To appear in Pattern Recognition, 2009

On Estimating Performance Indices for Biometric Identification[†]

Jay Bhatnagar, Ajay Kumar

Abstract

*This paper investigates an information theoretic approach for formulating performance indices for the biometric authentication. Firstly, we formulate the constrained capacity, as a performance index for biometric authentication system for the finite number of users. Like Shannon capacity, constrained capacity is formulated using signal to noise ratio which is estimated from known statistics of users' biometric information in the database. Constrained capacity of a user and of biometric system is fixed, given the database and the matching function. Experimental analysis using real palmprint and hand geometry images illustrates use of constrained capacity to estimate: (i) performance gains from the cohort information, (ii) the effective number of user-specific cohorts for a user and for the biometric system, (iii) information content of biometric features, and (iv) the performance of score level fusion rules for multimodal biometric system. Secondly, this paper investigates a rate-distortion framework for formulating false random correspondence probability as performance of a generic biometric. Our analysis concludes that constrained capacity can be a promising addition to performance of a biometric system. Similarly, individuality expressed as false random correspondence probability can be the performance index of a biometric trait***.**

I. Introduction

I (a) Motivation

There has been a significant interest in large scale applications of biometrics for secure personal authentication systems. The superiority in using biometrics over the conventional methods for user authentication, such as passwords and tokens, comes from the two vital attributes of biometric information [1],[3]: (i) it is unique to an individual and hard to duplicate when properly protected, and (ii) it validates the user and not the holder (given that (i) is true). Performance estimation is a key issue in the comparison

[†] A conference version of this paper was presented in Aug. 2007 during ICB 2007, in Seoul, and is referred as [25].

of biometric trait and biometric system for usage in large scale secure access technology. The challenge lies in evolving indices that can provide measures for security, reliability, privacy in relation to the biometric information content [1], [3]. Jain *et al*. [2] have discussed performance indices like *FAR, FRR, GAR, ROC* as measures for the imperfect accuracy in relation to signal capacity and representation limitations of the biometric template. The *ROC* of biometric system is a plot of *GAR* against *FAR*, which is defined from the genuine and imposter distributions for continuously varying thresholds. Error rates can be defined for a user as well as for all users in the database, for a continuously varying threshold [1]. Error rates, specified by thresholds on *ROC*, give some direction in choosing the operating point of biometric system. However, as a probability measure the error rate itself does not give much 'information' on the interplay between biometric signals (biometric matching scores, biometric features) of different users that define the probability space. Another limitation of error rate is that it specifies authentication performance as viewed by the decision block (classifier), at a given threshold and cost function. This means that performance measure based on the error rates firstly depend on thresholds and secondly, as probability measures it may be of limited use in describing the structure of signals constituting the sample space. We next discuss use of *EER* (equal error rate) as a performance measure [3] for comparing biometric systems and biometric traits.

The *EER* actually defines a fixed point on the *ROC* for which $(FAR = FRR)$, given a database and matching function. Therefore, *EER* that is fixed, for a given database and matching function, is widely used in the comparison and evaluation of biometric systems [1],[3],[21],[24]. However, a limitation of *EER* is that it does not give much insight on the achievable limit for performance of biometric system or of individual users in the database for reasons that shall now follow. Firstly, *EER* depends on the choice of training set and test set of biometric templates; hence, there may be a need for an average *EER* from different combinations of test and training sets. Secondly, a performance measure given by probability densities usually can aid in giving only point estimates of the qualitative or quantitative performance for the underlying sample space. *EER* is defined for a threshold on the *ROC*, and as discussed here it is evidently not a statistics generated directly from the complete 'signal space' comprising of users' biometric matching scores. Another performance measure of a biometric system can be given by *CMC* (cumulative match

characteristic) that ranks biometric templates of all users in the database based on their relative accuracy in authentication [3]. As a performance indicator *CMC* provides a qualitative comparison among biometric templates for all users and has been related to *ROC* [3], [21]. However, again *CMC* does not give much idea on inter-relation of users' based on their biometric information nor it provides a measure of the information content in biometric features or in matching scores.

In [4], Jain *et al.* have suggested the need to model information content of biometric template in relation to the number of users that can be 'reliably' identified. We point out that the terms 'reliable' and 'reliability' as used here correspond to error free biometric authentication. So, instead of the probability measures, it would be interesting to develop some statistics using the signal space itself to formulate performance bound of biometric system in terms of signals (users) themselves. A key role of information theory [5] has been in providing the achievable limit to the performance of a signal (alternately, transmitter) over noisy transmission channel which is given by the signal capacity (Shannon capacity). Conventionally, information capacity or Shannon capacity [5] is based on an average signal to noise ratio (*SNR*) which is a ratio of the second order statistics of the desired signal(s) and the noise signal(s). For our purpose, the biometric information that assists in reliable authentication can model the desired signal. Similarly, the biometric information that assists in loss of reliable authentication can model the noise. In this work, we propose that information capacity theorem [5], [20] can be applied to the biometric authentication channel given that matching scores (genuine and imposter) can be modeled in terms of signal plus noise. In doing so, information capacity of a biometric authentication channel can be given by *SNR* estimated from second order statistics of matching scores. This formulation for information capacity is referred here after as the constrained capacity of the biometric system as it is proposed for finite number of users. Like Shannon capacity, it is fixed for a database (alternately, channel state information) and matching function (alternately, receiver system), independent of the choice of threshold. The proposed constrained capacity of a user quantifies authentication performance of the user biometric in presence of noise, given a database and matching function. This noise mainly arises due to noisy acquisitions of a user's biometric template and also due to the variance in biometric templates of the imposters. We propose that constrained capacity of

the biometrics system will denote the number of users in the database that can be reliably authenticated, given a database and matching function. The information capacity of a signal transmission system is typically expressed in the units of bits/sec/hertz [13]. In this work, constrained capacity denotes the performance of a user or of biometric system expressed in terms of $(reliable\ users/user\ population)$. It is necessary to point that the information capacity does not provide a measure for the error rate of the channel nor an answer to how erroneous is the channel. Information capacity gives a measure for the information rate of the source that can be transmitted reliably on a noisy channel. Similarly, we can expect that *ROC* derived measures and the proposed constrained capacity present two different aspects on the performance of biometric authentication. Finally, we point that the constrained capacity proposed here can also be used to motivate information capacity of a binary hypothesis testing problem.

I (b) Contributions

The key contributions of this work can be summarized as follows: **(i)** in Sections III and IV, we formulate and quantify constrained capacity of a biometric system based on the information capacity theorem [5], [25], [27]. This framework is shown to be useful for estimating the effective number of cohorts for the performance improvement [19]; **(ii)** in Section V, we formulate constrained capacity for fixed fusion rules at score level [21], [22], **(iii)** in Section VI, we formulate false random correspondence of users based on their biometric features using a rate-distortion framework [8], [9], [10],[11], [20]; and **(iv)** in Section VII, we present experimental analysis to illustrate (i) and (ii) using a real database of 100 users for palmprint and hand geometry biometrics. We also give experimental analysis that illustrates use of constrained capacity to estimate the relevance of feature subsets [26] and to measure the individuality of biometric features.

II. Outline of Previous Work

II (a) Prior Work on Capacity of Biometric System

We now briefly discuss prior work in the literature pertaining to capacity of biometric system. Schmid *et al*. [6] outline a seminal approach in formulating performance indices like recognition capacity and reliability function of biometric system, refer Figure 1. Their work did not support numerical results to illustrate these performance indices or demonstrate other applications using these

performance indices. Some insights from [6] support the approach in [7] where authors have shown novel techniques to utilize the information in iris features for enhancing authentication.

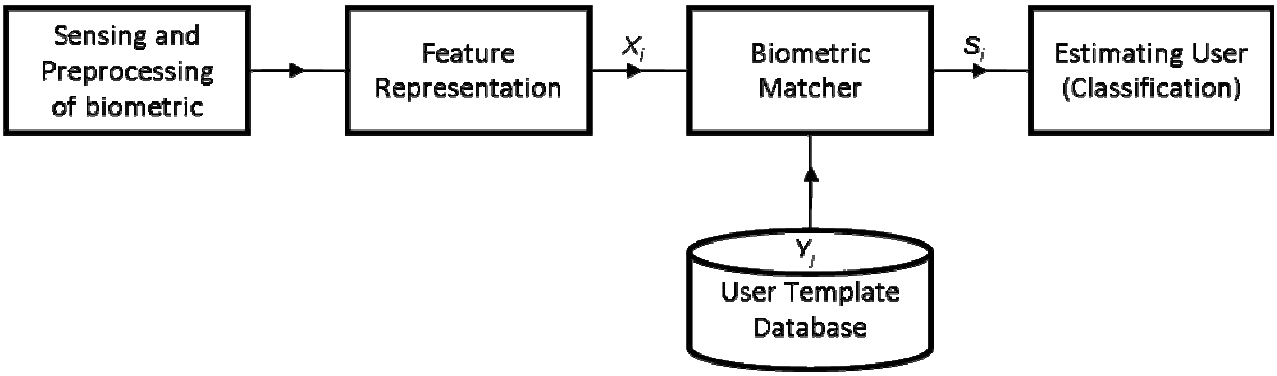

Fig. 1: Functional schematic for a biometric authentication system

Figure 1 can be used to summarize the approach employed by Schimd *et al.* [6] to compute the recognition capacity of a biometric system. They first formulate maximization of mutual information between a query template from any user *i* given by X_i and user templates Y_j in the database. Mutual information itself is formulated based on information density which is based on distribution functions of the biometric features. Finally, capacity is formulated as the ensemble average of the maximum mutual information for all users. It should be noted that this formulation for capacity is only one of the many approaches known in the information theory literature. For ensemble averages to represent true estimates of the capacity, the approach detailed in [6] finds limited applications for biometric systems employing very large user database. We point the following as regards the seminal work of authors in [6]: (i) it employs biometric features as the signal of interest, (ii) it formulates information density by averaging over an asymptotic user population to give mutual information and (iii) it uses Chernoff bound to formulate a reliability function of biometric system[‡]. The method presented in [6] has some additional concerns that can be summarized as follows: **(i)** the biometric template data usually comprises of feature vectors which have high dimensions. Computing distributions of highly dimensional signals such as features in order to formulate the information density is computationally complex [13]. Computing correlation matrices of features or matrix inversion employing such correlations is also computationally complex [13], [20]. The

[‡] Biometric authentication system is indeed some form of noisy recognition channel.

complexity is non-linear with the size of biometric features and user population [20], **(ii)** assumption of i.i.d (independent identically distributed) is essential to the approach shown in [6]. The following weakens the independence of features [3], [17]: (a) increase in the dimensionality of features, and (b) increase in number of users. Furthermore, the assumption of identical distribution of a feature for all users is not a realistic assumption, as this undermines the user-specific nature of a feature [18] and finally**, (iii)** in biometric authentication, a matching score (not the features in query template) is the signal used for authentication. Hence, it is unlikely to gain significant insights from a recognition capacity [6] based on biometric features for evaluating the authentication performance.

The approach formulating capacity in [6] is quite promising but subjective to the type of features and the technique used for feature representation. This approach may have limited practical utility due to a finite number of acquisitions and users available for validating its framework. This may be a possible reason as to why experimental results on the framework of [6] have not been reported, in our understanding of the literature. Furthermore, the approach developed in [6] when applied on an exhaustive set of users or templates will give performance of the biometric features. The performance expressed by capacity or reliability such as using [6] is independent of the choice of matcher which is a crucial block of the authentication system. Finally, we point out that a matching score is computed basically from the template features using a matching function. This means that capacity could be formulated for signals comprising of the biometric features as shown in [6] as also formulated for signals comprising of biometric matching scores as employed in this paper.

II (b) Prior work on Individuality of Biometrics

In this section we summarize prior work on individuality of biometrics from literature [8], [9], [11]. Individuality is the characteristic of a biometric that defines a property to discriminate the users' based on biometric features, for more details refer [4]. Authors in [8] have detailed a promising approach to ascertain the individuality of fingerprint which is represented by minutiae points. Such feature representation localizes the minutiae from its position and orientation on the fingerprint image. Authors [8] employ an empirical tolerance circle or overlap between minutiae to determine the minutiae correspondence between

fingerprints of different users in the population. The probability of false random correspondence (*FRC*) is obtained as Hyper geometric for minutiae position and binomial for minutiae orientation. The *FRC* for fingerprint is obtained as the product of these distributions. Zhu *et al.* [9] have incorporated mixture models with Gaussian distribution for minutiae position and Von-Mises distribution for minutiae orientation. They also employ an empirical tolerance for minutiae position and orientation to obtain an expression for *FRC*. The *FRC* obtained by them is Poisson distributed. Interestingly, individuality or discriminability offered by the biometrics [4] depends not only on the biometric features but also on the feature representation method. This latter property of individuality is not captured in these models, as also these models quantify *FRC* specifically for fingerprint.

In [11], Daugman suggested representing features of iris given by Iris Code. He also demonstrated that iris code was an optimal method for iris representation based on the distribution of a Hamming distortion (and degrees of freedom) computed between features of Iris Code. This distribution shows a binomial trend and it peaks at a normalized hamming distance of 0.5. Results on iris as highly individual biometric, outlined in the literature [11] strongly support that individuality (and *FRC*) is dependent on both the type of features and the method of its representation. Daugman's work [11] provides a fundamental basis for computing *FRC* that could be applicable to many biometric traits. Daugman shows an analysis only for Iris Code [11].

Other significant contributions that show individuality model for iris can also be seen in Bolle *et al.* [12]. In Section VI, we will show that the notion of distortion or tolerance can be connected to feature representation and finally to the individuality of the biometrics. We also point that individuality has been formulated based on feature level information.

III. Noisy Authentication Channel: Review of Shannon Capacity

In his seminal work Shannon showed communication model [5], [13] with a source which generates i.i.d (independent identically distributed) symbols S given by an entropy $h(S)$. In his basic model these symbols were transmitted on an additive white Gaussian noise channel (AWGN) given by N (Figure 2). Shannon formulated two powerful theorems, widely known as the source coding theorem and information

capacity or capacity theorem in the literature [5], [20]. Of these two, we first consider the information capacity theorem that in our framework is used to formulate a limit on information of user biometrics (some characteristic of the user biometric like matching score) from a noisy recognition channel for reliable authentication. The information in user biometric and its reliability measure has a maximum (sometimes referred as suprenum) as an achievable limit to authentication performance, given a database and matching function, which defines our framework for capacity of a biometric authentication system [5].

Consider the channel model with input to channel given by a Gaussian source $S \succ (0, \overline{S_G})$, where $\overline{S_G}$ denotes the average power per symbol in signal constellation [16]. Let *S* represent the statistics from desired signals or signals that aid reliable authentication. Similarly, additive white Gaussian noise (AWGN) representing the noisy channel is given as $N \succ (0, \sigma_N^2)$. Thus, *N* represents the noise statistics or the mechanisms that limit reliable authentication. Finally, capacity denoted by '*C*' can be expressed in terms of a signal to noise ratio (*SNR)* given by $\overline{S_G} / \sigma_N^2$ for the information model seen in Figure 2 [13]:

$$C = \frac{1}{2} \log_2 (1 + \frac{\overline{S_G}}{\sigma_N^2}) \quad (1)$$

The choice of Gaussian distribution is motivated by the following facts: (a) in the cumulative sense it tends to be a good reflection of most real world distributions [17], (b) it has the highest entropy for a given variance and is therefore useful in formulating bounds. This means that a Gaussian source under a power constraint results in the maximization of mutual information to give capacity. As also, Gaussian for noise offers a maximization of noise entropy which means transmitting on a very noisy channel [20], and (c) its closed form expression for entropy is analytically tractable [13].

Capacity, in Equation (1), is given in terms of bits/sec/hertz and defines a ceiling for the average information rate that can be reliably transmitted on the noisy channel. Capacity as given in Equation (1) gives an average performance measure [5] of the information rate bearing ability of the channel. It solely depends on the average statistics of the signal information and noise information to estimate the signal to noise ratio. In the crux, Shannon's classic work [27] edifies that error free transmission is possible by

operating/transmitting at information rates below the capacity (channel capacity). The operational usage of capacity as a performance measure in this work can be motivated by considering the capacity [5] under the following three settings: when capacity is less than one, equal to one and greater than one. In all these cases, transmitting information actually implies transmitting information in one use of the channel. When capacity is less than one then we can reliably transmit only fraction of the source rate on the channel. When capacity is equal to one then the information system is said to have achieved the channel capacity. It means that source information at this rate can be transmitted reliably on the channel. Finally, when capacity is greater than one then a source rate greater than the given information rate can be reliably transmitted on the channel. Once, *SNR* has been formulated from known statistics of the biometric then Equation (1) may be used to quantify capacity of the biometric system**.**

An important assumption underlying Shannon's formulation of capacity is the law of large numbers. Law of large numbers assumes long delays on the channel or a large number of transmitted symbols. One of the main motivations for this is that long observations ensure a stable channel or a stationary stochastic channel. However, when the channel is stationary for smaller delays then the requirement of large numbers can be relaxed. So for example, a BPSK (bipolar phase shift keying) transmission associates only two levels of phase information or symbols [14], [16] transmitted over AWGN [27] for which the capacity can be quantified by Equation (1). However, a large enough number of such 2-ary signals are assumed for transmission for which the noise model of the channel is stable [5]. At the same time, when channel is non-stationary such as when fading, shadowing, Doppler and so on, then an average *SNR* is employed to estimate the information capacity. Following Shannon's expository work alternate approaches were developed in the literature [13], [20] that derive results given in Shannon's theorems. These models formulate capacity using optimization and game theoretic approaches**.**

IV. Constrained Capacity of Biometric Authentication System: Proposed Framework

In this section we combine tools developed so far to propose a new model for formulating 'constrained' capacity of the biometric system. The term constrained refers to capacity formulated from matching scores for a fixed choice of matching function and database. Hereafter, capacity will mean the constrained capacity

unless specifically stated as otherwise. From signal theory perspective [14], [16], *M* users in the database could be represented by an *M*-ary signal probability space, with the signal space comprising of genuine matching scores $\{g_m\}$ and imposter matching scores $\{i_m\}$ for every user $m \in M$. This can be used to formulate the genuine and imposter distributions for user *m*. The matching scores (both genuine and imposter) so generated employ all templates of a user, without partitioning the database into test and training sets. The motivation for using all templates in the database is that it gives feasibility for formulating constrained capacity based on the complete state information of the noisy authentication channel. This is identical to estimating information capacity of a communication system with complete state information or complete knowledge of the noisy channel [15]. It may also be noted that authors in [6] have used all the biometric templates in the database for formulating the recognition capacity, without partitioning the biometric database into test and training sets. A challenging problem in developing constrained capacity based on matching scores is that different users will have different sample spaces of matching scores (genuine and imposter distributions). Therefore it is very difficult to define a unique probability space [17]. In the following we discuss an approach that resolves this issue and facilitate formulating *SNR* of the biometric system.

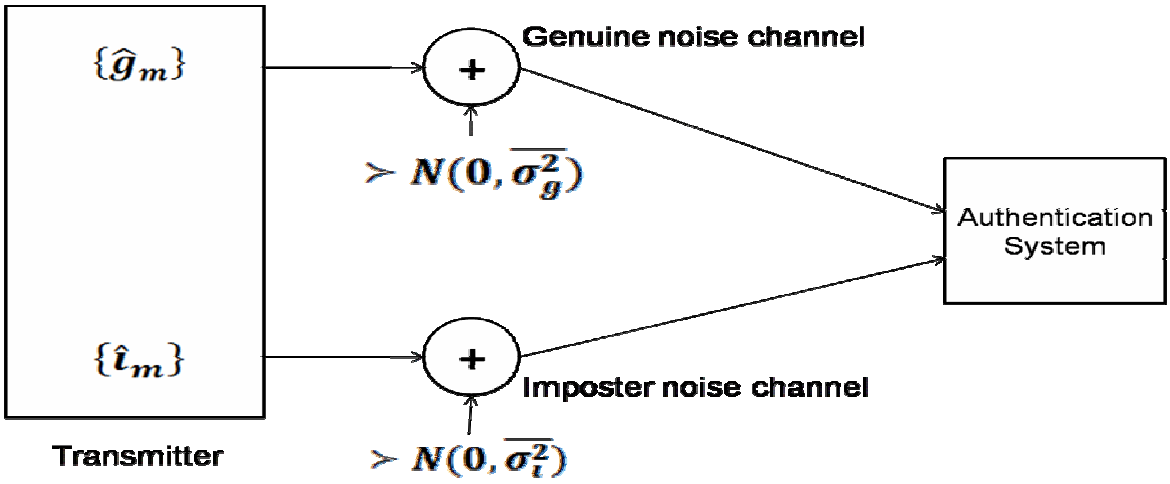

Fig. 2: A two channel model for biometric authentication system

Figure 2 depicts our approach for formulating capacity of biometric authentication [25]. Let $\{\hat{g}_m\}$ and $\{\hat{i}_m\}$ denote the median matching scores from genuine and imposter distributions respectively

for user $m \in M$. Let *A* be the desired source transmitting information sequence $\{\hat{g}_m\}$ and let *B* be the interfering source trying to corrupt these symbols by transmitting an information sequence $\{\hat{i}_m\}$. Thus, for each transmitted genuine signal from $\{\hat{g}_m\}$, *B* transmits an interference symbol from $\{\hat{i}_m\}$ with the same index. The next step is to model the noisy channel representing variability in genuine and imposter matching scores. In Section III we pointed out that Gaussian distribution can be justifiable from the central limit theorem [17] to characterize the noise/dispersion of scores due to cumulative effects. It also offers significant operational ease. In the same discussion some information theoretic properties of Gaussian were also cited that supports its wide applicability in many frameworks. Importantly, empirical observations based on the histograms of matching scores (genuine and imposter) for our database show a close agreement with Gaussian. So, we propose that information symbols $\{\hat{g}_m\}$ and $\{\hat{i}_m\}$ are subject to additive white Gaussian channels as these traverse the respective noise channels given by genuine and imposter noise, refer Figure 2. The noise channel for *A* is effectively a genuine Gaussian channel denoted by $n_{g(m)} \succ N(0, \sigma^2_{g(m)})$ and that for *B* is imposter Gaussian channel denoted by $n_{i(m)} \succ N(0, \sigma^2_{i(m)})$. The variances from genuine and imposter distributions are denoted by $\sigma^2_{g(m)}$ and $\sigma^2_{i(m)}$ respectively. Their averages (using sample mean) are given by $\overline{\sigma^2_g}$ and $\overline{\sigma^2_i}$ respectively. Equation (2-a) gives signal model that describes signal received from genuine channel for user *m*. Similarly, equation (2-b) gives signal model for signal received from imposter channel of user *m*.

$$R_{g(m)} = \hat{g}_m + n_{g(m)} \tag{2-a}$$

$$R_{i(m)} = \hat{i}_m + n_{i(m)} \tag{2-b}$$

The authentication block in Figure 2 receives noisy versions for all *M* users (hence *M*-ary signal space). Here each symbol is a signal pair $\left\{\hat{g}_m, \hat{i}_m\right\}$ from the genuine and imposter distributions of *m*. This concludes formulating an *M*-ary signal space model for *M*-ary hypotheses testing problem [14]. This

model is similar to Neyman-Pearson type used in jamming/intrusion systems [16], with target *A* and interferer *B*.

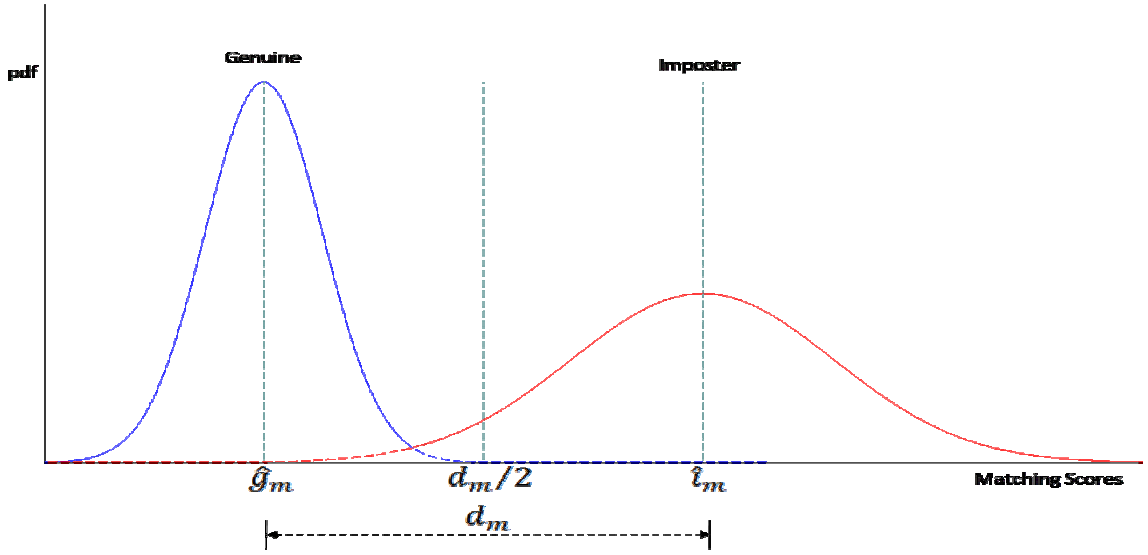

Fig. 3: Statistical distance measures for user *m*

Discernability, decidability index, *d*-prime are widely known in the biometrics literature [1],[3] and are similar to *SNR* as used here from information theory [5]. Figure 3 illustrates distance measure, between $\hat{g}_m$ and $\hat{i}_m$, denoted by d_m. As seen in Figure 4, the transmitter is a binary hypothesis transmitter that transmits a pair of signals for all users. Here, d_m is a random variable indexed by *m* [5], [14]; which is clearly dependent on values of $\hat{g}_m$ and $\hat{i}_m$. From this we define an average measure such as $\overline{d_m^2}$. The energy of signal constellation comprising of $\left\{\hat{g}_m, \hat{i}_m\right\}$ must be carefully modeled to satisfy the following requirements [16]: **(i)** signal pair with minimum energy allocation per signal to model worst case performance under noisy transmission, **(ii)** since binary hypothesis testing is essentially – two category classification, therefore we propose to choose antipodal signal set, with a correlation coefficient of -1. This mainly implies that a genuine hypothesis is negatively correlated with the imposter hypothesis for a given user / signal pair [5], and; **(iii)** the entries in $\left\{\hat{g}_m\right\}$ are user specific [18] and as such can be considered to be uncorrelated [17] for different users. A similar heuristic reasoning may be applied to the point- estimates

(median) comprising the imposter set $\left\{\hat{i}_m\right\}$. Figure 4 further simplifies the model by incorporating the worst case noise from genuine and imposter distributions [5], [7]. This aids in proposing an effective AWGN channel given by $n_e \succ N(0, \max(\overline{\sigma_g^2}, \overline{\sigma_i^2}))$. The final formulation for the linear additive channel model from this is given by equation (2-c), [7]:

$$R_m = \left\{\hat{g}_m, \hat{i}_m\right\} + n_e \tag{2-c}$$

For this the *SNR* can be given as:

$$\frac{\overline{S_G}}{\sigma_N^2} = \frac{\overline{d_m^2}}{4 \max(\overline{\sigma_g^2}, \overline{\sigma_i^2})} \quad . \tag{3}$$

Once the *SNR* has been characterized for the authentication channel, we can now apply the information capacity theorem [13], [15] to propose constrained capacity of the biometric authentication channel.

Constrained capacity of biometric authentication system: A real valued function of *SNR* (estimated from system database) which gives a suprenum on the average number of users that can be reliably authenticated. A symbol or signal pair in $\left\{\hat{g}_m, \hat{i}_m\right\}$ corresponds to a user; hence, the number of symbols transmitted in one use of the authentication channel is the same as the number of users. Therefore, the units of constrained capacity can be given as the number of users reliably authenticated per number of users in the database.

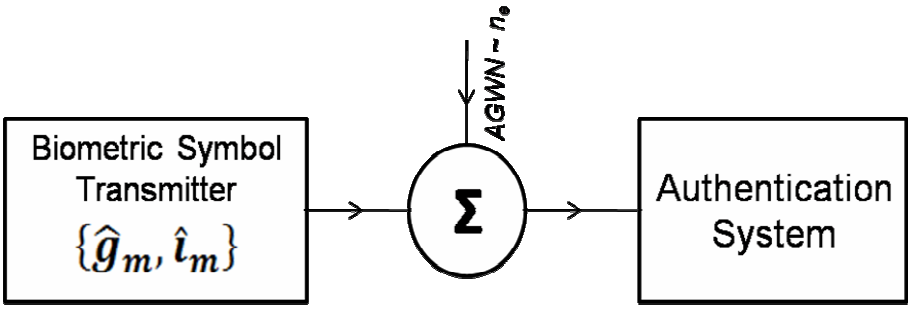

Fig. 4: Biometric authentication system depicted as Information model

The constrained capacity of the biometric authentication system is denoted as C_1 and is given as:

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

It may be an interesting point to discuss on the difference expected, between constrained capacity of biometric authentication system and the capacity of several other information systems. Observe that in Equation (4); values of the statistics $\overline{d_m^2}$ and $4\max\left(\overline{\sigma_g^2}, \overline{\sigma_i^2}\right)$ are intrinsic to the biometric trait and the database. For a given matching function, the numerator $\overline{d_m^2}$ is mainly dependent on the uniqueness attribute of the biometric trait. The denominator depends on the biometric trait as well as on the acquisition and processing of biometric features. So, the numerical range of *SNR* is mainly determined by the biometric trait followed by the database size and the quality of biometric templates. Therefore, we conjecture that constrained capacity of authentication system for a given biometric trait might not vary over many orders. In other information systems signal statistics can be controlled easily and dynamically by power variation, coding-modulation variation and other means to combat noise. This results in a large variation in capacity of other information systems owing to a flexible signal design. Finally, we observe that a variation of Equation (4) can be used to formulate the constrained capacity per user denoted as C_m. This gives the authentication performance for a user and can be given as:

$$C_m = \frac{1}{2}\log_2[1 + \frac{d_m^2}{4\max\left(\sigma_{i(m)}^2, \sigma_{g(m)}^2\right)}] \quad (5)$$

Constrained capacity can be useful to give the change in performance due to new information (signal or noise). Given that the signal information is uncorrelated to existing signals, second order statistics of this information can be added in the numerator of the *SNR* indicating additional power gains from desired signal. Similarly, additional noise statistics can be incorporated in the denominator. One such example is that of capacity enhancement that can result from incorporating user-quality indices at score level [18]. If we apply the fact that $\hat{g}_m$ is user-specific, as in it is different for different users, then the variability in $\left\{\hat{g}_m\right\}$ can be used to improve the authentication. The set $\left\{\hat{g}_m\right\}$ can be used to extract the

user-specific signal information which comprises of the difference between $\hat{g}_m$ for different users. The absolute value of the difference between the median genuine score of a user $l \in M$ with all other users' median genuine scores can give the user-specific signal set for this user. Performing this difference for every l gives the user-specific signal set for the database that can be given as $\left\{ \left| \hat{g}_n - \hat{g}_l \right| \right\}$ such that $l \neq n;$ and $l,n \in M$. It was earlier mentioned in this section that $\overline{d_m^2}$ is directly dependent on $\left\{ \hat{g}_m \right\}$. Now, if we can show that the user-specific signal set is uncorrelated with $\left\{ \hat{g}_m \right\}$ then it follows that the new signal statistics can be algebraically added in the numerator. We refer appendix B for a result which shows that the formulated user-specific signal statistics is uncorrelated with $\overline{d_m^2}$. The new constrained capacity C_2 can be formulated by incorporating the statistics of user-specific signal set denoted as $\hat{\sigma}_g^2$ in the numerator of *SNR*. Thus, the constrained capacity C_2 of the biometric system from incorporating user-specific quality index is given as:

$$C_2 = \frac{1}{2}\log_2[1 + \frac{\hat{\sigma}_g^2 + \overline{d_m^2}}{4\max\left(\overline{\sigma_g^2}, \overline{\sigma_i^2}\right)}] \tag{6}$$

We next investigate additional performance gains, in terms of capacity, achieved by incorporating the non-matching scores that are in the vicinity of a user's matching score given by $\hat{g}_m$. In the inception of cohorts, Authors in [19] have shown that an extended score template comprising of the 'neighborhood' non-matching scores (inter-user) along with a genuine matching score gives substantial improvement in the recognition accuracy. However, not all the non-matching scores (imposters) are useful for constructing the cohort set of a user. This means that the imposter users with *close* matching scores that lie in neighborhood of a genuine user, the *closeness* determined by the respective matching scores, can guarantee an increased confidence level. It can be expected that non-matching scores towards the tail of the distributions may not contribute significantly to the confidence level. Cohorts define the imposter users that constitute this non-matching neighborhood space of a genuine user, and as such may be regarded as 'signal' information [13],

[19]. We point that a biometric system is similar to multi-user detection system where a user's detection can be enhanced by incorporating the information of multi-user interference, refer Verdù [13]. The 'dominant interfering users' in a multi-user information system [13] are similar to the 'cohort users' that constitute an appended decision space in [19] and can be used to significantly improve in the authentication accuracy. We now pose two different problems to be tackled in our framework of constrained capacity: firstly, to formulate the constrained capacity of the biometric system that employs cohorts, and secondly, to determine the dominant interferers or close non-matching users to a given user.

We now outline the framework for formulating the constrained capacity with cohorts. It is proposed to show one possible selection of cohorts using 3σ bound which is a loose probability bound in Chebyschev sense [17]. This bound mainly considers the non-matching scores lying within 99.7% probability coverage about the median genuine score (but not including it) on the genuine distribution. The rule of 3σ will select $\pm\sigma, \pm 2\sigma, \pm 3\sigma$ points on the genuine distribution about the peak matching (genuine) score for a user to estimate cohort set of the user. Let the average variance obtained from the variances of the cohort set for *M*-users be denoted as $\overline{\sigma_c^2}$. We incorporate this as signal statistics to formulate 'C_3' which denotes constrained capacity due to cohorts, given as:

$$C_3 = \frac{1}{2}\log_2[1 + \frac{\overline{\sigma_c^2} + \overline{d_m^2}}{4\max\left(\sigma_g^2, \sigma_i^2\right)}] \tag{7}$$

So far, we have discussed formulation of constrained capacity for various settings. Next, constrained capacity is proposed for determining the dominant interferers or cohort users. Let *C* denote a generic constrained capacity of the biometric system. For *M* users in the database, based on the framework of constrained capacity the number of users that can be reliably identified (χ) can be given as:

$$\chi = CM \tag{8}$$

Let us consider the case when our authentication channel is unreliable, *i.e.* when the constrained capacity C is less than one. For this, let $\overline{C}$ be such that $\overline{C} = 1 - C$, $0 \le C < 1$, and $\overline{C} = 0$, $C \ge 1$. We define $\overline{C}$ as the maximum average unreliability of the biometric authentication system in terms of the users

in the database. Thus, $\overline{C}$ quantifies unreliability of the authentication channel in units of (unreliable users/ users in database). The number of unreliable users can be viewed as the number of dominant interferers or cohorts (*N*) and can be given as:

$$N = \overline{C}M \tag{9}$$

For a database of *M* users, greater the value of *C*, smaller is the effective number of cohort users estimated by Equation (9). This is particularly expected for highly individual or strong biometrics. For such biometrics, C is expected to get closer to 1 (also, validated by experimental results in Section VII). It may be noted that the actual number of effective cohorts vary from user to user.

V. Constrained Capacity of Multi modal Biometric Systems for Score level Fusion

In this section, we investigate use of constrained capacity to evaluate performance of fusion rules at score level. In biometrics literature several architectures have been shown that give fusion methodologies to justify superior performance achieved by multimodal biometric systems [4], [21], [22], [23], [24]. In fusion-algorithms authentication is performed after consolidating features or matching scores or decisions; thus resulting in three different levels of fusion. Multimodal biometrics systems achieve superior performance by exploiting information diversity which increases the distance measures (individuality) between users [21]. Score level fusion can be performed by consolidating matching scores from multiple representations of a biometrics or by consolidating outputs from multiple matchers [21], [22].

We now develop a framework to evaluate score level fusion methods. Score level fusion: Let $\{s_\alpha^p, s_\beta^p\}$ denote the matching scores generated from the modality α and modality β respectively, for p^{th} user, such that p = 1, 2,..., M. Let Ψ be the function which combines component scores from all different modalities or matchers to generate a fusion score such that the combined score S_C^p for *user p* is given by $S_C^p = \Psi\{s_\alpha^p, s_\beta^p\}$. The function Ψ denotes a fusion operation which could be weighted sum, difference, log product, max rule, and min rule on component scores [21]. Equation (4) can be used once we can formulate *SNR* following the fusion. The statistics using S_C^p to formulate *SNR* includes: distance separability ($\overline{d_\Psi^2}$),

average variance genuine ($\overline{\sigma^2_{g-\Psi}}$), and average variance imposter ($\overline{\sigma^2_{i-\Psi}}$). The resulting capacity using this statistics can be given as:

$$C_{\Psi} = \frac{1}{2}\log_2[1 + \frac{\overline{d^2_{\Psi}}}{4\max(\overline{\sigma^2_{g-\Psi}}, \overline{\sigma^2_{i-\Psi}})}] \quad (10)$$

Effectiveness of fusion rule can now be measured by computing capacity using Equation (10) and comparing this to capacity obtained from component modalities or matchers using Equation (4). Experimental results on score level fusion using hand geometry and palmprint biometrics are detailed in Section VII.

VI. False Random Correspondence Probability: Performance of the Biometric trait

Rate-distortion subsumes a major contribution in information theory for problems on source coding [20]. This was shown by Shannon in his treatise [5], [13] and is known as coding theorem for source. Source coding or feature representation often aids in finding efficient methods to represent the biometric source given its statistics. In this process, some attributes that constitute inherent properties of the biometric source can be formulated and quantified. Thus study of statistics at feature level leads to performance index of the biometric trait itself [16]. An attribute such as uniqueness and distinctiveness is important for evaluating the scalability of a biometrics system, given a method of authentication [1]. Individuality measures this uniqueness and distinctiveness offered by the biometric features. The type of features and the feature representation both influence the discriminability offered by a biometric trait. The source coding theorem [20] can be invoked to estimate the optimal number of bits required to 'uniquely' represent a source. The minimum rate condition at a distortion defines an inherent characteristic of an information source, as shown by information models in [20]. Similarly, the rate-distortion framework can be proposed for formulating a measure of the inherent information in biometrics or biometric individuality.

In biometrics literature, individuality is related to and expressed in terms of the probability of false random correspondence (*FRC*) [8], [9], [11], [24], [25]. The author's in [8] model *FRC* for the fingerprint biometric by employing an empirical tolerance which actually is tolerable distortion measure, as

we introduce later in this section. Similarly, Zhu *et al.* [9] have obtained *FRC* of fingerprint biometric using mixture models that generate minutiae features for an empirical estimate of tolerance or distortion. Section II (b) presents a brief review of prior work on *FRC*.

We define Individuality: As the minimum rate achievable for an information source comprising of biometric features from different users (one template per user), subject to a distortion given by δ. False Random Correspondence (*FRC*): *FRC* can be defined as the average probability that users based on their biometric trait are similar, under a given similarity criteria.

In documenting iris as a highly individual biometric, Daugman [11] showed variability measures in iris features represented by two-dimensional Gabor features. The distribution of normalized hamming distance or degrees of freedom of features, the features given by phase information of iris biometric, was shown to fit a binomial distribution. A binomial distribution of distortion measure given by the degrees of freedom with peak at 0.5 depicts strong independence among features of iris biometric. This also signifies a maximum entropy code for the binomial distribution which can be realized as sum of independent Bernoulli features (given by 1 or 0) [2]. However, after [11] there has been little work on random correspondence on a rate distortion framework. A plausible explanation for this could be that Daugman's approach used an exhaustive database. Thus work in [11] illustrated an empirical model for uniqueness of iris which represents an accurate estimate for very large population. By encoding phase variations using hamming distance, [11] applies a hamming distortion metric to locally source code partitions of the template along trajectory in the iris. This condition also provides a 'maximally spaced' code book since entropy of this distribution (which is binomial distributed) peaks at 0.5, a clairvoyant choice in favor of the Gabor representation. It is therefore reasonable to infer that uniqueness and individuality are dependent on the choice of feature representation. Applying rate-distortion concepts [20] it is proposed that approach employed in [8], [9] as well as [11] are two related manifestations of a more general approach. The rate-distortion frame work (refer appendix) will be revived to propose that: (i) individuality is a property of biometrics that depends on feature representation, and (ii) individuality can be formulated in terms of *FRC* for a distortion constraint δ. Condition of a minimum information rate for an average distortion constraint is

equivalent to condition of average information rate for minimum distortion, as shown in [20]. For a choice of feature representation (PCA, LDA, ICA, Gabor phase information, etc.) and for a given distortion constraint δ, smaller the rate-distortion more efficient is the representation. In that sense, rate-distortion is a direct measure of individuality. We propose to use minimum distortion condition on a source code book comprising of biometric features for formulating *FRC* of generic biometrics. This approach is rationally similar to that discussed in [20] and is used to develop the corollary here.

Corollary: Let the complete set of biometric features given by *F* comprise of *K* different features for *M* users, the feature representation (code book of features) is given by $\hat{F}$ for a distortion $\overline{D}$ in mean square sense (m.s.s). If the distances between corresponding features in $\hat{F}$, for different users, are denoted by $\{d_k\}$, $k \leq K$; and if $\{d_k\}$ has a binomial distribution with probability $= 0.5$, then a minimum rate condition is achievable.

Note: If $\{d_k\}$ are hamming distances then, $\overline{D} = \sum_{k \leq K} d_k$

Proof:

We use result given in appendix A**.** For expression of rate-distortion refer [20]. Expressing information rate (mutual information) for $f_k \in F$ likewise $\hat{f}_k \in \hat{F}$ in terms of entropies given by variable *h*.

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

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

We use the following inequality for biometric source coding model based on minimizing the rate-distortion constraint as given in [20]:

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

Where $h(d_k)$ denotes the entropy of hamming distance distribution with hamming distance computed among corresponding binarized feature vectors for all users. The variability/uncertainty in d_k denotes

average distortion. Constraint $h(d_k) \geq h(f_k - \hat{f}_k)$ is important to tighten the inequality further in Equation (12). Thus, *k* independent observations (appendix)

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

Given $d_k \in (0,1,....,N)$ is a discrete random variable, $\{d_k\}$ gives independent realizations of *K* random variables for *K* different features. We define $\overline{D} = \sum_{\forall k \in K} d_k$ that consolidates realizations from *K* independent partitions or independent random variables. Equation (14) shows addition of entropies from all partitions or feature sets to give the rate for $\hat{F}$. Hamming distortion $\overline{D}$ comes from independent features and can be characterized by binomial distribution parameterized by $\succ (p, N)$ [10]. Equation (14) can be minimized if the entropy $h(\overline{D}) = \sum_{k \in K} h(d_k)$ is maximized. Entropy for binomial $\succ (p, N)$ maximizes for $p = 0.5$, proving our claim.

We point that the minimum rate condition using Equation (14) can also be viewed as a test for a feature representation method that optimally extracts uniqueness from biometric source, given a distortion constraint. Conversely, feature representation for which the above condition is achieved gives optimal rate-distortion. As a direct illustration, we can cite the Iris Code and its optimality in capturing uniqueness of iris biometric using a Gabor phase representation [11].

The corollary with propositions presented in this section can easily lead to formulating the false random correspondence probability of biometrics. If, we approximate $h(F) \cong h(\overline{D})$. Define, $\delta > 0$; where δ denotes tolerable distortion which can be the Hamming distance for binarized features. An empirical tolerance threshold (as also used in [8], [9], [11]) given by δ will correspond to a conditional entropy given by $h(\overline{D} / \overline{D} \leq \delta)$. An exponent of conditional entropy to base 2 (assuming binarization of features) gives the average total number of sequences that lie within a Hamming radius of δ around the given features. This leads to formulating the expression for *FRC* as follows:

$$P\ (\text{false correspondence}) = \frac{2^{h(\overline{D} \,/\, \overline{D} \le \delta)}}{2^{h(\overline{D})}} \tag{15}$$

Clearly, from Equation (15), an increase in δ reflects greater distortion, thereby increasing the numerator with conditional entropy in the exponent. This will result in a higher false correspondence. The basic form of entropy function will depend on the choice of feature representation and the biometric trait. An interesting observation from this section is in noting that the Equation (15) gives the zero noise (genuine and imposter noise) reliability of the biometric system. Thus, Equation (15) defines a minimum (best) error rate of a noiseless biometric system, for a user population *M*. It may also be noted that Equation (15) can give an accurate estimation of *FRC* if distance distributions were computed for large user population with $M \to \infty$ (asymptotically), as also indicated in [11].

It may be noted that the prior approaches [8], [9] have regarded intra-class variance of biometric features as a limiting factor for individuality of the biometric. We may not have directly employed this in our formulation. Our formulation employs one biometric template per user each time, for *M* users, and then estimates entropies based on the template features. We estimate *FRC* using Equation (15), each time for a different set of templates for the *M* users while still keeping the one template per user condition. We average the *FRC* computed in this manner from different combinations of templates for *M* users. This finally gives us an estimate of *FRC* which characterizes performance of a given biometric trait.

VII. Experiments and Discussion

(A) Unimodal Capacity

In order to estimate real values of biometric capacity formulated in section 4, we perform experiments on real biometric samples. The hand images from 100 users (10 images from each) were acquired from the digital camera using unconstrained peg-free setup in indoor environment. The extraction of palmprint region and hand-shape images from each of the acquired images is similar as detailed in [26]. The Discrete Cosine Transform (DCT) is used for the characterization of unique palmprint texture. Each of the 300 × 300 pixels palmprint image is divided into 24 × 24 pixels overlapping blocks. The extent of this overlapping has

been empirically selected as 6 pixels. Thus we obtain 144 separate blocks from each palmprint image. The standard deviation of DCT coefficients, obtained from each of the overlapping blocks, is used to characterize the region. Thus we obtain a feature vector of 144 values for the palmprint biometric. The simultaneously extracted hand-shape image is used to extract 23 hand geometry features. The details on the feature extraction methodology are available in [26]. The class genuine and imposter score distributions, using Euclidean distance, were generated to extract the following parameters: $\overline{\sigma_i^2}$, $\overline{d_m^2}$, $\overline{\sigma_g^2}$, $\hat{\sigma}_g^2$ and $\overline{\sigma_c^2}$. Table 1 summarizes the system parameters useful for estimating constrained capacity. The texture features extracted from the palmprint and the hand geometry features are expected to be highly uncorrelated.

Table 1: Parameters from the experiments

Biometric Modality	$\overline{d_m^2}$	$\overline{\sigma_i^2}$	$\overline{\sigma_g^2}$	$\hat{\sigma}_g^2$	$\overline{\sigma_c^2}$
Palmprint	148530	63585	15545	11346	72542
Hand Geometry	7.71×10^9	5.37×10^9	4.00×10^8	9.70×10^7	1.87×10^9

Table 2: Constrained capacity and number of cohorts

Biometric Modality	C_1	C_2	C_3	N_1	N_2
Palmprint	0.33	0.35	0.47	67	65
Hand Geometry	0.2214	0.223	0.268	79	78

Table 2 shows the constrained capacity given by Equations (4), (5) and (6). From Table 2, we observe that C_1 for palmprint is much higher than the C_1 for hand geometry. Similar trend can also be observed for C_2 and C_3 of palmprint biometric which are also higher than C_2 and C_3 of hand geometry biometric. From this we can conclude that palmprint biometric gives a superior authentication performance than hand geometry biometric.

Table 2 also summarizes the number of cohorts based on Equation (9). In this table, N_1 and N_2 represent the cohorts computed using constrained capacities C_1 and C_2 respectively. Since, C_2 is greater than C_1 for a given biometric trait, this justifies the improvement in authentication performance as a result of incorporating user-specific statistics. Hence, the cohorts representation from C_2 (given by N_2) is smaller as compared to the cohorts representation from C_1 as given by N_1. Next, in this table, we observe that the increase in the constrained capacity C_3 from incorporating cohorts is about 50 % more than C_1 which denotes the constrained capacity without cohorts. Authors in [19] have also suggested significant increase in classification accuracy by using cohort information, thus verifying the proposed framework. .

(B) Constrained Capacity with Ranked Features

The objective of our experiments in this section is to ascertain change in constrained capacity with the addition of relevant features. In this experiment, 144 features from each of the real palmprint images were firstly ranked in the order of their merit using correlation-based feature selection algorithm as illustrated in [26]. Each of these 144 ranked features were partitioned into 12 features per partition, and used to generate statistics to compute the constrained capacity by augmenting more and more ranked features. The increase in constrained capacity is gradual but slow in the first few partitions of adding features. The gradient becomes zero after about 72 features (Figure 5) and is influenced by: (i) signal to noise ratio of the chosen feature partition, and (ii) loss of relevancy from the redundant features. It may be noted that Authors in [26] have shown 69 irrelevant features from the same dataset, thus verifying the proposed framework. This experiment suggests that irrelevant (redundant) features do not contribute in increasing the constrained capacity or performance of biometric authentication. From Figure 5, we also point that the constrained capacity finally stabilizes at 0.33 after 69 relevant features, which is in agreement with value obtained in Table 2.

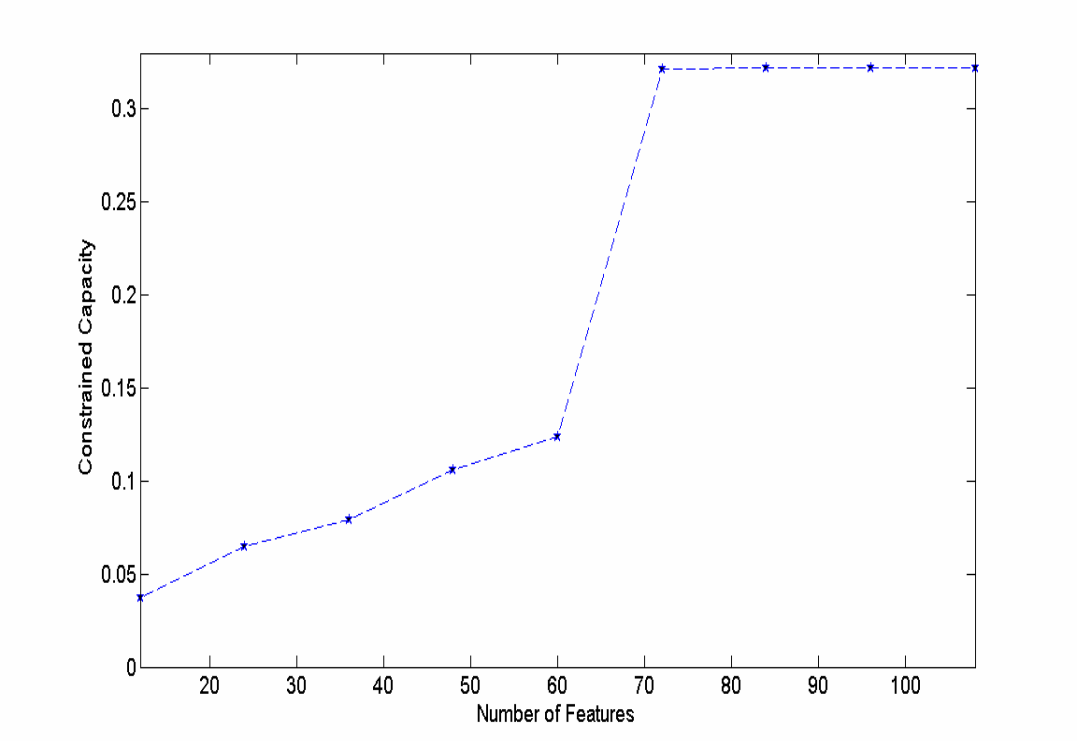

Fig. 5: Variation of constrained capacity with number of features for palmprint

The efficacy of biometric features in providing discriminatory information for authentication in literature [4], [8], [9] has been given in terms of the *FRC*; discussed with respect to fingerprint biometric. Authors in [8] use an empirical model to estimate overlap of minutiae based features and employ an empirical tolerance to estimate probability of random correspondence of fingerprint biometric. Authors in [9], give a more accurate model that captures the second order statistics for variability of minutiae based features to train a mixture model used in estimating probability of random correspondence.

We now show that constrained capacity can also give a measure for the discriminability or individuality [4] of biometric features. Hand geometry features used in this experiment are basically hand shape features extracted from the same hand image used to extract the palmprint biometric [26]. Individuality of biometric features can be measured by first selecting a set of features from biometric templates in the database and computing the constrained capacity for this selection. Figure 6 depicts plot

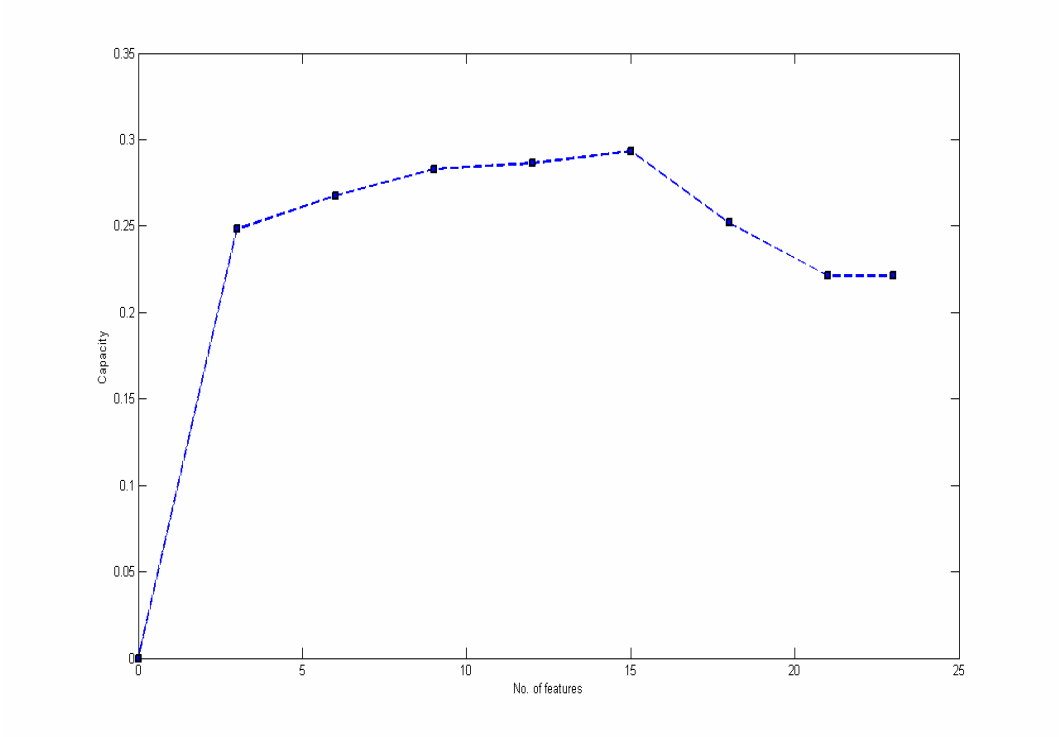

Fig. 6: Variation of constrained capacity with number of features for hand geometry

of constrained capacity as a function of hand geometry features, where we append 3 new features and then compute the corresponding constrained capacity. The slope of plot, after fixing two points on the x-axis, gives a measure of the differential information available from the additional features. For instance between two points given by 3 features and 9 features on the x-axis, constrained capacity is seen to change (increase) by 0.04. This means that the 6 additional hand geometry features correspond to a differential information measure of 0.04 (user per population). This works out to give a performance gain of 4 users for 100 users in our case and can be interpreted as follows. For a database of 100 users and using Equation (9), we may conclude that adding 6 hand geometry features shall maintain the authentication performance as before, even after adding 4 users in the database. Thus, the differential information of 0.04 is equivalent to and can be expressed by an alternate gain given in terms of 4 additional users for the same authentication performance. Let us now consider the change in capacity at two points given by 9 features and 20 features as can be seen in Figure 6, the change in capacity is (- 0.061). This indicates a loss of individuality quantified by (- 0.061) as a result of adding the 11 features.

Constrained capacity depends on *SNR* as seen in Equation (4); therefore, we can conclude that a change in capacity basically arises due to change in *SNR*. It may be further noted that *SNR* is ratio of $\overline{d_m^2}$ and the noise variance given by $\max\left(\overline{\sigma_g^2}, \overline{\sigma_i^2}\right)$. Hence, for a given biometric database, when the effective noise power or variance increases more than signal gain in $\overline{d_m^2}$ then the capacity dips. From Figure 6 it can be seen that authentication performance is optimum for 15 hand geometry features. We also see that the constrained capacity from hand geometry features is constant at 0.224 at 23 features; which is in agreement with Table 2. A similar analysis can be done from experiments that discuss relevance of features for palmprint biometric, Figure 5. From Figure 5 and Figure 6, it may be concluded that given a database, constrained capacity conveys a measure of individuality or discriminability of biometric features. It may be noted that Authors in [26] have also shown 15 relevant features from the same dataset, thus verifying the proposed framework. We can therefore conclude that constrained capacity as seen in Figures 5, 6 can be used to: (i) quantify the information in feature subsets, and (ii) sort the biometric features in their order of relevance.

(C) Constrained Capacity for Score level Fusion (Palmprint and Hand geometry)

Table 3: Constrained capacity and score distribution parameters from fixed fusion rules

System Parameters	Sum Rule	Product Rule	Log - Product Rule	Max Rule
C_Ψ	**0.435**	**0.599**	**0.732**	**0.382**
$\overline{d_m^2}$	0.053	0.003	6.831	0.037
$\overline{\sigma_i^2}$	0.003	0.0007	0.742	0.002
$\overline{\sigma_g^2}$	0.016	2.66×10^{-5}	0.976	0.013

Table 3 summarizes experimental values of constrained capacity using different score level fusion rules. Fixed combination rules such as sum, product, weighted sum, logarithm of product were investigated to observe the variation in the constrained capacity from the score level fusion of scores from hand geometry

and palmprint biometric. The constrained capacity is computed from the distribution of genuine and imposter scores as formulated in Section V. The experimental results illustrated in Table 3 suggest that the log-product rule gives best performance among the investigated fusion rules. Essentially, log-product function will nonlinearly scale scores from each modality, so that smaller values of score are boosted in each set to enhance the distance measures $\overline{d_m^2}$, this seems a plausible reason for its superior performance. Sum rule is known to perform better in score level fusion while consolidating matching scores from the relatively correlated feature set [23]. Product rule gives a superior performance over weighted sum rule, in our case when combining matching scores from the independent feature sets. Weighted sum rule was employed with weights that are varied continuously in the interval [0, 1]. Constrained capacity is used in this experiment to determine optimal weights. As seen in Figure 7, the maximum constrained capacity of 0.46 occurs for the weights 0.6 (palmprint) and 0.4 (hand geometry). In order to compare these results we examine an alternate method to determine the optimal choice of weights. Firstly, we partition biometric templates of each user into 8 training templates and 2 test templates. The matching scores from both modalities are normalized so in the range [0, 1]. Fusion of scores from the two modalities is then performed using weights that vary in [0, 1], in steps of 0.1. For a given combination of weights, all matching scores after fusion are used to generate the genuine and imposter matching scores. Finally, to determine an equal error point, a threshold is observed for which the number of genuine scores (false rejects) lying above this threshold is equal to the number of imposter scores (false accepts) lying below the threshold. This condition gives equal error count as a unique value, for a given combination of weights, using all matching scores for which the number of false accepts are same as the number of false rejects. It can be observed from Table 4 that the smallest equal error count occurs for weights 0.6 on palmprint and 0.4 on hand geometry. This concurs with our results using constrained capacity which attains maximum at 0.46 for these weights, refer Figure 7.

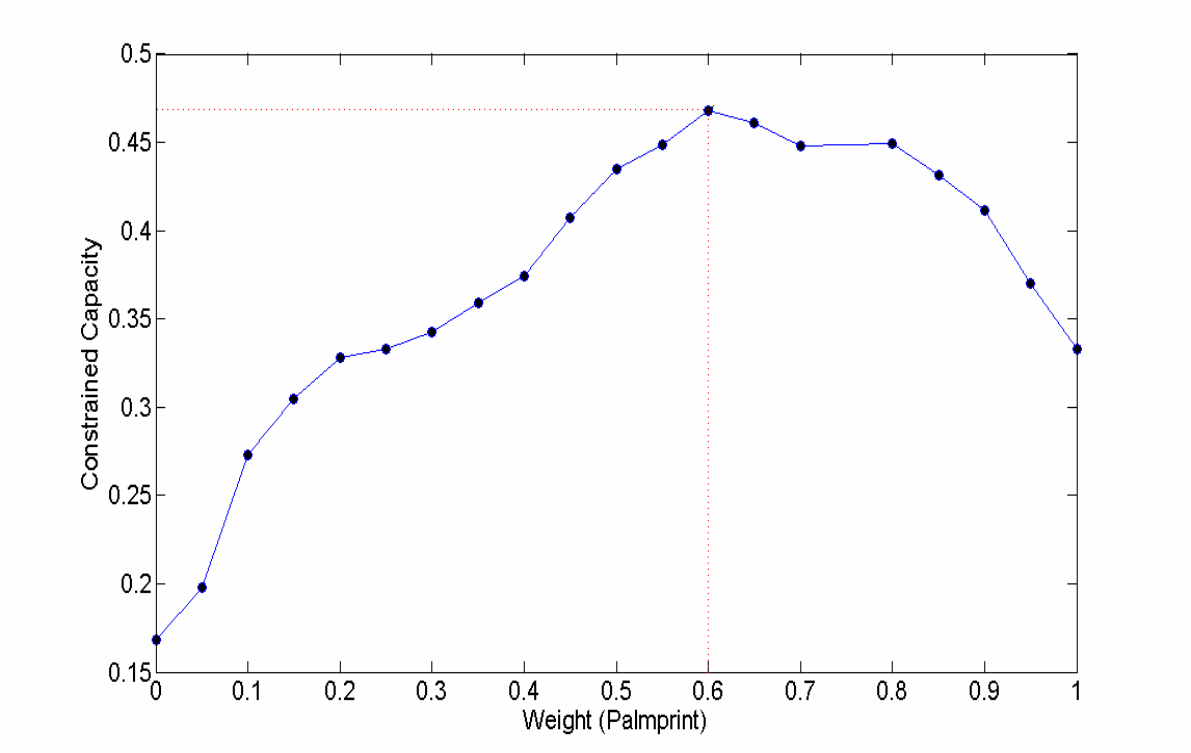

Fig. 7: Evaluation of Constrained capacity with palmprint weights for weighted sum rule

Table 4: Variation of number of false rejects and number of false accepts for weighted sum rule

Weight (Palmprint)	Weight (Hand Geometry)	Threshold (matching score)	Number of false rejects	Number of false accept
0.4	0.6	0.05905	2120	2120
0.5	0.5	0.0691	1937	1932
0.6	**0.4**	**0.07834**	**1744**	**1742**
0.7	0.3	0.08672	1825	1826
0.8	0.2	0.094234	1753	1752
0.9	0.1	0.10017	1816	1816
1	0	0.10123	2058	2058

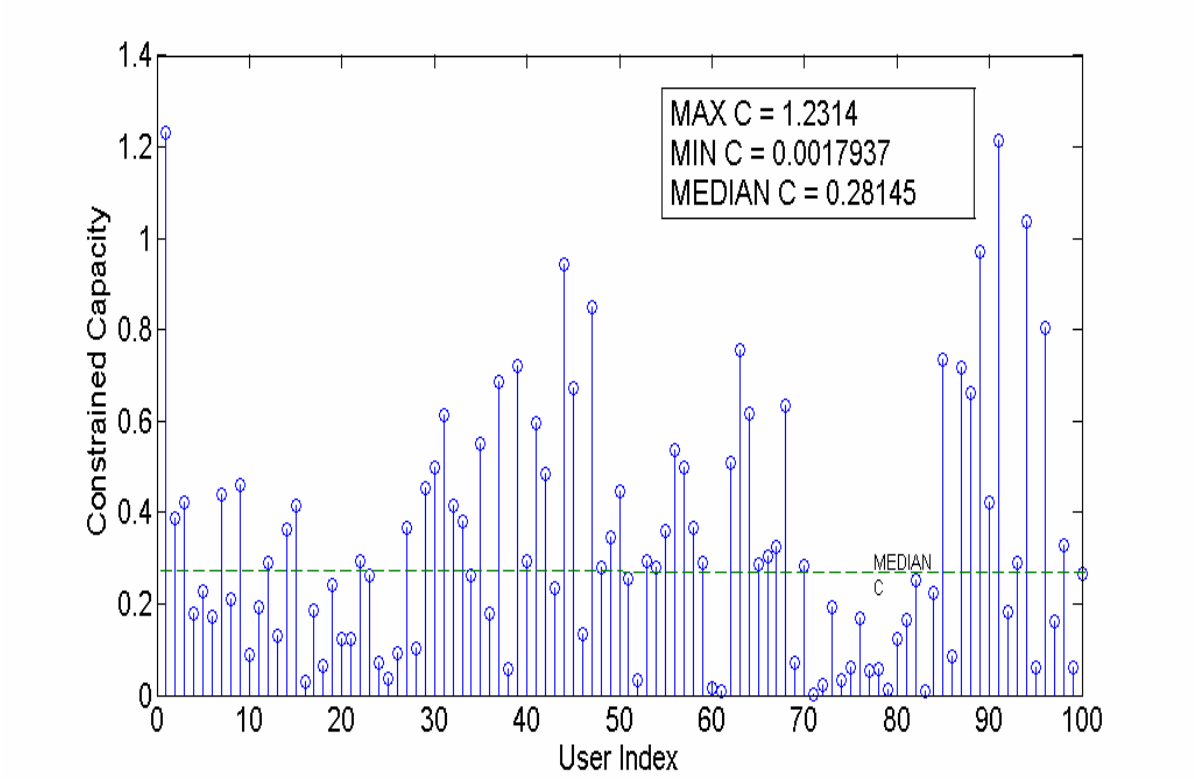

Fig. 8: Constrained capacity per user of palmprint biometric

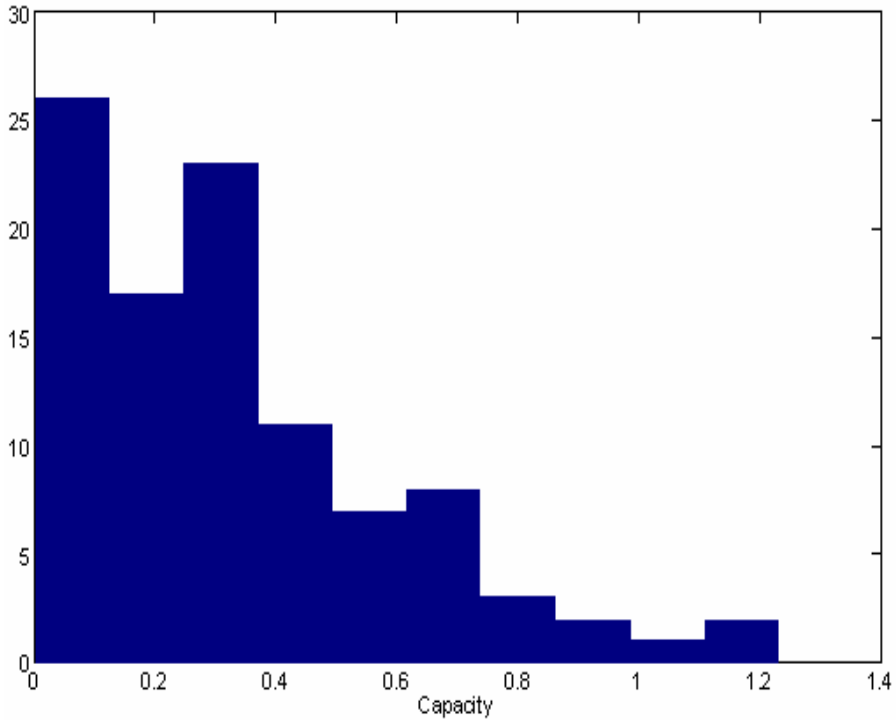

Fig. 9: Histogram plot for constrained capacity of palmprint biometric

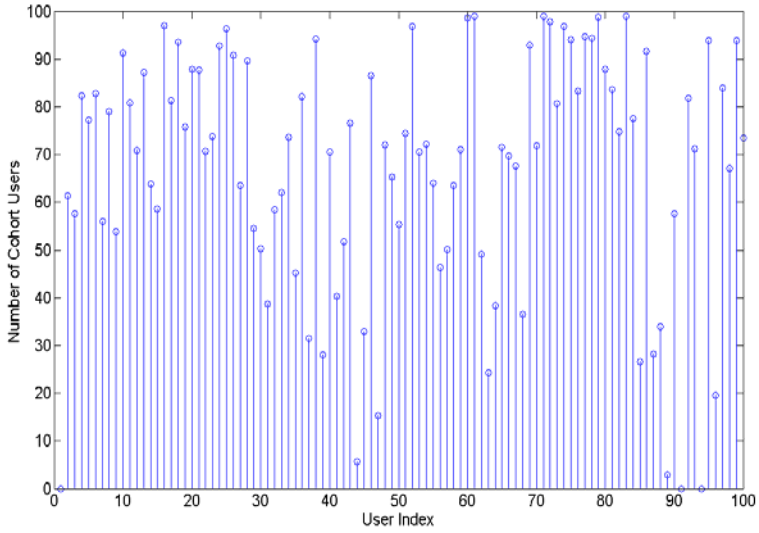

Fig. 10: Number of cohort users per user for palmprint biometric

Constrained capacity can be computed for individual users using Equation (5) and is illustrated in Figure 8 and 11. It is utilized for (a) estimating effective number of cohort users per user as seen in Figures 10 and 13, (b) stability of the database in terms of the variability in constrained capacity for users. Notable are the distributions (histograms) of constrained capacity over user population depicted in Figures 9 and 12. In this, a biometrics with a higher individuality can be expected to have location parameters of its distribution (median values) close to one. This can be observed by comparing Figures 9 and 12 which suggest that hand geometry is a weak biometrics, while palmprint can be considered as a stronger biometrics.

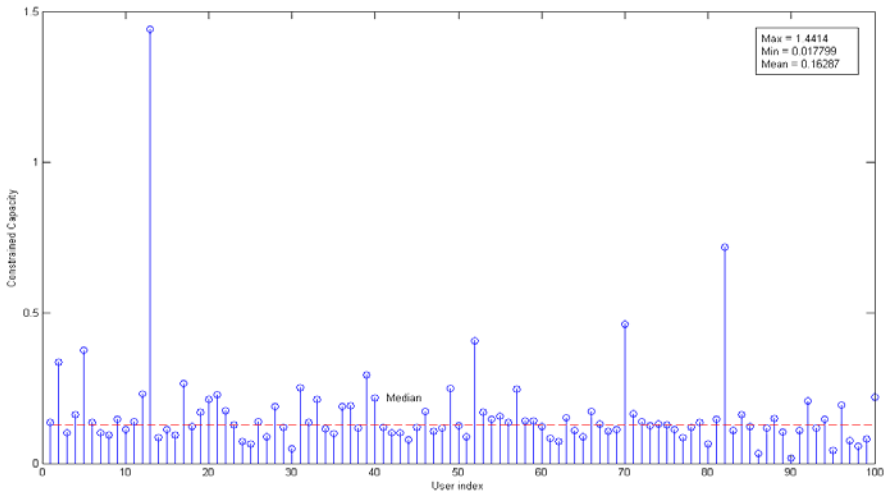

Fig. 11: Constrained capacity per user for hand geometry biometric

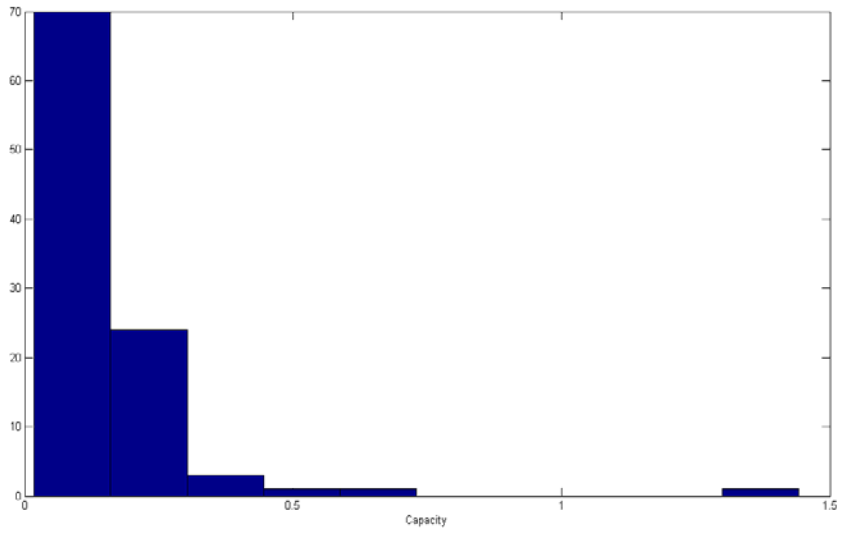

Fig. 12: Histogram plot for constrained capacity of hand geometry biometric

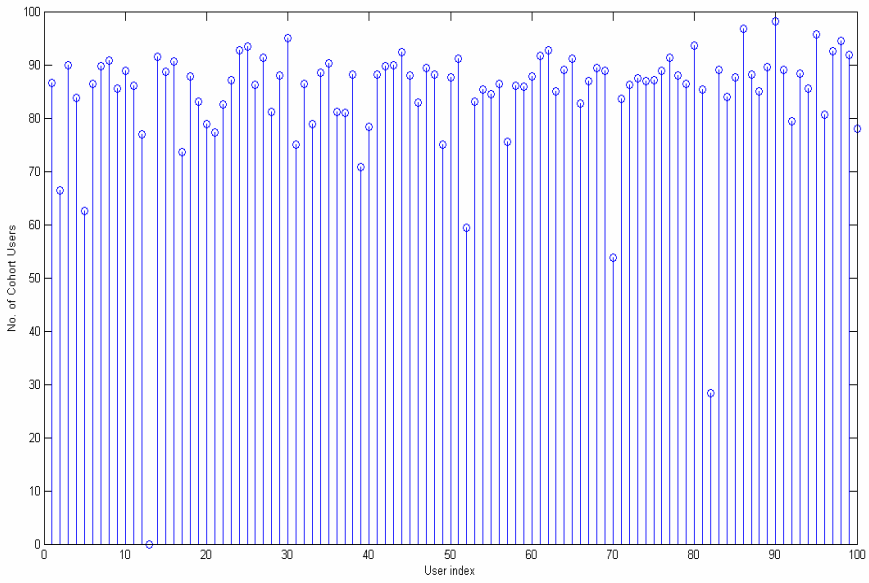

Fig. 13: Number of cohort users per user for hand geometry biometric

VIII. Conclusions

This paper has developed a new framework for constrained capacity as performance index of the user as well as the biometric system, given the database and matching function. In sections III-IV, constrained capacity of biometric system was formulated using second order statistics of the biometric information given in terms of a *SNR*. For this, the available biometric information must be categorized as signal or noise information, as seen in the statistics listed in Table 2. In real biometric applications we come across a finite

population of users and their biometric templates, and hence the information capacity theorem gives a constrained capacity of the biometric system. A significant advantage of constrained capacity developed in this work is that it gives a unified framework to evaluate authentication performance in terms of users, where as the traditional analysis employs probabilistic measures in terms of matching scores. Experiments in Section VII illustrate the use of constrained capacity in evaluating the authentication performance of biometric system and for a given user as seen in Figure 8 and Figure 11. The histogram plots in Section VII substantiate that palmprint biometric can give a superior authentication performance than hand geometry, as seen in Figure 9 and Figure 12. This paper has illustrated the usage of constrained capacity in estimating the relevance of biometric features [26] as well as in quantifying a measure of information or individuality of biometric features [4], as seen in Figure 5 and Figure 6. Finally, this work has also demonstrated use of constrained capacity for comparing some widely used score level fusion strategies, as seen in Table 3 and Figure 7. In conclusion, we can expect that constrained capacity may serve as computationally simple to handle and a useful performance tool in gaining important information theoretic insights for the design and evaluation of biometrics systems in the future.

In Section VI of this work, the probability of false correspondence of users based on their biometric features was developed using the source coding theorem. This paper has formulated a unifying framework to estimate *FRC* of a generic biometric trait given the intra-class biometric variance, and feature representation technique. This work has investigated the role of feature representation in characterizing the individuality of users' based on the biometric information content. In our opinion, biometric features constitute a natural code from the physical and behavioral characteristics of humans. Thus, a realistic model for quantifying the biometric variance and *FRC* will need detailed investigation in areas of evolution and sciences. We therefore feel that a holistic study of individuality is apparently closer to epigenetic and information theory than to information theory alone.

Appendix -A

Definitions:

$R(\overline{D})$: Rate distortion function is the infinum of achievable rates *R*, such that $R(\overline{D})$ is in the rate-distortion region of the source (*K*-feature statistics for all *M* users) for the corresponding distortion constraint $\overline{D}$ in mean square sense. Where, $k \in K$ denotes the number of different features per template.

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 { $\hat{f}$ }: feature representation, is optimal estimator of { f }: observed feature array; where $f \in F, \hat{f} \in \hat{F}$. Furthermore, mean square distortion $\overline{D}$ as the distortion measure in Gaussian frame work is effective for following reasons:

(i) Generally useful for real image data, (ii) Gives the minimum rate for representation error (tightened by the fact that rate at minimum distortion is equivalent to distortion at minimum rate), which is required in formulating individuality. Though the population statistics for *F*: Super set of all observed features of size *M* (one template per user), need not actually be Gaussian, we employ Gaussian approximation to *M*-user template statistics in order to deduce the achievable lower bound for rate-distortion $R(\overline{D})$.

Problem: To formulate rate distortion $R(\overline{D})$ 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 average distortion measure given by $\overline{D} \geq E\left|F - \hat{F}_k\right|^2 ; k \in K$

Hence, to show that: $R(\overline{D}) = \min \sum_{k=1}^{K} \frac{1}{2} \log_e \frac{\sigma_k^2}{D_k}; \sum_{\forall k \in K} D_k = \overline{D}$

Solution:

Let the *K*-different features per template be uncorrelated (for every $k \in K$, features are of different type hence we employ the uncorrelated approximation). 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 some k, which can be easily generalized under uncorrelatedness to all K features.

$$I(f_k; \hat{f}_k) = h(f_k) - h(f_k / \hat{f}_k)$$

$$= \frac{1}{2}\log_e(2\pi e)\sigma_k^2 - h(f_k - \hat{f}_k / \hat{f}_k)$$

$$\geq \frac{1}{2}\log_e(2\pi e)\sigma_k^2 - h(f_k - \hat{f}_k)\text{; (Conditioning reduces entropy)}$$

$$= \frac{1}{2}\log_e(2\pi e)\sigma_k^2 - h(N(0, E(f_k - \hat{f}_k)^2);$$

$$\geq \frac{1}{2}\log_e(2\pi e)\sigma_k^2 - \frac{1}{2}\log_e(2\pi e)D_k$$

$$R(D_k) = \frac{1}{2}\log_e[\frac{\sigma_k^2}{D_k}] = \inf\ I(f_k; \hat{f}_k)$$

For *K* uncorrelated variables (feature sets) the result can be generalized as below:

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

Appendix-B

Definitions: Let the random variable $\hat{g}_m$ with index $m \in M$ denote the complete set of median genuine scores for the database. Let the random variable $\hat{g}_l$ denote selection of a median genuine score such that $l \in M$. For this selection, let the random variable $\hat{g}_n$ denote a complement set of the median genuine scores such that $n \neq l, n \in M$.

Problem: We are interested to show that the random variables $\hat{g}_m$ and $\hat{g}_n - \hat{g}_l$ are uncorrelated.

Proof: To prove this we will need to show that the covariance of random variables $\hat{g}_m$ and $\hat{g}_n - \hat{g}_l$ is zero.

Let us assume that the expected value of $\hat{g}_m$ exists and is denoted as $\overline{g}$.

If '*E*' denotes the expectation operator then the covariance can be given as follows:

$$Cov(\hat{g}_m, \hat{g}_n - \hat{g}_l) = \mathrm{E}\left[\left(\hat{g}_m - \bar{g}\right)\left(\hat{g}_n - \hat{g}_l\right)\right]$$

$$= \mathrm{E}\left[\left(\hat{g}_m \hat{g}_n - \hat{g}_m \hat{g}_l - \bar{g}\hat{g}_n + \bar{g}\hat{g}_l\right)\right]$$

$$= \left[E\left(\hat{g}_m \hat{g}_n\right) - E\left(\hat{g}_m \hat{g}_l\right) - E\left(\bar{g}\hat{g}_n\right) + E\left(\bar{g}\hat{g}_l\right)\right]$$

We note that this is true in general for all $m \in M$ since there is no constraint on the choice of the index *m*.

The first order moments of $\hat{g}_l$ and $\hat{g}_n$ can be given as $\bar{g}$. Based on this, the covariance can be given as:

$$Cov(\hat{g}_m, \hat{g}_n - \hat{g}_l) = \left[E\left(\hat{g}_m^{\,2}\right) - E\left(\hat{g}_m^{\,2}\right) - \bar{g}^2 + \bar{g}^2\right] = 0.$$

This proves the claim.

It therefore follows that $\hat{g}_n - \hat{g}_l$ is uncorrelated with any function of $\hat{g}_m$.

IX. Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

X. References

[1] J. L. Wayman, A. K. Jain, D. Maltoni, and D. Maio (Eds.), *Biometric Systems: Technology, Design and Performance Evaluation*. New York: Springer Verlag, 2005.

[2] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman, "*Biometrics: A Grand Challenge*," *Proc. ICPR*, vol. II, UK, pp. 935-942, 2004.

[3]R M. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior, *Guide to Biometrics*, Springer Science, New York, 2003.

[4]A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A Tool for Information Security," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 125-143, 2006.

[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[6] N. A. Schmid and J. A. O'Sullivan, “Performance Prediction Methodology for Biometric System Using Large Deviations Approach,” *IEEE Trans. Signal Processing: Supplement on Secure Media,* vol. 52, no. 10, pp. 3036-3045, 2004.

[7] N.A. Schmid; M.V. Ketkar, H. Singh, and B. Cukic, “Performance analysis of iris-based identification system at the matching score level,”*IEEE Trans. on Information Forensics and Security,* vol. 1, no. 2, pp. 154-168, 2006.

[8] S. Pankanti, S. Prabhakar, and A. K. Jain “On the Individuality of Fingerprints,” *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 8, pp. 1010-1025, 2002.

[9] Y. Zhu, S.C. Dass and A.K. Jain, “Statistical Models for Assessing the Individuality of Fingerprints,” *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 3, pp. 391-401, 2007.

[10] A. Adler, R. Youmaran, and S. Loyka, “Towards a measure of biometric information,” *Proc. Canadian Conf. Computer Elec. Eng*., CCECE, Ottawa, Canada, 2006. http://www.sce.carleton.ca/faculty/adler//publications.

[11]J. Daugman, “Probing the uniqueness and randomness of Iris Codes: Results from 200 billion iris pair comparisons,” *Proc. IEEE*, vol. 94, no. 11, pp. 1927-1935, 2006.

[12] R.M. Bolle, S. Pankanti, J. H. Connell, N. K. Ratha, “Iris Individuality: A Partial Iris Model,” *Proc. of ICPR*, pp. 927-930, 2004.

[13] S. Verdú, and S. McLaughlin, *Information Theory: 50 years of Discovery*. Wiley Series, 1999.

[14] R. G. Gallager, “Power Limited Channels: Coding, Multi-access, and Spread Spectrum,” Codes, Graphs, and Systems, Kluwer Academic Publishers, (Eds.) R. Blahut, and R. Koetter, 2002.

[15] S. Verdú, and T. S. Han, “A General Formula for Channel Capacity,” *IEEE Trans. on Information. Theory*, vol. 40, no. 4, pp. 1147-1157, 1994.

[16] M. Simon, S. Hinedi, and W. Lindsey, *Digital communication Techniques: Signal Design and Detection.* Prentice Hall – New Jersey, 1995.

[17] W.Feller, *An Introduction to Probability Theory and Applications*. John Wiley & Sons, 1971.

[18] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia and A. K. Jain, “Incorporating image quality in multi-algorithm fingerprint verification,” *Proc. ICB*, pp. 213-220, Hong Kong, 2006.

[19] G. Aggarwal, N. Ratha, and R. M. Bolle, “Biometric Verification: Looking Beyond Raw Similarity Scores,” *Workshop on Multibiometrics* (*CVPR*), New York, pp. 31 – 36, 2006

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley & Sons. 1991.

[21] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer Verlag, 2006

[22] S. Dass, K. Nandakumar, and A. K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems," *Proc. AVBPA*, pp. 1049-1058, New York, 2005.

[23] A. Kumar and D. Zhang, "Personal authentication using multiple palmprint representation," *Pattern Recognition*, vol. 38, pp. 1695-1706, Oct. 2005.

[24] A. Ross, and A. K. Jain, "Information Fusion in Biometrics," *Pattern Recognition Letters*, vol. 24, no.13, pp.2115-2125, 2003.

[25] J. Bhatnagar and A. Kumar, "On Some Performance Indices for Biometric Identification System," *Proc. ICB 2007, Lecture Notes in Computer Science, Springer-Verlag GmbH*, vol. 4642, pp. 1043-1056, Aug. 2007.

[26] A. Kumar and D. Zhang, "Personal recognition using shape and texture," *IEEE Trans. Image Process.*, vol. 15, no 8, pp. 2454-2461, Aug. 2006.

[27] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July-October, 1948.