

# On Some Performance Indices for Biometric Identification System

Jay Bhatnagar and Ajay Kumar

Biometrics Research Laboratory  
Department of Electrical Engineering  
Indian Institute of Technology Delhi, New Delhi, India  
jbbhatnagar@ee.iitd.ac.in, ajaykr@ieee.org

**Abstract.** This paper investigates a new approach to formulate performance indices of biometric system using information theoretic models. The performance indices proposed here (unlike conventionally used FAR, GAR, DET etc.) are scalable in estimating performance of large scale biometric system. This work proposes a framework for identification capacity of a biometric system, along with insights on number of cohort users, capacity enhancements from user specific statistics etc. While incorporating feature level information in a rate-distortion framework, we derive condition for optimal feature representation. Furthermore, employing entropy measures to distance (hamming) distribution of the encoded templates, this paper proposes an upper bound for false random correspondence probability. Our analysis concludes that capacity can be the performance index of a biometric system while individuality expressed in false random correspondence can be the performance index of the biometric trait and representation. This paper also derives these indices and quantifies them from system parameters.

**Keywords:** Identification capacity, Joint source-channel coding, Individuality, FRC (false random correspondence probability).

## 1 Introduction

There has been significant interest in large scale applications of biometrics for secure personal authentication. Proposing scalable performance indices for large scale identification systems using biometrics is a challenging problem in biometrics research. The challenge lies in devising indices and evolving their interdependence to convey security, accuracy, privacy and such other measurable characteristics of biometric systems. Authors in [1] discuss system specific indices such as FAR, FRR, GAR, ROC etc. as measures for the imperfect accuracy in relation to signal capacity and representation limitations. However, error rates are performance indices that are dependent on choice of: threshold, user at that threshold. Further, error rates are computed by integrating region from known distributions (genuine and imposter). It will be computationally efficient (for reasons of scalability) to develop average performance measure of biometric identification system. If such a measure can be computed incorporating constrained statistical information instead of complete distributions, then it leads to efficient tools in proposing performance models [2].

Information conveyed in biometric is considered to be inherent part of individuality of a person, and a natural code generated from user specific characteristics – physiological or behavioral. A population of user biometric templates which is uniquely indexed (uniqueness) and maximally separable (distinctiveness) is desirable for reliable identification, however this may not guarantee security [1]. It is well known that in practical applications, user biometric data does not provide error free (perfect) identification. Loss of uniqueness and distinctiveness can be attributed to: (1) statistical variations in biometric data of a user (in-class) and, (2) statistical variations in the biometric data after inter-user comparisons. Authors of [3] allude the need to model information content (statistical measure) of biometric template in relation to the number of users that can be reliably identified. In the same work [3], the above reference to information content is stated as template capacity or inherent signal capacity (and individuality) of a biometric. Authors in [4] use Kullback-Leibler` divergence or the relative entropy  $D(p \parallel u)$  between ( $p$ ) population distribution and ( $u$ ) user distribution per feature to model uniqueness and individuality for face biometric. The approach of [4] demonstrates one of the few important distance measures from information theory perspective. Capacity and error rates have been widely known in information theory literature [2]. For the framework proposed here, Capacity represents: how many users on average can be reliably recognized by the identification system, for a given data base and noise model. Restated, capacity ties the decision accuracy in terms of number of users to database quality and system noise. Similarly, individuality sets a limit to identification capacity in absence of noise (in-class).

The key contributions of this paper can be summarized as follows:

- (i) Formulate and quantify capacity of a biometric identification system (section 3)
- (ii) Estimate the average number of cohort users from the capacity derived in (i)
- (iii) Derive false random correspondence probability with optimality for feature representation (section 4)

Biometric signals of our interest could be at feature level or score level (class genuine and imposter). Several reasons motivate the choice of matching score and feature template, respectively, in the proposed analysis of capacity and individuality. Due to higher dimensionality, it is computationally difficult to extract accurate cross-user information (inter-class statistics) at feature level; though the same can be tractable using matching scores. Also, with number of users tending to infinity (in the asymptote), the i.i.d. (independent identically distributed) assumption for feature vectors is weakened due to correlation present in multiple biometric acquisitions. Traditionally, signal received at the channel output is used for making decision in communication system [2]. Analogously, for biometric system it is matching score which represents the signal space used for decision making and not the query template itself. Therefore, in the framework for capacity of biometric system it is proposed to employ statistics from matching scores. Uniqueness and distinctiveness are statistical properties of biometric feature template and the feature representation, which together define individuality in this work. Hence, discussion on individuality (section 4) entails feature level information.<sup>1</sup>

---

<sup>1</sup> Information: refers Shannon measure for statistical information, expressed in terms of entropy [2].

## 2 Motivation: Review of Noisy Source-Channel Model

In the basic information model given by Shannon [5], i.i.d (independent identically distributed) symbols  $S$  from source with self information or entropy given by  $H(S)$  are transmitted on a memoryless additive white Gaussian noise channel given by  $N$ . Information theorems define maximum and minimum bounds on information rates for achievable reliability and fidelity, respectively [2]. Reliability (capacity) gives an average measure of erroneous decisions due to noisy observations made over an unreliable channel. Reliability measures the accuracy in identifying source constrained by the channel noise. Fidelity, gives an average measure of error in representation of the biometric source. The information theorems on source coding and information capacity cover these concepts [2]. A generalized form of information theorems is the joint source-channel coding theorem [6], which marks need for a combined source-channel model. This result states that a loss in signal-to-noise ratio (alternately, loss of discernability) at the source coder has an equivalent measure in loss of channel reliability (increase in bit error rate of the channel), also known as the source-channel separation principle. The separation principle provides us with a framework to decompose class genuine (also, class imposter) separately as source statistics and noise statistics, which otherwise may be perceived only as source distribution. We now formulate capacity of biometric identification system. **Biometric system capacity:** is the maximum information (max. number of users) per use of the noisy identification channel (in-class, inter-class variations) that can be guaranteed near zero error rate for identification. **Capacity** (pattern recognition context): The term capacity is measured as the maximum average mutual information on observation space of all measurable patterns  $X$  containing  $k$ , which maximizes the occurrence of a pattern of class  $k$  given the distribution of  $k$ . Average mutual information gives an average measure of statistics for optimal decision rule (maximum likelihood) [2].

$$C =: \max \bar{I}(S; X), \forall p(S). \tag{1}$$

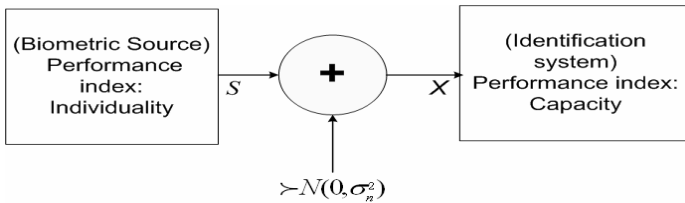


Fig. 1. Schematic of Biometric systems as identification channel

The linear channel model shown in above figure assumes Gaussian noise  $N \succ (0, \sigma_N^2)$  and desired signal modeled with statistics  $S \succ (0, \overline{S_G})$ . The resulting expression for capacity can be shown [2] as given by:

$$\bar{C} = \frac{1}{2} \log_2 \left( 1 + \frac{\overline{S_G}}{\sigma_N^2} \right). \tag{2}$$

The main reasons that influence choice of Gaussian are [4]: (a) Closed form expression is analytically tractable, (b) Tends to be a good reflection of real world distributions, and (c) Has the highest entropy for given variance and is useful in estimating bounds. In this paper *variance* has been used interchangeably for *statistics*. In context of the proposed approach signal statistics will cover all desired information or signals of interest for identification. Similarly, noise statistics will capture the mechanisms that restrict reliable identification.

### 3 A New Framework for Capacity

In this section we combine tools discussed so far to develop a model for estimating capacity of the biometric system. From a signal theory perspective [7],  $M$  registered users are required to be represented by an  $M$ -ary signal probability space. However, measures on the signal space of matching scores can be uniquely defined for a single user. A challenging problem in formulating capacity with score level information for  $M$ -users is that different users registered in the biometric system have different sample spaces (or class distributions) and a unique probability space cannot apparently be defined. The following model removes this shortcoming and facilitates use of score level signals to formulate a new framework for biometric system capacity.

Figure 2 depicts a model for estimating identification capacity. Let  $\{\hat{g}_m\}$  &  $\{\hat{i}_m\}$  denote the median scores taken from genuine and imposter distributions for user  $m \in M$ ; for which the respective distributions peak. Let  $A$  be the *desired user* transmitting information sequence  $\{\hat{g}_m\}$  and let  $B$  be the *interfering user* trying to corrupt these symbols by transmitting an information sequence  $\{\hat{i}_m\}$ .

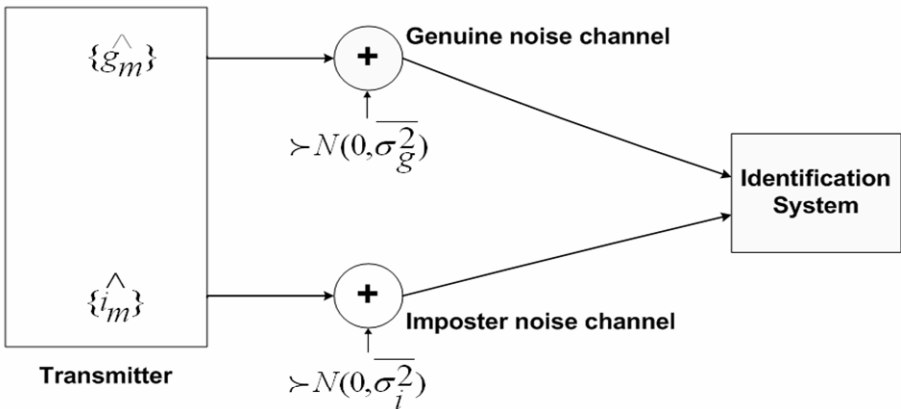


Fig. 2. A two channel biometric identification system

Thus for each transmitted genuine signal from  $\{\hat{g}_m\}$  in the indexing order  $m$ ,  $B$  transmits an interference symbol by transmitting the same indexed imposter signal from  $\{\hat{i}_m\}$ . Further, information symbols  $\{\hat{g}_m\}$  and  $\{\hat{i}_m\}$  are subject to random spreading given by additive white Gaussian channels. The spreading for  $A$  is from a genuine Gaussian channel  $\succ N(0, \sigma_{g(m)}^2)$  and that for  $B$  is from imposter Gaussian channel  $\succ N(0, \sigma_{i(m)}^2)$ . Variances given by  $\overline{\sigma_g^2}, \overline{\sigma_i^2}$  respectively are averages for genuine and imposters from all enrolled users  $M$ . (Figure 2) and denote the average trend to model the biometric system<sup>2</sup> noise. This almost completes formulating the model for  $M$ -ary hypotheses on a single probability space [8]. The resulting model looks Neyman-Pearson type used in jamming/intrusion systems, with target  $A$  and interferer  $B$ . Figure 3 is simplified form of Figure 2 incorporating a the worst case additive noise  $\succ N(0, \sigma_i^2)$  for both sets of information sequences. Figure 4 illustrates a useful statistical distance measure between  $\{\hat{g}_m\} = s_1$  and  $\{\hat{i}_m\} = s_2$  given by  $d_m$ . Discernability (decidability index) improves with increase in  $d_m$ . As observed in Figure 3, the transmitter is a 2-ary or binary hypothesis transmitter for  $M$  signal pairs that are randomly spaced, since  $d_m$  is a random variable. The resulting average signal energy for randomly chosen signal pair can be given:  $(s_1^2 + s_2^2)/2 \cong 2.5d_m^2$ , on the assumption that  $s_1 \cong d_m$ . The average for  $M$ -ary set of signal pairs can be given as  $2.5\overline{d_m^2}$ .

This completes formulating the model for biometric identification channel. Therefore, the signal to noise ratio as required in (2) can be expressed as:

$$\frac{\overline{S_G}}{\sigma_N^2} = \frac{2.5 \overline{d^2}}{\max(\overline{\sigma_g^2}, \overline{\sigma_i^2})} \quad (3)$$

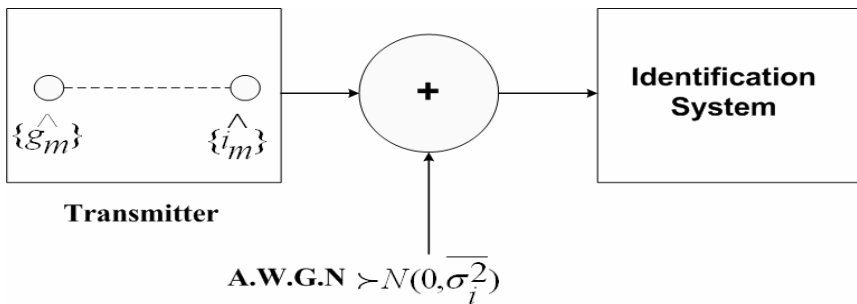


Fig. 3. A Generalized model for biometric identification system with random coding

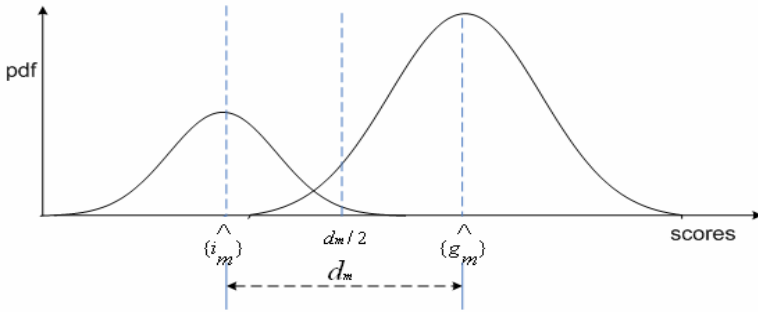


Fig. 4. Statistical distance measures and distributions for user  $m$

The achievable rate  $R$  of identification channel can be given using (2):

$$C_1 = \frac{1}{2} \log_2 \left[ 1 + \frac{2.5 \overline{d^2}}{\max(\sigma_g^2, \sigma_i^2)} \right]. \tag{4}$$

For model variations to incorporate new information/statistics (independent to & additional of existing signal statistics) and its effect on capacity, variance of new statistics can be added in the numerator of the log argument in (4), similarly any additions to noisy statistics can be accounted in the denominator. One such example is that of capacity improvement from incorporating quality indices at score level. If we apply the fact that the median (peak) genuine scores are different for different users, then the variability of the peaking of matching score from different users, offers a user-specific measure to improve classification performance [9]. Performance improvement using this signal statistics can be given by extracting the variance of the

peaks of genuine scores from all  $M$  users, denoted as  $\hat{\sigma}_g^2$ . Then under revised frame work new system capacity can be given by:

$$C_2 = \frac{1}{2} \log_2 \left[ 1 + \frac{\hat{\sigma}_g^2 + 2.5 \overline{d^2}}{\max(\sigma_g^2, \sigma_i^2)} \right]. \tag{5}$$

We will now propose the last setting formulating capacity with use of cohorts *score*. In the inception of cohorts [9], it was shown that an extended score template incorporating neighborhood non-matching scores and a matching score gives substantial improvement in classification performance. Not all non-matching scores do well in the cohorts set. Only close non-matching scores/users to the class matching score will constitute a complement set which gives an appended decision space or increased confidence level of decision. The notion of cohorts illustrates that loss of uniqueness (and, accuracy) due to system noise can be compensated by exploiting the structure of interfering bins [9] and augmenting close non-matching scores in the new decision space. It is proposed to employ one possible selection of cohorts using  $3\sigma$  bound.  $3\sigma$  rule is a loose probability bound in Markov sense [8], in that it includes all scores lying within 33 % probability about the mean/median genuine score. Markov inequality is a weak approximation as it gives the loosest bound in estimating

the spread of a random variable about its median. The rule of  $3\sigma$  will select  $\pm\sigma, \pm2\sigma, \pm3\sigma$  points about the peak matching (genuine) score for each user to estimate cohorts. Also towards the tails of the class genuine distribution, a non-matching score may not give accurate information on the class/ class neighbors.

Denote the average of such  $M$ -variances as  $\overline{\sigma_c}^2$ .

$$C_3 = \frac{1}{2} \log_2 \left[ 1 + \frac{\overline{\sigma_c}^2 + \hat{\sigma}_g^2 + 2.5d^2}{\max(\overline{\sigma_g}^2, \overline{\sigma_i}^2)} \right]. \tag{6}$$

From the definition of identification capacity the number of users ( $\chi$ ) that can be identified reliably.

$$\chi = CM. \tag{7}$$

If,  $C \in [0,1]$ , then  $1 - \overline{C} = C \in [0,1]$ . This gives the complement set  $\overline{C}$ , of the maximum average unreliability of the identification channel. In terms of average number of close/unreliable users, it can be formulated that;

$$N = \overline{CM}, N \leq M. \tag{8}$$

For a fixed number of users  $M$ , greater the value of  $C$ , smaller is the useful cohort set given by (8). This is particularly the case with highly individual biometrics. For highly individual biometrics,  $C$  is expected to be closer to 1 than 0.5 (also, validated by results for capacity in section 5). However, the actual number of (useful) close cohorts for users varies with user. A variation of (4) as given below can provide per user capacity. The user specific cohort requirement can be obtained as follows:

$$C_m = \frac{1}{2} \log_2 \left[ 1 + \frac{2.5d_m^2}{\max(\sigma_{g(m)}^2, \sigma_{i(m)}^2)} \right]. \tag{9}$$

Similarly, extrema of identification capacity for biometric system can be proposed to give peak capacity and minimum capacity based on the following expressions obtained from system parameters.

$$\left(\frac{S_G}{\sigma_N^2}\right)_{\max} = \left(\frac{2.5d_m^2}{\max(\sigma_{g(m)}^2, \sigma_{i(m)}^2)}\right)_{\max}, \quad \left(\frac{S_G}{\sigma_N^2}\right)_{\min} = \left(\frac{2.5d_m^2}{\max(\sigma_{g(m)}^2, \sigma_{i(m)}^2)}\right)_{\min} \tag{10}$$

### 4 Individuality and False Random Correspondence Probability

In biometrics literature, the individuality is related to and expressed in terms of probability of false random correspondence [10]. The prototype proposed in [10] for random correspondence of fingerprint defines a tolerance area which actually is tolerable distortion measure, as we introduce later in this section. Random correspondence was introduced in [10] as the limiting error rate arising from intra

class variability given noisy versions of a user template. Infact, intra class variability contains some measure of random correspondence along with system noise. For a really noisy biometric system, attributing intra class variability to random correspondence may be quite inaccurate. **Individuality:** As minimum information computed on population statistics comprising of templates from all users (one per user), subject to hamming distortion given by  $\delta$ . **FRC (false random correspondence):** Probability that two different user templates are  $\delta$  similar, for which the rate distortion and distortion rate exist and are equivalent.

In documenting iris as a highly individual biometric, Daugman [11] used variability measures at feature level characterized for Gabor features. However, after [11] not much work on random correspondence in the literature employed entropy or entropy rate (rate distortion theory). A plausible explanation for the above could be that the Daugman's approach in [11] used an exhaustive database on which probability measures and variability were shown to fit a binomial distribution for normalized hamming distances (scores) owing to encoding scheme from the Gabor features. Thus [11] illustrated an empirical model for uniqueness of iris which was an accurate asymptote for population statistics. We point here that encoding refers to feature representation method and its binarization. Author in [11] showed that encoding iris based on Iris code of Gabor features gives code with binomial distribution (independent realizations of many Bernoulli) with peak at 0.5 which signifies maximum source entropy code for the binomial family. By encoding variations based on hamming distance, source coding used in [11] applies a hamming distortion metric to locally source code partitions of the template along a trajectory in the iris. This condition also provides a maximally spaced code book as entropy of hamming distance (which is binomial distributed) peaks at 0.5 for best uniqueness, a clairvoyant choice in favor of the Gabor representation. It is therefore good to infer that uniqueness/individuality is dependent on the choice for representation of feature statistics. Applying rate-distortion concepts [12] it is proposed that approach employed in [11] and the approach in [10] are two related manifestations of a more general approach. Furthermore, it is analytically difficult to accurately evolve parametric models for larger number of features and/or larger population that can be generalized for all biometrics [12]. In this section the rate-distortion frame work will be revived to propose that: 1) individuality is a property of the source that can be measured in representation statistics (source coding features) given some minimum distortion, 2) individuality is related to false random correspondence probability for a distortion constraint  $\delta$ . The purpose is to seek the minimum information rate for an average distortion constraint or the average information rate for minimum distortion (both are equivalent approaches in giving optimal source representation [12]). For a choice of representation method (PCA, LDA, Gabor etc.) and given a distortion constraint, smaller the rate-distortion more efficient is the representation technique. In that sense, rate-distortion is a direct measure of individuality. We propose to use a minimum limiting distortion to generate a source code book and study the distance distribution on this code book to formulate false random correspondence. This approach is rationally similar to that discussed in [2]. The following corollary is useful in providing the analytical basis to false random correspondence based on distance measures.



**Corollary 1.** Given a source code book  $\hat{F}$  that lies in the rate-distortion region for avg. minimum distortion  $\bar{D}$ , If the distribution of hamming metric defined on source code book  $\hat{F}$  denoted by  $\{d_i\}, \forall i \in I \geq KM$  follows a binomial distribution with probability = 0.5, then minimum of achievable rate condition is implied.

**Proof:** We use result proved as bound for rate-distortion in **appendix A**

$$I(f_k; \hat{f}_k) \geq h(f_k) - h(f_k - \hat{f}_k). \tag{11}$$

$$I(f_k; \hat{f}_k) \geq \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - h(f_k - \hat{f}_k). \tag{12}$$

We use the following inequality for any source coding model [14]

$$h(f_k) \geq h(\hat{f}_k) \approx h(d_k) \geq h(f_k - \hat{f}_k). \tag{13}$$

$h(d_k)$ , denotes the entropy of hamming distance distribution, hamming distance computed between code vectors for feature  $k \in K$ . The variability/uncertainty in  $d_k$  denotes average hamming distortion, for equivalent to variance of  $f_k$ .

$h(d_k) \geq h(f_k - \hat{f}_k)$  is important to tighten the inequality further in (12).

Also, for  $k$  independent observations (appendix B).

$$\sum_{\forall k \in K} I(f_k; \hat{f}_k) \geq \sum_{\forall k \in K} \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - h(d_k) = R(\bar{D}) \tag{14}$$

Given  $d_k \in (0,1,\dots,N)$  is a discrete random variable,  $\{d_k\}$  gives independent realizations of  $K$  random variables for  $K$  different features.

We define,  $d_i = \sum_{\forall k \in K} d_k; i \in I$ . This gives realizations from independent  $K$  partitions and hence independent random variables. (14) shows addition of hamming metrics from  $K$  partitions (features) of  $M$ -templates to give the hamming distance distribution for  $\hat{F}$ . We characterize the independence of hamming metrics  $\{d_i\}$  by binomial distribution parameterized by  $\succ (p, I)$  [8]. Minimization of (14) is possible if entropy  $h(d_i) = \sum_{k \in K} h^{(m)}(d_k)$  is maximum, entropy for binomial  $\succ (p, I)$  maximizes for  $p = 0.5$ , proving the corollary. (Indexing by  $m$  denotes a template- $m$  which can have a maximum  $I$  as hamming metric in distance computations across the  $M$  template population). Therefore, a minimum rate achievability using (14) can also be viewed as the condition for accurately transferring

uniqueness from source statistics onto the space of representation statistics, given the distortion constraint. Conversely, a source representation for which the above condition is achieved with equality gives optimal rate-distortion. As a simple illustration, we cite the Iris code and its optimality in capturing uniqueness of iris biometric using a Gabor representation [11].

The corollary with propositions presented in this section help in formulating the false random correspondence probability of a biometric. If, we approximate  $H(F) \cong H(\hat{F}) = h(d)$  Since,  $h(d) \geq H(\bar{D})$ , Define,  $d_i \leq \delta; \delta > 0$  ; where  $\delta$  denotes tolerable hamming distortion which is binary hamming metric for tolerable distortion given by  $\bar{D}$ . The tolerable hamming threshold given by  $\delta$  corresponds to conditional entropy given by  $h(d / d \leq \delta)$ . An exponent of conditional entropy to base 2 gives the average total number of sequences that lie within a hamming distance of  $\delta$ . This gives an analytical expression for the average false random correspondence probability as:

$$P(\text{false correspondence}) = \frac{2^{h(d / d \leq \delta)}}{2^{h(d)}} \tag{15}$$

Clearly, from (15), increase in  $\delta$  reflects greater distortion, thereby increasing the numerator with conditional entropy in the exponent. This results in a higher false correspondence. The basic form of entropy function will depend on the *choice* of feature representation. Thus, uniqueness is dependent on the choice of feature representation. An interesting observation from this section is in noting that (15) is the zero noise reliability of the biometric system. Thus, (15) defines a minimum (best) error rate of noiseless biometric system, for a target population.

## 5 Experiments

In order to estimate the capacity formulated in section 3, we perform experiments on real biometric samples. Hand images of 100 users (10 images of each user) employed in [13] were used to simultaneously extract palmprint and hand geometry features. The feature extraction for each of the modalities is same as detailed in [13]. Class genuine and imposter score distributions, using Euclidean distance, were generated to

extract the following parameters:  $\overline{\sigma_i^2}$ ,  $\overline{d^2}$ ,  $\overline{\sigma_g^2}$ ,  $\hat{\sigma}_g^2$  and  $\overline{\sigma_c^2}$ . Table-1 as given below shows system parameters.

Some significant insights can be acquired from Table 2 which shows capacity and number of cohort users based on formulae (6), (7), (8) and (9). Capacity increase from incorporating user specific statistics and cohorts is about twice the capacity prior to the additional statistics, this explains improved identification performance of the biometric system as a result of incorporating additional information gains. Clearly,

**Table 1.** System parameters from the experiments

Biometric Modality	$\overline{\sigma_g^2}$	$\overline{\sigma_i^2}$	$\overline{d^2}$	$\hat{\sigma}_g^2$	$\overline{\sigma_c^2}$
Palmprint	1.5e+4	6.3e+4	2.2e+4	1.4e+4	3.5e+4
Hand geometry	2.988	7.652	1.295	0.893	2.55

**Table 2.** System capacity and required number of cohorts

Biometric Modality	$C_1$	$C_2$	$C_3$	$N_1$	$N_2$
Palmprint	0.45	0.53	0.7	55	47
Hand geometry	0.25	0.31	0.45	75	69

palmprint gives a superior identification performance, based on measurements for the database. The difference in identification capacity between palmprint and hand geometry remains approximately the same for every additional statistic.  $N_1$  is smaller than  $N_2$ , and indicates the number of cohorts users for two distinct cases when (1) raw capacity, (2) capacity with user specific statistics.

## 6 Conclusions

This paper formulates a new frame work for performance indices of biometric identification system. Section 3 proposed system level performance index namely the identification capacity. It was also illustrated in section 3, that boosting capacity is possible using extra information/ variability from statistics such as user quality, and cohorts. Capacity is useful to compare noisy performance at systems level and hence can be considered as a local performance index. Experimental results in section 5 gave real values to capacity for the input statistics and inference on number of cohort users. Section 4 formulates a generalized approach for false random correspondence (and, individuality) using rate-distortion and hamming distortion measure. Individuality is a global performance index, in that it gives the upper limit to performance of the biometric trait for a noiseless biometric system. It is also important to point here that information theoretic approach relies mainly on statistical averages; hence performance indices such as capacity & false random correspondence indicate average trends.

Several open problems spring from this analysis, the most natural is in asking how capacity can be related to a minimum test sample size and training sample size to

guarantee a stable ROC, since both capacity & reliability are system performance indices like ROC. Study of individuality is apparently closer to the theory of evolution than information theory. Our future work will consider extending the current analysis to other biometric traits such as fingerprint, iris, face, and gait and study the resulting performance measures from experimental work. A more challenging problem will be to evolve performance indices for multi biometric system that employ various types of fusion principles.

## Acknowledgement

This work is partially supported by the research grant from Ministry of Information and Communication Technology, Government of India, grant no. 12(54)/2006-ESD.

## References

1. Jain, A.K., Pankanti, S., Prabhakar, S., Hong, L., Ross, A., Wayman, J.L.: Biometrics: A Grand Challenge. In: Proc. ICPR, UK, vol. II, pp. 935–942 (2004)
2. Gallager, R.G.: Information Theory and Reliable Communication. John Wiley, Chichester (1968)
3. Jain, A.K., Ross, A., Pankanti, S.: Biometrics: A Tool for Information Security. IEEE Trans. Information Forensics and Security 1(2), 125–143 (2006)
4. Adler, A., Youmaran, R., Loyka, S.: Towards a Measure of Biometric Information (February 2006), <http://www.sce.carleton.ca/faculty/adler/publications>
5. Slepian, D.: Key Papers in the Development of Information Theory. IEEE Press, New York (1974)
6. Vembu, S., Verdú, S., Steinberg, Y.: The Source-Channel Separation Theorem Revisited. IEEE Trans. Information Theory 41(1), 44–54 (1995)
7. Simon, M., Hinedi, S., Lindsey, W.: Digital communication Techniques: Signal Design and Detection. Prentice-Hall, NJ (1995)
8. Feller, W.: An Introduction to Probability Theory and Applications. John Wiley & Sons, Chichester (1971)
9. Aggarwal, G., Ratha, N., Bolle, R.M.: Biometric Verification: Looking Beyond Raw Similarity Scores. In: Workshop on Biometrics (CVPR), New York, pp. 31–36 (2006)
10. Pankanti, S., Prabhakar, S., Jain, A.K.: On the Individuality of Fingerprints. IEEE Trans. PAMI 24(8), 1010–1025 (2002)
11. Daugman, J.: Probing the uniqueness and randomness of Iris Codes: Results from 200 billion iris pair comparisons. Proc. of the IEEE 94(11), 1927–1935 (2006)
12. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Chichester (1991)
13. Kumar, A., Zhang, D.: Feature selection and combination in biometrics. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 813–822. Springer, Heidelberg (2005)

## Appendix A

### Definitions:

$R(\bar{D})$ : Rate distortion function is the infimum of achievable rates  $R$ , such that  $R(\bar{D})$  is in the rate-distortion region of the source ( $K$ -feature statistics for all  $M$  users) for the corresponding distortion constraint  $\bar{D}$ .

We rephrase the need of Gaussian assumption to source coding (based on estimation theory) - For Gaussian distribution under mean sq. error, the conditional mean of  $\{\hat{f}\}$  is optimal estimator of source  $\{f\}$ ;  $f \in F, \hat{f} \in \hat{F}$ . Furthermore, mean square distortion  $\bar{D}$  as the distortion measure in Gaussian frame work is adopted for following reasons:1) generally useful for image information, 2) gives the minimum rate for representation error(tightened by the fact that rate at min. distortion is equivalent to distortion at min. rate), which is required in formulating individuality. Though the population statistics for  $F$  need not actually be Gaussian, we employ Normal approximation to  $M$ -user statistics to deduce the achievable worse case lower bound for rate-distortion  $R(\bar{D})$ .

*Problem:* To formulate rate distortion  $R(\bar{D})$  in source coding Feature statistics of  $K$ -different features per template, for  $M$  user population, approx. by Gaussian (for large number of users); and for avg. distortion measure given by

$$\bar{D} \geq E \left| F - \hat{F}_k \right|^2; k \in K$$

Hence, to show that:

$$R(\bar{D}) = \min \sum_{k=1}^K \frac{1}{2} \log_e \frac{\sigma_k^2}{D_k}; \sum_{\forall k \in K} D_k = \bar{D}$$

### Solution:

Let the  $K$ - different features per template be i.i.d (For non-i.i.d we subtract a template self-covariance from  $\sigma_k^2$ , in the numerator of the result to prove). Idea is to source code each  $k \in K$  features and then concatenate the  $K$  code blocks to generate the template code. We prove the stated result for  $k = 1$ , which can be easily generalized under uncorrelated ness to  $k = K$ .

$$\begin{aligned} I(f_k; \hat{f}_k) &= h(f_k) - h(f_k | \hat{f}_k) \\ &= \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - h(f_k - \hat{f}_k | \hat{f}_k) \\ &\geq \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - h(f_k - \hat{f}_k); \text{(Conditioning reduces} \\ &\qquad \qquad \qquad \text{entropy)} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - h(N(0, E(f_k - \hat{f}_k)^2)); \\
&\geq \frac{1}{2} \log_e (2\pi e) \sigma_k^2 - \frac{1}{2} \log_e (2\pi e) D_k \\
R(D_k) &= \frac{1}{2} \log_e \left[ \frac{\sigma_k^2}{D_k} \right] = \inf I(f_k; \hat{f}_k)
\end{aligned}$$

For  $K$  independent variables (feature sets) can be generalized as below:

$$R(\bar{D}) = \sum_{\forall k} R(D_k) = \sum_{\forall k} \frac{1}{2} \log_e \left[ \frac{\sigma_k^2}{D_k} \right] = \inf I(F; \hat{F})$$