L-024

虹彩情報の Renyi エントロピー推定に関する一考察
A Study on Estimation of Renyi Entropy of Iris Information

披田野 清良 * **赤尾 直彦** * **市野 将嗣** † **小松 尚久** * **高橋 健太** ‡
Seira HIDANO Naohiko AKAO Masatsugu ICHINO Naohisa KOMATSU Kenta TAKAHASHI

1. まえがき

近年，生体認証の識別性能や安全性を生体情報の情報量に基づき評価する試みが注目されており，筆者らは 2 次の Renyi エントロピー（以下，Renyi エントロピー）を用いた生体情報の情報量評価手法を提案している [1]．Renyi エントロピーは 2 つの生体情報が一致する可能性を情報量で表現した尺度であり，他人間照合実験を通して得られる生体情報間の距離分布から導出できる．しかし，これまでの検討において，距離分布の推定方法については必ずしも十分に議論されていない．そこで，本稿では，異なる複数の推定手法を用いて虹彩情報の Renyi エントロピーを定量的に評価し，推定手法の相違による評価結果の差異について考察する．

2. 生体情報の情報量評価手法

Renyi エントロピーを用いた生体情報の情報量評価手法について述べる．生体情報 B の Renyi エントロピー $H_2(B)$ は，B の取り得る値の集合を $\mathcal{B}$，確率関数を $p_B(b), b \in \mathcal{B}$ とすると，次式で表される．

$$H_2(B) = -\log_2 \sum_{b \in \mathcal{B}} p_B(b)^2 \tag{1}$$

(1) 式の $\sum_{b \in \mathcal{B}} p_B(b)^2$ は 2 つの生体情報が同一の値を取る確率を示しており，このとき，それらの間の距離は 0 となる．このため，$H_2(B)$ は，生体情報間の距離 D の確率関数 $p_D(d), d \in \mathbb{R}$ を用いて，次式で表される．

$$H_2(B) = -\log_2 p_D(0) \tag{2}$$

$p_D(0)$ は $p_D(d)$ が既知であれば容易に導出できる．Daugman の虹彩認証モデル [2] のように $p_D(d)$ の形状が十分に検討された生体情報であれば，$p_D(d)$ は D のサンプルを用いてパラメトリックに推定する．一方，$p_D(d)$ の形状が未知でモデル化が困難な場合は，分布の形状を仮定しないノンパラメトリックな手法を用いて推定する．

3. 二項分布を用いた Renyi エントロピー評価

虹彩情報 C の Renyi エントロピー $H_2(C)$ を評価するに際して，本稿では，Daugman の虹彩認証モデル [2] を採用する．C は虹彩画像より生成可能な虹彩コード $\{0,1\}^n$ で記述する．また，虹彩は瞼や睫毛，光の反射などの影響を受けるため，これらの環境要因の影響の有無を示すマスクコード W を用いる．W の各ビットは，対応する C のビットに環境要因の影響がある場合は 0，それ以外は 1 を返す．2 つの虹彩情報 C と C' の間の距離 D_C は，それぞれのマスクコード W, W' を用いて，次式で表される．

$$D_C = \frac{g_{HD}(C \cap W \cap W', C' \cap W \cap W')}{\| W \cap W' \|} \tag{3}$$

ただし，$g_{HD}(C, C')$ は C と C' の間のハミング距離を示し，$\| \cdot \|$ は 1 を返すビットの個数を示す．また，Daugman の虹彩認証モデルでは，虹彩領域を回転させながら複数の異なる C を生成し，それらの中で D_C が最小となる場合の結果を照合スコア D_C^* として採用している．以下，回転補正を行う照合をベストマッチ，補正を行わない照合を非ベストマッチと呼ぶ．

非ベストマッチの場合，D_C の確率関数 $p_{D_C}(d)$ は，C の各ビットの一致確率を θ，C の中で識別に有効なビット数を $\hat{n}$ と仮定した場合，次式に示す二項分布でモデル化される．

$$p_{D_C}(d) = \frac{\hat{n}!}{(\hat{n}d)!(\hat{n}(1-d))!} \theta^{\hat{n}(1-d)} (1-\theta)^{\hat{n}d} \tag{4}$$

このとき，D_C の期待値 $E(D_C)$ と分散 $V(D_C)$ はそれぞれ次のように表される．

$$E(D_C) = 1 - \theta \tag{5}$$
$$V(D_C) = \theta(1-\theta)/\hat{n} \tag{6}$$

(4) 式より，2 つの生体情報が一致する確率は $p_{D_C}(0) = \theta^{\hat{n}}$ となり，$H_2(C)$ は次式で表される．

$$H_2(C) = -\log_2 \theta^{\hat{n}} \tag{7}$$

$\hat{n}$ は，環境要因の影響やビット間の相関により，C のコード長 n とは一致せずに大きく減少すると考えられる．そこで，C のサンプルを用いた他人間照合実験を通して得られる D_C の平均および分散，(5) 式，(6) 式より推定する．

一方，ベストマッチの場合の距離 D_C^* の確率関数 $p_{D_C^*}(d)$ は，回転補正による回転数を r とすると，$p_{D_C}(d)$ を用いて，次式で表される [2]．

$$p_{D_C^*}(d) = r p_{D_C}(d) \left[1 - P_{D_C}(d)\right]^{r-1} \tag{8}$$

ただし，$P_{D_C}(d)$ は $p_{D_C}(d)$ の累積分布を示す．しかし，(8) 式でモデル化した場合，$p_{D_C^*}(d)$ は D_C^* のサンプルから得られる統計量を用いて直接パラメトリックに推定することができない．そこで，本章では，ベストマッチの場合も非ベストマッチの場合と同様に，$p_{D_C^*}(d)$ を (4) 式の二項分布でモデル化し，$H_2(C)$ は (7) 式より導出する．

以下，実際に虹彩画像データベースを用いて他人間照合実験を行い，虹彩情報の Renyi エントロピーを定

*早稲田大学, Waseda University
†電気通信大学, The University of Electro-Communications
‡東京大学, The University of Tokyo

量的に評価した結果について述べる．本実験では，虹彩データベースとして，CASIA-IrisV3 に収録さている Lamp を使用した．Lamp は，異なる照明状況下で撮影された虹彩画像を収録しており，他人間照合実験では，異なる虹彩から 2 つの画像をランダムに選択し，100,000 回の照合を行った．ただし，虹彩コードの生成には Telecom Management Sud Paris が公開している OSIRIS version 2.01 を使用し，照合はベストマッチの場合のみ OSIRIS を使用した．

非ベストマッチの場合，D_C の平均は 0.463，分散は 0.0013 であった．(5) 式，(6) 式より，θ の推定値は 0.537，$\hat{n}$ の推定値は 191 となり，$H_2(C)$ は 172 bit であった．このとき，(8) 式より，ベストマッチの場合の $H_2(C)$ の理論値は 169 bit となる．一方，実際にベストマッチで照合した場合，D_C^* の平均は 0.438，分散は 0.0012 であった．(5) 式，(6) 式より，θ の推定値は 0.562，$\hat{n}$ の推定値は 212 となり，$H_2(C)$ は 177 bit であった．

4. ノンパラメトリックな手法を用いた Renyi エントロピー評価

本章では，ベストマッチにより照合を行った際の距離 D_C^* の確率関数 $p_{D_C^*}(d)$ を核関数に基づくノンパラメトリックな手法を用いて推定し，虹彩情報 C の Renyi エントロピー $H_2(C)$ を評価する．

まず，核関数として，$p_{D_C^*}(0)$ において推定誤差を抑えることができる特殊な関数 Discrete Triangular Kernel (以下，DTK)[3] を採用する．DTK $K_{N,h,a}(x)$ は，D_C^* の N 個のサンプルのヒストグラムにおける階級値をそれぞれ X_i $(i = 1, \ldots, N)$，階級値の取り得る値を $x \in \mathbb{R}$ とすると，次式で表される．

$$K_{N,h,a}(x) = \frac{(a+1)^h - |X_i - x|^h}{(2a+1)(a+1)^h - \sum_{k=0}^{a} k^h} \quad (9)$$

ただし，$h \in \mathbb{R}, h > 0$ は $K_{N,h,a}(x)$ の平滑化パラメータとし，$a \in \mathbb{N}$ は $K_{N,h,a}(x)$ の x からの広がりを示すパラメータとする．また，a は，$x = 0$ における推定誤差を抑えるために，次のように x に応じて値を変化させる．

$$a = \begin{cases} j & \text{if } x = j,\ j \in \{0, \ldots, k-1\} \\ k & \text{if } x \in \{k, k+1, \ldots\} \end{cases} \quad (10)$$

X の確率関数 $p_X(x)$ は，$K_{N,h,a}(x)$ を用いて，次式で表される．

$$p_X(x) = \frac{1}{N} \sum_{i=1}^{N} K_{N,h,a}(x) \quad (11)$$

以下，DTK に基づくノンパラメトリックな手法を用いて $H_2(C)$ を定量的に評価した結果について述べる．ただし，虹彩データベースおよび虹彩認証アルゴリズムは 3 章と同じものを使用した．また，h, a は，D_C^* のサンプルを用いた Cross-Validation 法により決定しており，それぞれ $a = 5, h = 0.987$ であった．図 1 に，D_C^* の実験値と推定分布を示す．図 1 より，ノンパラメトリックな手法を用いて推定した距離分布が実験値に近似していることが分かる．このとき，$H_2(C)$ は 15 bit であった．

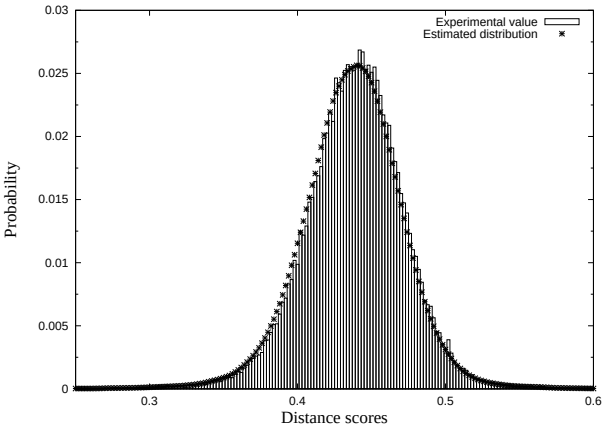

図 1: ノンパラメトリックな手法による推定距離分布

5. 推定手法間の比較

ベストマッチの場合の虹彩情報の Renyi エントロピー $H_2(C)$ は，3 章にて二項分布を用いた際に 177 bit，4 章にてノンパラメトリックな手法を用いた際に 15 bit であったことから，(8) 式より理論的に導出した 169 bit と比較すると，二項分布を用いた方が理論値と近い値を取ることが分かる．一方，ノンパラメトリックな手法を用いた場合，$H_2(C)$ を過小に評価する可能性がある．また，筆者らは，文献 [1] において，正規分布を用いた場合についも言及しており，$H_2(C)$ は 125 bit であった．したがって，この場合も，二項分布を用いた場合に比べて，$H_2(C)$ を過小に評価する可能性がある．

以上より，非ベストマッチの場合の Daugman の二項分布による距離分布のモデル化が十分に正しいと仮定するならば，ベストマッチの場合も，$H_2(C)$ は，距離分布を二項分布でモデル化することにより，正確に評価できる可能性があると言える．

6. まとめと今後の課題

本稿では，パラメトリックな手法およびノンパラメトリックな手法を用いて，虹彩情報の Renyi エントロピーを定量的に評価した．その結果，虹彩情報間の距離分布を二項分布でモデル化した際に，虹彩情報の Renyi エントロピーを正確に評価できる可能性があるという知見を得た．今後は，Renyi エントロピーを用いた虹彩情報の Shannon エントロピー推定の妥当性を検討する．

参考文献

[1] 赤尾 直彦，披田野 清良，小松 尚久，“ 最小距離エントロピーを用いた虹彩情報の情報量推定に関する一考察，”2011 年暗号と情報セキュリティシンポジウム予稿集，2011.

[2] J. Daugman,“ The Importance of Being Random: statistical principles of iris recognition, ”Pattern Recognition, Vol.36, No.2, pp.279-291, 2003.

[3] C.C. Kokonendji, T.S. Kiesse and S.S. Zocchi,“ Discrete triangular distribution and nonparametric estimation for probability mass function, ”Journal of Nonparametric Statistics, Vol.19, pp.241-254, 2007.