A Generalization of the Theory of Biometric System Entropy

Kenta Takahashi* and Takao Murakami†
* Hitachi, Ltd., Yokohama, Japan
E-mail: kenta.takahashi.bw@hitachi.com
† National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
E-mail: takao-murakami@aist.go.jp

***Abstract*—Takahashi and Murakami introduced a theoretical framework to define and evaluate a measure of information gained through a biometric matching system, called the *Biometric System Entropy* or BSE. The BSE enables us to understand and evaluate the personal identification capability of biometric information from an information theoretical point of view. However, there are limitations when evaluating the BSE for actual systems and biometric information; (1) the BSE cannot be applied to evaluation of biometric information entropy of individuals, (2) it requires a strong and unrealistic assumption regarding statistical distributions of biometric information. In this paper, we generalize the theory of the BSE and give a new measure of biometric information so that we can evaluate both individual and average entropy of any kind of biometric information and verification system without unrealistic assumptions.**

I. Introduction

Biometric verification technologies for identifying individuals based on measurement data of physical or behavioral characteristics such as fingerprints, faces, irises, veins, gaits, etc., are becoming popular. The most widely used measures for quantitatively evaluating the individual identification performance of a certain biometric verification system are the false rejection rate (FRR) and the false acceptance rate (FAR). However, these error rates change depending on the threshold parameter t for the matching score (i.e., distance or similarity between the biometric information for enrollment and verification). Thus, it is standardized to describe the accuracy using the DET (Decision Error Tradeoff) curve [1]. The DET curve is defined as the locus of the set of points $(FAR(t), FRR(t))$ parametrized by t. Although the DET curve can describe the accuracy of the biometric verification system in detail, it is not easy to understand intuitively. For example, when the DET curves of two systems intersect each other, it will not be straightforward to tell which is better.

On the other hand, several attempts have been made to evaluate individual identification performance of biometric verification systems based on the entropy of biometric information [2], [3], [4], [7]. These approaches enable quantitative comparison not only between different biometric systems but also with passwords, personal identification numbers, etc., so that the discrimination ability can be more intuitively understood.

In particular, the BSE (Biometric System Entropy) proposed by Takahashi and Murakami [7] has various "natural" properties as the entropy of biometric information, such as nonnegativity, subadditivity, relationship with Bayes classification error and with the DET curve. In addition, the BSE can be practically evaluated for an arbitrary biometric verification system (although it is an asymptotic approximation). In this sense, it can be said that the BSE combines practicality and theoretical support.

However, there are limitations when evaluating the BSE for actual systems and biometric information. Firstly, the BSE of a system S is defined for a user set $\mathcal{U}$ (with probabilistic distribution over it) and thus it represents the average entropy for personal identification over $\mathcal{U}$. Therefore, although the BSE can be used as a information theoretical measure to evaluate the identification performance of S itself, it cannot be used to evaluate entropy of individual biometric information. Furthermore, in deriving a practical evaluation measure and procedure of the BSE, a too ideal assumption is made in [7] where all users follow the same matching score distribution, and there is room for discussion about the validity thereof.

In this paper, we generalize the theory of the BSE and provide a new framework to evaluate a measure of biometric information entropy of each individual and of average entropy of any kind of biometric system without unrealistic assumptions. Our contributions are summarized as follows:

1) We propose a new definition, called the *Biometric System Entropy of Individual (BSEI)*, as a measure of biometric information of each individual to be consistent theoretically with the conventional definition of the BSE. We show useful properties to understand the BSEI.
2) Since it is difficult to evaluate and calculate the BSEI directly according to our new definition in practice, we derive an approximate measure of the BSEI that can be easily evaluated and calculated for actual systems.
3) We provide a practical evaluation protocol for the approximated BSEI.
4) We provide a new formula for approximated evaluation of the (conventional) BSE in consideration of individual distributions without unrealistic assumptions.

II. Biometric System Entropy (BSE)

In this section we overview the definition and properties of the BSE and point out several issues.

A. Modeling of a Biometric Verification System

In the theory of the BSE a biometric verification system S is modeled as a (mathematical) function which takes a pair of biometric information $(\boldsymbol{b}_1, \boldsymbol{b}_2) \in \mathcal{B}^2$ as inputs and outputs a real number $x \in \mathbb{X}$ as a score, i.e.,

$$S: \ \mathcal{B}^2 \to \mathcal{X} \subseteq \mathbb{R} \ (\ S(\boldsymbol{b}_1, \boldsymbol{b}_2) = x \ ), \tag{1}$$

Here the biometric information $\boldsymbol{b}_i \in \mathcal{B}$ may be a "raw data" such as an image or a "feature data" extracted from the row data, and the score $x \in \mathcal{X}$ may be a similarity, a distance or a decision result such as $x = 1$ (OK) and $x = 0$ (NG).

Note that the above model does not require any assumption on the type of biometric data, the structure of the feature space or the feature distribution on the space. Since the BSE is defined for a system modeled above, it does not depend on the internal structure of the system S and thus can be evaluated only based on the distribution of the output score x, so that it enables "black-box evaluation" and can be widely applied to any biometric verification system.

In contrast, most existing attempts to define and evaluate the entropy of biometric information mentioned above limit the feature data to specific types and formats such as a minutia set of a fingerprint or an iris code. Furthermore, these attempts assume that the feature distributions of actual biometric systems can be represented and estimated explicitly, despite it is extremely difficult or impossible in practice.

B. Definition of the BSE

Consider the amount of information for individual identification that can be gained by measuring biometric information $\boldsymbol{b} \in \mathcal{B}$. Let $\mathcal{U} = \{u_1, \cdots, u_N\}$ be the set of users of the biometric authentication system S, and let U be the random variable representing the user ID. Assume that the biometric information (template) $\boldsymbol{b}_i$ of each user $u_i \in \mathcal{U}$ is already enrolled on the system. In the identification phase, the only and best way to identify an unidentified user U based only on his/her biometric information $\boldsymbol{b}$ and the verification system S is to match $\boldsymbol{b}$ and each user's template $\boldsymbol{b}_i$, and make decision based on the score sequence:

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_N), \ x_i = S(\boldsymbol{b}, \boldsymbol{b}_i). \tag{2}$$

Let $\boldsymbol{X}$ be a random variable and $\boldsymbol{x}$ be a realization of $\boldsymbol{X}$. The BSE of a biometric verification system S with respect to the user set $\mathcal{U}$ is defined as follows,

$$BSE(\mathcal{U}, S) = I(U; \boldsymbol{X}). \tag{3}$$

Since $I(U; \boldsymbol{X}) = H(U) - H(U|\boldsymbol{X})$, the BSE defined above can be interpreted as the decrease of ambiguity (entropy) about the identity of the (unidentified) user U caused by observing the score sequence $\boldsymbol{x}$; in other words the BSE can be viewed to represent how much identification information about U was gained through S.

C. Properties of the BSE

The BSE has various interesting properties that can be naturally and intuitively interpreted as the amount of "personal identification information" gained through a biometric verification system. In this section, we briefly overview these properties.

1) Nonnegativity:

Theorem 1.

$$BSE(\mathcal{U}, S) = I(U; \boldsymbol{X}) \geq 0, \tag{4}$$

with equality if and only if U and $\boldsymbol{X}$ are independent.

2) Subadditivity: Let us consider a multimodal biometric system S consisting of two subsystems S^1, S^2 that uses different biometric information $\boldsymbol{b}^1, \boldsymbol{b}^2$ such as a fingerprint and an iris. Such a system S will be represented as follows:

$$S((\boldsymbol{b}^1, \boldsymbol{b}^2), (\boldsymbol{b}_i^1, \boldsymbol{b}_i^2)) = T(S^1(\boldsymbol{b}^1, \boldsymbol{b}_i^1), S^2(\boldsymbol{b}^2, \boldsymbol{b}_i^2))$$

where $T(x^1, x^2)$ is a score fusion function.

As for the BSE of the multimodal biometric system S and the subsystems S^1, S^2, the following theorem holds.

Theorem 2 ([7])**.** *If the two score sequences $\boldsymbol{x}^i = (S^i(\boldsymbol{b}, \boldsymbol{b}_1), \cdots, S^i(\boldsymbol{b}, \boldsymbol{b}_N))$ $(i = 1, 2)$ are statistically independent for each fixed user u (i.e., $p(\boldsymbol{x}^1, \boldsymbol{x}^2|u) = p(\boldsymbol{x}^1|u)p(\boldsymbol{x}^2|u)$), then the following inequality holds:*

$$BSE(\mathcal{U}, S) \leq BSE(\mathcal{U}, S^1) + BSE(\mathcal{U}, S^2), \tag{5}$$

with equality if and only if $\boldsymbol{X}^1$ and $\boldsymbol{X}^2$ are independent and $\boldsymbol{X} = T(\boldsymbol{X}^1, \boldsymbol{X}^2)$ is a sufficient statistic of U.

3) Other Properties: The following properties related to the BSE are also shown in [7].

- If two biometrics verification systems have the same DET curve, they also have the same BSE.
- Let ϵ_U and $\epsilon_{U|\boldsymbol{X}}$ be Bayes error rates with respect to user identification prior and posterior to the observation of $\boldsymbol{X}$. Then the following inequality holds:

  Proposition 1 ([8])**.**

$$\epsilon_{U|\boldsymbol{X}} \geq 1 + \frac{I(U; \boldsymbol{X}) + \log 2}{\log(1 - \epsilon_U)} \tag{6}$$

$$= 1 - \frac{I(U; \boldsymbol{X}) + \log 2}{\log N} \tag{7}$$

(*when U is uniformly distributed*)

D. Asymptotic approximation of the BSE

There are two problems to use the $BSE(\mathcal{U}, S) = I(U; \boldsymbol{X})$ directly as a measure. Firstly, it depends on the set of users $\mathcal{U} \subset \Omega$. Secondly, it is hard to estimate the distribution of the score vector $\boldsymbol{X}$ and to calculate the mutual information $I(U; \boldsymbol{X})$, especially when N is large. To make the measure easy to evaluate, in [7], it is shown under certain assumptions that $BSE(\mathcal{U}, S)$ can be asymptotically approximated by the Kullback-Leibler divergence

$$D(f_G \parallel f_I) = \sum_x f_G(x) \log \frac{f_G(x)}{f_I(x)}, \tag{8}$$

where $f_G(x)$ is the *genuine score distribution* and $f_I(x)$ is the *impostor score distribution* of the system S with respect to the user set $\mathcal{U}$.

Theorem 3. *If the prior distribution U over $\mathcal{U}$ is uniform (i.e., $p(u_i) = 1/|\mathcal{U}|$ for all $u_i \in \mathcal{U}$) and each score $x_i = S(\boldsymbol{b}, \boldsymbol{b_i})$ $(i = 1, 2, \cdots, N)$ follows $f_G(x)$ if $U = u_i$ and follows $f_I(x)$ if $U \neq u_i$ independently, the following convergence holds:*

$$BSE(\mathcal{U}, S) \to D(f_G \parallel f_I) \ (|\mathcal{U}| \to \infty).$$

III. A Measure of Individual Biometric Information

As described above, the BSE is defined for a pair of user set $\mathcal{U}$ and a biometric verification system S. In this section, we provide a new definition of a measure of biometric information with respect to each individual $u_i \in \mathcal{U}$ to be consistent with the definition of BSE.

A. Definition

We define *Biometric System Entropy of Individuals (BSEI)* of a biometric verification system S with respect to a user u_i as follows:

$$\begin{aligned} BSEI(u_i, S) &= D(p(\boldsymbol{X}|u_i) \parallel p(\boldsymbol{X})) \\ &= \sum_{\boldsymbol{x} \in \mathcal{X}^N} p(\boldsymbol{x}|u_i) \log \frac{p(\boldsymbol{x}|u_i)}{p(\boldsymbol{x})}. \end{aligned} \tag{9}$$

B. Properties of the BSEI

The average of the BSEI of a system S with respect to the users $u_i \in \mathcal{U}$ is equal to the BSE with respect to the set $\mathcal{U}$, i.e.,

Proposition 2.

$$\begin{aligned} & E_U[BSEI(U, S)] \\ = & \sum_{u_i \in \mathcal{U}} p(u_i) \sum_{\boldsymbol{x} \in \mathcal{X}^N} p(\boldsymbol{x}|u_i) \log \frac{p(\boldsymbol{x}|u_i)}{p(\boldsymbol{x})} \\ = & \sum_{u_i \in \mathcal{U}} \sum_{\boldsymbol{x} \in \mathcal{X}^N} p(\boldsymbol{x}, u_i) \log \frac{p(\boldsymbol{x}, u_i)}{p(\boldsymbol{x}) p(\boldsymbol{u_i})} \\ = & I(U, \boldsymbol{X}) nonumber \qquad (10) \\ = & BSE(U, S) \qquad (11) \end{aligned}$$

This proposition is obvious from the definition.

The following is another property:

Proposition 3. *If $H(\boldsymbol{X}|U) = 0$, for each $u_i \in U$, there exists only one $\boldsymbol{x}$ satisfying $p(\boldsymbol{x}|u_i) = 1$. Furthermore, if we denote the above $\boldsymbol{x}$ as $\boldsymbol{x}_i$, then $BSEI(u, S) = -\log p(\boldsymbol{x}_i)$.*

This property means that if $\boldsymbol{x}$ is decisional information such as a password of an individual, and not probabilistic (or "fuzzy") data such as a set of fingerprint minutiae or an iris code, then $BSEI(u, S)$ is equal to the self-entropy of $\boldsymbol{x}$. In the concept of *password entropy* [6], the strength of each password can be evaluated as its own self-entropy. In this sense, the BSEI can be interpreted as an extension of the password entropy to probabilistic information such as biometric data.

C. Asymptotic approximation of the BSEI

It is difficult to evaluate the BSEI directly according to the definition (9) in practice since it requires the explicit knowledge of the distribution $p(\boldsymbol{x}|u)$ of score vector $\boldsymbol{x}$ for each user $u \in \mathcal{U}$. To solve this problem, we derive an approximate measure of the BSEI that can be easily evaluated for actual systems.

Let a user set $\mathcal{U}$ of a system S be an arbitrary subset of the user population Ω with cardinality N (i.e., $|\mathcal{U}| = N$). Let $\boldsymbol{x} = (x_1, \cdots, x_N)$ be an output of S when identifying an unknown user U.

We make the following two assumptions:

(i) $p(\boldsymbol{x}|U) = \Pi_{j=1}^N p(x_j|U)$. This means that the scores $x_j = S(\boldsymbol{b}, \boldsymbol{b}_j)$ $(j = 1, 2, \cdots, N)$ are conditionally independent given U.

(ii) Conditional distribution $p(x_j|u_i)$ of the score x_j can be described as follows:

$$p(x_j|u_i) = \begin{cases} f_G^{(i)}(x_j) & (i = j) \\ f_I^{(i)}(x_j) & (i \neq j). \end{cases} \tag{12}$$

(iii) $p(u_i) = 1/N$ $(i = 1, \cdots, N)$.

Note that the above assumption (ii) takes into account the difference of distributions among the users, whereas in [7] it is assumed that the genuine and impostor distributions $f_G(x), f_I(x)$ are identical regardless of the user u_i, i.e., $f_G^{(1)} = \cdots = f_G^{(N)}$ and $f_I^{(1)} = \cdots = f_I^{(N)}$.

Now the following theorem holds:

Theorem 4.

$$BSEI(u_i, S) \to D(f_G^{(i)}(x) \parallel f_I^{(i)}(x)) \ (N \to \infty). \tag{13}$$

(*Proof*)

$$\begin{aligned} BSEI(u_i, S) &= D(p(\boldsymbol{X}|u_i) \parallel p(\boldsymbol{X})) \\ &= \sum_{\boldsymbol{x} \in \mathcal{X}^N} p(\boldsymbol{x}|u_i) \log \frac{p(\boldsymbol{x}|u_i)}{p(\boldsymbol{x})} \\ &= \sum_{\boldsymbol{x} \in \mathcal{X}^N} p(\boldsymbol{x}|u_i) \log \frac{p(\boldsymbol{x}|u_i)}{\sum_{i=1}^N p(\boldsymbol{x}|u_i) p(u_i)} \end{aligned} \tag{14}$$

From the assumption (ii),

$$p(\boldsymbol{x}|u_i) = \prod_{j=1}^N p(x_j|u_i) = \frac{f_G^{(i)}(x_i)}{f_I^{(i)}(x_i)} \prod_{j=1}^N f_I^{(i)}(x_j). \tag{15}$$

Let $F(\boldsymbol{x}) = \prod_{j=1}^N f_I^{(i)}(x_j)$ and $g(x) = \frac{f_G^{(i)}(x_i)}{f_I^{(i)}(x_i)}$. Then from

eq. (14) (15),

$$\begin{aligned} BSEI(u_i,S) &= \sum_{\boldsymbol{x}\in\mathcal{X}^N} F(\boldsymbol{x})g(x_i)\log\frac{F(\boldsymbol{x})g(x_i)}{\sum_j F(\boldsymbol{x})g(x_j)/N} \\ &= \sum_{\boldsymbol{x}\in\mathcal{X}^N} F(\boldsymbol{x})g(x_i)\log\frac{g(x_i)}{\sum_j g(x_j)/N} \\ &= \sum_{\boldsymbol{x}\in\mathcal{X}^N} F(\boldsymbol{x})g(x_i)\log g(x_i) \\ &\quad - \sum_{\boldsymbol{x}\in\mathcal{X}^N} F(\boldsymbol{x})g(x_i)\log\left(\frac{1}{N}\sum_j g(x_j)\right). \end{aligned} \tag{16}$$

Here, the first term of the right side can be expanded as follows:

$$\begin{aligned} &\sum_{\boldsymbol{x}\in\mathcal{X}^N} F(X)g(x_i)\log g(x_i) \\ =& \sum_{x_i\in\mathcal{X}} f_I^{(i)}(x_i)g(x_i)\log g(x_i)\prod_{j\neq i}\sum_{x_j\in\mathcal{X}} f_I^{(i)}(x_j) \\ =& \sum_{x_i\in\mathcal{X}} f_G^{(i)}(x)\log\frac{f_G^{(i)}(x_i)}{f_I^{(i)}(x_i)}\prod_{j\neq i}1 \\ =& \sum_{x\in\mathcal{X}} f_G^{(i)}(x)\log\frac{f_G^{(i)}(x)}{f_I^{(i)}(x)} \\ =& D(f_G^{(i)} \parallel f_I^{(i)}). \end{aligned} \tag{17}$$

On the other hand, if we let

$$y_i = g(x_i),\ \bar{y}_{-i} = \frac{1}{N-1}\sum_{j\neq i}^{N} y_j, \tag{18}$$

$$\boldsymbol{x}_{-i} = (x_1,\cdots,x_{i-1},x_{i+1},\cdots x_N), \tag{19}$$

$$F_{-i}(\boldsymbol{x}) = \prod_{i\neq j} f_I(x_i) \tag{20}$$

then the second term of the right side of eq (16) can be transformed as follows:

$$\begin{aligned} &-\sum_{\boldsymbol{x}\in\mathcal{X}^N} F(\boldsymbol{x})g(x_i)\log\left(\frac{1}{N}\sum_j g(x_j)\right) \\ =& -\sum_{x_i\in\mathcal{X}} f_G(x_i)\sum_{\boldsymbol{x}_{-i}\in\mathcal{X}^{N-1}} F_{-i}(\boldsymbol{x})\log\left(\frac{y_i}{N}+\bar{y}_{-i}\right) \\ =& -E_{f_G(x_i)}\left[E_{F_{-i}(\boldsymbol{x}_{-i})}\left[\log\left(\frac{y_i}{N}+\frac{N-1}{N}\bar{y}_{-i}\right)\right]\right] \end{aligned} \tag{21}$$

From the *law of large numbers*, if x_i follows the distribution $f_I(x)$, then the following stochastic convergence holds:

$$\begin{aligned} \bar{y}_{-i} \xrightarrow{p} E_{f_I(x)}[g(x)] &= \sum_{x\in\mathcal{X}} f_I(x)g(x)dx \\ &= \sum_{x\in\mathcal{X}} f_G(x) = 1\ (N\to\infty). \end{aligned} \tag{22}$$

Therefore the term in the $\log()$ of the right side of eq. (21) also converges stochastically as follows:

$$\frac{y_i}{N} + \frac{N-1}{N}\bar{y}_{-i} \xrightarrow{p} 0+1 = 1\ (N\to\infty). \tag{23}$$

Since $\log y$ is continuous at $y = 1$, from the *continuous mapping theorem*,

$$\log\left(\frac{y_i}{N}+\bar{y}_{-i}\right) \xrightarrow{p} \log 1 = 0\ (N\to\infty), \tag{24}$$

and thus

$$E_{F_{-i}(\boldsymbol{x}_{-i})}\left[\log\left(\frac{y_i}{N}+\bar{y}_{-i}\right)\right] \to 0\ (N\to\infty). \tag{25}$$

Therefore the right side of eq. (21) also converges to 0 as follows:

$$-E_{f_G(x_i)}\left[E_{F_{-i}(\boldsymbol{x}_{-i})}\left[\log\left(\frac{y_i}{N}+\bar{y}_{-i}\right)\right]\right] \to -E_{f_G(x_i)}[0] = 0. \tag{26}$$

From eq. (16) (17) (21) (26),

$$D(p(\boldsymbol{X}|u_i) \parallel p(\boldsymbol{X})) \to D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))\ (N\to\infty). \tag{27}$$

□.

As mentioned before, it is hard to experimentally evaluate the BSEI defined as $D(p(\boldsymbol{X}|u_i) \parallel p(\boldsymbol{X}))$ directly for a certain system S and set $\mathcal{U}$. In contrast, the asymptotic approximation formula $D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))$ of the BSEI can be evaluated experimentally without difficulties, as shown in the following subsection. Thus, we propose to use $D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))$ as an approximation of the BSEI.

D. Evaluation Protocol of the Approximated BSEI

In this section, we provide an evaluation protocol for the approximated BSEI $D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))$ for a system S and a user u_i. Since the true distributions $f_G^{(i)}(x), f_I^{(i)}(x)$ are unknown, we collect samples of biometric information, calculate the scores using S and the sample set, estimate the distributions based on the score set and calculate the divergence $D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))$.

1) Sample Collection and Verification: Firstly, we collect multiple biometric samples (or "shots") $\boldsymbol{x}_i^k$ $(k = 1,\cdots,M)$ of the same biometric characteristic (such as the right index fingerprint) of a user u_i and perform round-robin verification to calculate a genuine score set X_G ($|X_G| = M$). Next, we collect biometric samples $\boldsymbol{x}_j$ $(j = 1,\cdots,N)$ from N users other than u_i, and verify them to each $\boldsymbol{x}_i^k$ to calculate an impostor score set X_I ($|X_I| = MN$). In order to increase the statistical reliability of the evaluation result, it is desirable to make M, N as large as possible. If different biometric characteristics of a user such as index and middle fingerprints can be considered to be independent, we can treat them as of different users.

2) Estimation of Distributions and KL-divergence: Here we describe several methods to estimate the KL-divergence $D(f_G^{(i)}(x) \parallel f_I^{(i)}(x))$ from the sample sets X_G, X_I.

When the score x is discrete, the histogram method can be applied. For example when x is a binary score with $x = 1$ ("matched") and $x = 0$ ("unmatched"), the $f_G^{(i)}(x)$ and $f_I^{(i)}(x)$ are Bernoulli distributions and can be estimated as follows:

$$\tilde{f}_G^{(i)}(1) = p,\ f_G(0) = 1 - p,$$
$$\tilde{f}_I^{(i)}(1) = q,\ f_I(0) = 1 - q,$$

where $p = n_G/|X_G|$, $q = n_I/|X_I|$ and n_G, n_I are the numbers of 1 contained in X_G, X_I respectively. In this case the KL-divergence is calculated as follows:

$$\begin{aligned} D(\tilde{f}_G^{(i)} \parallel \tilde{f}_I^{(i)}) &= \sum_{x=0,1} \tilde{f}_G^{(i)}(x) \log \frac{\tilde{f}_G^{(i)}(x)}{\tilde{f}_I^{(i)}(x)} \\ &= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \end{aligned} \quad (28)$$

On the other hand, if x is continuous, the KL-divergence estimator can be classified into two approaches, i.e., parametric and non-parametric methods.

A simple example of the parametric method is to assume $f_G^{(i)}(x)$ and $f_I^{(i)}(x)$ to follow parametric distribution models (such as Gaussian distributions) and estimate the parameters via e.g., maximum likelihood estimation. Then the KL-divergence can be numerically calculated. Although this method is simple and easy to implement, the statistical reliability of the results may be degraded when the true distribution of $f_G^{(i)}(x), f_I^{(i)}(x)$ largely deviate from the models.

A representative non-parametric KL-divergence estimator is the *Nearest Neighbor (NN) estimator*. For example, the following k-NN estimator introduced by Wang et.al., [9] is known to be asymptotically unbiased and consistent.

$$\begin{aligned} \hat{D}(f_G \parallel f_I) = \quad & \frac{1}{N} \sum_{i=1}^{N} \left\{ \log \frac{\nu_{k_i}(i)}{\rho_{l_i}(i)} + \psi(l_i) - \psi(k_i) \right\} \\ & + \log \frac{M}{N-1}, \end{aligned} \quad (29)$$

where $\rho_{l_i}(i)$ ($\nu_{k_i}(i)$) denotes the distance (i.e., absolute value of difference) from x_i to the l_i-th (k_i-th) nearest neighbor of the set $X_G \setminus \{x_i\}$ (X_I). l_i (k_i) denotes the number of values contained in $X_G \setminus \{x_i\}$ (X_I) and the interval $[x_i - \epsilon, x_i + \epsilon]$ where $\epsilon = \max\{\rho_1(i), \nu_1(i)\}$. Here $\psi(n)$ is the Digamma function [10] defined as follows:

$$\psi(n) = -\gamma + \sum_{k=1}^{n-1} \frac{1}{k}, \quad (30)$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

There are advantages and disadvantages to these estimation methods, and it is necessary to choose an appropriate method depending on the size and statistical properties of the score sets.

IV. Approximated Evaluation of the BSE in Consideration of Individual Distributions

In this section, based on the discussion so far, we derived a new approximate evaluation measure of the BSE considering individual score distributions.

As described in Sec. II-D, the previous work requires a strong assumptions that the genuine (impostor) distribution $f_G(x)$ ($f_I(x)$) is identical regardless of the users u_i, is required to derive the approximate expression of the BSE in [7]. Let (A1) be the above (conventional) assumption.

On the other hand, in this paper, we relaxed the assumption to more realistic and general one as eq. (12) that means the genuine (impostor) distribution $f_G^{(i)}(x)$ ($f_I^{(i)}(x)$) can be different for each user u_i. Let (A2) be the above (new) assumption.

Since the average of the BSEI with respect to the users $u_i \in \Omega$ is equal to the BSE with respect to the set $\mathcal{U}$ (Proposition 2), we can say that the average of the approximated BSE,

$$\sum_i \frac{1}{N} D(f_G^{(i)} \parallel f_I^{(i)}), \quad (31)$$

is a better approximation of the BSE than the conventional one (i.e., $D(f_G \parallel f_I)$) in the sense that it is based on the more realistic and general assumption (A2) instead of (A1).

In the followings, we discuss the relationship between the conventional approximations of the BSE and the proposed one (eq. (31)).

In the evaluation protocol according to the conventional approximation, we collect biometric samples from N users, perform round-robin verification to calculate genuine and impostor score sets, and then estimate the KL-divergence $D(\tilde{f}_G \parallel \tilde{f}_I)$ under the assumption (A1).

In the case of new approximation (eq. (31)), the evaluation protocol is the same as conventional one until calculating the score sets. However, unlike the conventional protocol, the KL-divergence $D(\tilde{f}_G^{(i)} \parallel \tilde{f}_I^{(i)})$ is estimated under the assumption (A2) and the approximated BSE is calculated as the average of the KL-divergence.

Notice that which of (A1) and (A2) (or neither) is established does not matter to perform evaluation. The difference between the two protocols is whether to define the score distributions of each user as the average distributions over the user set or as individual distributions. It will be clear that the following relationship holds for the two ways of definition: [1]

$$f_G(x) = \frac{1}{N} \sum_i f_G^{(i)}(x),\ f_I(x) = \frac{1}{N} \sum_i f_I^{(i)}(x). \quad (32)$$

Then, the following inequality holds between the two approximations:

Theorem 5.

$$\frac{1}{N} \sum_i D(f_G^{(i)} \parallel f_I^{(i)}) \geq D(f_G \parallel f_I). \quad (33)$$

[1] Although this equation does not necessarily holds for estimated distributions, the error between both sides will become smaller as the estimation accuracy of the distributions increases.

(*Proof*) From the *log sum inequality*,

$$\frac{1}{N}\sum_{i=1}^{N} f_G^{(i)}(x)\log\frac{f_G^{(i)}(x)}{f_I^{(i)}(x)} \quad \geq \quad f_G(x)\log\frac{f_G(x)}{f_I(x)}. \quad (34)$$

By taking summation of both sides of the above inequality with respect to x, we can obtain eq.(33). □.

The theorem suggests that it will be possible to extract more information for personal identification from biometrics by designing the decision logic considering the score distribution of each user rather than of common one. In addition, we can see that the theorem explains the reason, from the viewpoint of information entropy, why normalization of a score by considering the distribution for each user improve the total accuracy, as empirically known to be effective [12].

V. Conclusion

In this paper, we extended and refine the theory of the BSE (Biometric System Entropy) which is an information theoretic measure of personal identification performance of a biometric verification system S. Specifically, we newly defined the BSEI (Biometric System Entropy of Individuals) representing the entropy of individual biometric information whereas the BSE represents the average entropy over the user set $\mathcal{U}$, and showed several properties. We also proved that the BSEI asymptotically converges to $D(f_G^{(i)} \parallel f_I^{(i)})$ under some natural assumption so that it can be used as an approximate evaluation measure of the BSEI, and proposed a practical protocol to evaluate the BSE of a system S experimentally. Finally, we derived a more sophisticated approximate evaluation measure of the BSE considering individual score distributions.

Experimental evaluation of the BSEI for an actual biometric verification system and individual biometric information is one of the future works. In addition, further studies are needed in order to investigate the optimal decision algorithm which derives information for identification from various kind of biometric data to the limit.

References

[1] A. J. Mansfield and J. L. Wayman, "Best Practices in Testing and Reporting Performance of Biometric Devices," National Physical Laboratory, Center for Mathematics and Scientific Computing, Tech. Rep., Version 2.01, 2002.

[2] S. Pankanti, S. Prabhakar and A. K. Jain, "On the individuality of Fingerprints," IEEE Trans. on PAMI, Vol. 24, No. 8, pp. 1010–1025, 2002.

[3] Daugman, J., "The importance of being random: Statistical principles of iris recognition," Pattern Recognition, Vol. 36, No. 2, pp. 279-291, 2003.

[4] A. Adler, R. Youmaran, and S. Loyka, "Towards a measure of biometric information," in Proc. Can. Conf. Comp. Elec. Eng. (CCECE), 2006.

[5] E. Henry, "Classification and Uses of Fingerprints," Routledge, London, 1900.

[6] W. Burr, D. Dodson, and W. Polk, "Electronic authentication guideline," NIST Special Publication 800-63, 2004.

[7] K. Takahashi and T. Murakami, "A Measure of Information Gained through Biometric Systems," Elsevier Image and Vision Computing, Vol.32, No.12, pp.1194-1203, 2014.

[8] Te Sun Han and Sergio Verdú, "Generalizing the Fano inequality," IEEE Transactions on Information Theory, Vo.40, No.4, pp.1247-1251, 1994.

[9] Q.Wang, S. Kulkarni and S. Verdu, "Divergence estimation for multi-dimensional densities via k-nearest-neighbor distances," IEEE International Symposium on Information Theory (ISIT2009), 2009.

[10] C. M. Bishop, "Pattern Recognition and Machine Learning," Springer, 2006.

[11] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004.

[12] T. Murakami, K. Takahashi, and K. Matsuura, "Toward Optimal Fusion Algorithms with Security against Wolves and Lambs in Biometrics," IEEE Transactions on Information Forensics and Security, Vol.9, No.2, pp.259-271, 2014.