototypes and decision boundaries for a verification problem having only positive samples," *Pattern Recognition Letters*, vol. 17 pp. 691-697, 1996.
- [75] Kim, S. H., Park, M. S., and Kim, J. Applying personalized weights to a feature set for on-line signature verification. [Proc. 3rd Int. Conf. on Document Analysis and Recognition], 882-885. 1995. Montréal, Canada, IEEE Computer Society.
- [76] Kimura, F., Kayahara, N., Miyake, Y., and Shridhar, M. Machine and human recognition of segmented characters from handwritten words. [Proc. 4th Int. Conf. Doc. Anal. & Recogn.], 866-869. 1997. Ulm-Germany, IEEE Computer Society.
- [77] Kolers, P. A., "The role of shape and geometry in picture recognition," in Lipkin, B. S. and Rosenfeld, A. (eds.) *Picture Processing and Psychopictorics* Academic Press, Inc., 1970, pp. 181-202.
- [78] Leclerc, F. and Plamondon, R., "Automatic signature verification: The State of the Art - 1989-1993," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 3-19.
- [79] Lee, L. L., Berger, T., and Aviczer, E., "Reliable on-line human signature verification systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 643-647, 1996.
- [80] Lee, S. and Pan, J. C., "Offline tracing and representation of signatures," *IEEE Trans. Syst. Man Cybern.*, vol. 22, no. 4, pp. 755-771, 1992.
- [81] Lejtman, D. Z. and George, S. E. On-line handwritten signature verification using wavelets and back-propagation neural networks. [Proc. 6th Int. Conf. on Document Analysis and Recognition]. 2001. Seattle. 2001.

- [82] Leung, L. W., King, B., and Vohora, V. Comparison of image data fusion techniques using Entropy and INI. [Proc. 22nd Asian Conf. on Remote Sensing, Singapore]. 2001. CRISP, SISV, AARS.
- [83] Lipkin, B. S., "Introduction: Psychopictorics," in Lipkin, B. S. and Rosenfeld, A. (eds.) *Picture processing and psychopictorics* New York: Academic Press Inc., 1970, pp. 3-36.
- [84] Liu, C. N., Herbst, N. M., and Anthony, N. J., "Automatic signature verification and field test results," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, no. 1, pp. 35-38, 1979.
- [85] Loncaric, S., "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983-1001, 1998.
- [86] Lorette, G., "Handwriting recognition or reading? What is the situation at the dawn of the 3rd millenium?," *Int. J. Doc. Anal. & Recogn.*, vol. 2, no. 1, pp. 2-12, 1999.
- [87] Lorette, G. and Plamondon, R., "Dynamic approaches to handwritten signature verification," in Plamondon, R. and Leedham, C. G. (eds.) *Computer Processing of Handwriting* World Scientific, 1990, pp. 21-47.
- [88] Machado, A. M. C., Campos, M. F. M., Siqueira, A. M., and De Carvalho, O. S. F. An iterative algorithm for segmenting lanes in gel electrophoresis images. [Proc. X Brazilian Symposium on Computer Graphics and Image Processing], 140-146. 1997. USA, IEEE Press.
- [89] Martens, R. and Claesen, L. Dynamic programming optimisation for on-line signature verification. [Proc. 4th Int. Conf. Doc. Anal. & Recogn.], 653-656. 1997. Ulm-Germany, IEEE Computer Society.
- [90] Martens, R. and Claesen, L. On-line signature verification: discrimination emphasised. [Proc. 4th Int. Conf. on Document Analysis and Recognition], 657-660. 1997. Ulm-Germany, IEEE Computer Society.
- [91] Martín, M. A. and Rey J.M., "On the role of Shannon's entropy as a measure of heterogeneity," *Geoderma*, vol. 98 pp. 1-3, 2000.
- [92] McClave, J. T., Dietrich, F. H., and Sincich, T., *Statistics* New Jersey: Prentice-Hall, Inc., 1997.
- [93] Meeks, M. L. and Kuklinski, T. T., "Measurement of dynamic digitizer performance," in Plamondon, R. and Leedham, C. G. (eds.) *Computer Processing of Handwriting* World Scientific, 1990, pp. 89-110.
- [94] Mehtre, B. M., Kankanhalli, M. S., and Lee, W. F., "Shape measures for content based image retrieval: a comparison," *Information Processing & Management*, vol. 33, no. 3, pp. 319-337, 1997.
- [95] Munich, M. E. and Perona, P. Camera-based id verification by signature tracking. [Proc. 5th Europ. Conf. Comput. Vision], 782-796. 1998. Springer-Verlag. LNCS-Series Vol. 1407-1408. Burkhardt, H. and Neumann, B.
- [96] Munich, M. E. and Perona, P. Continuous Dynamic Time Warping for translation-invariant curve alignment with applications to signature verification. [Proc. 8th IEEE Int. Conf. on Computer Vision], 108-115. 1999.
- [97] Nagel, R. N. and Rosenfeld, A., "Computer Detection of Freehand Forgeries," *IEEE Transactions on Computers*, vol. C-26, no. 9, pp. 895-905, 1977.
- [98] Nalwa, V. S. Automatic on-line signature verification. 85 (2)[Proc. of the IEEE], 215-239. 1997.

- [99] Nelson, W. and Kishon, E. Use of dynamic features for signature verification. [Proc. IEEE Conf. on Systems, Man and Cybernetics]. 1991. Charlottesville, VA.
- [100] Nelson, W., Turin, W., and Hastie, T., "Statistical methods for on-line signature verification," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 109-129.
- [101] Nouboud, F., "Handwritten signature verification: A global approach," in Impedovo, S. (ed.) *Fundamentals in Handwriting Recognition* Springer-Verlag, 1994.
- [102] Parizeau, M. and Plamondon, R., "A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 710-717, 1990.
- [103] Parizeau, M. and Plamondon, R. Machine vs humans in a cursive script reading experiment without linguistic knowledge. vol. II [Proc. 12th Int. Conf. on Pattern Recognition, Jerusalem], 93-98. 1994. 9.
- [104] Park, J. S. and Chang, D. H. 2-D invariant descriptors for shape-based image retrieval. <http://citeseer.nj.nec.com/cs> . 2003. NEC Research Institute 2002.
- [105] Pavlidis, I., Mavuduru, R., and Papanikolopoulos, N. P. Off-line recognition of signatures using revolving active deformable models. [Proc. of the 1994 IEEE Int. Conf. on Systems, Man and Cybernetics], 771-776. 1994. San Antonio. 1994.
- [106] Pavlidis, I., Papanikolopoulos, N. P., and Mavuduru, R., "Signature identification through the use of deformable structures," *Signal Processing*, vol. 71 pp. 187-201, 1998.
- [107] Penagos, J. D., Prabhakaran, N., and Wunnavu, S. V. An efficient scheme for dynamic signature verification. [IEEE SOUTHEASTCON '96. Bringing Together Education, Science and Technology], 451-457. 1996. USA, IEEE. 11-4-1996.
- [108] Plamondon, R. A model-based segmentation framework for computer processing of handwriting. [Proc. 11th Int. Conf. on Pattern Recognition], 303-307. 1992. The Hague, Netherlands, IEEE Comput. Soc. Press.
- [109] Plamondon, R., "The design of an on-line signature verification system: from theory to practice," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 155-171.
- [110] Plamondon, R., "A kinematic theory of rapid human movements. Part I: Movement representation and generation," *Biological Cybernetics*, vol. 72 pp. 295-307, 1995.
- [111] Plamondon, R. and Guerfali, W., "The generation of handwriting with delta-lognormal synergies," *Biological Cybernetics*, vol. 78, no. 2, pp. 119-132, 1998.
- [112] Plamondon, R. and Lorette, G., "Automatic signature verification and writer identification - The State of the Art," *Pattern Recognition*, vol. 22, no. 2, pp. 107-131, 1989.
- [113] Plamondon, R. and Lorette, G. On-line signature verification: How many countries are in the race? [ICCST, Zurich], 183-191. 1989.
- [114] Plamondon, R., Lorette, G., and Sabourin, R., "Automatic processing of signature images: Static techniques and methods," in Plamondon, R. and Leedham, C. G. (eds.) *Computer Processing of Handwriting* World Scientific, 1990, pp. 49-63.
- [115] Plamondon, R. and Maarse, F. J., "An evaluation of motor models of handwriting," *IEEE Trans. Syst. Man Cybern.* , vol. 19, no. 5, pp. 1060-1072, 1989.

- [116] Plamondon, R., Yergeau, P., and Brault, J.-J., "A multi-level signature verification system," in Impedovo, S. and Simon, J. C. (eds.) *From Pixels to Features III: Frontiers in Handwriting Recognition* Elsevier Science Publishers B.V., 1992.
- [117] Qi, Y. and Hunt, B. R., "Signature verification using global and grid features," *Pattern Recognition*, vol. 27, no. 12, pp. 1621-1629, 1994.
- [118] Ramesh, V. E. and Murty, M.N., "Off-line signature verification using genetically optimized weighted features," *Pattern Recognition*, vol. 32, no. 2, pp. 217-233, 1999.
- [119] Randolph, D. and Krishnan, G. Off-line machine recognition of forgeries. [Proc. SPIE- Int. Soc. for Optical Engineering], 255-264. 1990. San Diego. Machine Vision Systems Integration in Industry.
- [120] Rhee, T. H., Cho, S. J., and Kim, J. H. On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries. <http://icdar.djvuzone.org/> [ICDAR 2001 online]. 2001.
- [121] Sabourin, R., Drouhard, J.-P., and Wah, E. S. Shape matrices as a mixed shape factor for off-line signature verification. [Proc. 4th Int. Conf. on Document Analysis and Recognition], 661-665. 1997. Ulm-Germany, IEEE Computer Society.
- [122] Sabourin, R., Genest, G., and Prêteux, F. J., "Off-line signature verification by local granulometric size distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 976-988, 1997.
- [123] Sabourin, R. and Plamondon, R. Observability and similarity in spatial relations in the structural interpretation of handwritten signature images. [Proc. 7th Scandinavian Conf. on Image Analysis], 477-485. 1991. Arlborg, Dannemark.
- [124] Sabourin, R., Plamondon, R., and Beaumier, L., "Structural interpretation of handwritten signature images," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 69-107.
- [125] Sabourin, R., Plamondon, R., and Lorette, G., "Off-line identification with handwritten signature images: Survey and perspectives," in Baird, H. S. e. al. (ed.) *Structured Document Image Analysis* Springer-Verlag, 1992, pp. 219-234.
- [126] Sanderson, S. and Erbetta, J. H. Authentication for secure environments based on iris scanning technology. [Colloquium on Visual Biometrics]. 2000. Savoy Place, London WC2R 0BL, the IEE.
- [127] Sato, Y. and Kogure, K. Online signature verification based on shape, motion, and writing pressure. 2[Proc. 6th Int. Conf. on Pattern Recognition], 823-826. 1982. IEEE.
- [128] Schmidt, C. and Kraiss, K.-F. Establishment of personalized templates for automatic signature verification. [Proc. 4th Int. Conf. Doc. Anal. & Recogn.], 263-267. 1997. Ulm-Germany, IEEE Computer Society.
- [129] Schmidt, C. and Olschewski, F., "Signature verification using a self-organizing map -A connectionist approach based on dynamic features of the signature," in Simner, M. L. (ed.) *Basic and Applied Issues in Handwriting and Drawing Research*. 7th Biennial Conf. of the International Graphonomics Society. ed. The Netherlands: London, International Graphonomics Society, 1995, pp. 138-139.
- [130] Schomaker, L., "From handwriting analysis to pen-computer applications," *Electronics & Communication Engineering Journal*, vol. 10, no. 3, pp. 93-102, 1998.

- [131] Schomaker, L. and Segers, E., "Finding features used in the human reading of cursive handwriting," *International Journal on Document Analysis and Recognition*, vol. 2, no. 1, pp. 13-18, 1999.
- [132] Schomaker, L. R. B. and Plamondon, R., "The relation between pen force and pen-point kinematics in handwriting," *Biological Cybernetics*, vol. 63 pp. 277-289, 1990.
- [133] Seiler, D. A. and Zusne, L., "Judged complexity of tachistoscopically viewed random shapes," *Percept.Mot.Skills*, vol. 24 pp. 884-886, 1967.
- [134] Shannon, C. E., "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27 pp. 379-423, 1948.
- [135] Siegel, S., *Nonparametric statistics for the behavioral sciences* Tokyo: McGraw-Hill Kogakusha Ltd., 1956.
- [136] Stenson, H. H., "The physical factor structure of random forms and their judged complexity," *Percept.Psychophys.*, vol. 1, no. 303, pp. 310, 1966.
- [137] Stenson, H. H., "The psychophysical dimensions of similarity among random shapes," *Percept.Psychophys.*, vol. 3 pp. 201-214, 1968.
- [138] Tanabe, K., Yoshihara, M., Kameya, H., Mori, S., Omata, S., and Ito, T. Automatic signature verification based on the dynamic feature of Pressure. <http://icdar.djvuzone.org/> [ICDAR 2001 Online]. 2001.
- [139] Taylor, J. R., *An introduction to error analysis: the study of uncertainties in physical measurements*, 2nd ed. ed. Sausalito, California: University Science Books, 1997, 1939.
- [140] Taza, A. and Suen, C. Y., "Discrimination of planar shapes using shape matrices," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 5, pp. 1281-1289, 1989.
- [141] Tolba, A. S., "GloveSignature: A virtual-reality-based system for dynamic signature verification," *Digital Signal Processing*, vol. 9 pp. 241-266, 1999.
- [142] Van Gemmert, A. W. A. and Van Galen, G. P., "Dynamic features of mimicking another person's writing and signature," in Simner, M. L., Leedham, C. G., and Thomassen, A. J. W. M. (eds.) *Handwriting and Drawing Research: Basic and Applied Issues* Amsterdam: IOS Press, 1996, pp. 459-471.
- [143] Vanderplas, J. M. and Garvin, E. A., "Complexity, association value, and practice as factors in shape recognition following paired-associates training," *J.Exp.Psychol.*, vol. 57, no. 3, pp. 155-163, 1959.
- [144] Vielhauer, C., Steinmetz, R., and Mayerhöfer, A. Transitivity based enrollment strategy for signature verification systems. <http://icdar.djvuzone.org/> [ICDAR 2001 Online]. 2001.
- [145] White, B. W. Complexity and heterogeneity in the visual recognition of two-dimensional forms. Wulfeck, J. W. and Taylor, J. H. Proc. Symp. sponsored by Armed Forces -NRC Committee on Vision[Form discrimination as related to military problems], 158-161. 1957. Washington, D. C., Nat. Acad. Sci.-Nat. Res. Coun.
- [146] Williams, C. S., *Designing Digital Filters* Prentice-Hall Inc., 1986.
- [147] Wirtz, B. Stroke-based time warping for signature verification. [Proc. 3rd Int. Conf. on Document Analysis and Recognition], 179-182. 1995. Montréal, Canada, IEEE Computer Society.

- [148] Wirtz, B. Average prototypes for stroke-based signature verification. [Proc. 4th Int. Conf. Doc. Anal. & Recogn.], 268-272. 1997. Ulm-Germany, IEEE Computer Society.
- [149] Yasuhara, M. and Oka, M., "Signature verification experiment based on nonlinear time alignment: A feasibility study," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-17 pp. 212-216, 1977.
- [150] Yoo, J. H., Nixon, M. S., and Harris, C. J. Extracting gait signatures based on anatomical knowledge. [BMVA Symposium on Advancing Biometric Technologies]. Royal Statistical Society, London. 2002.
- [151] Yoshimura, I. and Yoshimura, M., "On-line signature verification incorporating the direction of pen movement -An experimental examination of the effectiveness," in Impedovo, S. and Simon, J. C. (eds.) *From Pixels to Features III: Frontiers in Handwriting Recognition* Elsevier Science Publishers B.V., 1992, pp. 353-361.
- [152] Yoshimura, I. and Yoshimura, M., "Off-line verification of Japanese signatures after elimination of background patterns," in Plamondon, R. (ed.) *Progress in automatic signature verification* World Scientific, 1994, pp. 53-68.
- [153] Yoshimura, M. and Yoshimura, I. An Application of the Sequential Dynamic Programming Matching Method to Off-Line Signature Verification. [BSDIA], 299-310. 1997. Springer.
- [154] Zusne, L., *Visual perception of form* Academic Press Inc., 1970.

Appendixes

Appendix A

Human Perception Experiments

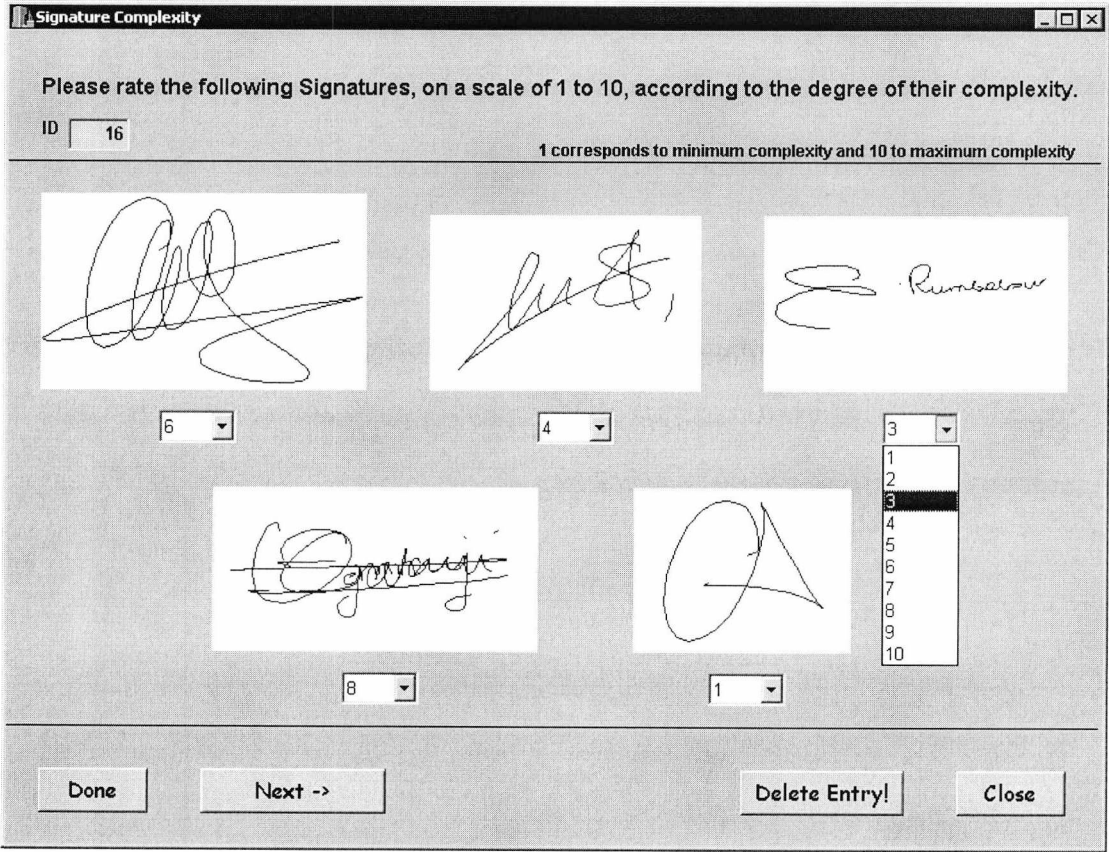


Figure A.1: Interface of the perceived signature complexity experiment.



Figure A.2: Interface of the human perception Experiment 1.

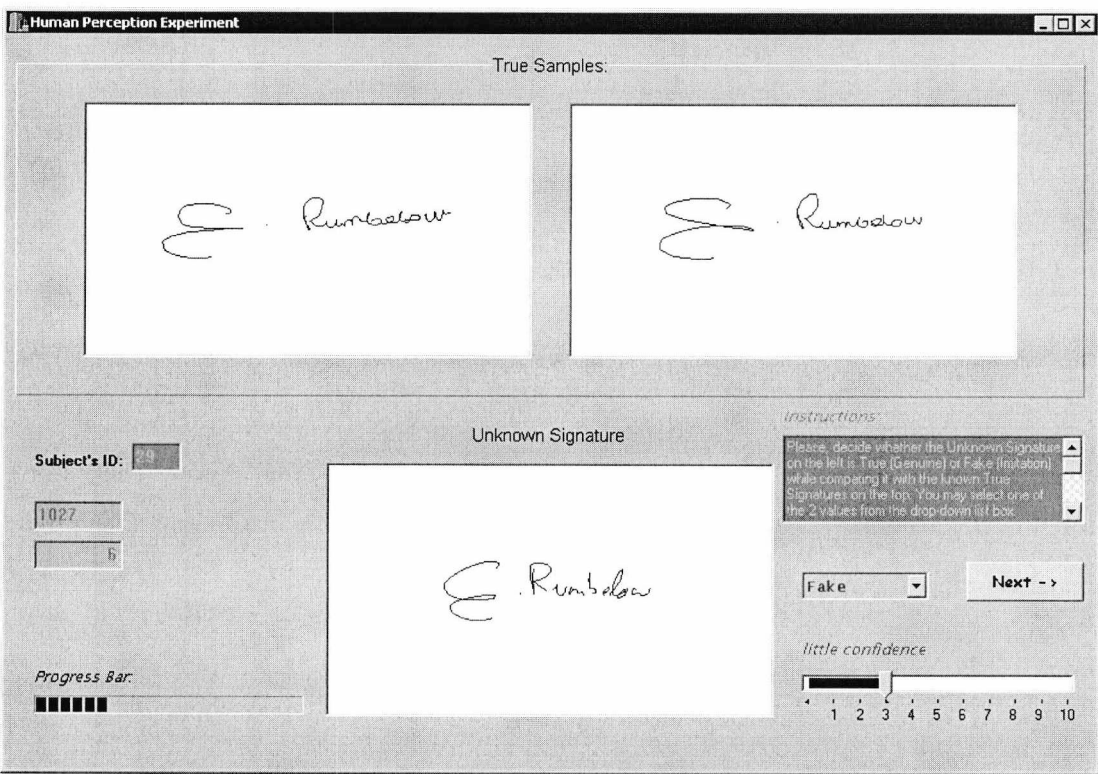


Figure A.3: Interface of the human perception Experiment 2.

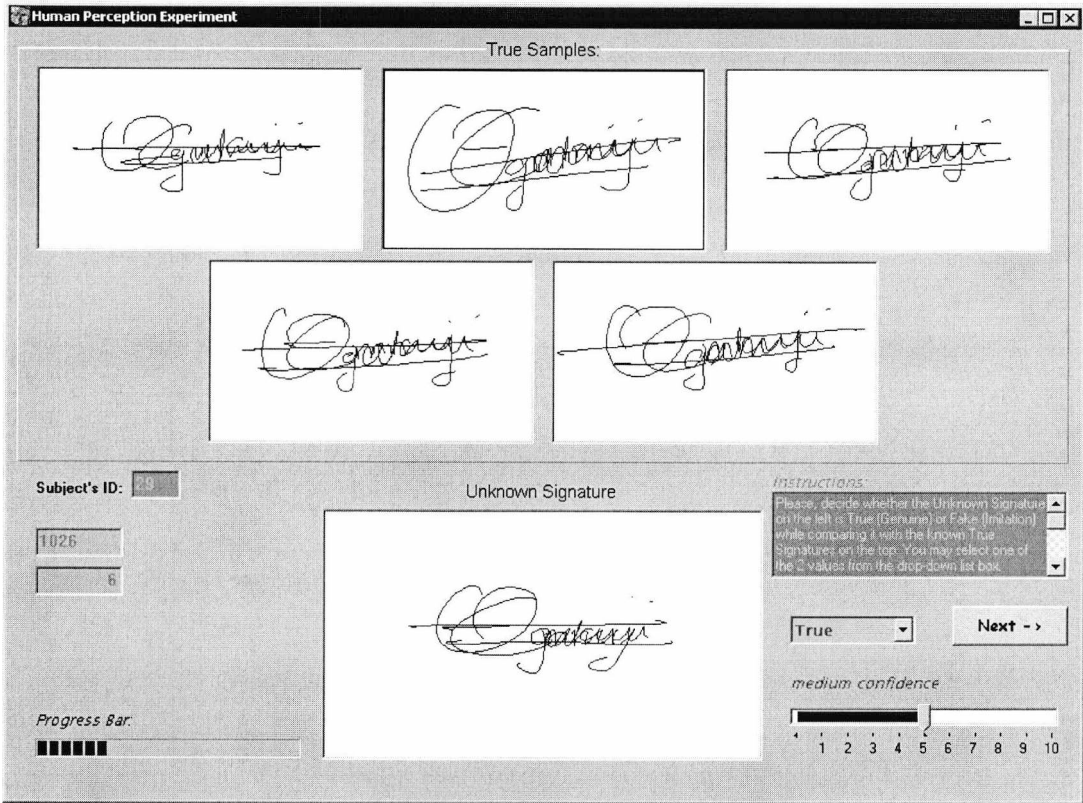


Figure A.4: Interface of the human perception Experiment 3.

Appendix B

Automatic Signature Verification

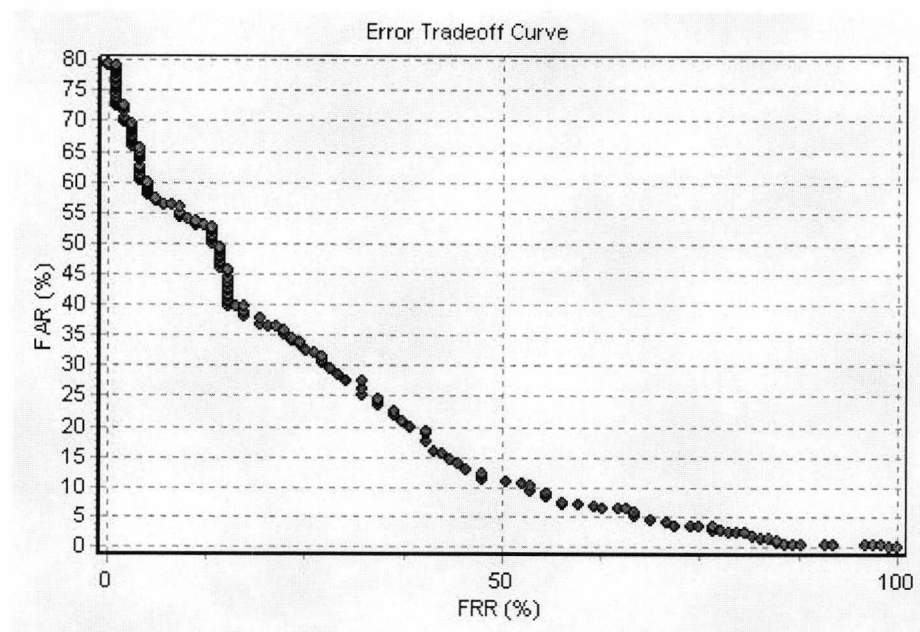
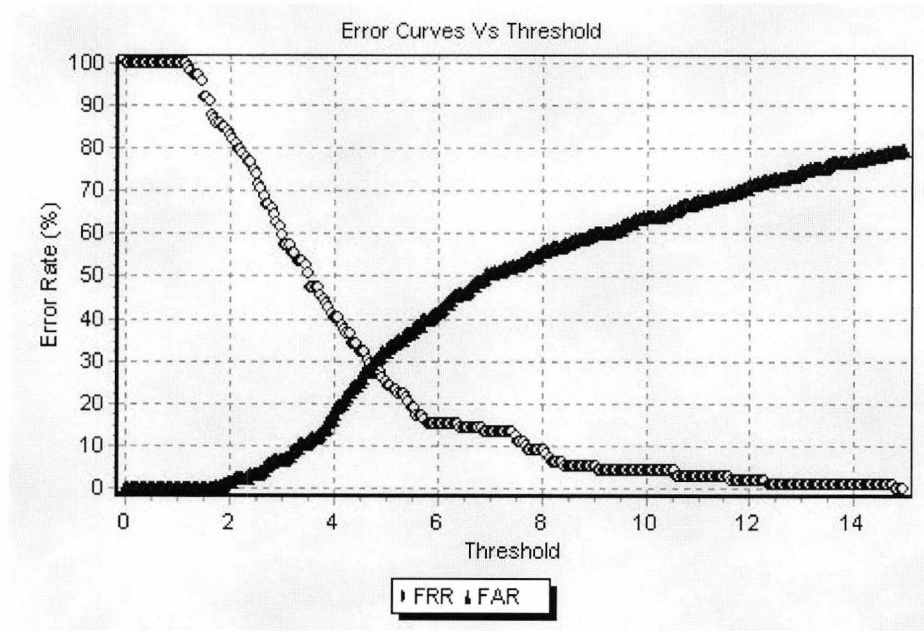
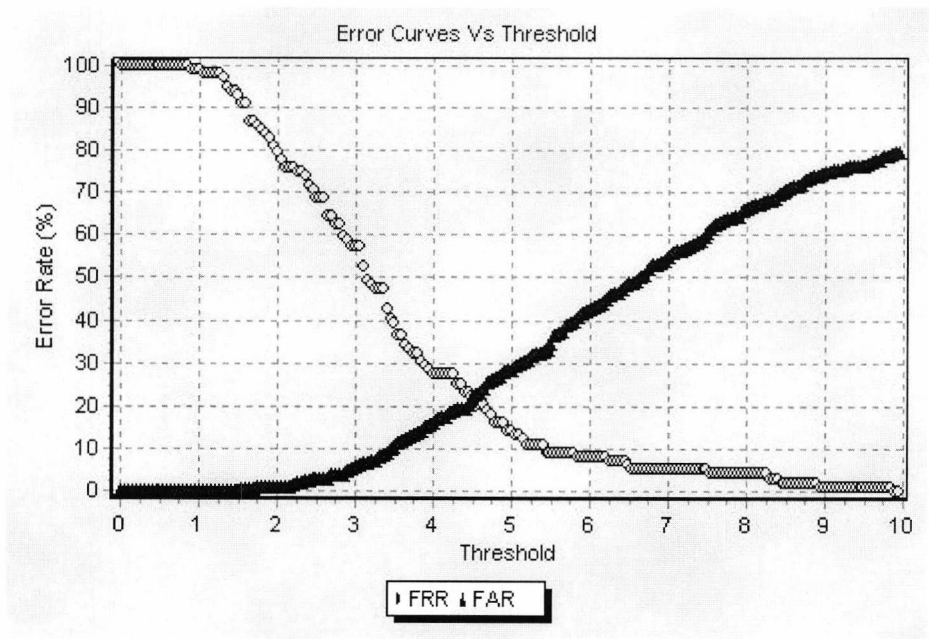
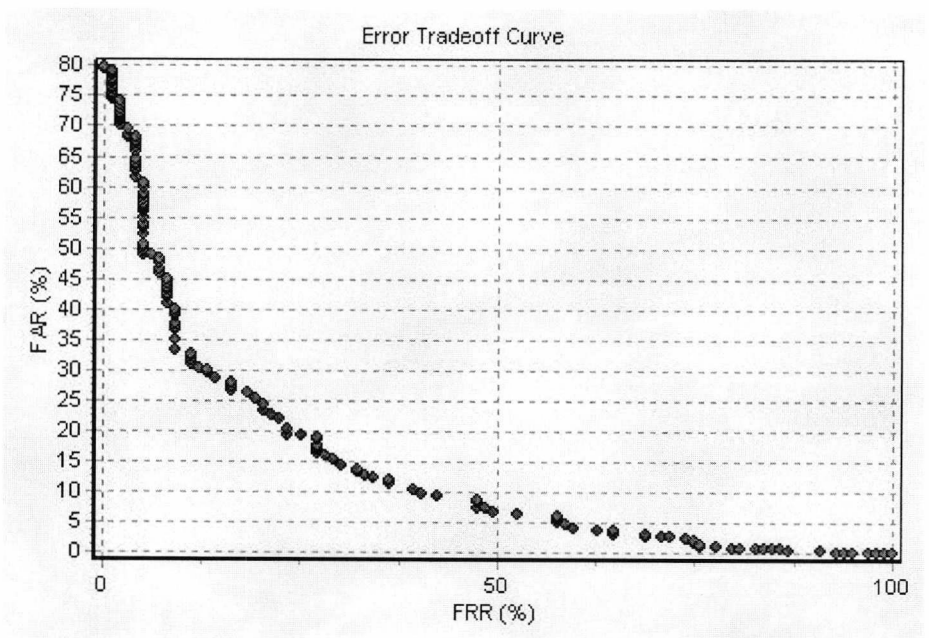


Figure B.1: Static verification performance for the set of seven central moments (a) error rates vs threshold and (b) error tradeoff curve.

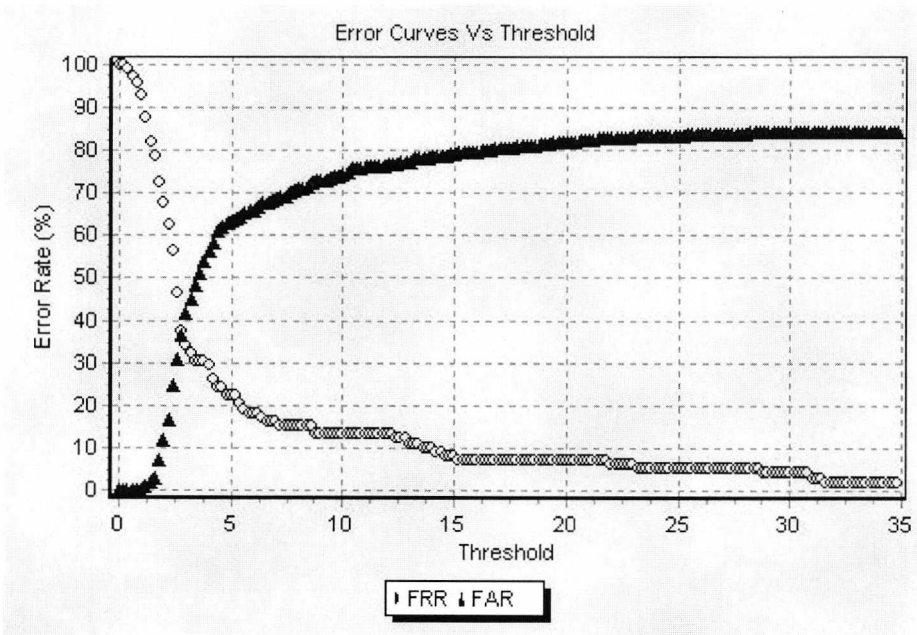


(a)

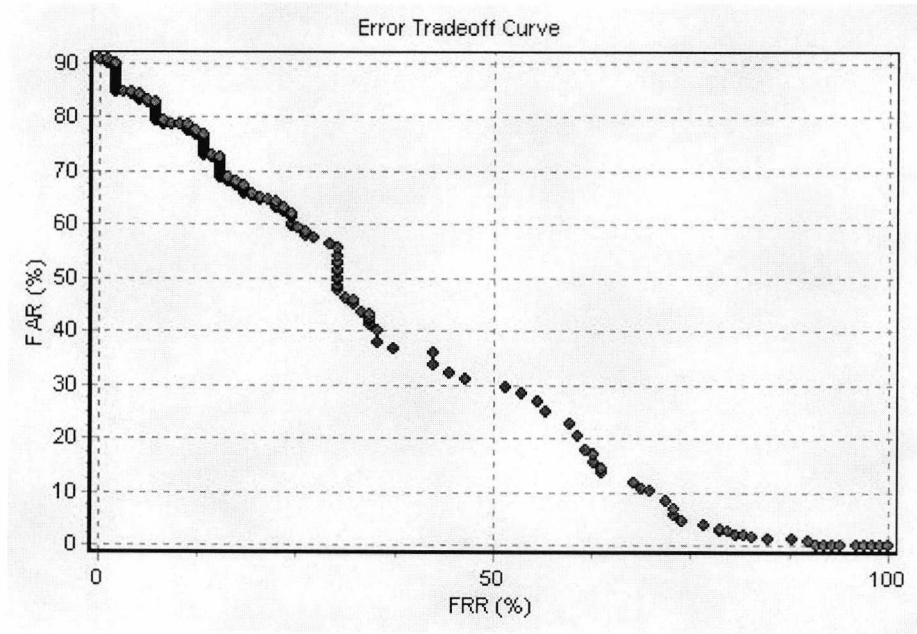


(b)

Figure B.2: Static verification performance for the set of seven normalised central moments (a) error rates vs threshold and (b) error tradeoff curve.



(a)



(b)

Figure B.3: Static verification performance for the set of seven Hu moments (a) error rates vs threshold and (b) error tradeoff curve.

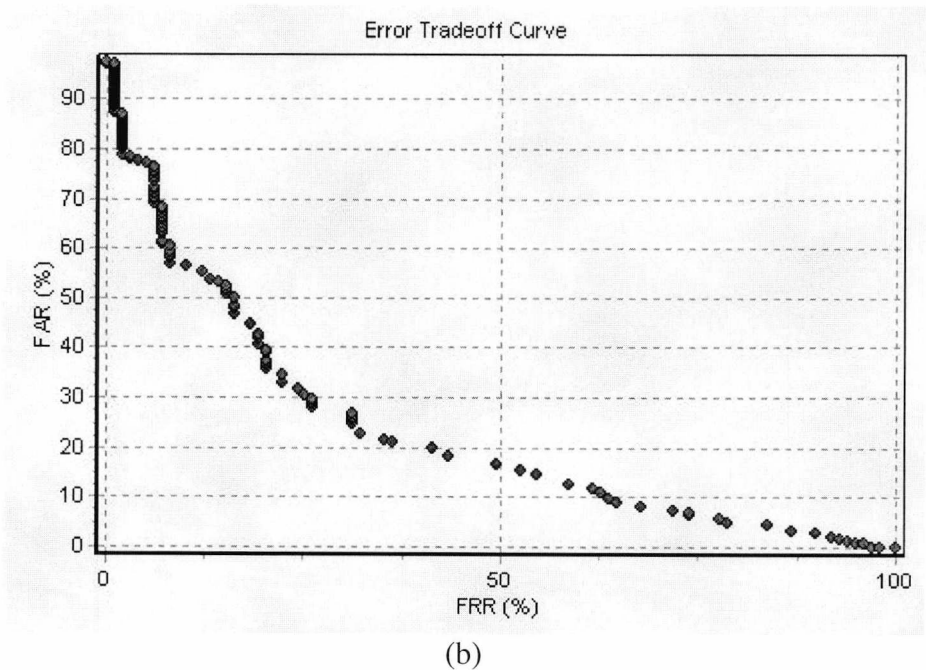
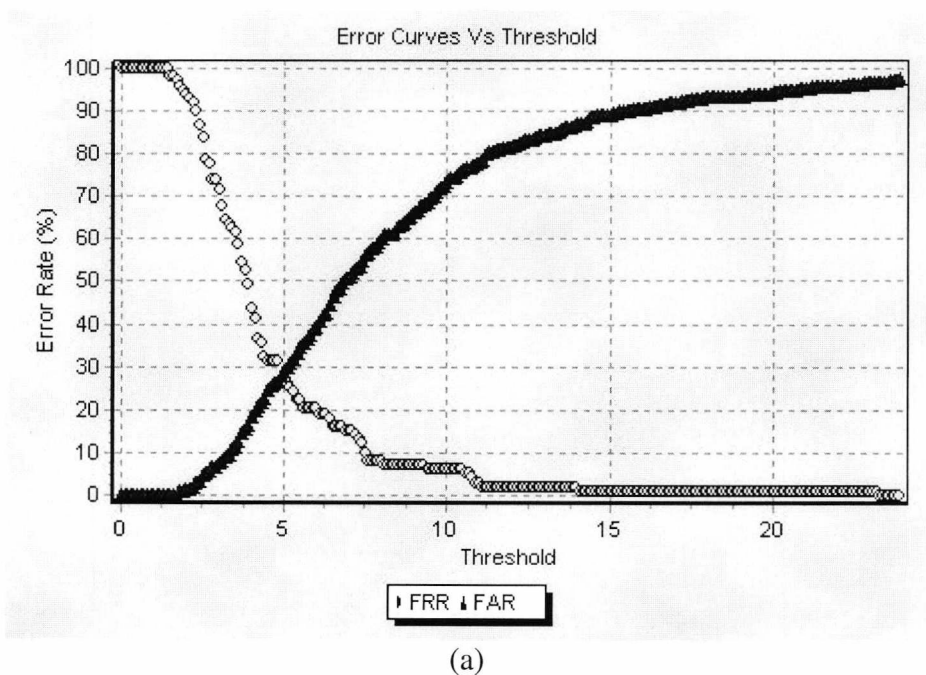


Figure B.4: Static verification performance for the set of seven Global features (a) error rates vs threshold and (b) error tradeoff curve.

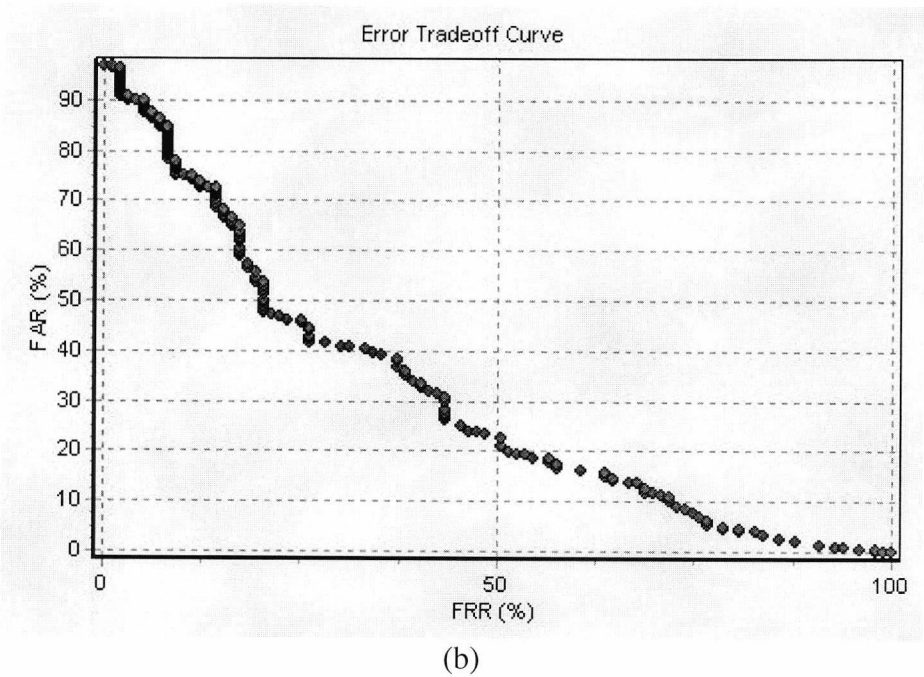
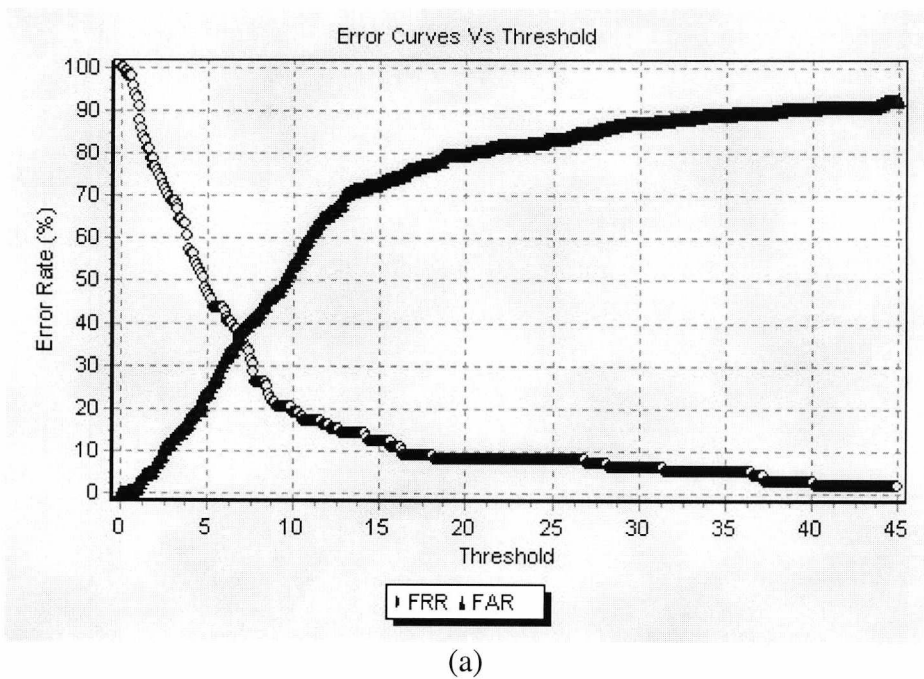


Figure B.5: Static verification performance for the Limited Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.

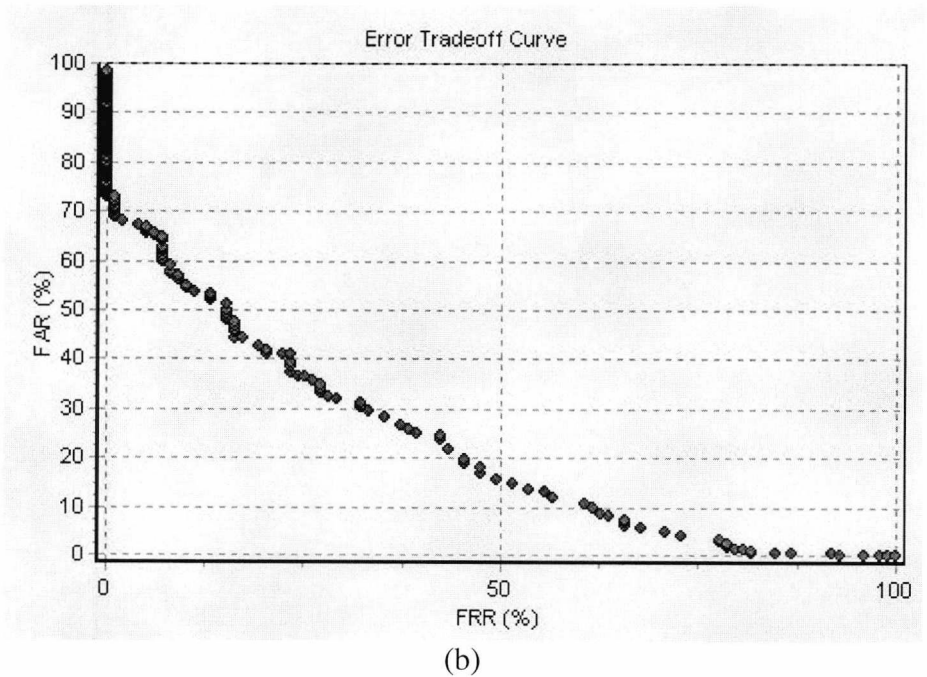
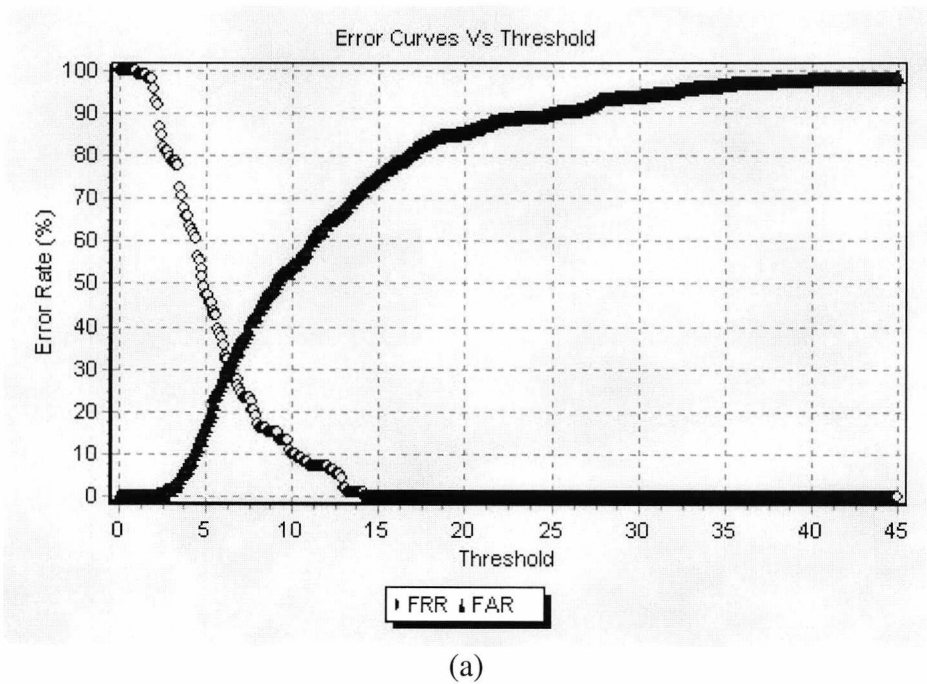


Figure B.6: Static verification performance for the Small Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.

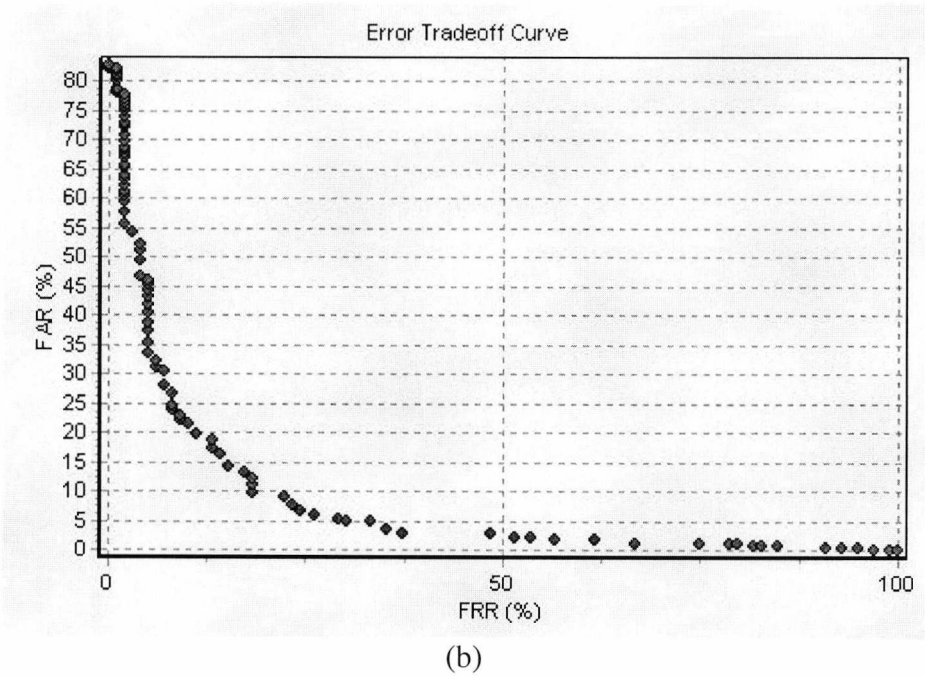
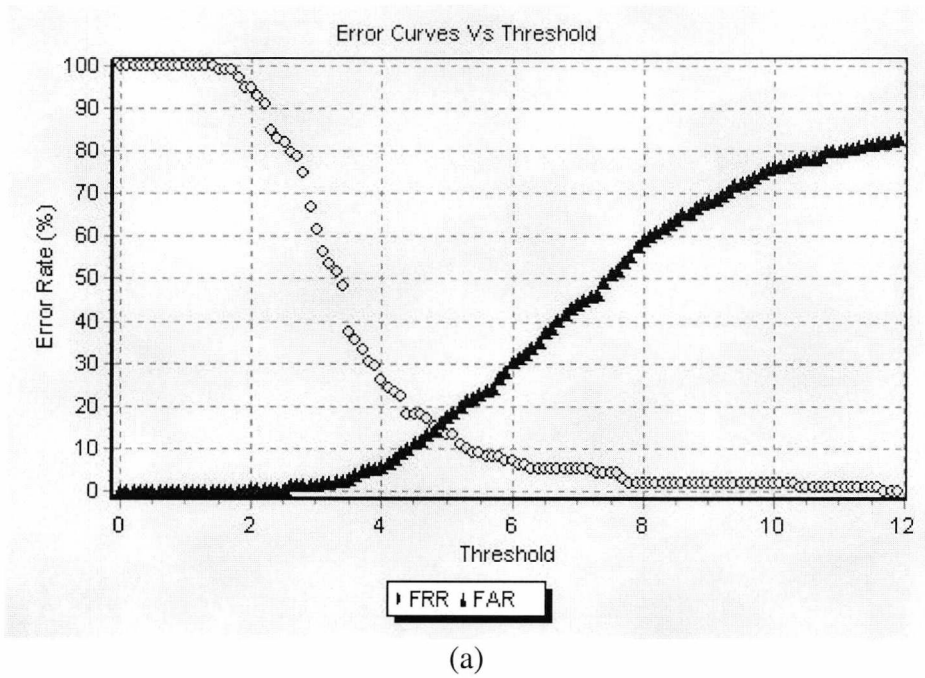


Figure B.7: Static verification performance for the Large Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.

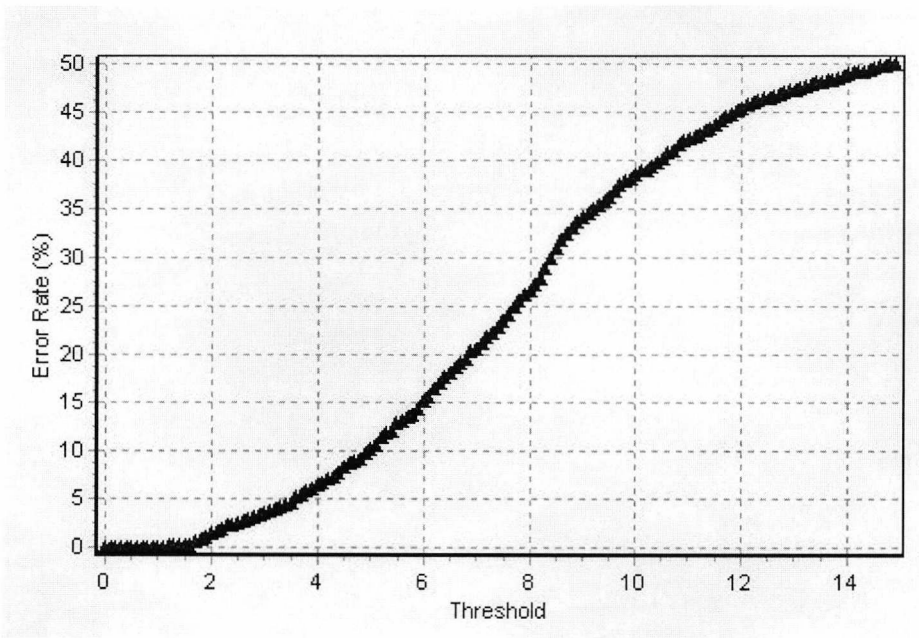


Figure B.8: Static verification –FAR vs threshold for random forgeries and the Limited Reference Set.

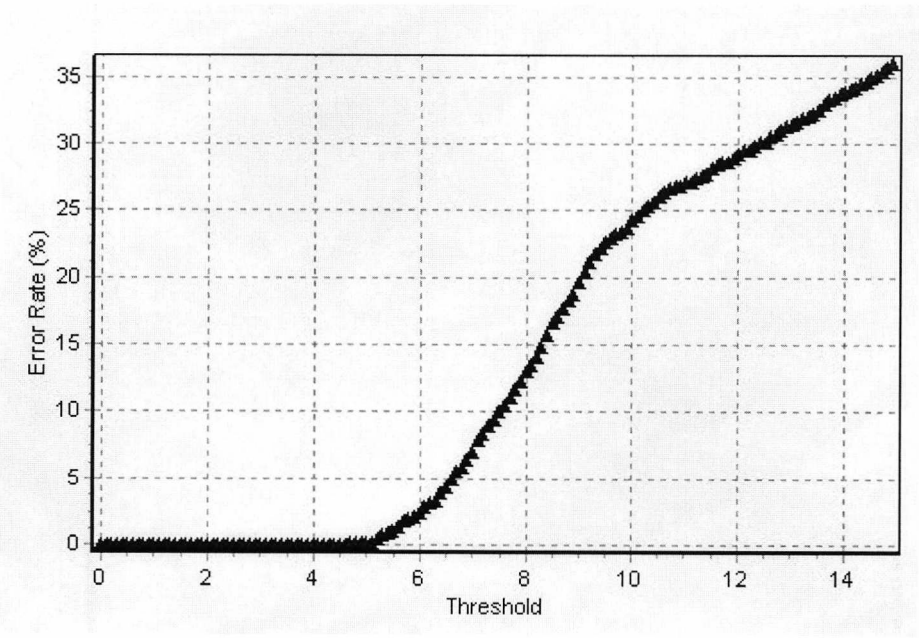


Figure B.9: Static verification –FAR vs threshold for random forgeries and the Small Reference Set.

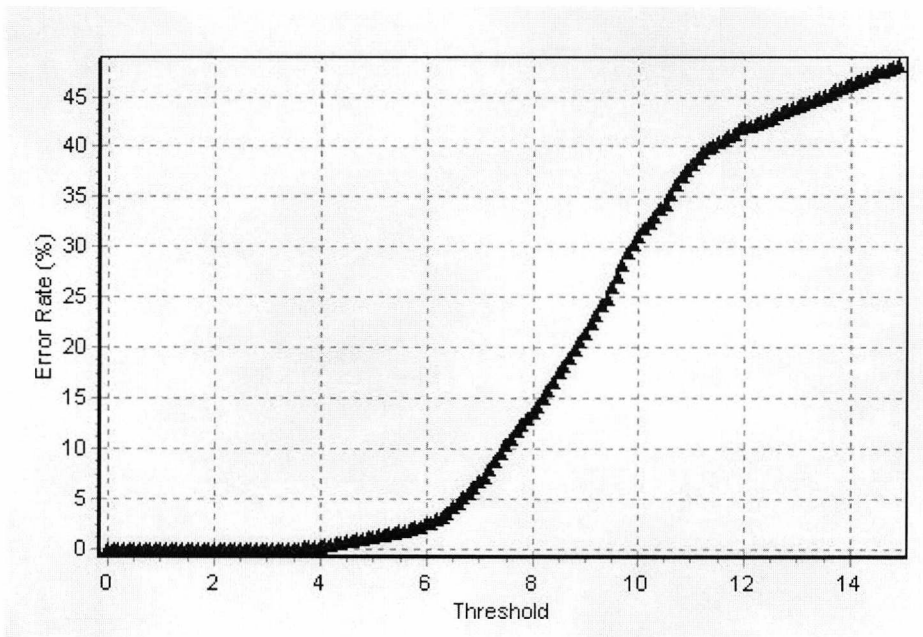
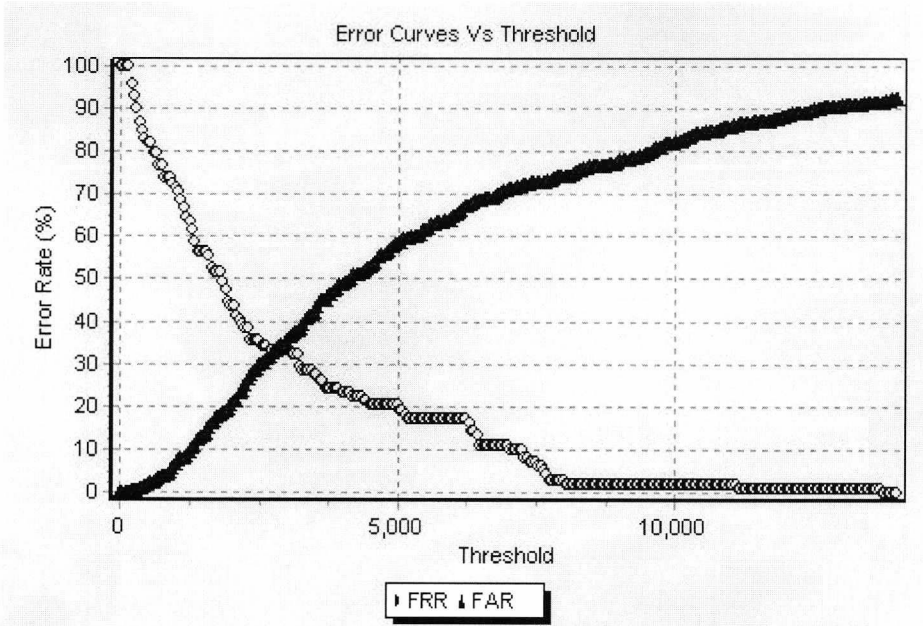
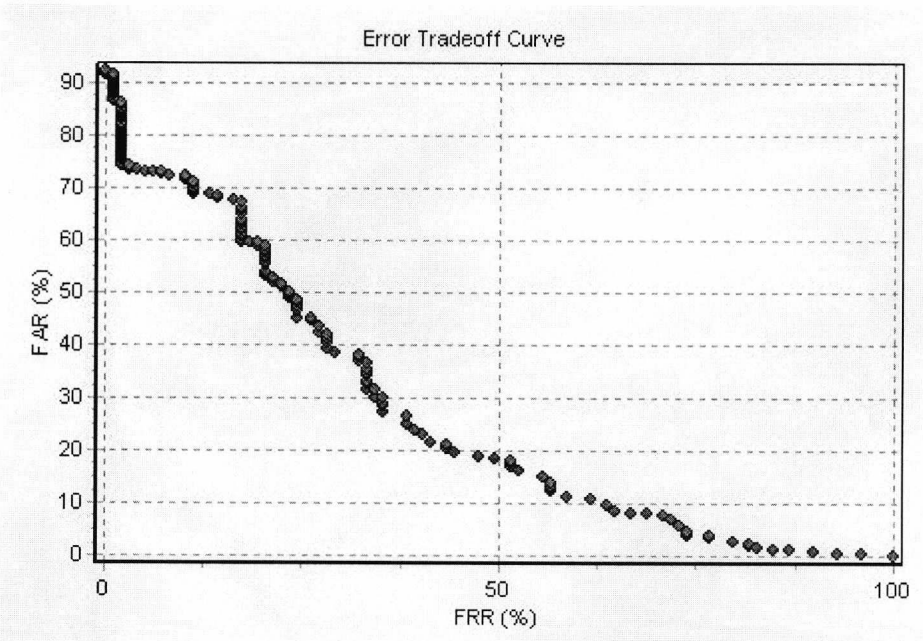


Figure B.10: Static verification –FAR vs threshold for random forgeries and the Large Reference Set.

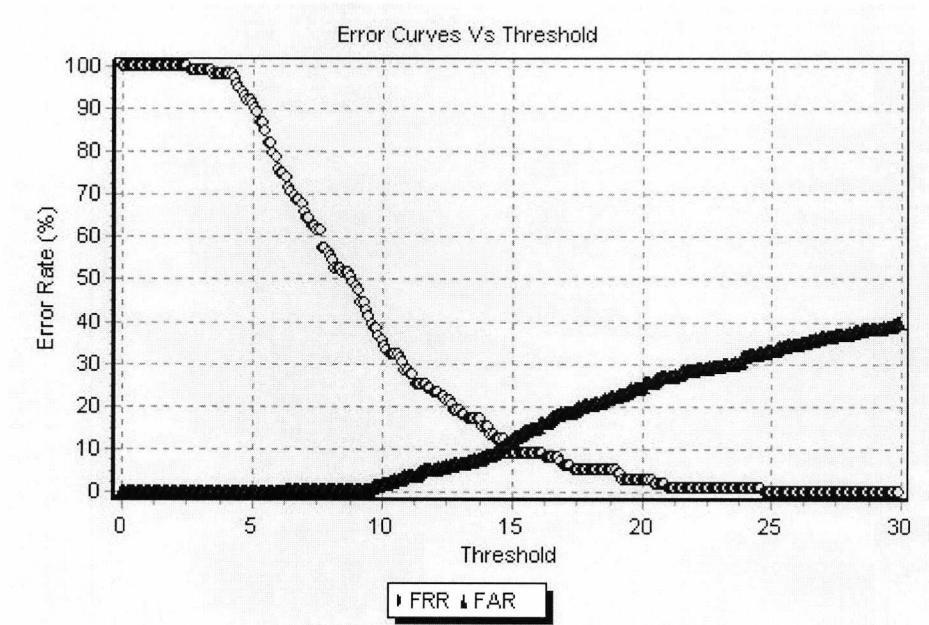


(a)

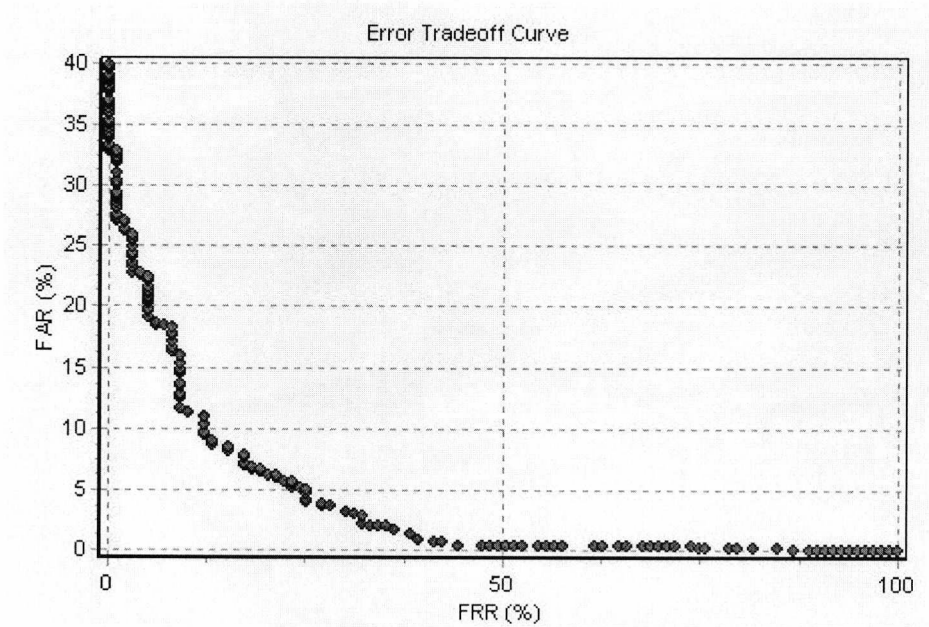


(b)

Figure B.11: Dynamic verification performance for the Limited Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.

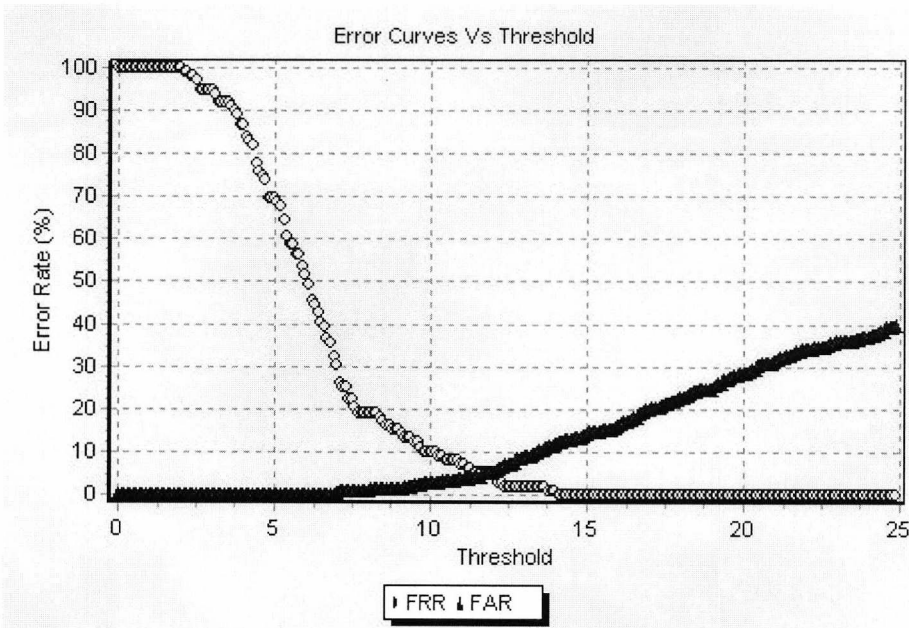


(a)

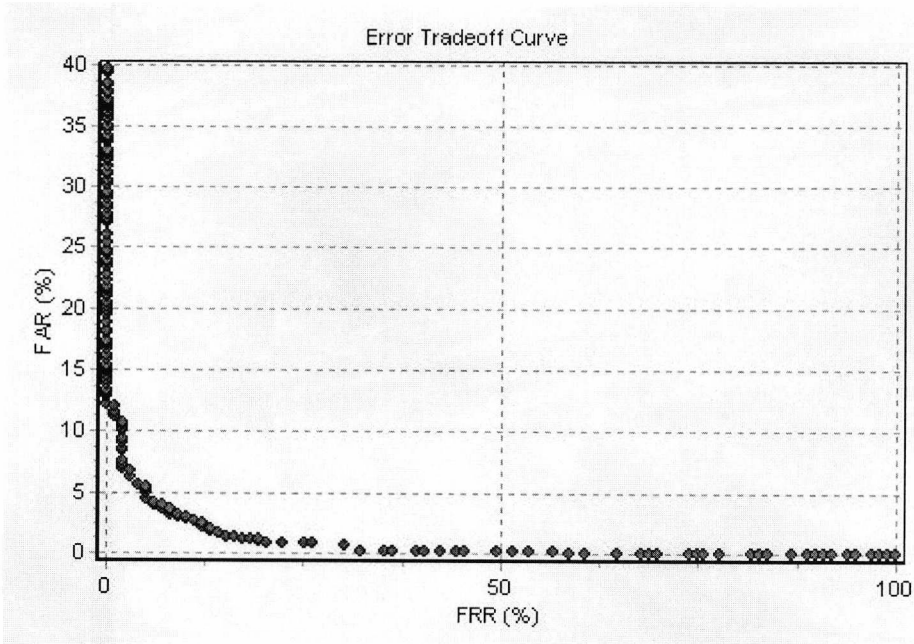


(b)

Figure B.12: Dynamic verification performance for the Small Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.



(a)



(b)

Figure B.13: Dynamic verification performance for the Large Reference Set (a) error rates vs threshold and (b) error tradeoff curve, using skilled forgeries.

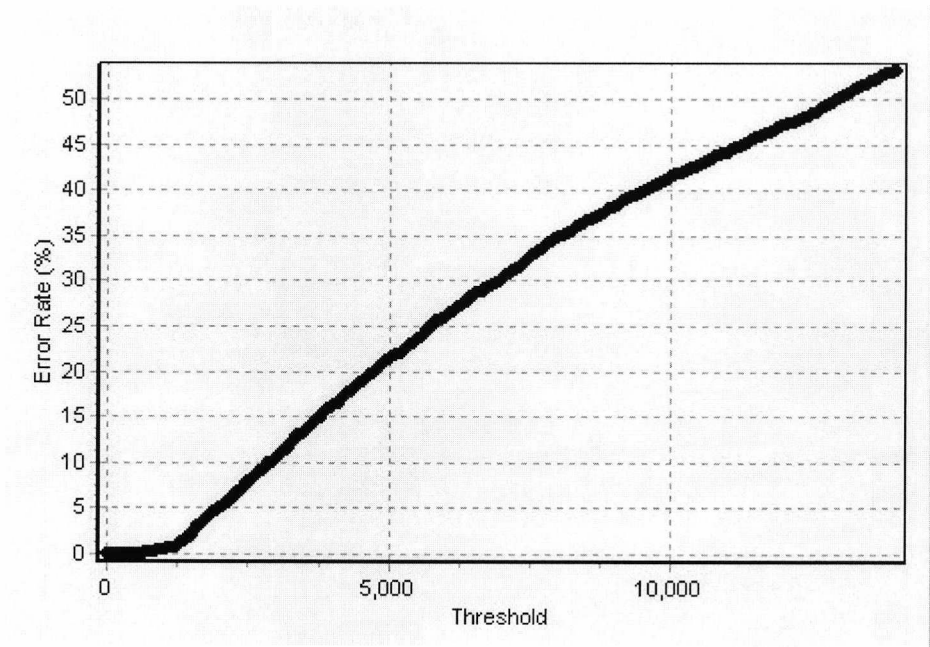


Figure B.14: Dynamic verification –FAR vs threshold for random forgeries and the Limited Reference Set.

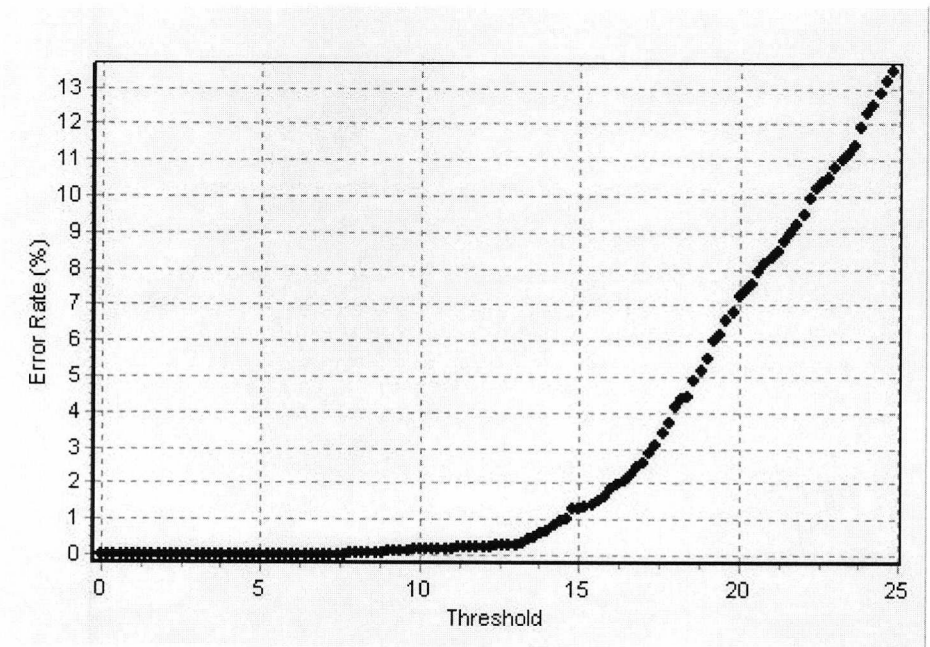


Figure B.15: Dynamic verification –FAR vs threshold for random forgeries and the Small Reference Set.

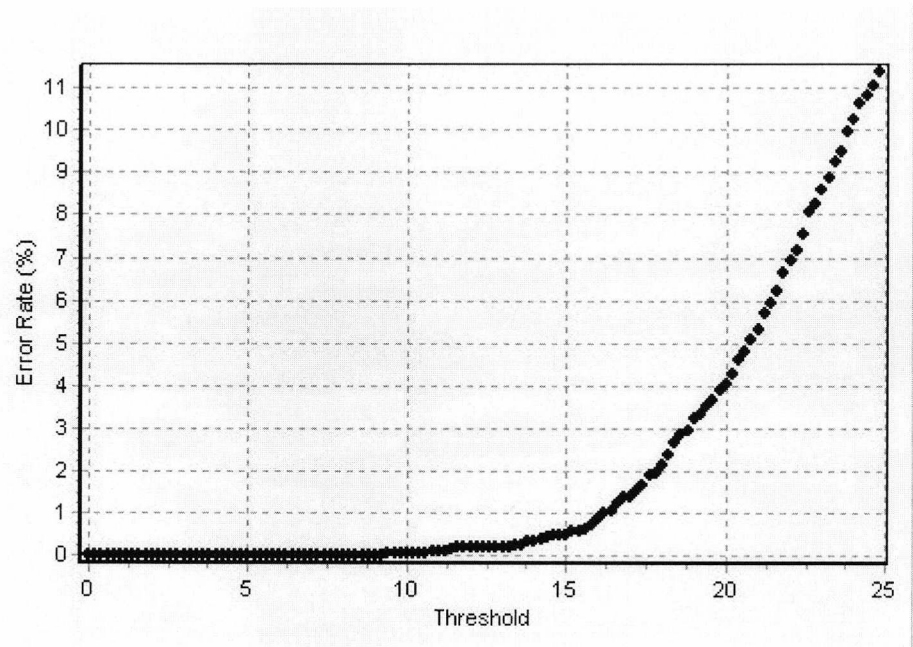


Figure B.16: Dynamic verification –FAR vs threshold for random forgeries and the Large Reference Set.

Appendix C

Handwritten Signature Variability

Table C.1a: Intra-class similarity of target signatures, using the method proposed in [57].

T1	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.92	0.94	0.93	0.92	0.93
RS2	0.92	1	0.92	0.92	0.91	0.92
RS3	0.94	0.92	1	0.93	0.92	0.93
RS4	0.93	0.92	0.93	1	0.92	0.93
RS5	0.92	0.91	0.92	0.92	1	0.92

T2	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.96	0.95	0.96	0.94	0.95
RS2	0.96	1	0.95	0.96	0.94	0.95
RS3	0.95	0.95	1	0.95	0.94	0.95
RS4	0.96	0.96	0.95	1	0.94	0.95
RS5	0.94	0.94	0.94	0.94	1	0.94

T3	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.98	0.98	0.98	0.98	0.98
RS2	0.98	1	0.98	0.98	0.98	0.98
RS3	0.98	0.98	1	0.98	0.98	0.98
RS4	0.98	0.98	0.98	1	0.98	0.98
RS5	0.98	0.98	0.98	0.98	0.98	0.98

T4	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.93	0.92	0.92	0.91	0.92
RS2	0.93	1	0.91	0.91	0.91	0.92
RS3	0.92	0.91	1	0.93	0.93	0.92
RS4	0.92	0.91	0.93	1	0.93	0.92
RS5	0.91	0.91	0.93	0.93	1	0.92

T5	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.97	0.97	0.97	0.97	0.97
RS2	0.97	1	0.97	0.97	0.97	0.97
RS3	0.97	0.97	1	0.97	0.97	0.97
RS4	0.97	0.97	0.97	1	0.97	0.97
RS5	0.97	0.97	0.97	0.97	1	0.97

Table C.1b: Inter-class similarity of target signatures, using the method proposed in [57].

Target	RS	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8
T1	RS1	0.94	0.95	0.92	0.95	0.94	0.95	0.91	0.95
		X	X	X	X	X	X		X
T2	RS1	0.94	0.92	0.97	0.96	0.93	0.94	0.96	0.97
		X		X	X		X	X	X
T3	RS1	0.96	0.97	0.94	0.95	0.97	0.97	0.94	0.94
T4	RS3	0.94	0.91	0.94	0.94	0.93	0.92	0.94	0.94
		X	X	X	X	X	X	X	X
T5	RS1	0.96	0.95	0.97	0.93	0.96	0.96	0.97	0.93
				X				X	

Table C.2a: Intra-class similarity of target signatures, using the method proposed in [140].

T1	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	-1.14	-0.81	-1.08	-1.08	-1.03
RS2	-0.71	1	-0.71	-0.91	-0.87	-0.8
RS3	-0.71	-1.01	1	-1.09	-1.08	-0.97
RS4	-0.58	-0.81	-0.68	1	-0.84	-0.73
RS5	-0.64	-0.84	-0.74	-0.91	1	-0.78

T2	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	-0.64	-1.14	-1.32	-1.42	-1.13
RS2	-0.52	1	-1.01	-1.2	-1.3	-1.01
RS3	-0.56	-0.58	1	-0.84	-0.9	-0.72
RS4	-0.57	-0.6	-0.71	1	-0.85	-0.68
RS5	-0.58	-0.62	-0.71	-0.78	1	-0.67

T3	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	-0.89	-0.89	-0.83	-0.86	-0.87
RS2	-0.78	1	-0.67	-0.64	-0.64	-0.69
RS3	-0.71	-0.62	1	-0.56	-0.7	-0.65
RS4	-0.77	-0.69	-0.66	1	-0.53	-0.66
RS5	-0.71	-0.6	-0.72	-0.46	1	-0.62

T4	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	-0.71	-0.57	-0.54	-0.68	-0.62
RS2	-0.51	1	-0.6	-0.55	-0.62	-0.57
RS3	-0.63	-0.9	1	-0.67	-0.69	-0.72
RS4	-0.54	-0.76	-0.6	1	-0.57	-0.62
RS5	-0.59	-0.75	-0.54	-0.49	1	-0.59

T5	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	-1.08	-0.99	-1.08	-1	-1.04
RS2	-0.75	1	-0.85	-0.94	-0.87	-0.85
RS3	-0.85	-1.05	1	-0.97	-0.77	-0.91
RS4	-0.71	-0.89	-0.74	1	-0.85	-0.8
RS5	-0.86	-1.07	-0.76	-1.09	1	-0.94

Table C.2b: Inter-class similarity of target signatures, using the method proposed in [140].

Target	RS	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8
T1	RS4	-0.43	-0.26	-0.82	-0.37	-0.28	-0.26	-1	-0.38
		X	X	X	X	X	X		X
T2	RS5	-1.6	-1.58	-0.69	-0.41	-1.38	-1.59	-0.71	-0.48
				X	X			X	X
T3	RS5	-1.02	-0.63	-1.88	-1.65	-0.75	-0.6	-2.07	-2.03
			X				X		
T4	RS2	-0.18	-0.78	-0.31	-0.22	-0.29	-0.53	-0.31	-0.22
		X		X	X	X	X	X	X
T5	RS4	-1.56	-2.98	-1.23	-3.34	-1.05	-2.37	-1.3	-3.03

Table C.3a: Intra-class similarity of target signatures, using the method proposed in [121] without weights.

T1	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.07	0.07	0.06	0.06
RS2	0.03	1	0.04	0.04	0.04	0.04
RS3	0.07	0.04	1	0.04	0.03	0.05
RS4	0.07	0.04	0.04	1	0.05	0.05
RS5	0.06	0.04	0.03	0.05	1	0.05

T2	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.15	0.08	0.06	0.04	0.08
RS2	0.15	1	0.06	0.07	0.03	0.08
RS3	0.08	0.06	1	0.06	0.07	0.07
RS4	0.06	0.07	0.06	1	0.04	0.06
RS5	0.04	0.03	0.07	0.04	1	0.05

T3	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.05	0.06	0.06	0.05
RS2	0.03	1	0.09	0.09	0.09	0.08
RS3	0.05	0.09	1	0.12	0.08	0.09
RS4	0.06	0.09	0.12	1	0.14	0.1
RS5	0.06	0.09	0.08	0.14	1	0.09

T4	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.23	0.13	0.12	0.09	0.14
RS2	0.23	1	0.13	0.14	0.11	0.15
RS3	0.13	0.13	1	0.15	0.14	0.14
RS4	0.12	0.14	0.15	1	0.17	0.14
RS5	0.09	0.11	0.14	0.17	1	0.13

T5	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.02	0.05	0.02	0.03
RS2	0.03	1	0.02	0.03	0.01	0.02
RS3	0.02	0.02	1	0.04	0.05	0.03
RS4	0.05	0.03	0.04	1	0.01	0.03
RS5	0.02	0.01	0.05	0.01	1	0.02

Table C.3b: Inter-class similarity of target signatures, using the method proposed in [121] without weights.

Target	RS	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8
T1	RS1	0.06	0.03	0.08	0.02	0.05	0.06	0.07	0.03
		X	X	X		X	X	X	X
T2	RS1	0.07	0.07	0.03	0.03	0.03	0.1	0.02	0.05
		X	X				X		X
T3	RS4	0.02	0	0.03	0.02	0.02	0.01	0.03	0.01
T4	RS2	0.07	0.03	0.02	0.02	0.1	0.07	0.02	0.03
T5	RS3	0.02	0.03	0.01	0.04	0.05	0.02	0.01	0.03
		X	X		X	X	X		X

Table C.4a: Intra-class similarity of target signatures, using baseline adjustment.

T1	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.06	0.04	0.04	0.04
RS2	0.04	1	0.04	0.03	0.04	0.04
RS3	0.06	0.04	1	0.03	0.04	0.04
RS4	0.04	0.04	0.04	1	0.05	0.04
RS5	0.05	0.04	0.05	0.04	1	0.05

T2	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.1	0.05	0.03	0.03	0.05
RS2	0.09	1	0.04	0.04	0.03	0.05
RS3	0.03	0.04	1	0.04	0.08	0.05
RS4	0.04	0.06	0.04	1	0.04	0.05
RS5	0.03	0.04	0.06	0.06	1	0.05

T3	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.04	0.05	0.03	0.04	0.04
RS2	0.05	1	0.08	0.1	0.07	0.08
RS3	0.06	0.1	1	0.12	0.08	0.09
RS4	0.04	0.06	0.1	1	0.12	0.08
RS5	0.04	0.06	0.08	0.15	1	0.08

T4	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.07	0.12	0.11	0.1	0.1
RS2	0.09	1	0.08	0.09	0.08	0.09
RS3	0.1	0.08	1	0.11	0.11	0.1
RS4	0.1	0.1	0.14	1	0.12	0.12
RS5	0.11	0.08	0.09	0.13	1	0.1

T5	RS1	RS2	RS3	RS4	RS5	AVG
RS1	1	0.03	0.06	0.02	0.03	0.04
RS2	0.02	1	0.01	0.02	0.02	0.02
RS3	0.04	0.01	1	0.02	0.02	0.02
RS4	0.02	0.04	0.04	1	0.01	0.03
RS5	0.01	0.01	0.03	0.01	1	0.01

Table C.4b: Inter-class similarity of target signatures, using baseline adjustment.

Target	RS	TS1	TS2	TS3	TS4	TS5	TS6	TS7	TS8
T1	RS5	0.03	0.01	0.06	0.01	0.03	0.02	0.06	0.02
				X				X	
T2	RS4	0.03	0.04	0.02	0.01	0.03	0.04	0.02	0.02
			X				X		
T3	RS3	0.02	0.02	0.04	0.03	0.01	0.01	0.04	0.01
T4	RS4	0.04	0.03	0.04	0.02	0.04	0.06	0.02	0.02
T5	RS1	0.02	0.02	0.02	0.02	0.03	0.02	0.01	0.03
		X	X	X	X	X	X		X

Appendix D

Human Visual Inspection of Signatures

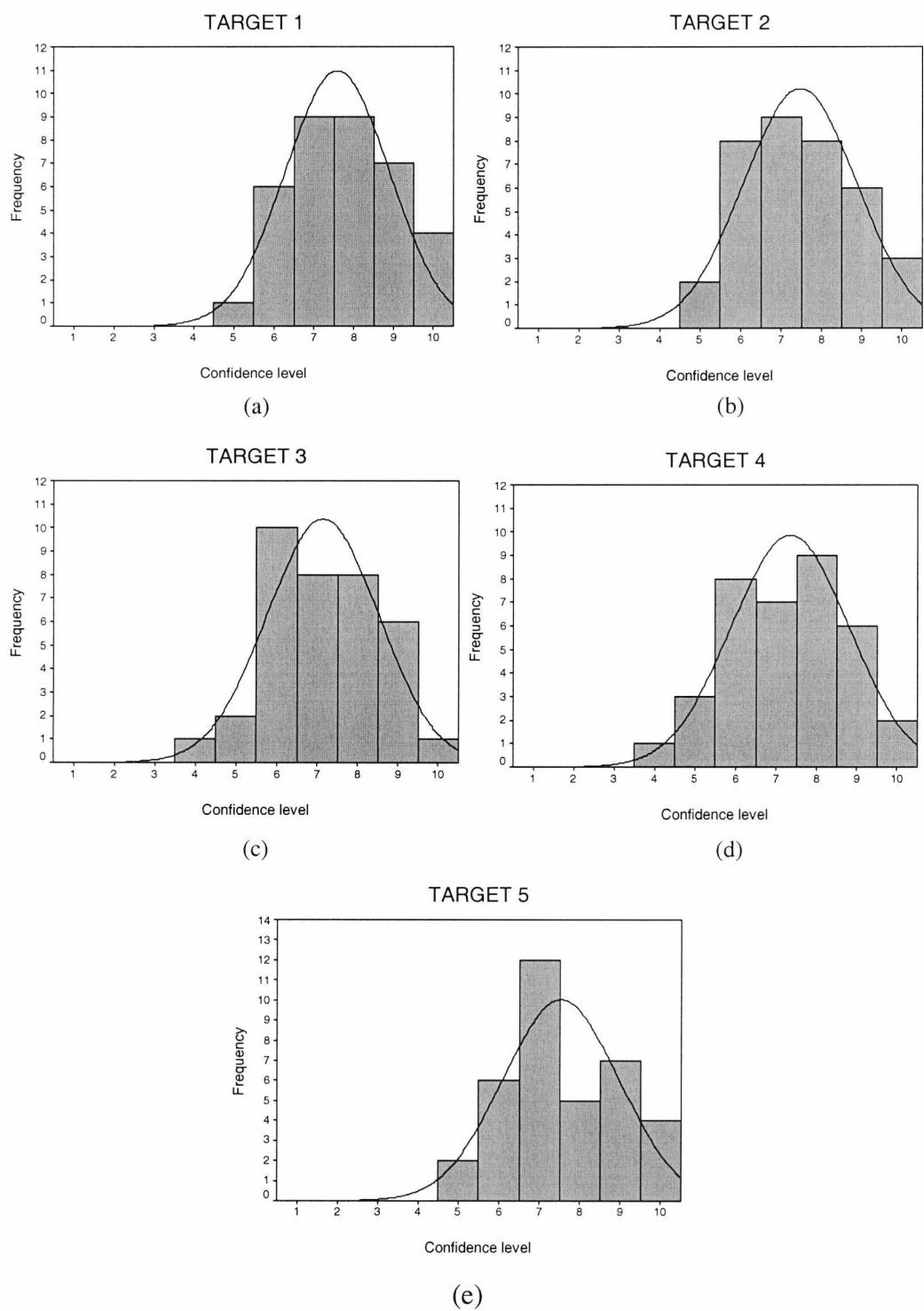


Figure D.1: Frequency distribution of human confidence in verifying the authenticity of target signatures: (a) 1, (b) 2, (c) 3, (d) 4 and (e) 5.

Table D.1: Mean confidence values for the target signatures.

	Target 1	Target 2	Target 3	Target 4	Target 5	Average
Experiment 2	6.7	7.0	7.3	7.0	6.8	7.0
Experiment 3	7.0	7.2	7.3	7.0	7.1	7.1

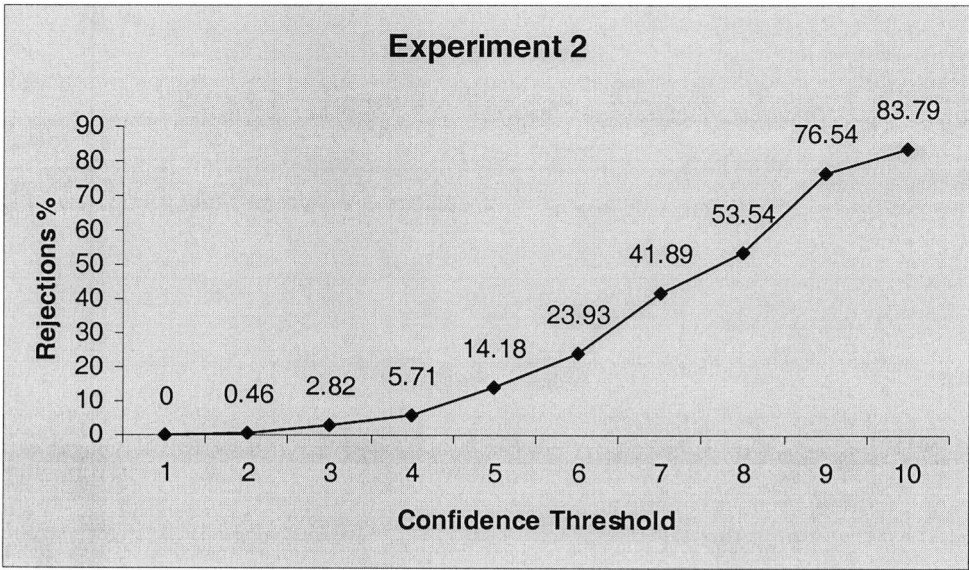


Figure D.2: Rejection rate vs confidence threshold for Experiment 2.

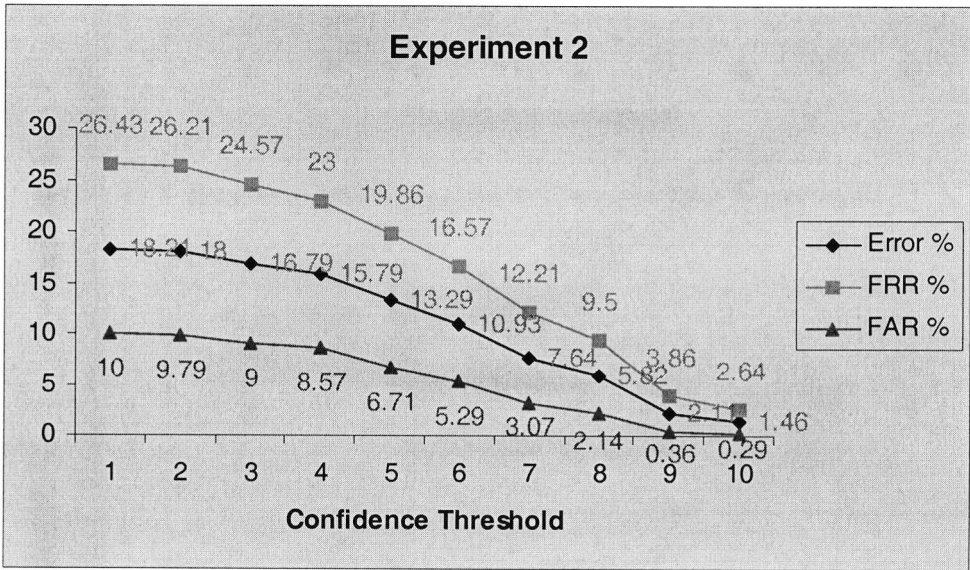


Figure D.3: Error rates vs confidence threshold for Experiment 2.

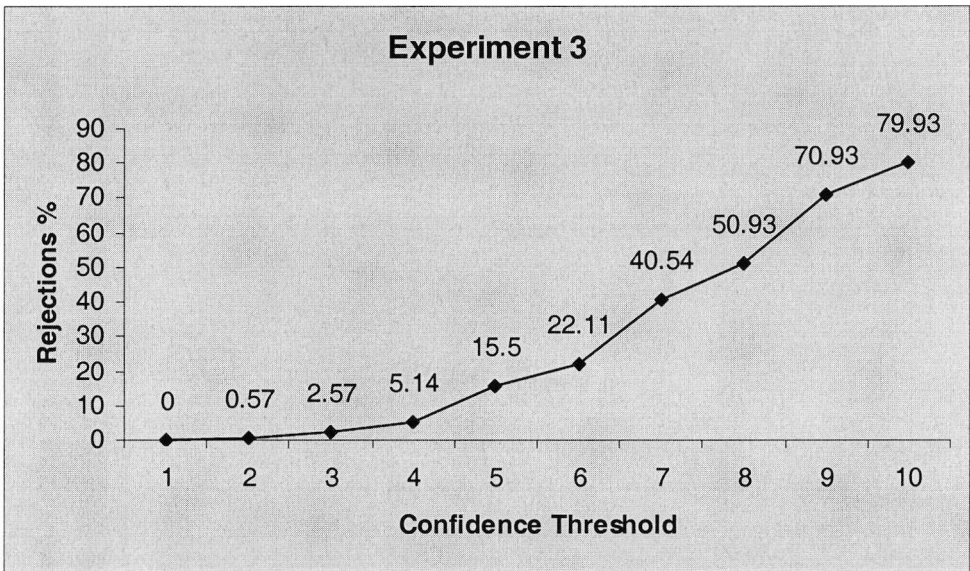


Figure D.4: Rejection rate vs confidence threshold for Experiment 3.

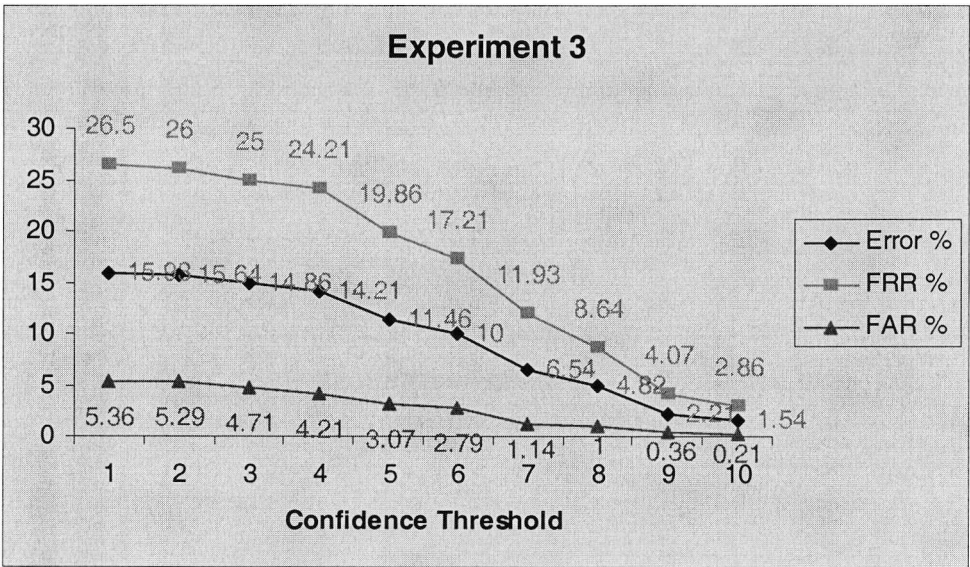


Figure D.5: Error rates vs confidence threshold for Experiment 3.

