Estimating Sample Size Requirements for Reliable Personal Authentication Using User-Specific Samples

Jay Bhatnagar, Ajay Kumar
Department of Electrical Engineering, Indian Institute of Technology, New Delhi, India
jbhatnagar@ee.iitd.ac.in, ajaykr@ee.iitd.ac.in

Abstract

The goal of this paper is to determine bounds for estimating minimum sample size requirement for reliable biometric identification. A new approach for the reliable estimation of the minimum sample size is proposed for arbitrary ensemble of subjects. A bound on number of acquisitions/samples per subject is arrived through an iterative procedure that tests sequences for user-specific sequences. The approach proposed in this paper is supported by information theoretic measures. These results are fundamental to the integration of concepts from statistics, complexity and probabilistic (Borel) measure spaces. We evolve a novel concept of information equivalence in comparing random sequences for its information content. Furthermore, the problem of missing or lost/corrupted matching scores is also investigated. The solution for these missing biometric matching scores is based on completeness of certain typical space and these scores can be estimated using proposed iterative algorithm.

I. Introduction

Reliability in personal authentication is the key to stringent security requirements in many application domains ranging from airport surveillance to electronic banking. Many physiological characteristics of humans, *i.e.*, biometrics, are typically invariant over time, easy to acquire, and unique to each individual. There has been increasing interest in the assessment of quality of biometric images for variety of applications [2]-[3]. However, there has not been any attempt to reliably extract/tune the system performance when biometric features cannot be reliably extracted (missing /lost) from low quality images.

The first objective of this paper is to propose succession of weak bounds to later justify a strong upper bound on minimum biometric sample size requirement per subject. This paper also investigates how system improvement can be made under poor quality of input biometric samples from subject(s). Strong bound is based on user-specific sequences for a given biometric modality. A bound on minimum sample size is strong when it can accurately describe a biometric source for the claimed sample size from the bound, under information equivalence. All other non-AEP (Asymptotic Equi-partition Property) based conditions [1], analytically also give estimates of minimum sample size [10]. Information measures are defined on biometric samples/sequences in measure space. It is shown that the space of these random biometric sequences has *measure* and *topological* properties that naturally lead to the concept of information measures on distributions, covers, $\varepsilon^{(k)}$ closed covers or $\varepsilon^{(k)}$- typical sequences, homeomorphism, equivalence relations etc. Based on these measures, a system is proposed for the estimation of minimal sample size using an iterative algorithm. Suitable choice of tunable thresholds and knowledge of $\varepsilon^{(k)}$- sets greatly improves the estimation accuracy for minimum sample size/subject on subject ensemble. Towards concluding sections, we propose architecture for missing/lost score estimation using successive refinements from cached $\varepsilon^{(k)}$ typical sequences, iteratively combined with input samples.

II. Problem Background

Let T_j be statistical test functions of each stage $j \le J$ that sequentially transform space of input sequences $\{X_i\}$, until final decision stage $j = J$. The system design is optimal if it is jointly optimal. This will mean that all $\{T_j\}$, jointly optimize identification over the complete ensemble of all K users under J processing blocks. Since input to each T_j is a statistical set the following can be said:

(i) A Biometric system works with a sequential set of statistical test functions at each block of data processing from feature extraction to decision making.

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

(ii) We aim at reduction in noise at the front end input stage of biometric system followed by outputs $T_j\{X_i\}$. The function space constituted by $\{T_j\}$ will *search* on the *complexity* of observed space $\{X_i\}$ or equivalently it will partition the observed space of $\{X_i\}$ into $\varepsilon^{(k)}$ typical for some subject *k*, here $\varepsilon^{(k)}$ is typical set which is information theoretically defined.

(iii) There is *no* such *unique* sequence to consider as *reliable* matching reference that can give enough conditional information from user k. Although some error control methods have been suggested [9] to identify the *typical error* vector set. We instead focus on *subject* $\varepsilon^{(k)}$ *typical* sequence, than typical error sets. There exist well defined structures given by AEP property (alternately, weak law of large numbers) for $\varepsilon^{(k)}$ typical sequences [5].

The AEP also induces that adding redundant or extra sequences to those already in a compact measurable $\varepsilon^{(k)}$ typical set, will not significantly add to information measures in making inference about or describing the biometric source. The $\varepsilon^{(k)}$ typical sets for subject *k* contain equiprobable sequences (equivalent in information measure) that give a minimum description complexity of a statistical source. The minimum complexity or description length of samples (number of biometric acquisitions) in the $\varepsilon^{(k)}$ typical sense is reached with probability close to 1. The probabilistic model of this source can be approximated by a normal distribution for large size of user typical sequences.

(iv) The performance metrics such as FAR (λ), GAR (λ) ~ ROC (λ) and minimum sample size per subject for given reliability (significance level) vary with quality of test inputs, choice of threshold in ROC *etc*.. Variation in quality of inputs is a random effect although threshold is an adjustable system parameter. A dense set such as $\varepsilon^{(k)}$ typical set characterizes a biometric subject with built-in probability measures that distribute randomness (uncertainty) on most likely $\varepsilon^{(k)}$ sequences for a subject *k*. The minimum realization of $\varepsilon^{(k)}$ space that can always extract a subject sequence string from presented sequence is the sequence space which maximizes average self information function over ensemble of *K* subjects. This does not include the inter subject correlation. With inter subject correlation or information dependence among subjects; it is required to change the $\varepsilon^{(k)}$ space per subject to give *minimal jointly* typical $\varepsilon^{(k,j)}$ description sets. Measure of *change* is given by information measures and inequalities. In such cases, knowledge of intra- and inter-sample dependence given by conditionals can be used to compute bound on minimum required samples per subject under the stated conditions.

III. Problem Definition

Estimate number of samples per subject (or acquisitions per subject) from the ensemble of *K* subjects, with the acquisitions for same subject being statistically correlated. Let $\{X_i\}^k$ denote time series of biometric sequences of subject *k*, where *k* is an arbitrary subject from population *K*. $\{X_i\}^k$ or simply $\{X_i\}$ is sequence of countable finite cardinality $i \leq N$. All $\{X_i\}$ have finite first and second moments. Thus implies that; $\|X\|^2 = \{\sum_{\forall i} | X_i |^2\}^{1/2} < \infty$, with norm defined as second order expectation in l^2 space of random sequences $\{X_i\}$. We do not prove the existence of such a space, knowing that all the second moments are finite.

Denote $\overline{\{X_i\}}$ as span of $\{X_i\}$ defined as: $\overline{\{X_i\}} = \{X_i\}_{\forall i} \cup \{X_\phi\}_{\forall \phi}$, where indexing ϕ is for null space with zero probability measure. Any further analysis is based for $\{X_i\}$ space, as we work with real biometric samples having non- zero measure. Again, it is trivial that biometric samples of $\{X_i\}$ constitute a non – empty sequence space.

Ergodic property of a sequential space $\{X_i\}^k$ with associated probability measures defines equality between ensemble average and time average of the *property*. Lastly, ergodicity can be used to test for convergence of the time series in the property.

We look for a sample size $N = \check{N}$, ergodic in first moment; formally stating the moment ergodic condition to computing size $\check{N}$:

$$\check{N}(k) = \{N : \operatorname{argmin}[P[\frac{1}{N}\sum_{\forall i} X_i^{(k)} - E(X_i^{(k)})] \geq \varepsilon]\}$$

$\varepsilon \geq 0$ in probability, (1)

$$\check{N}(k) = \{N : E(X_i^{(k)}) \rightarrow \frac{1}{N}\sum_{\forall i} X_i^{(k)} \; probability\}$$

(2)

For sequences $\{X_i\}$ with finite first moments, convergence in probability to sample mean is same as ergodicity in first moment. The resulting bound for $\check{N}$ is weak as explained now. The Information contained in first moment ergodic sequences suffers from non usage of higher order statistical information. Thus $\check{N}$ is a weak lower bound. Ergodic conditions on higher order moments can be used to estimating $\check{N}$. Based on

second-order moments we have better information set which accounts for the intra-subject correlation and hence give better bounds on estimate $\check{N}$.

Loeve's criteria/theorem: For random sequences $\{X_i\}$ defined on real space with non-zero Borel measures, under the condition that $\forall\, i$, and bounded second moments for all i : $\{\sum_{\forall i} \mid X_i \mid^2\}^{1/2} < \infty$, *there exists* a valid correlation function $\stackrel{\Delta}{=} R_x(\Gamma)$. If, $R_x(n,m) \to R_x(\Gamma)$, $\lim n,m < N = \check{N}$ then the sequences $\{X_i\}$ converge in correlation.

weakest bound -

$$\check{N} = \left\{ N : \mid \{\lim P[\|\{X_{\check{n}} - X_m\| > \varepsilon \to 0]\} \right\} \quad (3)$$

$\forall\ n, m \leq N = \check{N}$

If we can show existence of a correlation function such as $R_x(\Gamma)$ then convergence is defined for all such $\{X_i\}$ sequences that exist in $\varepsilon^{(k)}$ typical set with probability > 0.Let the source described (alternately, ε - spanned by) by $\{X_i\}$ be *ergodic in law* or distribution (weakest of ergodic behaviors).

Joint approach: Ergodic conditions combined with AEP can give tighter estimate of minimum number of required acquisitions, $i = \check{N}$ *than* ergodic conditions alone. However, the tighter AEP bound suffers from conditioning on weaker ergodic convergence. This may lead to poor estimate. Summarizing the choice of minimum $\grave{N}$ *per* subject below, for all subjects = 1, 2……., *K* there will be *K* such values for $\check{N}_K$ (some distinct and some repeated values).Choice of function such as

$$\grave{N} = \max[\check{N}_1, \check{N}_2 \ldots\ldots\ldots\ldots, \check{N}_K] \quad (4)$$

Equation (4) assures sufficient number of biometric samples $\grave{N}$ / *subject* so that for any subject *k* the estimate does not vary significantly from using inequalities in computing $\check{N}_k$. Summarizing the approach so far we have used ergodic properties of time series. The HOE (higher order ergodicity) also accounted for inter-sample correlation. All the preceding bounds are weak since convergence is merely that of random sequences to some random sequence. The above bounds do not account for the distribution of information spectrum characterizing the biometric subject. Applying AEP to algorithms above, we distribute probability metrics on sequence space spanned by $\{X_i\}$, such that probability of any ∂ - neighborhood is uniform for $\partial < \varepsilon; \varepsilon\ ,\ \partial > 0$. This means almost all sequences in the ε - ball convey the same amount of self information. This also can be interpreted as asking for say *M* sequences that make a source look like white sequence, with each sequence contributing almost equal measure of uncertainty (variance, self information) to the overall uncertainty of the source (total self information or total variance).

IV. Generating $\varepsilon^{(k)}$- Typical Sequence Space: Huffman argument

Huffman's argument [1] best characterizes a useful algorithm to describe a source using ε typical space. The algorithm is also useful in constructing ε typical sets for boot strap generator that will be used later. Choose large number of low probable sequences and small number of high probable sequences. The resulting information spectrum looks equidistant with respect to the subject *k*, distributing randomness uniformly. Alternately, the ε typical sets comprise sequences with equivalent information measures. *K* such source $\varepsilon^{(k)}$- typical balls can be generated based on knowledge of distribution functions of sources. Alternately, inverse Fano – Elias algorithm can also be used to construct the $\varepsilon^{(k)}$ source space [1]. We now define some useful information measures: Mutual -information is a *random variable* and quantifies information distribution per sequence of $\{X_i\}$ for subject *s=k;*

$$\mathrm{I}(x_i, s = k\ ) = \log \frac{p[s=k]}{p[x=x_i\ ]}; \forall i \in N \quad (5)$$

It is clear from equation (5) that any distribution $p[x = x_i]$ on ensemble of $p[s=k]$ that can maximize average mutual information gives an optimal estimation or testing rule on [7] biometric samples $\{X_i\}$.Also note that maximizing the log likelihood function is similar to MLE (Maximum Likelihood) rule. The numerator is a-priori probability. If numerator is uniform distributed then biometric source presents maximum uncertainty. It is well known that entropy for an arbitrary biometric source is smaller than the entropy of a uniform distribution of biometric samples.

Average self information:

$$\hat{\mathrm{I}}(X_i\ ) \stackrel{\Delta}{=} E_X[I(x_i)]; \forall i \in N$$

$$\hat{\mathrm{I}}(X_i) = \sum_{\forall i \in N} p[x=x_i] \log \frac{1}{p[x=x_i]} \quad (6)$$

Applying A.E.P to information space of large size N of biometric samples and using measures defined above, we obtain a tight upper bound on sample size per subject $\check{N}$. This leads to defining a mode of convergence based on Information measures, which we state below.

$$\check{N} = \left\{ \min N : \left| \frac{1}{N} I(X_i) \rightarrow \hat{I}(X_i), \forall i \in N \right\} \quad (7)$$

probability

Restating the above equation:

$$\check{N} = \{N \mid \operatorname{argmin} \left\{ p\left[\left| \sum_{\forall i} p(x = x_i) \log \frac{1}{p(x = x_i)} - \frac{1}{N} \sum_{\forall i} \frac{1}{\log p(x = x_i)} \right| \right] > \varepsilon \right\} \} \quad (8)$$

Hence, the resulting bound on $N = \hat{N}$

V. Bound on Minimum Sample Size

On this note we revisit our problem of information measures and equivalence in $\varepsilon^{(k)}$ typical set. Clearly, from definition information equivalence can be achieved only by staying inside the ε typical set. This suggests that we can query for multiple samples or acquisitions for input sequences enough to cover a ε typical set maximally. From homeomorphism again we can see a linear mapping from $\varepsilon^{(k)}$ typical set sequence space into a $\varepsilon^{(k)}$ typical set of scores.

Setting, $\mathbf{N} = \check{\mathbf{N}}$ samples or acquisitions per subject to ascertain that enough of these samples are information sequences in a closed cover C $[\varepsilon^{(k)}]$ or $\varepsilon^{(k)}$ typical set. The value $\check{N}$ will vary from subject to subject**.** A rule for making *multiple acquisitions* will be to acquire samples $\mathbf{N} = \check{\mathbf{N}}$ of $\{X_{(i)}\}^{(k)}$ to cover the $\varepsilon^{(k)}$ typical space *maximally* (Appendix A). This result will be used for missing score problem. Let us look at a case where we have a much weaker input information sequence space such as $\left\{ \tilde{Y}_{(i)} \right\}$

Estimating bound of minimum samples per subject for ensemble of K subjects under correlated multiple acquisitions for a subject input data set $\left\{ \tilde{Y}_{(i)} \right\}$.The information space $\left\{ \tilde{Y}_{(i)} \right\}^{(k)}$ is weaker than $\{X_{(i)}\}^{(k)}$ (also, refer Appendix A). We can model effect of inter sample correlation by a Markov model, using a first order homogenous Markov chain, the choice made for simplicity of analysis. Higher order and more structured Markov chains can better model interdependence of samples / acquisitions. The order reflects the memory requirement or correlative behavior for dependence among samples [6].As a result of dependence the average information measure will be reduced per length of input data $\left\{ \tilde{Y}_{(i)} \right\}$.This asks for **at least** (lower bound since the correlated samples may not maximally cover ε-typicality) $\hat{E}$ extra samples. The increased number of min. samples improve the $\varepsilon^{(k)}$ typicality of $\left\{ \tilde{Y}_{(i)} \right\}$,this in turn minimizes distortion function in describing a certain subject $s = k$. Information measure in the new correlated space can be related to that of original uncorrelated case.

$$I\left[\left\{ \tilde{Y}_{(i)} \right\} \right] = \left[\sum_{\forall i,j} \log \frac{1}{p(y_i)} - \log \frac{1}{p(y_i) \big/ p(y_j)} \right] \quad (9)$$

Note the first term and second terms, in (9) are random variables, describing self information and conditional self information (from dependence). Expectation of $I\left[\left\{ \tilde{Y}(i) \right\} \right]$ makes more sense as a non-negative ensemble information measure. Another useful measure is:

$$\Omega\left[\left\{ \tilde{Y}_{(i)} \right\} \right] \stackrel{\Delta}{=} \min \arg\left(E\left[I\left\{ \tilde{Y}_{(i)} \right\} \right] \right), \forall (i, j) \in N \quad (10)$$

Such source will be a poor or weak information set to cover the required ε - typical set, for which as suggested before we will need at least $\hat{E}$ extra samples. We now look at quantifying the variable $\hat{E}$.

$$\hat{E} = \sum_{\forall i,j} p\left[y_{(i),} y_{(j)} \right] \log \frac{1}{p[y_{(i)}] \big/ p\left[y_{(j)} \right]} \quad (11)$$

This is the statistical average of variable in (9) that measures loss of information from dependence (please refer Appendix A). The $\hat{E}$ bounds the maximum loss or minima condition of information measure for input $\left\{ \tilde{Y}_{(i)} \right\}$.So, for a correlated set of acquisitions $\left\{ \tilde{Y}_{(i)} \right\}$ the number of acquisitions required to achieve an information set equivalence (strong) between $\left\{ \tilde{Y}_{(i)} \right\}$ and $\{X_{(i)}\}$ for arbitrarily low distortion measure is $= \hat{E} + \check{N}$.

VI. Problem of Missing Features

Often the quality of samples and therefore the scores may be below acceptable threshold giving poor quality of information measures. The quality of scores is a random variable over all the acquisitions made. Sorting input samples of user sequence in maximizing information measures is not as tractable. The *intractability* is from NP (non polynomial) hardness of input sequences. A theoretical model that gives locations of reliable (or poor quality) sequences in acquisitions made is also analytically intractable requiring complex combinatorial optimization procedures for locator functions.

A possible solution will be over sample the input space or work in space of large acquisitions. We propose an estimate of the over sampling required based on information measures and equivalence. Due to poor quality of samples, the resulting scores may be missing or erased. We have not quantified but a natural consequence of an insufficient score set will be incomplete set of points for the classifier to base a decision (alternately, choose a directed branching for search based algorithms). We now formulate a model to the case of missing score or erased score.

Let p denote penalty or probability that one out of i scores is erased. For simplicity we can assume that loss (/erasure) of any score has same penalty as any other score on the total information measure. If k of i scores are erased, where k is a random variable then there will be some k_c = critical k, such that any more erasures $k > k_c$ will substantially increase the error rate of the classifier.

$P(k = k_c) = {}^{i}C_{k_c} p^{k_c} q^{i-k_c}$; Binomial Distribution – B (i , p) where $q = 1 - p$

We assume independence of erasure events with binomial distribution for simplifying the analysis. Binomial is *asymptotically* good fit distribution of number of erasures under assumption of locating an erasure with probability p uniformly over sample size $i \to N$ .The equal penalty condition for missing data applies over all input samples. From above it follows that each of ${}^{i}C_{k_c}$ permutations contributes to the same information loss because a finite number of such combinations will have zero measure.
(The ensemble average probability of having k_c erasures) $\geq \frac{1}{{}^{i}C_{k_c}} \sum_{\forall l} p_l[k = k_c]$; $l \leq {}^{i}C_{k_c}$

Where, $p_l[k = k_c] = p^{k_c}.q^{i-k_c}$

Re-arranging, $p_l[k = k_c] = \left(\frac{p}{q}\right)^{k_{cc}} q^i$ (12)

Expressing average loss of information below for the erasure distribution such as proposed above;

$$I_E(k_c) = \sum_{\forall k_c} p_l[k = k_c] \log \frac{1}{p_l[k = k_c]} \quad (13)$$

We conclude by proposing architecture and corresponding algorithm on problems of minimum sample size estimation. Another algorithm is proposed using bootstrap ε - *user typical* sequence generator for problem on missing score.

VII. Estimating Minimum Sample Size Requirements

The block diagram of the algorithm for the sample size requirements is shown in figure 1. The proposed algorithm can be summarized as follows:

(i) Generate a user specific ε typical set, for K subject population there will be K such $\mathcal{E}$ typical sets

(ii) Define an initial threshold S_{th} *(j=0)* for every user $s = k$, this threshold can be set from several criteria, one of which we use here is a-priori probability for users, $s = k$. Additional criteria such as a-priori on quality of samples will also be useful

(iii) Accept input samples a_j for say user $s = k$, this starts the iterative procedure for iteration j

(iv) Feed a_j into each of the ε^k - *user matchers*. The matcher outputs are then ordered and compared with respect to threshold = S_{th} *(j).* (*Note: that* we quality of estimate for bound can be improved adjusting threshold on every iteration j. As a first step, we simplify that S_{th} *(j)* is same for all iterations. This is valid for equal probability of poor quality samples independent of input sample size)

(v) From the last step we get a set of survivors that is put in memory. This list will have ordered survivors the ordering defined by threshold. [1], [8]

(vi) Next iteration $j = 1$ will start and we follow the steps 3 to 6, With every iteration the input sample size will increase as we query for more samples in every new iteration

(vii) As j is successively incremented, $j \to J$, where J is finite valued integer, the algorithm proceeds iteratively to smallest size of survivors which will in most cases be singleton set for large J. This claim is argued on nature of user ε- typical sequences. Majority voting will declare the winner or final survivor. In case of more survivors some intrinsic information can be used. One such intrinsic measure is total metric of ordering index given to user $s = k$,

over all iterations.The order number designates a value to each survivor coming from that iteration. It is an intrinsic measure since it is incorporated in the iterations from one to next and must be used to improve the estimate further. The algorithm comes out of step 6 and terminates.

The algorithm will give K outputs to the minimum sample size requirement for K users. We can either use a mean value of the above spread based on simple average.

$$\hat{N} = \max(N_1, N_2 .....N_k) \quad (14)$$

$$\hat{N} = \frac{1}{K}\sum_{\forall k} N_k \quad (15)$$

$$\hat{N} = \frac{1}{K}\sum_{\forall k} N_k + \sigma_N \quad (16)$$

where, σ_N denotes sample variance computed on $\{N_i\}$

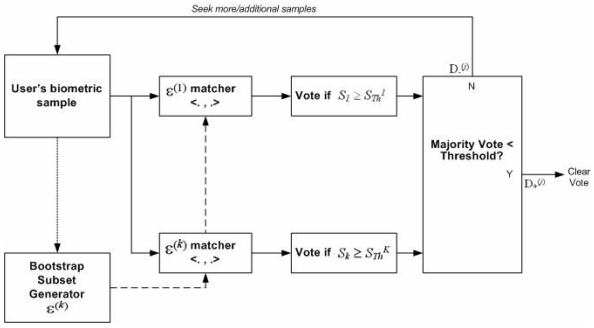

Figure 1: Estimating minimum sample requirements for K user ensemble.

Many more functions can be used to get a uniformly reliable minimum sample size for ensemble of K subjects. The maximum function gives an overestimate or greatest $\hat{N}$ for minimum sample size requirement. A simple sample mean function can be considered as the least value below which a fraction of total users $\xi > \frac{K}{2}$, will be under-sampled. Under-sampling as we refer from loss function equations will give unfaithful representation of user ε- typical sequences. The next bound of minimum sample requirement which has some σ_N more samples than the sample mean size will perform sub-optimally and give closer information measures to the ε- typical without over sampling.

Algorithm for missing sample/ features set compensation

(i) In iteration $j = 0$, the input sequence a_j is fed into combiners which have nil user ε – typical elements, so the combination is with null space [1], [8]

(ii) The inputs sequence a_j are available at input of matchers

(iii) Score values from matchers are then ordered and compared to some threshold

(iv) A majority decision is made on the outputs and a list of survivors stored

(v) Increment j, and now combine ε_j –typical from each user $s = k$, such that each typical sequence is innovations (set complement) for the a_j input samples presented. The combination takes place in the summer and algorithm moves to step ii)

(vi) Repeat steps (ii) – (v), until a final survivor is announced with majority voting from candidate survivors at some iteration $j = J$. Algorithm terminates. If a_j is already close to certain user typical set at any iteration $j \geq 0$, this information is used as priori intrinsic in setting the thresholds.

VIII. Conclusion

This paper proposes new approach for the reliable estimation of the minimum sample size for the random ensemble of subjects. A bound on number of acquisitions/samples per subject is arrived through an iterative procedure that tests sequences for user-specific sequences. The emphasis was not on a certain family of distributions or specific tests but on much broader information model perspective and user-specific estimation from these data models. The achieved bounds in section V and VI form a good basis for estimating both approximate processes from the models and in prediction of their performance. One may choose alternate pre-classifier over majority voting under obvious dependence of estimated size on input correlation. The novelty of our approach lies in employing information equivalence in comparing random sequences for its information content. Furthermore, this paper has also suggested a new approach to the problem of missing or lost/corrupted matching scores.

Appendix A

Definition 1: **Weak information space -** Given information equivalent sets, Sequences $\left\{\tilde{Y}_{(i)}\right\}$ is information weaker than $\{X_{(i)}\}$, if conditioned on some event ($s = k$) the information measure of former is uniformly smaller under all information similarity, with the cardinality of $\left\{\tilde{Y}_{(i)}\right\} \geq \{X_{(i)}\}$, for each set of similarity of information.

Definition 2: **Distortion** – Information loss
Given sequences $\{X_i\}$ and $\{Y_i\}$ under information equivalence. Distortion or loss of information = $D = E_{x,y}\left\|X_i - Y_i\right\|^2$

We aim to construct with high probability $\varepsilon^{(k)}$ typical sequences from $\{X_i\}$. Any sequences $\{Y_i\}$ will lead to information loss (under-sampled source representation). A good estimate of size $N = \check{N}$ will minimize D for joint ensemble (X_i, Y_i).

Definition 3: **Information Equivalence**
Given, $\{X_i\}^k \in \varepsilon^{(k)}$ typical set and if, $\exists$ another $\{Y_i\}^k \in \varepsilon^{(k)}$ typical set. $\{X_i\}^k$ is information equivalent to $\{Y_i\}^k$ denoted by, $\{X_i\}^k \sim \{Y_i\}^k$ if mutual information from $\{X_i\}^k$ in denoting a subject $s = k$ is similar to $\{Y_i\}^k$ i.e.,
$I[s = k/\{X_i\}] \sim I[s = k/\{Y_i\}]$ Implies that $\{X_i\}^k$ and $\{Y_i\}^k$ have non-null intersection with the same $\varepsilon^{(k)}$ typical set. Stronger the similarity more is the point-wise convergence of a sequence from $\{X_i\}^k$ into $\{Y_i\}^k$ or alternately stronger is dependence on $\varepsilon^{(k)}$ typical set. Information measure equivalence also has a weaker connotation to that of **Similarity**: When augmenting sequences $\{X_i\}^k$ and $\{Y_i\}^k$ from arbitrary subjects, by say φ new elements (sequences) if the information measures from the two sequences before augmenting and after augmenting show similar variation (increase or decrease) then the information measures are similar[4].

• A simple construction for similarity results when $\{Y_i\}$ is improper subspace of $\{X_i\}$ for which, $I[s = k/\{Y_i\}] \leq I[s = k/\{X_i\}]$ and,
$I[s = k/\{X_i\} \cup \varphi] \leq I[s = k/\{Y_i\} \cup \varphi]$, where U denotes set union operation for a (innovation) new sequence φ, where $\varphi \neq \{0\}$ with probability = 1.

• Given, $\{X_i\}^k = \{Y_i\}^k$, similarity implies equality implies similarity.

Lemma: The total sequence space $\{Z_{(i)}\}$ for some subject s = k can be partitioned into equivalence classes under information equivalence. Each equivalence class has at least one point which is accumulation point of $\varepsilon^{(k)}$ typical set of subject s = k.

Corollary (state without proof): If a family of sequence $\{X_i\}$ for subject k is information equivalent to any other Z sequence, and if the information measures of random sequence $\{X_i\}$ converge point wise to that of sequence, *implies that* , $\{X_i\}$ is nominally covered by $\varepsilon^{(k)}$ typical set and Z is contained in $\varepsilon^{(k)}$ typical set.

Corollary or definition of ε typical space: It is the smallest closed topological cover that contains information measures from all sub-covers, where each sub-cover is topological space defined by a unique equivalence class (*information class*).

The above simple results follow *naturally* from elementary topology of real spaces and concept of class as ordered sets. We define some more basic equivalence relations. Consider a proper space (alternately, closed sets) A, for which $a \in A$ and $b \in A$. Choose any $a,b \in A$. [a] ~ [b]; is equivalence relation under some operation O, if for any a, all of these exist: inverse, identity element and composition properties are satisfied under O.

Proof for existence of Information equivalence: Information equivalence relations are defined on a domain which is Borel measurable space of (0,1). Borel space (ß) is always closed under finite intersections and unions on measure sets, other than for those sequences that lie in null space. Thus, the sequences or subsequences on such ß -space are closed groups under set theoretic operations of unions, intersections and also composition. We know that ß -space has equivalence relations and classes to be decomposed for each proper space or set of equivalence classes [b] ∈ ß. The transformation from space of I (I can be self or mutual information measures) with information measures on sequences, to ß- space is onto and continuous, other than for a finite null space with measure zero. Therefore, we can define a homeomorphism from space I to space ß for

all sequences $\{X_i\}$ This proves our claim of existence of an equivalence relation in information sense.

IX. **References**

[1] Robert G. Gallager, *Information theory and reliable communication*, John Wiley, 1968.

[2] Y. Chen, S. Dass and A K. Jain, "Localized Iris Image Quality Using 2-D Wavelets", *Proc. of International Conference on Biometrics (ICB)*, pp. 373-381, Hong Kong, Jan. 2006.

[3] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia and A. K. Jain, *"Incorporating image quality in multi-algorithm fingerprint verification," Proc. of International Conference on Biometrics (ICB)*, Jan. 2006

[4] Z. Zhang and R.Yeung, "*On Characterization of Entropy Function via Information Inequalities*", *IEEE Trans. on Inform. Theory,* pp. 1440-1452, Jul., 1998

[5] T. Cover and P. Hart, "*Nearest neighbor pattern classification*", *IEEE Trans. Inform. Theory*, IT-13, pp. 21-27, 1967.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 2.* New York: Wiley, 1966

[7] N. A. Schmid and J. A. O'Sullivan, "Performance Prediction Methodology for Biometric Systems Using a Large Deviations Approach," *IEEE Transactions on Signal Processing*, Oct. 2004.

[8] G. D. Forney, B. J. Frey, R. Koetter, F. Kschischang, R. J. McEliece, D. Spielman, "Codes on Graphs and Iterative Algorithms", special issue of the *IEEE Transactions on Information Theory*, Feb. 2001

[9] U. Uludag, S. Pankanti and A K. Jain, "Fuzzy Vault for Fingerprints", *AVBPA 2005*, pp. 310-319, Jul. 2005.

[10] Andy Adler, Richard Youmaran, Sergey Loyka, "Information content of biometric features", Invited talk - *Biometrics Consortium Conference,* Arlington 2005.