How to Measure Biometric Information?

Yagiz Sutcu  
Computer Science & Eng. Dept.  
Polytechnic Institute of NYU  
Brooklyn, NY, 11201, USA  
Email: yagiz@isis.poly.edu

Husrev T. Sencar  
Dept. of Computer Engineering  
TOBB University of Economics and Technology  
Ankara, TURKEY  
Email: htsencar@etu.edu.tr

Nasir Memon  
Computer Science & Eng. Dept.  
Polytechnic Institute of NYU  
Brooklyn, NY, 11201, USA  
Email: memon@nyu.edu

Abstract—Being able to measure the actual information content of biometrics is very important but also a challenging problem. Main difficulty here is not only related to the selected feature representation of the biometric data, but also related to the matching algorithm employed in biometric systems. In this paper, we propose a new measure for measuring biometric information using relative entropy between intra-user and inter-user distance distributions. As an example, we evaluated the proposed measure on a face image dataset.

Index Terms—biometric information; relative entropy

I. INTRODUCTION

Computer-verifiable biometrics, such as fingerprints and iris scans, provide an attractive alternative to classical authentication/verification schemes like passwords. Biometrics have the advantage that, unlike passwords, they do not have to be remembered and they are difficult to forge. However, they pose new challenges. A key characteristic is that each time a biometric is measured, the observation differs slightly. Biometric systems must be robust to such variations and most of the biometric systems deal with such variability by relying on pattern recognition.

One of the important problems in biometrics is measuring the amount of information in biometrics. More specifically, how much identifying information is available from a given biometric modality? Main difficulty here is not only related to the selected feature representation of the biometric data, but also related to the matching algorithm employed. This becomes more relevant question when designing/evaluating biometric template protection methods [1].

There are various efforts addressing this issue. In [2], a significantly large database is used for estimating the randomness contained in iris images. It is shown that inter-user distance distribution forms a perfect binomial distribution with $p = 0.5$ and $N = 249$ degrees of freedom which generates a discrimination entropy of about 3.2 bits/mm$^2$. In [3], the individuality of fingerprints is investigated using statistical models of minutiae distributions. Estimates of fingerprint individuality are obtained using the probability of random correspondence (PRC). PRC is defined as the probability of an arbitrary impostor fingerprint from a population of individuals will have corresponding features similar to that of the query fingerprint. While the concept of effective keyspace of a biometric is introduced in [4], relative entropy is proposed for measuring the information content of the biometrics in [5]. However, we still do not have a general unifying approach for measuring biometric information.

This paper is organized as follows. In Section II, we provide the details of relative entropy-based approach proposed in [5] for measuring the information content of biometrics and discuss its merits and limitations. We also describe our proposed biometric information measure at the end of this section. In Section III, we present our experimental results obtained from a face database as an example application. We conclude in Section IV.

II. MEASURING BIOMETRIC INFORMATION

Statistical entropy is the measure of uncertainty in a random variable [6]. More specifically, entropy of a discrete random variable $X$ with probability mass function $p$ is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log(p(x_i))$$ (1)

Kullback-Leibler divergence (also called information divergence, information gain, or relative entropy), on the other hand, is a non-symmetric measure of the difference between two probability distributions $P$ and $Q$. KL-divergence measures the expected number of extra bits required to code samples from $P$ when using a code based on $Q$, rather than using a code based on $P$, and defined as

$$D(p||q) = \sum_{x=1}^{n} p(x)\log(p(x)/q(x))$$ (2)

In [5], authors proposed using KL-divergence for measuring biometric information. The rationale behind this choice is the fact that, they defined the biometric information as the decrease in uncertainty about the identity of a person due to a set of biometric features measurements. Therefore, KL-divergence between the distribution $p_k(x)$ of a given user, $k$, and the distribution $q(x)$ of the population, (i.e. all users enrolled to the system), that is,

$$D_k(p_k||q) = \sum_{x=1}^{n} p_k(x)\log(p_k(x)/q(x))$$ (3)

gives the biometric information of user $k$.

The average biometric information (ABI) of the system in this case is defined as,
\[ ABI = \frac{1}{N_q} \sum_{k=1}^{N_q} D_k(p_k || q) \]  

(4)

where \( N_q \) is the total number of users enrolled to the system.

In [5], authors consider PCA-based eigenfaces [7] approach for feature representation and tested their proposed measure using a subset of Aberdeen face database [8]. By assuming PCA features follow a multivariate Gaussian distribution, it becomes very simple to estimate the biometric information of each user and the average biometric information of the system as well. However, there are some issues and limitations of this approach.

First of all, the assumption of the PCA features follows a multivariate Gaussian distribution only allows us to reliably calculate \( N_k - 1 \) eigenvalues of the covariance matrix of user \( k \). Here, \( N_k \) is the total number of available biometric samples collected from user \( k \). In reality, total number of samples of a given user is much smaller than the dimension of the raw biometric data (i.e., face image in this case). That is, \( N_k \ll M \). As a result, calculation of more than \( N_k - 1 \) eigenvalues, hence the relative entropy \( D_k(p_k || q) \) are not reliable.

There are several ways to deal with this issue and one of them is to use masking as suggested in [5]. Similarly, some of the many other regularization techniques can be employed [9]. However, most of those techniques require extra parameters which need to be estimated depending on the dataset considered.

Secondly, the assumption of the PCA features follows a multivariate Gaussian distribution is a questionable assumption. In literature, there are many normality tests exist and in our experiments, we evaluated the normality of the PCA features using Lilliefors test [10] which is an adaptation of the Kolmogorov-Smirnov test and is used to test the null hypothesis that data come from a normally distributed population without specifying the expected value and variance.

Thirdly, and more importantly, proposed biometric information measure does not allow us to evaluate the actual amount of information considered.

Consider for instance, PCA features as the representation of the biometric data at hand. Increasing the dimension of the feature vector of PCA features does not increase the separation between the inter-user and intra-user distributions while the average biometric information measured in terms of KL-divergence (as proposed in [5]) increases with the dimensionality of the feature vectors. (See Figure 2(c), red curve.)

Therefore, we propose to use between-user and within-user distributions for estimating the average biometric information instead. That is \( D(p_w \| p_b) \) where \( p_w \) and \( p_b \) are the within-user and between-user distance distributions, respectively. This measure is not only threshold independent, but also takes the similarity measure used for matching into account.

### III. Experiments and Results

In our experiments, we used the Olivetti Face Database (ORL database[11]) and considered PCA-based Eigenface method [7]. ORL face database consists of 10 different images of 40 distinct subjects and we simply considered euclidian distance based similarity measure for comparing biometric data. Sample images from this database are given in Figure 1.

We first calculated the average feature vector for each user (using all available data from that user) and then those average feature vectors are considered as the population distribution for the estimation. Hence, the inter-user distance distribution is calculated as the pairwise distances between those average feature vectors. Intra-user distribution, on the other hand, is simply calculated as combining the distributions calculated as the pairwise distances between the samples of the same user.

#### A. Estimating the Entropy

Estimation of the differential entropy from observations of a random variable is not only great importance for a wide range of signal processing applications but is also a difficult problem. The largest category of entropy estimators is the so-called plug-in estimators. These estimators are based on an initial, parametric or nonparametric, density estimation followed by numerical integration.

In practice, many continuous generating distributions are modeled as Gaussian distributions. However, such estimates are usually not reliable especially when the number of samples is smaller than the number of dimensions. Furthermore, Gaussian assumption for PCA features may not be an acceptable assumption at all. To test the validity of this assumption, we considered Lilliefors test for normality [10] and the result of this test is depicted in Figure 2(a). As can be easily seen from the Figure 2(a), a significant subset of the PCA features actually failed to pass the normality test. While for the first
PCA features, almost half (21) of those features could not pass the normality test, number of non-Gaussian features in the first 100 and 150 PCA features are 29 and 40 respectively.

An alternative to plug-in entropy estimators is formed by entropy estimators that utilize the observations of the stochastic variable directly. Such methods are based on the nearest-neighbor estimation and data-dependent sample-spacing. For a comprehensive overview of entropy estimators see [12].

The nearest neighbor (NN) estimate of entropy for d-dimensional space is calculated as follows. Let \( \rho_n(i) \) be the nearest neighbor distance of \( X_i \) and the other \( X_j \) such that; \( \rho_n(i) = \min_{j \neq i, j \leq n} ||X_i - X_j|| \). Then the nearest neighbor estimate of entropy is

\[
\hat{H}_{NN}(X) = \frac{1}{n} \sum_{i=1}^{n} \log(n \rho_n(i)) + \log(2) + C_E \quad (5)
\]

where \( C_E \) is the Euler constant defined as:

\[
C_E = -\int_{0}^{\infty} e^{-t \log(t)} dt \quad (6)
\]

In Figure 2(c) (blue curve), the variation of the NN estimate of the entropy of PCA features with increasing number of feature vector dimensionality is given. All of the available data are considered as the population distribution for the estimation. Estimated entropy of PCA features increases with the dimensionality at first and then exhibits a saturation-like behavior (around 14.5 Nats) which is similar to the behavior of the total percent variance explained by increasing number of principal components (Figure 2(b)).

It is worth mentioning the fact that, under Gaussian assumption, entropy of PCA features that is estimated directly using covariance matrix diverges extremely fast. While the estimated entropy is about 68 Nats for \( d = 10 \), for \( d = 20 \) and \( d = 50 \) those estimations are about 128 and 293 Nats respectively.

**B. Estimating the KL-Divergence**

Similar to entropy estimation approaches, beside plug-in type estimators, there are estimators which consider nearest neighbor (NN) distances in order to bypasses the difficulties associated with partitioning in a high-dimensional space [13], [14]. In our approach, we considered the NN-based estimator proposed in [14] and here we briefly describe its details.

Let \( \{X_1, ..., X_n\} \) and \( \{Y_1, ..., Y_m\} \) be i.i.d. \( d \)-dimensional samples drawn independently from the densities \( p \) and \( q \) respectively. And let the distance of \( X_i \) to its nearest neighbor in \( \{X_j\}_{j \neq i} \) is defined as

\[
\rho_n(i) = \min_{j \neq i, j \leq n} ||X_i - X_j|| \quad (7)
\]

In addition, also let the distance of \( X_i \) to its nearest neighbor in \( \{Y_j\} \) is defined as

\[
\nu_m(i) = \min_{j} ||X_i - Y_j|| \quad (8)
\]

Hence the NN estimator of the KL-Divergence \( \hat{D}(p||q) \) is given as:

\[
\hat{D}(p||q) = \frac{d}{n} \sum_{i=1}^{n} \log \frac{\nu_m(i)}{\rho_n(i)} + \log \frac{m}{n-1} \quad (9)
\]

Our proposed biometric information measure is estimated using NN-based estimator proposed in [14] and depicted in Figure 3 (red). Similarly, the entropy of inter-distance distribution is also estimated using NN-based estimator (blue).

It is easy to see that, the estimated entropy of the inter-distance distribution is around 8.5 Nats, which is almost 6 Nats less than the entropy estimated using the PCA features distribution directly. This difference is most likely due to the fact that, euclidean distance which is used to compare the similarity between biometric samples only considers a limited amount of information contained in the feature vectors and causes an entropy loss of approximately \%40. Furthermore,
like the biometric information measure (proposed in [5]), the proposed measure (after a slight increase) decreases with the increasing dimensionality of the PCA-based feature vectors. That is mainly due to the fact that, our proposed biometric information measure not only captures the decreasing distinguishing power of additional features to the biometric data, but also provides an insight for choosing an optimal dimensionality for the feature vectors.

IV. CONCLUSIONS

Measuring the information content of biometric data has significant importance. In this paper, we proposed a measure to estimate the biometric information based on KL-divergence between intra-user and inter-user distance distributions. We also evaluated our proposed measure using a face database.

Our current efforts include; (1) testing our proposed measure with larger datasets and with some other biometric modalities and (2) developing a simple and systematic way of choosing features for maximizing the ratio of effectively considered amount of information for a given similarity measure/matching algorithm.

REFERENCES


