PREFACE

In recent years, numerous research papers have considered the problem of privacy protection in biometric authentication systems. When someone wants to authenticate himself, the fresh biometric data he presents has to match with the reference related to the one he claims to be. Many proposals have been made to ensure the confidentiality of the biometric data involved during this verification. Also discussed in this compilation, is the problem of finding a face recognition system that works well both under variable illumination conditions and under strictly controlled acquisition conditions. Information fusion techniques are also explored and have been widely developed in the community of voice biometrics along with several uni-modal speaker recognition algorithms.

Experimental clinical trial settings are now often extended to community entities beyond academic research centers. In such settings, a cluster randomized clinical trial (cluster-RCT) design can be useful to rigorously test the effectiveness of a new intervention. Investigators are most commonly interested in assessing the following three types of intervention effects: overall intervention effect, change in intervention effect over time, and local intervention effect at the end of the study. At the design stage of the cluster-RCT, it is essential to estimate a sample size sufficient for adequate statistical power to evaluate the different intervention effects. However, the sample size estimation must account for the multilevel data structure that is necessitated by the nature of the cluster-RCT design. In Chapter 1, the authors consider a three-level data structure and summarize sample size approaches for testing intervention effects within a unified framework of mixed-effects linear models which offer flexibility in the analysis of multilevel data and hypotheses testing in a cluster-RCT. The sample size methods are presented in closed form and

have been validated by simulation studies. Important features of sample size determination for each primary hypothesis are also discussed.

One important application of gene expression profiles data is tumor classification. Because of its characteristics of high dimensionality and small sample size problem, and a great number of redundant genes not related to tumor phenotypes, various dimensional reduction methods are currently used for gene expression profiles preprocessing. Manifold learning is a recently developed technique for dimensional reduction, which are likely to be more suitable for gene expression profiles analysis. Chapter 2 will focus on using manifold learning methods to nonlinearly map the gene expression data to low dimensions and reveal the intrinsic distribution of the original data, so as to classify the tumor genes more accurately. Based on Locally Linear Embedding (LLE) and modified maximizing margin criterion (MMMC), a modified locally linear discriminant embedding (MLLDE) is proposed for tumor classification. In the proposed algorithm, the authors design a novel geodesic distance measure, and construct a vector translation and distance rescaling model to enhance the recognition ability of the original LLE from two aspects. One is the property that the embedding cost function is invariant to translation and rescaling, the other is that the transformation to maximize MMMC is introduced. To validate the efficiency, the proposed algorithm is applied to classifying seven public gene expression datasets. The experimental results show that MLLDE can obtain higher classification accuracies than some other methods.

Voice biometrics, also called speaker recognition, is the process of determining who spoke in a recorded utterance. This technique is widely used in many areas e.g., access management, access control, and forensic detection. On the constraint of the sole feature as input pattern, either low level acoustic feature e.g., Mel Frequency Cepstral Coefficients, Linear Predictive Coefficients or high level feature, e.g., phonetic, voice biometrics have been researched over several decades in the community of speech recognition including many sophisticated approaches, e.g., Gaussian Mixture Model, Hidden Markov Model, Support Vector Machine etc. However, a bottleneck to improve performance came into the existence by only using one kind of features. In order to break through it, the fusion approach is introduced into voice biometrics. The objective of Chapter 3 is to show the rationale behind of using fusion methods. At the point of view of biometrics, it systematically classifies the existing approaches into three fusion levels, feature level, matching-score level, and decision-making level. After descriptions of the fundamental basis, each level fusion technique will be described. Then several

experimental results will be presented to show the effectiveness of the performance of the fusion techniques.

In Chapter 4 the problem of finding a face recognition system that works well both under variable illumination conditions and under strictly controlled acquisition conditions is considered. The problem under consideration has to do with the fact that systems that work well (compared with standard methods) with variable illumination conditions often suffer a drop in performance on images where illumination is strictly controlled. In this chapter the authors review existing techniques for obtaining illumination robustness and propose a method for handling illumination variance that combines different matchers and preprocessing methods. An extensive evaluation of the authors' system is performed on several datasets (CMU, ORL, Extended YALE-B, and BioLab). The authors' results show that even though some standalone matchers are inconsistent in performance depending on the database, the fusion of different methods performs consistently well across all tested datasets and illumination conditions. The authors' experiments show that the best result are obtained using gradientfaces as a preprocessing method and orthogonal linear graph embedding as a feature transform.

Physical activity (PA) is a modifiable lifestyle factor for many chronic diseases and its health benefits are well known. PA outcomes are often measured and assessed in many clinical and epidemiological studies. Chapter 5 first reviews the problems and issues regarding the analysis of PA outcomes. These include outliers, presence of many zeros and correlated observations, which violate the statistical assumptions and render standard regression analysis inappropriate. An alternative two-part generalized linear mixed models (GLMM) approach is proposed to analyze the heterogeneous and correlated PA data. At the first part, a logistic mixed regression model is fitted to estimate the prevalence of PA and factors associated with PA participation. A gamma mixed regression model is adopted at the second part to assess the effects of predictor variables among those with positive PA outcomes. Variations between clusters are accommodated by random effects within the GLMM framework. The proposed methods are demonstrated using data collected from a community-based study of elderly PA in Japan. The findings provide useful information for targeting physically inactive population subgroups for health promotion programs.

In the last years, numerous research papers have considered the problem of privacy protection in biometric authentication systems (1-to-1). When someone wants to authenticate himself, the fresh biometric data he presents has to match with the reference related to the one he claims to be. Many

proposals have thus been made to ensure the confidentiality of the biometric data involved during this verification.

Biometric identification enables to identify an individual among many others without requiring any prior claim of identity (1-to-many). Typically, it consists in checking the belonging of a user to a database. Paradoxically, today, identification constitutes the main application of biometric systems whereas privacy protection in such identification systems has received relatively little attention in the literature. In Chapter 6, the authors show how to extend the previous works on biometric authentication in order to also cover biometric identification while maintaining the privacy of the individuals. This is not a trivial task as the authors want to go much faster than performing comparisons with all the biometric references in the system, i.e. an exhaustive search. The main difficulty comes here from the fact that as they must deal with data that are protected, the authors cannot merely rely on traditional biometric identification methods. This can be overcome successfully thanks to new techniques that have been suggested recently in a few papers.

As discussed in Chapter 7, systematic designs (where the allocation of varieties to plots is according to some fixed, non-random, pattern or scheme) are often used for the controls in large early generation variety trials (EGVTs). These are field trials which compare a large number of new varieties (or lines) with standard/control ones, and which are used to select some high yielding lines for further investigation. Usually the new lines are unreplicated at a site. The efficiency of different types of systematic unreplicated EGVT designs compared to optimal or efficient designs is investigated. Two design optimality criteria are used, and are compared for their association with the probability of selecting the highest yielding varieties. The data are assumed to be spatially dependent, and three models are considered for variety effects: both test and control effects fixed, controls effects fixed but variety effects random; and both test and control variety random.