13 Performance Prediction Methodology for Multibiometric Systems

Natalia A. Schmid and Joseph A. O’Sullivan

13.1 Introduction

Secure personal authentication and identification are challenging problems for modern society. While the rapid development of new information technologies makes life more organized and convenient, it also necessitates the development of new security systems to prevent unauthorized access and abuse. Biometrics-based authentication and identification systems are often designed for controlling access to facilities, personal computing and communication devices, financial accounts, and information databases.

For our purposes, study of biometrics is a study of physiological and behavioral characteristics of an individual [432]. To be qualified as a biometric, it must satisfy a number of requirements. The four most important are (1) universality: everyone must possess it; (2) uniqueness: no two individuals should have the same value of the characteristic; (3) permanence: it must not change with time; (4) measurability: it must be easy to measure. Several characteristics have been proposed, researched, and implemented for personal identification, including voice, infrared facial and hand vein thermograms, fingerprints, face, iris, ear, gait, keystroke dynamics, DNA, signature, acoustic emissions, odor, retinal scan, and hand and finger geometry. Most of these characteristics are believed to be unique. The uniqueness of physiological characteristics is a result of both genetic conditions and random physical processes occurring during the stage of embryonic development. Here, the physiological processes are modeled mathematically.

An engineering approach to the problem of identification is to state it as a recognition problem. Most information available to engineers about biometrics is in the form of images (iris scan, fingerprint, hand geometry) or signals (voice). Often systems (such as iris recognition based on IrisCode [433]) extract a robust set of features selected for good recognition performance. After features are extracted, the most common approach is to treat extracted features as deterministic parameters. For instance in a fingerprint identification problem, a set of minutia points is considered as one of the best deterministic representations in terms of performance, provided that the number of minutiae points is large enough. To make a fingerprint identification system based on minutiae points robust, bounding boxes used for counting the number of matching minutia points in two fingerprints are adjusted adaptively [434]. In face recognition, principal components analysis (PCA) and linear discriminant analysis (LDA) have been leading techniques [435, 436].
recognition techniques based on a three-dimensional representation are the current state-of-the-art techniques (see for example [437]). In hand geometry, features are Euclidean distances between different points on a hand [438]. An alternative to a deterministic approach is to model images, signals, or extracted features as realizations of stochastic processes. While for physical signatures, such as the distribution of magnetic particles on magnetic tape or ink, or the distribution of woodgrains on a wooden surface, estimation of stochastic models is more straightforward, for most biometric signatures it remains a challenging task. Common models of signatures as linear combinations of components may be considered to be stochastic models; the distributions of the coefficients are often closely approximated using standard parameterized distributions. The estimation of these models from training data is beyond the scope of this chapter. However, we note that the number of parameters, features, or components used in a representation must not be too large in order to avoid over-fitting the training data; which ones to use may be determined based on the information provided [439]. In this chapter, we proceed with an analysis assuming the stochastic models are known, providing the analytical tools to predict performance given these models.

A typical identification system is designed to operate in two modes: the off-line mode and the on-line mode. During the off-line mode, data are collected for each individual, and a template is extracted and stored in a database. The template may comprise a vector of features, coefficients in a components model, or parameters in a probability distribution for the data. During the on-line mode, input data are acquired, its template is extracted, and a match is performed against each template stored in the database. Performance of a recognition system is usually measured in terms of the probabilities of error as a function of design parameters (receiver operating characteristic (ROC) curves for a binary problem). Performance of practical biometrics-based identification systems is often evaluated by modeling the decision function as a Bernoulli distributed random variable with its mean estimated using training data and then evaluating the confidence intervals for the estimated probabilities of error [440–443].

Major challenges of biometric system design today include high-performance, efficient strategies for biometric fusion at the matching score level [444, 445]; adjustment of thresholds in decision functions with the goal of achieving optimal performance; fusion at the feature level [446]; defining level of individuality for a unimodal biometric [447]; collection of data under conditions of environmental and demographic variations for the purpose of system testing and performance prediction [432, 448, 449]. In this chapter, we address two major challenges. They are performance prediction and capacity analysis for multimodal biometrics-based identification systems. For our analysis, we model multimodal templates stored in the database of an identification system as noisy realizations of independent and identically distributed stationary ergodic random processes. The unknown input data presented for identification are matched against every template in the database. Here matching is equivalent to finding a noisy template that has a known joint distribution with the input data. The problem of identification is then a statistical hypothesis