EMPIRICAL CAPACITY OF A BIOMETRIC CHANNEL UNDER THE CONSTRAINT OF GLOBAL PCA AND ICA ENCODING

Francesco Nicolo and Natalia A. Schmid

Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV 26506-6109
fnicolo@mix.wvu.edu, Natalia.Schmid@mail.wvu.edu

ABSTRACT

The ability of practical biometric systems to recognize a large number of subjects is constrained by a variety of factors that include a choice of a source encoding technique, quality of images, complexity and variability of underlying patterns and of collected data. Given a source encoding technique, the remaining factors can be attributed to distortions due to a biometric recognition channel. In this work, we define empirical mutual information and recognition rate and evaluate empirical recognition capacity of biometric systems under the constraint of two global encoding techniques: Principal Component Analysis (PCA) and Independent Component Analysis (ICA). The empirical capacity of biometric systems is numerically evaluated as a point of intersection of the empirical mutual information rate curve plotted as a function of the recognition rate and the diagonal line bisecting the first quadrant. The developed methodology is applied to find the empirical capacity of different recognition channels formed during acquisition of different iris and face databases.

Index Terms— Biometrics, Information theory, Stochastic Model, Capacity

1. INTRODUCTION

In many large scale biometric-based recognition problems, knowledge of the limiting capabilities of underlying recognition systems is critical. These limits, however, are determined by a variety of factors including source coding techniques used to process data, quality, complexity, and variability of the collected data. Given an encoding technique, the remaining factors can be attributed to a recognition channel introduced and characterized by Schmid and O'Sullivan in [1]. Similar to a communication channel, a recognition channel is characterized by its capacity, with the difference being recognition capacity. In a biometric-based recognition problem, recognition capacity can be thought as being the maximum number of classes that can be successfully recognized asymptotically with probability of recognition error close to zero when the number of informative samples gets large. Thus, capacity can be viewed as a measure of performance that can be used to evaluate capabilities of large scale recognition systems. Also, since the maximum number of biometric classes that can be successfully recognized is directly related to distortions and noise present in the images or signals submitted for recognition, we propose to treat capacity as a measure of overall quality of data in a given database.

In this work, we introduce the concepts of empirical mutual information rate and constrained empirical capacity. We further develop and evaluate stochastic models for data encoded using global PCA and ICA. These models are applied to evaluate the empirical capacity of iris and face-based recognition systems. The evaluation is performed using data from six public databases.

2. RECOGNITION CAPACITY

This section summarizes results from [1]. Suppose that a biometric database is composed of templates (processed and encoded images) $\mathbf{X}(1), \ldots, \mathbf{X}(M)$ of M distinct biometric classes. Each template, $\mathbf{X}(m)$, is a column vector of length n. Assume that $\mathbf{Y}$ is a template containing information about a biometric class to be identified. If the templates in the database are modeled as realizations of M independent and identically distributed (i.i.d.) random vectors and the template submitted for recognition is viewed as a noisy realization of one of database templates, a template in the database and the template submitted for identification will have some information in common and thus can be described by a joint probability density function $p_{\mathbf{X},\mathbf{Y}}$. Otherwise, the templates in the database and the query template do not have information in common and thus can be described by the product probability density function $p_{\mathbf{X}} p_{\mathbf{Y}}$ with $p_{\mathbf{X}}$ and $p_{\mathbf{Y}}$ being marginals of $p_{\mathbf{X},\mathbf{Y}}$.

Acquired encoded data often allow probabilistic description. Provided that encoded templates are independent or weakly dependent and can be treated as almost identically distributed, the evaluated joint and marginal probability distributions for a biometric template to be recognized and for a

This work was supported by a grant from NSF IUCRC Center for Identification Technology Research.

template from a database can then be used to form the information density,

$$i_n = \frac{1}{n} \log \frac{p_{\mathbf{X},\mathbf{Y}}}{p_{\mathbf{X}} p_{\mathbf{Y}}}. \quad (1)$$

When the template distributions are known, the constrained recognition capacity is the mutual information rate defined as

$$\bar{I}(X,Y) = \lim_{n \to \infty} E[i_n], \quad (2)$$

where the expected value is with respect to the joint probability density function.

In practical cases, given encoded data (templates), their probability densities can be empirically evaluated using classical parametric and modern nonparametric estimation techniques. Then the expression under the expected value in (2) will contain estimated parameters and will not present a deterministic sequence any more. Thus, in practice, we deal with random sequences.

3. PCA-BASED EMPIRICAL RECOGNITION CAPACITY

In this section we provide an expression for the empirical mutual information rate characterizing a noisy biometric channel under the constraint of Principal Component Analysis (PCA)-encoded data and describe how to evaluate the empirical capacity of the channel.

3.1. Gaussian Model for PCA encoded data

Consider a biometric database with M classes. We assume that PCA templates $\mathbf{X}(m),\ m = 1, 2, \ldots, M$, stored in the database are realizations of i.i.d. vector processes. The processes are Gaussian with zero mean and unknown diagonal covariance matrix Λ. The elements along the diagonal are the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ of a scatter matrix Σ estimated using images of M biometric classes (training data).

We model the query PCA-encoded image as a realization of one of database templates contaminated with a realization of zero mean Gaussian noise with independent components having the unknown variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$. Thus $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{N} \sim \mathcal{N}(0, \Sigma_N)$. Σ_N is an unknown diagonal covariance matrix. In our case the noisy candidate $\mathbf{Y}$ is one of the encoded noisy images from the testing set. During the identification the noisy template will be compared against all the templates $\mathbf{X}(m),\ \ m = 1, 2, \ldots, M$, in the database. Since the templates from a biometric library and a query template are realizations of the processes with unknown parameters, the parameters are estimated using training data. Then in place of the expected value in the right side of (2) we introduce the empirical mutual information rate:

$$\bar{I}_n = \frac{1}{n} E_{\mathbf{X},\mathbf{Y}}\{i_n(\Lambda, \Sigma_N)\} = \frac{1}{2n} \sum_{k=1}^{n} \log\left(1 + \frac{\lambda_k}{\sigma_k^2}\right), \quad (3)$$

where the expected value is with respect to the joint probability density function and σ_k^2 and λ_k, $k = 1, \ldots, n$, are the estimated variances of the noise and estimated eigenvalues.

We define the recognition rate as $R = \log(M)/n$, where M is the number of biometric classes to recognize and n is the template length. If we had a sequence of PCA codes $(n, 2^{nR})$ with the recognition rate R, we would be able to evaluate empirically the trend of the sequence of $\bar{I}_n$ as a function of the rate R. Then the empirical recognition capacity can be obtained as a point of intersection between the empirical mutual information curve plotted as a function of the recognition rate and the diagonal line bisecting the first quadrant. This strategy is valid provided that the empirical sequence in (3) is ergodic.

In this work, to find the empirical recognition capacity at a given image resolution, we analyze the dependence of the number of essential PCA components on the resolution of unwrapped interpolated iris images and on face images.

3.2. Databases

All experiments are performed using data from six publicly available datasets: (1) CASIA-III iris database provided by the Chinese Academy of Sciences, (2) an iris database of images collected at West Virginia University (WVU), (3) BATH iris database provided by the University of Bath, (4) ICE, Phase-I dataset used by the NIST in Phase-I of the Iris Challenge Evaluation 2005, (5) FRGC version 1 face database used in the 2006 Face Recognition Grand Challenge (FRGC) Competition, and (6) Yale database, a database of black-and-white 2D frontal view face images.

3.3. Case I: High Pixel Count

In this case, the resolution of a biometric image r is large compared to the number of classes M. Therefore, the matrices Λ and Σ_N are not well estimated. The results obtained in this subsection are not the limiting capacity values. We find values of the empirical capacity per component by analyzing the joint trend of eigenvalues and noise variances as the number of biometric classes and template length increase.

In our experiments, we follow the traditional PCA method that produces estimates of n largest eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and n corresponding eigenvectors of the empirical covariance matrix formed using images of M objects.

The values of the empirical capacity per component $\bar{I}_n$ for the case $r >> M$ are summarized in the fourth column of Table 1. The data can be interpreted as characteristics of the overall quality of different biometric databases.

3.4. Case II: Low Pixel Count

In this case the number of biometric classes M is much larger than the resolution r of images. Therefore, we obtain reliable estimates of the matrices Λ and Σ_N, and thus can appeal

Table 1. Empirical Capacity per component of Biometric Channels using PCA encoded data for the case $M << r$.

Dataset	M	r	$\bar{I}_n$, nats/pc
ICE 2005 Iris	108	11520	0.4305
WVU Iris	108	46080	0.3198
BATH Iris	50	11520	1.1284
CASIA-III Iris	59	46080	0.5030
FRGC 2006 Face	108	32256	0.3537
Yale Face	38	32256	0.4784

to the concept of empirical capacity. To form a large number of classes, we involve a dataset of synthetic iris images generated at WVU [3]. The total number of classes currently supported by this dataset is 10,000. To establish the needed relationship between the number of classes and resolution of data, that is, to ensure that $M >> r$, we downsample segmented transformed iris images to the size 8×90, 8×45, and 2×22. We calculate the PCA-based empirical mutual information rate for different increasing values of the recognition rate, R (see Fig. 1). There are three plots, each parameterized by a specified resolution. The empirical capacity is calculated as the point of intersection between the empirical mutual information rate parameterized by a resolution and a diagonal curve bisecting the first quadrant. The empirical capacity values at the resolutions 8×90, 8×45, and 2×22 are 1.3, 1.5, and 0.4, respectively. As expected, the capacity of PCA-based recognition system evaluated using the images at resolution 8×45 and 8×90 is higher compared to the capacity of PCA-based recognition system at resolution 2×22.

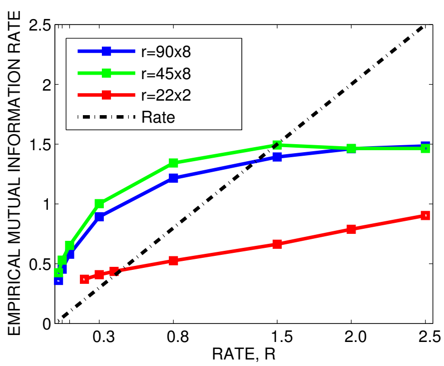

Fig. 1. PCA-based empirical mutual information rate at three resolution levels (2×22, 8×45, 8×90) as a function of the recognition rate, R for the case $M >> r$.

4. ICA-BASED RECOGNITION CAPACITY

In this section we evaluate the empirical constrained capacity of a noisy biometric-based recognition channel under the constraint of ICA-encoded data.

4.1. Bessel K Model for ICA encoded data

Let $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(M)$ be n-dimensional vectors of ICA components. Each vector is a projection of an image from an individual biometric class onto the space formed by the columns of mixing matrix [2]. Suppose that the vectors are independent and identically distributed each described by a Bessel K distribution. To be more specific,

$$\mathbf{X}(m) = \sqrt{G(m)}\mathbf{Z}(m) + \mu(m), \quad m = 1, 2, \ldots, M, \quad (4)$$

where $\mathbf{Z}(m)$ is the Gaussian distributed vector with zero mean and unknown covariance matrix Σ_z, $G(m)$ is a gamma-distributed random variable with unknown parameters α and θ and $\mu(m)$ is the mean vector of $\mathbf{X}(m)$. The vector $\mathbf{Z}(m)$ and the scalar $G(m)$ are independent.

A noisy ICA template presented for identification is modeled as a Bessel K distributed vector augmented with independent Gaussian noise with zero mean and diagonal covariance matrix Σ_N with unknown variances $\sigma_1^2, \sigma_2^2, ..., \sigma_n^2$, that is, $\mathbf{Y} = \mathbf{X}_Y + \mathbf{N}$. Since for a given gain $G = g$, Bessel K distribution is a scaled Gaussian distribution, the empirical mutual information rate can be evaluated using the method of iterated expectations. Under the hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ have a signature in common, the combined vector $[\mathbf{X}^T, \mathbf{Y}^T]^T$, given gain values $G = g$ and $G_Y = g_Y$, is Gaussian distributed with zero mean and covariance matrix $\mathbf{R}_1$ with unknown Σ_z and Σ_N

$$\mathbf{R}_1 = \begin{bmatrix} g\Sigma_Z & g\Sigma_Z \\ g\Sigma_Z & g\Sigma_Z + \Sigma_N \end{bmatrix}.$$

Under the hypothesis that $\mathbf{X}$ and $\mathbf{Y}$ do not have a signature in common, the vector $[\mathbf{X}^T, \mathbf{Y}^T]^T$, conditioned on gain vectors $G = g$ and $G_Y = g_Y$, is Gaussian distributed with zero mean and unknown block diagonal covariance matrix $\mathbf{R}_0$. The matrix $\mathbf{R}_0$ is the matrix $\mathbf{R}_1$ with off diagonal blocks set to zeros.

Taking in account that G and G_Y are i.i.d. the general expression for the empirical mutual information rate is then

$$\bar{I}_n = \frac{1}{2n}\sum_{k=1}^{n}\left\{\left[\mathrm{E}_G[g] + \alpha\theta\sigma_k^2\right]\, \mathrm{E}_{G_Y}\left[\frac{1}{g_Y + \alpha\theta\sigma_k^2}\right]\right\} + \frac{1}{2n}\sum_{k=1}^{n}\left\{\mathrm{E}_{G_Y}\left[\log\left(1 + \frac{g_Y}{\alpha\theta\sigma_k^2}\right)\right]\right\} - \frac{1}{2}, \quad (5)$$

where the parameters α, θ, are the estimated parameters of a Bessel K distribution fitted in a set composed of all coefficients of M concatenated ICA templates, obtained during the training stage. In this case, we practically consider $\sigma_k^2 = \sigma_0^2$ for $k = 1, 2, \ldots, n$. The noise power σ_0^2 is estimated involving additional training templates. Since in a large scale database the number of classes M is very large, the number of independent components n is determined by the resolution of images.

For a fixed image resolution $\mathrm{E}_G[g] = \alpha\theta$, the expression (5) reduces to

$$\bar{I}_n = \frac{1}{2}\left[\frac{S_1(1+\sigma_0^2)\alpha}{\Gamma(\alpha)} + \frac{S_2}{\Gamma(\alpha)} - 1\right], \qquad (6)$$

where S_1 and S_2 are given by

$$S_1 = \int_0^{+\infty} \frac{t^{\alpha-1}}{t+\alpha\sigma_0^2}\exp(-t)\,dt,$$

$$S_2 = \int_0^{+\infty} \ln\left(1+\frac{t}{\alpha\sigma_0^2}\right) t^{\alpha-1}\exp(-t)dt.$$

These two integrals cannot be written in closed form and are evaluated numerically. Note from (6) that the empirical mutual information rate is a function of the shape parameter α of the gamma distributed variables G or G_Y and of the noise level σ_0^2. However, it does not depend on the parameter θ.

4.2. Case I: High Pixel Count

We first consider the case of $r >> M$. M images (one per class) are selected to train the global ICA-based encoding system. As a first step we perform PCA encoding. Then the estimated matrices of eigenvalues and eigenvectors are used to whiten the matrix representing concatenated training images. Later FastICA algorithm [2] is applied to the new matrix resulting in M ICA templates. The values of the empirical capacity (in nats per independent component) are summarized in the fourth column of Table 2. Similarly to the PCA case high values of $\bar{I}_n$ indicate good quality dataset, low values of $\bar{I}_n$ indicate low quality dataset.

Table 2. Empirical Capacity per component of Biometric Channels using ICA encoded data for the case $M << r$.

Dataset	M	r	$\bar{I}_n$, nats/ic
ICE 2005 Iris	108	11520	0.7650
WVU Iris	108	46080	0.5301
BATH Iris	50	11520	2.9483
CASIA-III Iris	59	46080	0.8102
FRGC 2006 Face	108	32256	0.5147
Yale Face	38	32256	0.5952

4.3. Case II: Low Pixel Count

Similarly to evaluating the empirical capacity of PCA-based recognition systems we consider the synthetic iris dataset composed of 10,000 iris classes. To ensure that $M >> r$, we downsample segmented transformed iris images. The global ICA algorithm is trained on a single image per class. We calculate the empirical mutual information rate using expression (6) for different increasing values of the recognition rate, R. Fig. 2 displays the plot of the empirical mutual information rate parameterized by a given image resolution as a function of the recognition rate R. From the figure we can find the empirical recognition capacity evaluated at different resolution levels. The empirical capacity at resolution 2×22 is approximately 0.4. We were unable to obtain the values of the empirical capacity at the other resolutions, since the integrals S_1 and S_2 diverge at high values of the recognition rate.

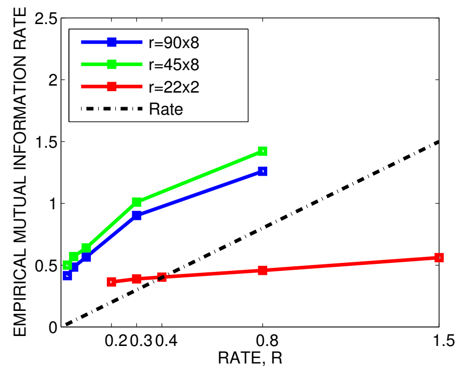

Fig. 2. Shown is the ICA-based empirical mutual information rate at three resolution levels (2×22, 8×45, 8×90) plotted as a function of the recognition rate R for the case $M >> r$.

5. SUMMARY

In this work, we proposed two stochastic models describing PCA and ICA encoded biometric data. These models were applied to evaluate the constrained empirical capacity of recognition channels formed during the collection of data. Four iris databases (CASIA-III, WVU, BATH, and ICE 2005), two face database (FRGC 2006 and Yale), and a medium size dataset of synthetic irises generated at WVU were involved.

6. REFERENCES

[1] N. A. Schmid and J. A. O'Sullivan, "Performance Prediction Methodology for Biometrics Systems Using a Large Deviations Approach," *IEEE Trans. on Signal Processing, Supplement on Secure Media*, vol. 52, no. 10, 2004, pp. 3036-3045.

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley and Sons, 2001.

[3] J. Zuo, N. A. Schmid, and X. Chen, "On Generation and Analysis of Synthetic Iris Images," *IEEE Trans. on Inform. Forensics and Security*, vol. 2, no. 1, 2007, pp. 77-90.