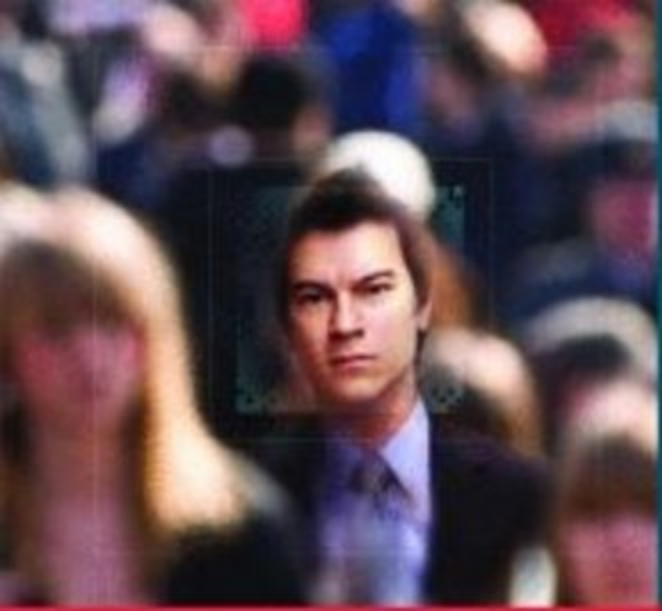

Ted Dunstone
Neil Yager

Biometric System and Data Analysis

Design, Evaluation, and Data Mining

Springer

Biometric System and Data Analysis

Design, Evaluation, and Data Mining

Biometric System and Data Analysis

Design, Evaluation, and Data Mining

by

Ted Dunstone
Neil Yager

Eveleigh, NSW, Australia

Springer

Editors:
Ted Dunstone
Biometix
National Innovation Centre
The Australian Technology Park
Eveleigh, NSW 1430
Australia
ted.dunstone@biometix.com

Neil Yager
Biometix
National Innovation Centre
The Australian Technology Park
Eveleigh, NSW 1430
Australia
neil.yager@biometix.com

ISBN-13: 978-0-387-77625-5 e-ISBN-13: 978-0-387-77627-9

Library of Congress Control Number: 2008934649

Printed on acid-free paper

springer.com

Preface

Overview

Biometrics is the identification of an individual using a distinctive aspect of their biology or behavior. Biometric systems are now being used for large national and corporate security projects, and their effectiveness rests on an understanding of biometric systems and data analysis.

Books on biometrics tend to focus on the details of biometric systems and their components, and distinguish between the various biometric modalities. This book is different. It presents a unified view by focusing on the common aspects of all biometric systems - the input (biometric images and person metadata) and the output (similarity scores). The matching algorithms and sensing technologies may change with each new advance, however the decision process for matches is not affected. Building on this base, this book brings together core aspects of statistics and probability to provide a comprehensive guide to evaluating, interpreting and understanding biometric data.

In doing so, we paint a coherent and intuitive picture which will be equally useful to novices and advanced readers. All theoretical concepts are grounded with practical examples and techniques for the evaluation and set-up of real-world, operational biometric systems. Along the way, other relevant topics are introduced - including machine learning, data mining and prediction - which have been widely applied to other fields, but not yet rigorously applied to biometrics. Case studies and examples from several major biometric modalities are covered.

The focus of biometric research over the past four decades typically has been on the bottom line of driving down system-wide error rates, and as a result powerful recognition algorithms have been developed for a number of biometric modalities. These algorithms operate extremely well under laboratory conditions, but their performance can fall short in the real-world. One reason has been the focus on algorithmic performance as measured by collective statistics. A major theme of this book is the importance of data quality and the accuracy variation between individual users.

These factors form the basis of system performance, and understanding their effects will be necessary for the development of the next generation of biometric systems.

Before putting any biometric system into operation, it needs to be evaluated for accuracy, security and effectiveness. A variety of performance measures are available, and selecting an inappropriate measure can result in highly misleading statistics. This book places an emphasis on the various biometric performance measures, what they mean, and when they should and should not be applied. The evaluation techniques are presented rigorously. However, they are accompanied by intuitive explanations that can be used to convey the essence of the statistical concepts to a general audience.

Objectives

This book aims to provide a comprehensive treatment of understanding and improving biometric systems at various levels. Our practical experience in evaluating and working with a wide range of real-world biometric systems and data is used to inform newcomers and experienced practitioners through clear diagrams and carefully prepared examples. A number of novel techniques for biometric analysis are also presented. The book provides a solid understanding of:

- The fundamentals of biometric systems
- Organization and structure of biometric information
- The basics of multimodal systems
- Biometric data quality issues and their impact on system performance, in particular pictorial examples for fingerprint, face, iris and speech
- The assessment of individual user and user group accuracy
- Setting up and conducting biometric evaluations (there are dedicated chapters on identity document and surveillance systems)
- ISO-consistent vocabulary for all descriptions
- Theoretical methods to create standard system level statistical measures
- Techniques to establish the existing level of fraud detection in identity systems
- Assessing biometric vulnerabilities

Audience

This book represents an effort to bridge the divide between biometric researchers and the biometric industry. It aims to appeal to people with an operational responsibility for biometric systems, as well as technical and academic readers. In general, Part I is directed towards people with operational responsibility and Part II towards technical experts and academics. However, it is hoped that the book as a whole will be of benefit to both audiences. Introductory readers can use the more detailed

aspects in Part II as a reference and a source of information once they have a firm understanding of the basics. Readers with a strong technical background can quickly gain a high-level, intuitive view of the field from Part I. All readers will gain valuable insight into the nature of biometric systems from the examples throughout the book, which are the result of many years of practical experience.

The more technically challenging chapters in Part II - the biometric performance hierarchy - have simple and clear introductory sections. Sections marked with a '*' include more complex, mathematically-oriented material which can be skipped without losing continuity. Diagrams and examples have been extensively used to illustrate the techniques. The emphasis is on practical analysis techniques that are useful for analyzing a variety of real biometric systems over different modalities.

Appealing to introductory and advanced technical readers in the same volume risks making the scope so wide that neither group's needs are adequately addressed. We hope to have avoided this pitfall by drawing from a variety of sources, both academic and industrial.

The following are specific audiences that have been identified, along with the chapters they are likely to find particularly relevant:

- *System Implementers and Designers*: A detailed understanding of techniques helps make sure a system has been configured and tuned correctly. The detection and diagnosis of common problems is a major theme of this book. In particular Chaps. 1, 2, 3, 6 and 12 contain useful information.
- *Researchers and Students*: Every effort has been made to ensure students are guided through both the introductory and advanced topics. For new students, working through Part I should give a good basic grounding in biometrics and biometric analysis. Researchers will be particularly interested in the techniques presented in Chaps. 8 and 9, which present a novel, user-centric framework for system analysis.
- *Security Consultants and System Evaluators*: Accuracy evaluation of system performance and the explanation of results to stakeholders are difficult and costly. This book aims to ensure that the outcomes from an evaluation actually measure expected performance and can be communicated clearly to a non-technical audience. Chaps 1, 2, 3, 5 and 6 are particularly relevant.
- *Forensic Investigators*: For anyone using a biometric system for forensic investigation or legal purposes, it is vital to have a solid understanding of the likelihood of a match being the actual person concerned in order to provide a statistical justification for a decision. In addition to the introductory chapters of Part I, the specialized chapters on identity systems (Chap. 10) and surveillance (Chap. 11) will be valuable.
- *Surveillance Operators*: Setting up and evaluating biometric surveillance systems is more difficult than for other biometric systems. Chapter 11 presents techniques for designing these systems and running trials for their evaluation.
- *Vendors and Algorithm Developers*: This book should allow those developing biometric systems to undertake more realistic testing and evaluation, and hence gain a better understanding of factors that impact on real-world performance.

The measures introduced in Parts II and Part III should be of primary interest in providing new tools and insights into how to improve the accuracy of existing algorithms.

- *Auditors:* As biometrics become part of larger and more critical systems, the need to independently audit their performance is increasing. Traditional IT security auditing techniques do not adequately encompass the non-deterministic nature of the security outcomes for biometric systems. The techniques outlined in Part I will give any auditor a good idea of the issues that need to be considered for auditing a biometric system. In addition, Chap. 12 will be of assistance in assessing a risk management plan.

Organization

This book is in three parts: a general introduction to biometric fundamentals, a detailed technical treatment of analysis techniques and special topics in data analysis.

Many core concepts are presented twice. They are introduced in Part I, with an emphasis on examples and intuitive understanding. Part II provides a deeper, more technical and statistically grounded analysis. Each chapter can be read on its own, but there is a clear progression of ideas from introductory to advanced and from generalized to specific. The terminology in Chap. 6 has been written to be consistent with the ISO 19795.1 standard.

The content is:

Part I: Preliminaries provides a step-by-step introduction to biometric systems, data and analysis that does not require any statistical background. Chapter 1 introduces biometric systems, how they work and related issues such as privacy and vulnerability. Chapter 2 provides the preliminaries of the analysis of biometric systems and data through simple diagrams and worked examples. This chapter has been written to ensure that even readers who are not familiar with statistical techniques can follow the descriptions and gain an appreciation of how biometric systems make decisions. Chapter 3 looks at the biometric data. Practical examples are given for fingerprint identification, face recognition, speaker verification, iris recognition, vascular recognition, keystroke dynamics and others. The implicit structure and organization of biometric matching results are illustrated, along with references on how they can be efficiently stored and retrieved. Multimodal systems, the use of more than one biometric matching algorithm or characteristic, are examined in Chap. 4. In Chap. 5, general requirements for running a biometric evaluation are introduced along with practical considerations and guidelines. Chapter 6 introduces the standard terminology used throughout the book, which is consistent with standard ISO definitions and is a convenient reference.

Part II: The biometric performance hierarchy examines in detail the analysis of biometric systems, from the highest level of system evaluation to the performance of individual users. Chapter 7 details the statistical basis of biometric systems. Starting from the fundamental building block of a single match score, collective error

rates for the system as a whole are developed. Verification, closed-set identification and open-set identification systems are differentiated, with an emphasis on the specific measures that are appropriate for each type of system. Almost always, some users of biometrics systems perform better than others. These differences are the subject of Chap. 8. There is a description of a framework which allows the detection and characterization of problem users, based on individual evaluation and the members of the biometric menagerie. Chapter 9 presents the novel application of data mining techniques to biometric systems to extract knowledge from masses of information. In particular, machine learning is used to automatically detect groups of problem users with common attributes.

Part III: Special topics in biometric data analysis introduces specific topics on the analysis of biometric data and systems. Chapter 10 is on the topical subject of identity document identification systems, identity theft and proof of identity. Techniques to estimate the level of fraud in existing biometric databases are presented, as well as a discussion on the use of biometrics in a legal setting. The set-up and analysis of covert surveillance systems is examined in Chap. 11. The analysis of data from surveillance systems is complicated by the less structured and less controlled environments where they are often required to operate. Chapter 12 provides an introduction to detecting and mitigating vulnerabilities in biometrics systems.

Subjects Not Covered in Detail

This book is different from other books on biometrics. By placing an emphasis on biometric data and analysis, some subjects have been covered in less detail:

- The emphasis is on the input and the output of biometric algorithms; not what happens in between. In other words, the actual matching engine is treated as a "black box". The book does not cover in detail the underlying image-processing and pattern-recognition algorithms which form the core of biometric matching. Also, the hardware and sensor mechanisms used for the acquisition of biometric samples are not detailed, except where it is relevant to the analysis. The analysis is picked up once a similarity score has been generated and it is followed all the way to high-level performance measures.
- There is no attempt to comprehensively review emerging biometrics. However, the techniques introduced are applicable to any new biometric type and, indeed, could be applied to many other areas of pattern classification, from medical imaging to number-plate recognition.
- In general, the focus is on quantitative rather than qualitative analysis. Qualitative aspects of a biometric system include considerations such as cost and ease of use, and these are not discussed in detail.
- In some areas, such as confidence intervals, there are still open questions and academic debate about the best techniques. When this is the case, we present only an overview of the most common techniques. In other areas, such as multimodal

analysis, techniques still are emerging and are treated only at an introductory level in this edition.

Original Research

This book presents a number of topics which have not been comprehensively considered in previously publications. Some insights are the results of original research we have conducted into the analysis of biometric systems, while others have grown naturally from our practical experience in the field.

Original research covers:

Zoo Analysis Through our efforts in evaluating real-world systems, we have gained an appreciation of the importance of the performance of individual users of biometric systems. The original members of the biometric menagerie (sheep, goats, lambs and wolves) are well known. However, we have recently extended this group of animals to characterize other problem users. Chapter 8 contains a detailed analysis of the subject, as well as presenting a user-centric framework for system evaluation. The zoo analogy is used extensively and is an integral component of any biometric evaluation.

Data Mining Although user-level analysis has achieved a level of acceptance in the biometric community, the existing techniques for group-level analysis are primitive. In general, trends in the data are currently discovered through a manual process of directly computing and comparing group accuracy (i.e. men vs women). Chapter 9 draws from the fields of machine learning and data mining, and presents the novel application of intelligent techniques for automated biometric knowledge discovery.

Fraud Level Identification One of the most promising applications of biometric matching is detecting fraud in identity databases. For example, driver's license and passport authorities often have little concrete evidence for their estimates of existing levels of fraud. Chapter 10 presents the results of our initial investigations in this area. In particular, it outlines tests that need to be run, and how the results can be used to establish the extent of fraudulent activity.

Surveillance Systems We have particular expertise in the analysis of biometric surveillance systems. This is an emerging application and there is little published information on the best ways to design and assess these systems. In Chap. 11 we highlight the major issues and provide practical tips for setting up systems and running evaluation trials. This chapter will be an invaluable resource and time-saver for anyone involved in surveillance evaluations.

Vulnerabilities Accuracy in biometric systems has traditionally been based on the probability of success for a random, non-motivated impostor attempting to gain system access. The discovery and reporting of vulnerabilities in biometric systems has not been particularly well analyzed, even though it has a crucial role in ensuring a secure system.

Acknowledgments

This book is the culmination of many years working with, and evaluating, biometric systems. Along the way, correspondence with leading practitioners such as Jim Wayman and Tony Mansfield has contributed significantly to our knowledge, and for this we are greatly indebted. Other colleagues who have shared valuable ideas with us over the years include Johnathon Phillips of NIST; Michael Petrov, Michael Brauckmann and Jonathan Wells of L1; Alfredo Herrera, Frank Weber and Raphael Villedieu of Cognitec; Geoff Poutlon, formerly of CSIRO; Aidan Roy of the Institute for Quantum Information Science; Terry Hartmann of UNISYS; Dijana Petrovska of Biosecure; Terry Aulich of Aulich & Co; Valorie Valencia and Roger Cottam of Authenti-Corp; Brett Minifie and Ian Christofis of HP; and Stephen J. Elliott and Eric Kukula of Purdue University.

Over the years, the clients of our consulting company, Biometix, have challenged us to be clear in our thinking and explanations - our discourse with them has increased our own understanding, ultimately allowing us to write this book with clarity. In particular, it has been a pleasure to work with Ross Summerfield of Centrelink; Kenneth Beaton of KAZ; Ossie Camenzuli, formerly with the RTA; Dominique Estival, Stephen Anthony, Stephen Norris and Son Bao Pham of Appen; Jason Prince, Johnathan Moulds and Karen Shirely of the AFP; and Rick Hyslop and Derek Northrope of UNISYS. Thanks also to Barry Westlake for his inspiring vision of the future of Biometix, and to Fabrice Lestideau, Teewoon Tan and Navin Keswani for their help in building the foundations. A great many others from the research, commercial and government communities, particularly those who are part of the Biometrics Institute, have shaped our appreciation for the many different aspects of biometrics, and to all those we express our deep gratitude.

Min Sub Kim has provided invaluable feedback, and a thorough proof reading, for Part II of this book. Furthermore, editing done by Patrick Weaver, and the huge efforts by Coralie and Simon Dunstone, have contributed greatly to ensuring a high standard and consistent quality throughout. Despite the monumental efforts of all those who helped in the preparation of this manuscript, the authors take full responsibility for any mistakes, inconsistencies or omissions.

Data sets, examples and diagrams were kindly provided by Johnathon Phillips, Tony Mansfield, David Cho, Sammy Phang and Patrick Flynn.

Ted would like to thank the support and forbearance of his friends, the people from the Institute, Isabelle Moeller and Michelle Turner, and in particular the patience (and fabulous food) of Susan Crockett while this book was written - often at odd hours. Neil would like to thank his family for being by his side, even from a world away. He also thanks Fjóla Dögg Helgadóttir, who sparkles and shines, for helping in ways that words cannot measure. This book, and much besides, would not have been possible without her smile.

Ted Dunstone
July 2008 *Neil Yager*

There are a number of sections of this book where the reader may find it beneficial to perform some worked examples using computer software.

The authors have made data and a limited license to a computer program available, free of charge, for readers and their associates, that will make it easy to perform these examples.

This data and the limited license to PerformixPC can be downloaded from www.biomet.org using the special code "PRMDEMO".

This code is not unique so it may be given to others, including students, to carry out these examples.

The limited license for PerformixPC provides virtually all of the functions of the software and does not expire, but is only able to operate on small data sets such as those used in these examples.

This software and data is supplied without any guarantees of its fitness for use and the authors and the publisher accept no liability for its use.

Contents