Chapter 17
Towards Measuring the Amount of Discriminatory Information in Finger Vein Biometric Characteristics Using a Relative Entropy Estimator

Vedrana Krivokuća, Marta Gomez-Barrero, Sébastien Marcel, Christian Rathgeb and Christoph Busch

Abstract This chapter makes the first attempt to quantify the amount of discriminatory information in finger vein biometric characteristics in terms of Relative Entropy (RE) calculated on genuine and impostor comparison scores using a Nearest Neighbour (NN) estimator. Our findings indicate that the RE is system-specific, meaning that it would be misleading to claim a universal finger vein RE estimate. We show, however, that the RE can be used to rank finger vein recognition systems (tested on the same database using the same experimental protocol) in terms of their expected recognition accuracy, and that this ranking is equivalent to that achieved using the EER. This implies that the RE estimator is a reliable indicator of the amount of discriminatory information in a finger vein recognition system. We also propose a Normalised Relative Entropy (NRE) metric to help us better understand the significance of the RE values, as well as to enable a fair benchmark of different biometric systems (tested on different databases and potentially using different experimental protocols) in terms of their RE. We discuss how the proposed NRE metric can be used as a complement to the EER in benchmarking the discriminative capabilities of different biometric systems, and we consider two potential issues that must be taken into account when calculating the RE and NRE in practice.

V. Krivokuća (✉) · S. Marcel
Idiap Research Institute, Martigny, Switzerland
e-mail: vedrana.krivokuca@idiap.ch

S. Marcel
e-mail: sebastien.marcel@idiap.ch

M. Gomez-Barrero · C. Rathgeb · C. Busch
Hochschule Darmstadt, Darmstadt, Germany
e-mail: marta.gomez-barrero@h-da.de

C. Rathgeb
e-mail: christian.rathgeb@h-da.de

C. Busch
e-mail: christoph.busch@h-da.de

507

Keywords Finger veins · Relative entropy · Nearest neighbour estimator ·
Biometric template protection · Security · Privacy · Discriminatory information ·
Kullback–Leibler divergence · VERA · UTFVP · Wide Line Detector · Repeated
Line Tracking · Maximum Curvature

17.1 Introduction

There is no doubt that biometrics are fast becoming ubiquitous in response to a grow-
ing need for more robust identity assurance. A negative consequence of this increasing
reliance on biometrics is the looming threat of serious privacy and security concerns
in the event that the growing biometric databases are breached.[1] Fortunately, the past
decade has seen notable efforts in advancing the field of biometric template protec-
tion, which is dedicated to protecting the biometric data that is collected and used
for recognition purposes, thereby safeguarding the privacy of the data subjects and
preventing "spoofing" attacks using stolen biometric templates. Unfortunately, we
are still lacking solid methods for evaluating the effectiveness of the proposed solu-
tions. An important missing ingredient is a measure of the amount of discriminatory
information in a biometric system.

A few approaches, for example, [1–3], have focused on estimating the "individu-
ality" (or discrimination capability) of biometric templates in terms of the inter-class
variation alone (i.e. the False Match Rate or False Accept Rate). Along the same
lines, the best-known attempt to measure the amount of information in a biometric
system is probably the approach proposed by Daugman [4]. This method computes
the Hamming distance between every pair of non-mated IrisCodes, and the resulting
distance distribution is then fitted to a binomial distribution. The number of degrees
of freedom of the representative binomial distribution approximates the number of
independent bits in each binary IrisCode, which in turn provides an estimate for the
discrimination entropy of the underlying biometric characteristic. This approach was
adopted to measure the entropy of finger vein patterns in [5]. However, as explained
in [5], while this method of measuring entropy is correct from the source coding point
of view, the issue with calculating the entropy in this way is that it only provides
a reasonable estimate of the amount of biometric information if there is no vari-
ation between multiple samples captured from the same biometric instance. Since
this intra-class variation is unlikely to be zero in practice, the discrimination entropy
would probably overestimate the amount of available biometric information [6, 7].

In an attempt to extend the idea of using entropy as a measure of biometric
information while more practically incorporating both inter- and intra-class variation,
several authors have adopted the *relative entropy* approach. Adler et al. [8] defined
the term "biometric information" as *the decrease in uncertainty about the identity*

---

[1]For a real-life example, see: http://money.cnn.com/2015/09/23/technology/opm-fingerprint-hack.

of a person due to a set of biometric measurements. They proposed estimating the biometric information via the *relative entropy* or *Kullback–Leibler (KL) Divergence* between the intra-class and inter-class biometric feature distributions. Takahashi and Murakami [6] adopted a similar approach to [8], except that they used *comparison score* distributions instead of *feature* distributions, since this ensures that the whole recognition pipeline is considered when estimating the amount of discriminative biometric information in the system. Around the same time, Sutcu et al. [9] adopted the same method as that employed in [6], with an important difference: they used a Nearest Neighbour (NN) estimator for the KL divergence, thereby removing the need to establish models for the comparison score distributions prior to computing the relative entropy.

This paper adopts the approach proposed in [9] to estimate the amount of discriminatory information in finger vein biometrics. We show that the Relative Entropy (RE) metric is equivalent to the Equal Error Rate (EER) in terms of enabling us to rank finger vein biometric systems according to their expected recognition accuracy. This suggests that the RE metric can provide a reliable estimation of the amount of discriminatory information in finger vein recognition systems. We additionally propose a Normalised Relative Entropy (NRE) metric to help us gain a more intuitive understanding of the significance of RE values and to allow us to fairly benchmark the REs of different biometric systems. The new metric can be used in conjunction with the EER to determine the best-performing biometric system.

The remainder of this chapter is structured as follows. Section 17.2 explains the adopted RE metric in more detail. Section 17.3 presents our results for the RE of finger vein patterns and shows how this metric can be used to rank finger vein recognition systems in comparison with the EER. Section 17.4 proposes the new NRE metric and presents NRE results on various finger vein recognition systems. Section 17.5 discusses how the NRE could be a useful complement to the EER in benchmarking the discrimination capabilities of different biometric systems, and we also present two issues that must be considered when calculating the RE and NRE in practice. Section 17.6 concludes this chapter and proposes a primary direction for future work.

17.2   Measuring Biometric Information via Relative Entropy

Let us say that $G(x)$ represents the probability distribution of *genuine* (mated) comparison scores in a biometric recognition system, and $I(x)$ represents the probability distribution of *impostor* (non-mated) comparison scores. The RE between these two distributions is then defined in terms of the KL divergence as follows:

$$D(G||I) = \sum_{i=1}^{n} G(x_i) \log_2 \frac{G(x_i)}{I(x_i)} \tag{17.1}$$
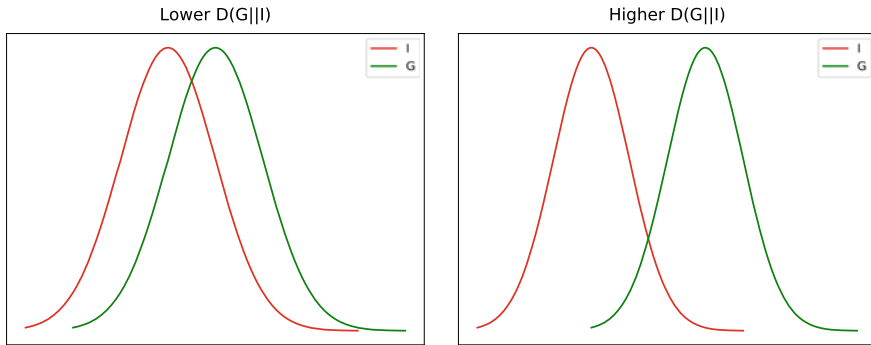
Fig. 17.1 Examples of G and I relationships producing lower and higher $D(G||I)$ values

In information-theoretic terms, $D(G||I)$ tells us the number of extra bits that we would need to encode samples from G when using a code based on I, compared to simply using a code based on G itself. Relating this to our biometric system, we can think of $D(G||I)$ as providing some indication of how closely our genuine score distribution corresponds to our impostor score distribution. The worse the match, the *higher* the $D(G||I)$ value and the easier it is to tell the two distributions apart. Consequently, the higher the RE, the easier it should be for our biometric recognition system to differentiate between genuine users and impostors based on their corresponding comparison scores, and thus the better the expected recognition accuracy. Figure 17.1 shows a simple illustration of what the relationship between G and I might look like for lower and higher $D(G||I)$ values.

One issue with using Eq. (17.1) to estimate the RE is evident when we consider what is represented by n. Technically, n is meant to denote the total number of comparison scores, and it is expected that the G and I distributions extend over the same range of scores. This, however, is not usually the case, since the overlap between the two distributions should only be partial. One consequence of this is that we will have at least one division by 0, for the range where $I(x) = 0$ but $G(x) \neq 0$. The result will be $D(G||I) = \infty$. This makes sense theoretically, since if a score does not exist in I then it is impossible to represent it using a code based on I. For our purposes, however, an RE of ∞ does not tell us much, since we already expect only partial overlap between G and I. So, we would like our RE metric to generate a finite number to represent the amount of information in our biometric recognition system.

Another issue with Eq. (17.1) is that this approach requires us to produce models for the genuine and impostor score distributions, G and I. Since the number of scores we have access to is generally not very large (this is particularly likely to be the case for genuine scores), it may be difficult to generate accurate models for the underlying score distributions.

In light of the issues mentioned above, Sutcu et al. [9] proposed approximating the RE using the NN estimator from [10]. Let $s_g^1, \ldots, s_g^{N_g}$ and $s_i^1, \ldots, s_i^{N_i}$ represent the comparison scores from the sets of genuine and impostor scores, respectively. Further,

let $d_{gg}(i) = \min_{j \neq i} ||s_g^i - s_g^j||$ represent the distance between the genuine score s_g^i and its nearest neighbour in the set of genuine scores, and let $d_{gi}(i) = \min_j ||s_g^i - s_i^j||$ denote the distance between the genuine score s_g^i and its nearest neighbour in the set of impostor scores. Then the NN estimator of the KL divergence is defined as

$$\hat{D}(G||I) = \frac{1}{N_g} \sum_{i=1}^{N_g} \log_2 \frac{d_{gi}(i)}{d_{gg}(i)} + \log_2 \frac{N_i}{N_g - 1} \tag{17.2}$$

Using Eq. (17.2), we can estimate the RE of a biometric system using the genuine and impostor comparison scores directly, without establishing models for the underlying probability densities. Moreover, using the proposed KL divergence estimator, we can circumvent the issue of not having complete overlap between the genuine and impostor score distributions. For these reasons, this is the approach we adopted to estimate the amount of information in finger vein patterns.

17.3 Relative Entropy of Finger Vein Patterns

We used the NN estimator approach from [9] to estimate the RE of finger vein patterns.[2] Section 17.3.1 describes our adopted finger vein recognition systems, and Sect. 17.3.2 presents our RE results for finger vein patterns.

17.3.1 Finger Vein Recognition Systems

We used two public finger vein databases for our investigation: VERA[3] [11] and UTFVP[4] [12]. VERA consists of two images for each of 110 data subjects' left and right index fingers, which makes up 440 samples in total. UTFVP consists of four images for each of 60 data subjects' left and right index, ring and middle fingers, which makes up 1,440 samples in total. Both databases were captured using the same imaging device, but with slightly different acquisition conditions. Figure 17.2 shows an example of a finger image from each database.
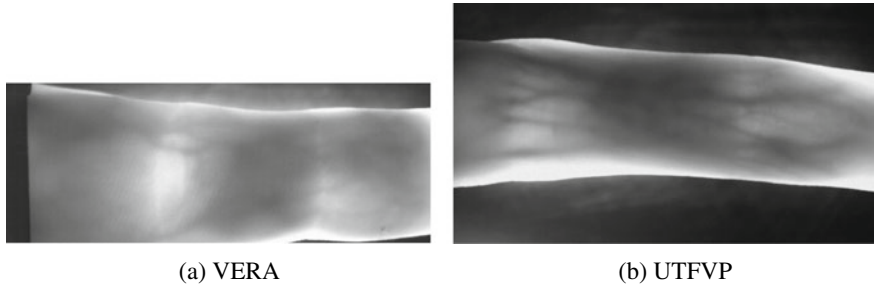
Finger vein patterns were extracted and compared using the `bob.bio.vein` PyPI package.[5] To extract the vein patterns from the finger images in each database, the fingers were first cropped and horizontally aligned as per [13, 14]. Next, the finger vein pattern was extracted from the cropped finger images using three well-known

---

[2]Code available at https://gitlab.idiap.ch/bob/bob.chapter.fingerveins_relative_entropy.

[3]https://www.idiap.ch/dataset/vera-fingervein.

[4]http://scs.ewi.utwente.nl/downloads/show,Finger%20Vein/.

[5]https://pypi.python.org/pypi/bob.bio.vein.

(a) VERA   (b) UTFVP

Fig. 17.2 Examples of finger images from the VERA and UTFVP databases. Note that the UTFVP images are larger in size, as shown in this figure

feature extractors: Wide Line Detector (WLD) [14], Repeated Line Tracking (RLT) [15] and Maximum Curvature (MC) [16].

The comparison between the extracted finger vein patterns was performed separately for each extractor, using the algorithm proposed in [15]. This method is based on a cross-correlation between the enrolled finger vein template and the probe template obtained during verification. The resulting comparison scores lie in the range [0, 0.5], where 0.5 represents maximum cross-correlation and thus a perfect match.

### 17.3.2 Relative Entropy of Finger Veins

We used Eq. (17.2) to calculate the RE of finger vein patterns[6] for each of the three feature extractors (WLD, RLT, and MC) on both the VERA and UTFVP databases. One issue we faced when implementing this equation was dealing with the case where the $d_{gg}(i)$ and/or $d_{gi}(i)$ terms were zero. If $d_{gi}(i) = 0$ (regardless of what value $d_{gg}(i)$ takes), this would result in $\hat{D}(G||I) = -\infty$, whereas $d_{gg}(i) = 0$ (regardless of what value $d_{gi}(i)$ takes) would result in $\hat{D}(G||I) = \infty$. This is one of the issues we wanted to circumvent by using the NN estimator in the first place! Neither the paper that proposed the NN estimator for KL divergence [10], nor the paper that proposed using this estimator to calculate the RE of biometrics [9], suggests how to proceed in this scenario. So, we decided to add a small value (ϵ) of 10^{-10} to every $d_{gg}(i)$ and $d_{gi}(i)$ term that turned out to be 0. The choice of ϵ was based on the fact that our comparison scores are rounded to 8 decimal places, so we wanted to ensure that ϵ would be smaller than 10^{-8} to minimise the impact on the original score distribution.[7]

---

[6]Note: $RE = \hat{D}(G||I)$.

[7]This choice of ϵ may not necessarily be optimal, but it seems sensible.

Table 17.1 Relative Entropy (RE) and Equal Error Rate (EER) for different extractors on the VERA and UTFVP databases. The RE and EER ranks refer to the rankings of the three extractors (separately for each database) in terms of the highest RE and lowest EER, respectively

DB	Extractor	RE	EER (%)	RE rank	EER rank
VERA	WLD	11.8	9.5	2	2
VERA	RLT	4.2	24.3	3	3
VERA	MC	13.2	4.3	1	1
UTFVP	WLD	18.9	2.7	2	2
UTFVP	RLT	18.0	3.2	3	3
UTFVP	MC	19.5	0.8	1	1
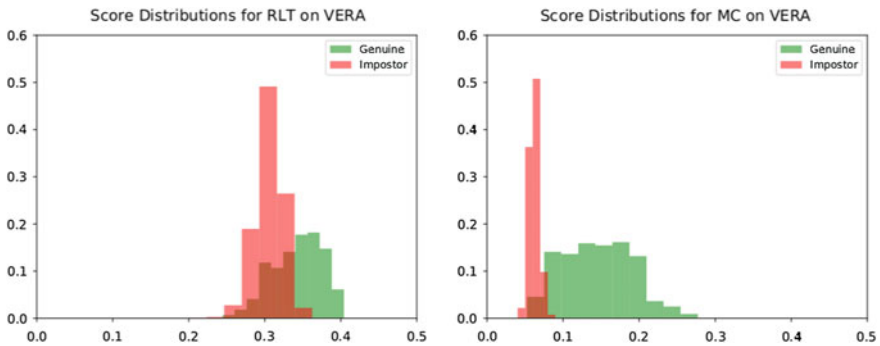
Fig. 17.3 Genuine and impostor score distributions corresponding to the lowest (left) and highest (right) RE values for the VERA database from Table 17.1

For this experiment, a comparison score was calculated between a finger vein template and *every other* finger vein template in the database. The resulting RE values are summarised in Table 17.1, along with the corresponding EERs.[8]

We can interpret the RE results in Table 17.1 as providing an indication of how many bits of discriminatory information are contained in a particular finger vein recognition system. For example, we can see that using the RLT extractor on the VERA database results in a system with only 4.2 bits of discriminatory information, while the MC extractor on the same database contains 13.2 bits of discriminatory information. Figure 17.3 illustrates the genuine and impostor score distributions for these two RE results.

Since our results show the RE to be dependent upon both the feature extractor and database adopted, it would be misleading to claim a universal finger vein RE estimate; rather, it makes more sense for the RE to be system-specific.

---

[8]Note that we have chosen to compare the RE to the EER, because the EER is a widely used metric for evaluating the overall recognition accuracy (in terms of the trade-off between the False Match Rate (FMR) and False Non-Match Rate (FNMR)) of a biometric recognition system. The comparison seems appropriate, since RE aims to provide us with an idea of a biometric system's overall discrimination capability.

Intuitively, we can see that, the higher the RE, the greater the amount of discriminatory information, and thus the greater the expected recognition capabilities of the underlying system. This intuition is confirmed when we compare the REs and EERs of the different systems in Table 17.1, in terms of the RE-based versus EER-based rankings. From this analysis, it is evident that the ranking of the three extractors for each database is the same regardless of whether that ranking is based on the RE or the EER. In particular, MC has the highest RE and lowest EER, while RLT has the lowest RE and highest EER. This implies that the most discriminatory information is contained in finger vein patterns that have been extracted using the MC extractor, and the least discriminatory information is contained in RLT-extracted finger veins. These results suggest the possibility of using the REs of different finger vein recognition systems to rank the systems according to the amount of discriminatory information and thus their expected recognition accuracies. Consequently, it appears reasonable to conclude that the RE estimator is a reliable indicator of the amount of discriminatory information in a finger vein recognition system.

While RE quantifies the amount of discriminatory information in a biometric system, it is difficult to gauge what exactly this number, on its own, means. For example, what exactly does x bits of discriminatory information signify, and is a y-bit difference in the REs of two biometric systems significant? Furthermore, benchmarking different biometric systems in terms of their RE is not straightforward, since the RE estimate depends on both the comparison score range as well as on the number of genuine (N_g) and impostor scores (N_i) for each database and experimental protocol. Consequently, REs reported for different biometric systems usually do not lie in the same $[RE_{\min}, RE_{\max}]$ range.[9] To help us better understand the meaning of the RE metric in the context of a biometric system, as well as to enable fair cross-system RE benchmarking, Sect. 17.4 adapts Eq. (17.2) to propose a *normalised* RE metric.

17.4   Normalised Relative Entropy

This section proposes a normalised version of the RE (NRE), based on the NN estimator in Eq. (17.2). The reason for this normalisation is to help us interpret the RE in a more intuitive way, and to enable fair benchmarking of different biometric systems in terms of their RE.

We propose using the well-known "min–max" normalisation formulated by Eq. (17.3):

$$NRE = \frac{RE - RE_{\min}}{RE_{\max} - RE_{\min}} \tag{17.3}$$

[9]For the finger vein systems we used, the comparison scores for both the VERA and UTFVP databases lie in the same range of [0, 0.5]. However, the N_g values across the two databases are different as are the N_i values. Consequently, the $[RE_{\min}, RE_{\max}]$ range is not the same for both databases, meaning that we cannot fairly compare the RE results across the two databases.

In Eq. (17.3), $RE_{\min}$ and $RE_{\max}$ refer to the minimum and maximum possible RE values, respectively, for a particular biometric system. Thus, we need to begin by establishing $RE_{\min}$ and $RE_{\max}$. In this formulation, we assume that comparison scores are similarity values, such that small scores indicate low similarity and large scores indicate high similarity. Keeping this in mind, the minimum RE would occur when all d_{gi} values are zero and all d_{gg} values are as large as possible. Therefore, for each genuine score, there would need to be at least one impostor score with exactly the same value, and all the genuine scores would need to be spread apart as far as possible. Let us say that all scores lie in the range $[s_{\min}, s_{\max}]$, and that the number of genuine scores for a particular database and experimental protocol is denoted by N_g. Then, the maximum possible d_{gg} value would be $\frac{s_{\max} - s_{\min}}{N_g}$. By adapting Eq. (17.2), our equation for the minimum RE thus becomes

$$RE_{\min} = \frac{1}{N_g} \sum_{i=1}^{N_g} \log_2 \frac{0}{\frac{s_{\max} - s_{\min}}{N_g}} + \log_2 \frac{N_i}{N_g - 1} \tag{17.4}$$

If we now tried to solve Eq. (17.4), we would get $RE_{\min} = -\infty$, because of the 0 d_{gi} term. Since this is an impractical result for measuring the (finite) amount of information in a biometric system, we replace the 0 with ϵ. Furthermore, we can see that the division by N_g gets cancelled out by the summation across N_g, so we can simplify Eq. (17.4) as follows:

$$RE_{\min} = \log_2 \frac{\epsilon}{\frac{s_{\max} - s_{\min}}{N_g}} + \log_2 \frac{N_i}{N_g - 1} \tag{17.5}$$

Equation (17.5) thus becomes the final $RE_{\min}$ equation.

The maximum RE would occur when all d_{gi} values are as large as possible and all d_{gg} values are zero. The only way this could occur would be if all the genuine scores took on the largest possible value, $s_{\max}$, and all the impostor scores took on the smallest possible value, $s_{\min}$. In this case, the genuine and impostor score sets would be as different as possible. By adapting Eq. (17.2), we thus get the following equation for the maximum RE:

$$RE_{\max} = \frac{1}{N_g} \sum_{i=1}^{N_g} \log_2 \frac{s_{\max} - s_{\min}}{0} + \log_2 \frac{N_i}{N_g - 1} \tag{17.6}$$

If we tried to solve Eq. (17.6), we would get $RE_{\max} = \infty$ due to the 0 term in the denominator. So, once again we replace the 0 term with ϵ. Furthermore, just like we did for Eq. (17.4), we can simplify Eq. (17.6) by removing the N_g division and summation. Our final equation for $RE_{\max}$ thus becomes

$$RE_{\max} = \log_2 \frac{s_{\max} - s_{\min}}{\epsilon} + \log_2 \frac{N_i}{N_g - 1} \tag{17.7}$$

We can now use Eq. (17.3), with Eq. (17.5) for $RE_{\min}$ and Eq. (17.7) for $RE_{\max}$, to calculate the NRE of a particular biometric system.

Due to the "min–max" operation in Eq. (17.3), the NRE will lie in the range [0.00, 1.00]. We can thus interpret the NRE as follows. An NRE of 0.00 would suggest that the system in question contains zero discriminative information (i.e. recognition would actually be impossible), whereas an NRE of 1.00 would indicate that the system contains the maximum amount of discriminative information possible for that system (i.e. the recognition accuracy would be expected to be perfect).

Figure 17.4 illustrates what the impostor and genuine comparison score distributions might look like for a minimum NRE system and a maximum NRE system, when the comparison score range is [0, 0.5] (i.e. the score range corresponding to our finger vein recognition systems).

In general, therefore, we can look at the NRE as providing an indication of the proportion of the maximum amount of discriminatory information that the corresponding biometric system contains. An NRE of 0.50, for example, would indicate that the biometric system achieves only 50% of the maximum attainable recognition accuracy. Therefore, the higher the NRE, the better the expected recognition accuracy of the biometric system we are measuring.

Table 17.2 shows the NRE results for our aforementioned finger vein recognition systems. Note that, for these finger vein systems: $s_{\min} = 0$; $s_{\max} = 0.5$; $N_g = 440$ for VERA; $N_g = 4,320$ for UTFVP; $N_i = 192,720$ for VERA; $N_i = 2,067,840$ for UTFVP.

Note that the first column of Table 17.2 refers to the finger vein recognition system constructed using the specified database and feature extractor. We have pooled the databases and extractors into "systems" now to indicate that the NRE values can be benchmarked *across systems* (as opposed to, for example, in Table 17.1, where the databases were separate to indicate that RE-based benchmarking of the different extractors should be *database-specific*).
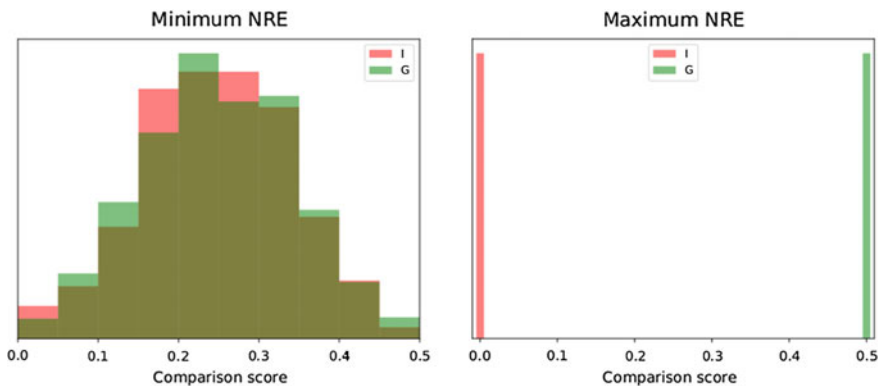
Fig. 17.4 Illustration of impostor and genuine score distributions for a minimum and a maximum NRE system, when the comparison score range is [0, 0.5]

Table 17.2 Relative Entropy (RE) and Normalised Relative Entropy (NRE) for different finger vein recognition systems

System	RE	NRE
VERA-WLD	11.8	0.48
VERA-RLT	4.2	0.34
VERA-MC	13.2	0.50
UTFVP-WLD	18.9	0.58
UTFVP-RLT	18.0	0.56
UTFVP-MC	19.5	0.59

As an example of how the NRE results from Table 17.2 can be interpreted, let us compare the NRE of VERA-RLT to that of UTFVP-MC. The NRE of 0.34 for VERA-RLT tells us that this system achieves only 34% of the maximum attainable discrimination capability. Comparatively, the UTFVP-MC system contains 59% of the maximum amount of discriminative information. So, we could conclude that the UTFVP-MC finger vein recognition system contains 25% more discriminatory information than the VERA-RLT system.

Using the NRE also helps us gauge the significance of the differences in the REs across different biometric systems. For example, if we look at the RE on its own for the UTFVP-WLD and UTFVP-MC systems in Table 17.2, we can see that the latter system's RE is 0.6 bits larger than the former system's RE. It is difficult to tell, however, whether or not this is a significant difference. If we then look at the NREs of the two systems, we can see that their difference is only 0.01. This indicates that the 0.6-bit difference between the two systems' REs is not too significant in terms of the proportion of the maximum discriminatory information the two systems contain. On the other hand, the 15.3-bit difference in the REs between the VERA-RLT and UTFVP-MC systems seems much more significant, and we may be tempted to conclude that the latter system contains about five times more discriminative information than the former system. Looking at the two systems' NREs, we do see a fairly significant difference, but we would have to conclude that the UTFVP-MC system contains not five times, but two times, more discriminative information than the VERA-RLT system.

In this section, we have shown how the NRE can be used for RE-based benchmarking of different finger vein recognition systems, for which comparison scores were evaluated on different databases. The main reason for using the NRE in our case was thus to conduct fair cross-database system benchmarking. Our proposed NRE metric, however, can also be used to fairly benchmark the REs of systems based on different biometric modalities, tested on different databases using different experimental protocols. For example, part of our future work will involve benchmarking the NRE of our best finger vein recognition system, UTFVP-MC, against NREs of systems based on different types of biometrics. This makes the proposed NRE metric a flexible tool for both quantifying and benchmarking the amount of discriminative information contained in different biometric systems.

17.5 Discussion

In this section, we begin by presenting a discussion on an important aspect of the NRE, which supports its adoption in the biometrics community. We then discuss two potential issues that may arise when calculating the NRE, and we suggest the means of dealing with them. Sections 17.5.1, 17.5.2 and 17.5.3, respectively, tackle these three discussion points.

17.5.1 NRE as a Complement to EER

So far, we have shown how the RE can be used to measure the amount of discriminatory information in finger vein recognition systems. We also proposed the NRE metric to fairly benchmark the REs across different biometric systems. In this section, we discuss how an NRE estimate could complement the EER to provide a more complete picture of the performance of a biometric recognition system.

In Sect. 17.2, we explained how, in the context of a biometric recognition system, the RE metric provides some indication of how closely our genuine score distribution matches our impostor score distribution. Let us explore the meaning of this by considering Eq. (17.2). Equation (17.2) tells us that we are attempting to estimate the relative entropy of a set of genuine comparison scores (G) in terms of a set of impostor comparison scores (I). In other words, we wish to quantify the "closeness" of these two sets[10] of scores. The d_{gi} and d_{gg} terms represent the distance between a genuine score and its closest score in the set of impostor and genuine scores, respectively. Larger d_{gi} values will result in *larger* RE results, whereas larger d_{gg} values will result in *smaller* RE results.[11] We can thus see that larger REs favour a larger inter-class variance (i.e. greater separation between genuine comparison trials and impostor trials) and a smaller intra-class variance (i.e. smaller separation between multiple biometric samples from the same biometric instance). This makes the RE suitable as a measure of the performance of a biometric recognition system: the larger the RE value, the better the recognition accuracy. The best (highest) RE would, therefore, be obtained in the case where all the d_{gi} values are as large as possible, while the d_{gg} values are as small as possible, and vice versa for the worst (lowest) RE.

The RE metric thus informs us about two things: how far genuine scores are from impostor scores, and how far genuine scores are from each other. Consider the case where we have a set of impostor scores, I, and a set of genuine scores, G. The larger the intersection between I and G, the smaller the d_{gi} values and thus the lower the RE. Conversely, the smaller the intersection between the two sets, the greater the d_{gi} values and thus the higher the RE. So far, the RE metric appears to tell us the same thing as the EER, since a smaller EER indicates less overlap between genuine and

---

[10]Note: We are purposely using the word "set" as opposed to "distribution", since the NN estimator in Eq. (17.2) works directly on the scores as opposed to distributions representing the scores.

[11]Assume constant N_g and N_i values.
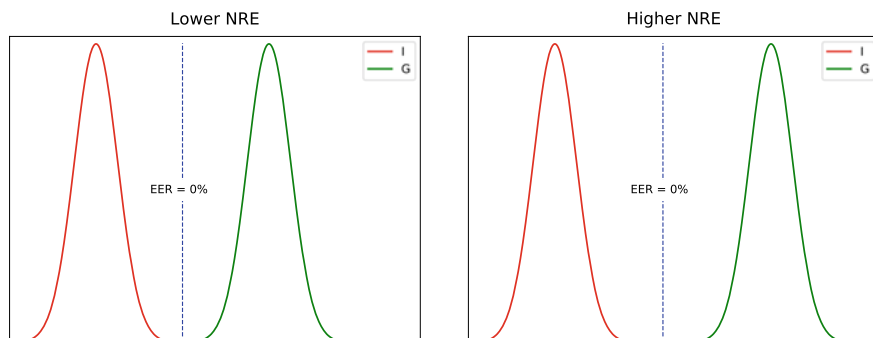
Fig. 17.5 Two biometric systems with the same EER of 0%, but where the system on the right has greater separation between the impostor and genuine comparison scores, and thus a higher NRE than the system on the left

impostor comparison scores, while a larger EER indicates more overlap. Where the two metrics differ, however, is in the scenario where I and G are completely separated. In this case, the further apart the two sets of scores are the higher the resulting RE. The EER, however, would be 0% regardless of whether the separation is small or large. Imagine if we had to benchmark two biometric systems, both of which had complete separation between the genuine and impostor comparison scores, but where for one system the separation was much larger than for the other, as illustrated[12] in Fig. 17.5. If we considered only the EER, it would indicate that the two systems are the same (i.e. both have an EER of 0%). The NRE,[13] however, would clearly indicate that the system with greater separation is better in terms of distinguishing genuine trials from impostors, since the NRE value would be higher for that system. In this case, complementing the EER with an NRE estimate would provide a more complete picture of the system comparison. This could come in useful particularly in situations where the data used for testing the biometric system was collected in a constrained environment, in which case an EER of 0% could be expected. The NRE, on the other hand, would provide us with more insight into the separation between the genuine and impostor score distributions.

Another example of a scenario in which the NRE metric would be a useful complement to the EER is when we have two biometric systems for which I is the same and the separation (or overlap) between I and G is the same, but G differs. In particular, in the first system the genuine scores are closer together, while in the second system the genuine scores are further apart from each other. Figure 17.6 illustrates

---

[12]Note: The only reason for using probability density plots in this figure is to present a cleaner illustration of our point. Probability density functions are *not* used to represent genuine and impostor score distributions for the NRE calculation.

[13]When benchmarking different biometric systems, the NRE should be used instead of the RE to ensure that the benchmarking is fair. The only exception to this rule would be in the case where the different systems had the same comparison score range, and the same N_g and N_i values, in which case the resulting REs would lie in the same $[RE_{min}, RE_{max}]$ range.
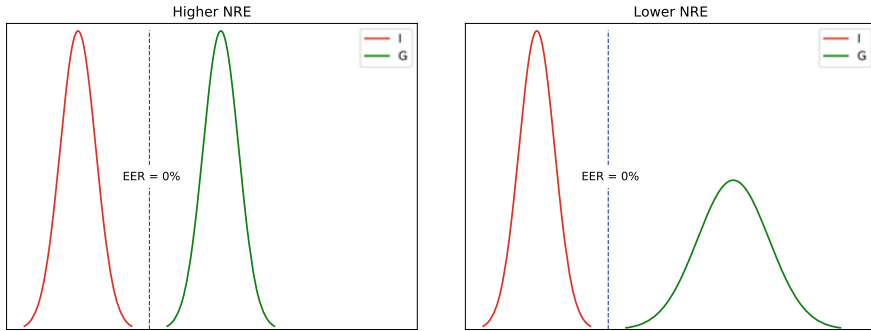
Fig. 17.6 Two biometric systems with the same *I*, the same separation between *I* and *G* and thus the same EER, but with different *G*. In particular, *G* for the system on the right has a larger variance, and thus the NRE is lower to reflect this

this scenario.[14] In this case, since the separation between *I* and *G* for both systems is the same, the EER would also be the same, thereby indicating that one system is just as good as the other. The NRE, however, would be smaller for the second system due to the larger d_{gg} values. The NRE would thus indicate that the larger intra-class variance in the second system makes this system less preferable in terms of biometric performance when compared to the first system, for which the genuine scores are closer together and thus the intra-class variance is smaller. Using both NRE and EER together, we could thus conclude that, although both systems can be expected to achieve the same error rate, the system with the smaller intra-class variance would be a superior choice.

When choosing between the EER and NRE metrics for evaluating the performance of a biometric system, we would still recommend using the EER as the primary one, since it is more practical in providing us with a solid indication of our system's expected error rate. The NRE, however, would be a useful complement to the EER when we are trying to decide on the best of *n* biometric systems that have the *same* EER.

17.5.2 Selecting the ε Parameter

As mentioned in the introductory paragraph of Sect. 17.3.2, ϵ is a parameter chosen to deal with zero score differences (i.e. $d_{gg} = 0$ or $d_{gi} = 0$) in order to avoid an RE of $\pm\infty$ (which would be meaningless in the context of measuring the amount of discriminatory information in a biometric system). It is clear from Eqs. (17.2), (17.3), (17.5) and (17.7), however, that the choice of ϵ could potentially have a significant effect on the resulting RE and, therefore, NRE, particularly if the number of zero score

---

[14]Note: In Fig. 17.6, the EER for both systems is 0%; however, it could also be possible for both systems to have the same non-zero EER. In this case, *I* and *G* would partially overlap.

differences is large. While the number of zero score differences will be dependent on the biometric system in question and this number is, therefore, difficult to generalise, we wished to see what effect the choice of ϵ would have on the RE and NRE of our best finger vein recognition system, that obtained when using MC-extracted finger veins from the UTFVP database. Figure 17.7 shows plots of the RE and NRE versus ϵ, when ϵ is selected to lie in the range $[10^{-12}, 10^{-8}]$. For convenience, Table 17.3 summarises the RE and NRE values from Fig. 17.7.

From Fig. 17.7 and Table 17.3, we can see that, while the choice of ϵ does affect the RE and NRE to some degree (more specifically, the RE and NRE decrease as ϵ decreases[15]), this effect does not appear to be significant. So, we may conclude that, as long as the ϵ parameter is sensibly chosen (i.e. smaller than the comparison scores, but not so small that it is effectively zero), then the RE and NRE estimates should be reasonable.
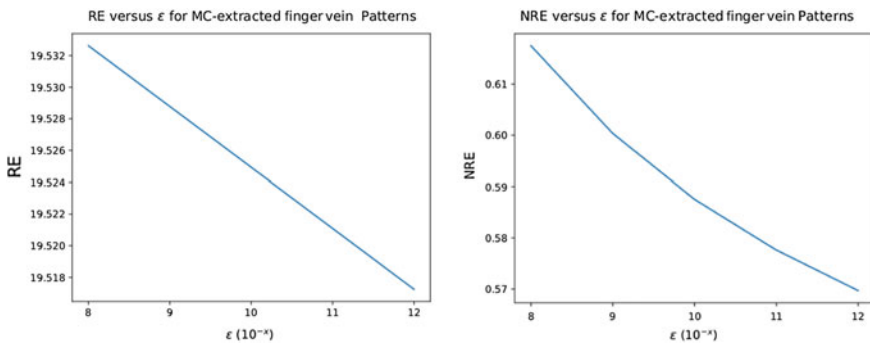
Fig. 17.7 RE versus ϵ and NRE versus ϵ, when ϵ takes on different values in the range $[10^{-12}, 10^{-8}]$, for MC-extracted finger vein patterns in the UTFVP database

Table 17.3 RE and NRE for MC-extracted finger veins from UTFVP, when ϵ is varied in the range $[10^{-12}, 10^{-8}]$. Note that, for consistency with Table 17.2, RE and NRE values are rounded to 1 d.p. and 2 d.p., respectively

ϵ	RE	NRE
10^{-8}	19.5	0.62
10^{-9}	19.5	0.60
10^{-10}	19.5	0.59
10^{-11}	19.5	0.58
10^{-12}	19.5	0.57

---

[15]In general, the RE, and thus the NRE, would be expected to *decrease* with a decrease in ϵ when there are more d_{gi} than d_{gg} zero score differences. Alternatively, the RE, and thus the NRE, would be expected to *increase* with a decrease in ϵ when there are more d_{gg} than d_{gi} zero score differences.

17.5.3 Number of Nearest Neighbours

The method proposed in [9] to estimate the RE of biometrics uses only the *first* nearest genuine and impostor neighbours of each genuine score. An issue with this approach is that it makes the RE estimate highly dependent on any single score, even if that score is an outlier. This might be particularly problematic if we do not have a large number of scores to work with, which is often the case.

It seems that a safer approach would be to use k nearest neighbours, where $k > 1$, then average the resulting $d_{gg}(i)$ and $d_{gi}(i)$ values over these k neighbours prior to estimating the RE. This would introduce some smoothing to the underlying score distributions, thereby stabilising the RE estimates. While the effect of k on the RE, and therefore NRE, is difficult to generalise since it would, in practice, be dependent on the biometric system in question, we wished to test the effect of the choice of k on the RE and NRE of our best finger vein recognition system, that obtained when using MC-extracted finger veins from the UTFVP database. Figure 17.8 shows plots of the RE and NRE versus k, when k increases from 1 to 5. For convenience, Table 17.4 summarises the RE and NRE values from Fig. 17.8. Note that, for this experiment, $\epsilon = 10^{-10}$, as for the RE and NRE experiments in Sects. 17.3 and 17.4.
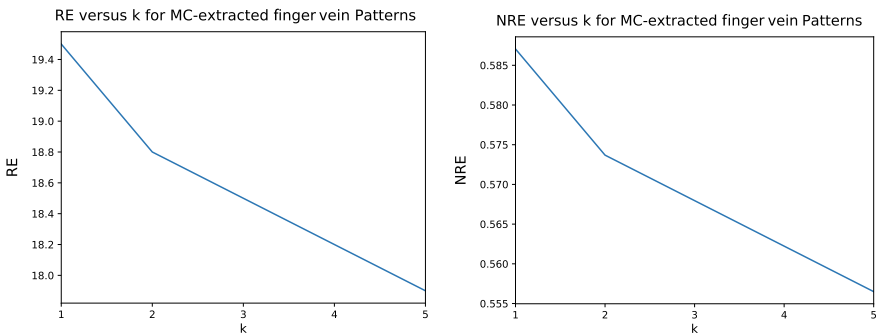
Fig. 17.8 RE versus k and NRE versus k, when k increases from 1 to 5, for MC-extracted finger vein patterns in the UTFVP database

Table 17.4 RE and NRE for MC-extracted finger veins from UTFVP, when k increases from 1 to 5. Note that, for consistency with Tables 17.2 and 17.3, RE and NRE values are rounded to 1 d.p. and 2 d.p., respectively

k	RE	NRE
1	19.5	0.59
2	18.8	0.57
3	18.5	0.57
4	18.2	0.56
5	17.9	0.56

From Fig. 17.8 and Table 17.4, it is evident that increasing k tends to decrease both the RE and NRE, but the decrease is not drastic for $k \leq 5$. This decrease makes sense, since a larger k means a greater degree of smoothing, which decreases the effects of individual comparison scores. Another consequence of using a larger k would be that the effect of the ϵ parameter on RE and NRE would be expected to be less pronounced. This is because a larger k means that a larger number of neighbouring scores are averaged when calculating the RE and NRE, so we are less likely to encounter zero average scores than in the scenario where only *one* nearest neighbouring score is considered. Keeping the aforementioned points in mind, it is important to sensibly tune the k and ϵ parameters depending on the biometric system in question (e.g. if there are outlier scores, use $k > 1$, and select ϵ based on the score precision, as discussed in Sect. 17.5.2). Furthermore, we urge researchers adopting the RE and NRE measures to be transparent about their selection of these parameters to ensure fair system comparisons across the biometrics community.

Note that the NN estimator on which Eq. (17.2) is based [10] *is* actually a k-NN estimator, where k denotes the number of nearest neighbours. It is not clear, however, whether the proposed k-NN estimator is based on *averaging* the k nearest neighbouring scores, as we have done for Fig. 17.8 and Table 17.4, or whether the authors meant that *only* the kth neighbour should be used. If their intention is the latter, then our averaging approach represents an effective new way of stabilising the k-NN estimator for RE measures.

17.6 Conclusions and Future Work

This chapter represents the first attempt at estimating the amount of information in finger vein biometrics in terms of score-based Relative Entropy (RE), using the previously proposed Nearest Neighbour estimator. We made five important contributions.

First, we showed that the RE estimate is system-specific. In our experiments, the RE differed across finger vein recognition systems employing different feature extractors and different testing databases. For this reason, we refrain from claiming a universal finger vein RE estimate, since this would be misleading.

Second, we showed that the RE can be used to rank different finger vein recognition systems, which are tested on the same database using the same experimental protocol (in our case, the difference was the feature extractor employed), in terms of the amount of discriminative biometric information available. The ranking was shown to be comparable to an EER-based ranking, which implies that the RE estimate is a reliable indicator of the amount of discriminatory information in finger vein recognition systems.

Third, we proposed a new metric, the Normalised Relative Entropy (NRE), to help us gauge the significance of individual RE scores as well as to enable fair benchmarking of different biometric systems (in particular, systems tested on different databases using different experimental protocols) in terms of their RE. The NRE lies in the range [0.00, 1.00] and represents the proportion of the maximum amount of

discriminatory information that is contained in the biometric system being measured. The higher the NRE, the better the system is expected to be at distinguishing genuine trials from impostors.

Fourth, we discussed how the NRE metric could be a beneficial complement to the EER in ranking different biometric systems in terms of their discrimination capabilities. The NRE would be particularly useful in choosing the best of n biometric systems that have the same EER.

Finally, we discussed two potential issues in calculating the RE and NRE, namely, the effects of the ϵ parameter and the number of nearest neighbours (k) used for computing the genuine–genuine and genuine–impostor score differences. We showed that, as long as ϵ is sensibly selected, its effect on the RE and NRE is unlikely to be significant. We also showed that increasing the number of nearest score neighbours may be expected to slightly decrease the RE and NRE, but the upside is that using a larger number of nearest neighbours would help to dilute the effects of outliers among the genuine and impostor comparison scores. We concluded by suggesting that ϵ and k be tuned according to the biometric system being evaluated and that researchers be transparent in terms of reporting their selection of these two parameters.

At the moment, our primary aim for future work in this direction is to use our proposed NRE metric to benchmark finger vein recognition systems against systems based on other biometric modalities, in terms of the amount of discriminatory information contained in each system.

References

1. Ye Y, Zheng H, Ni L, Liu S, Li W (2016) A study on the individuality of finger vein based on statistical analysis. In: 2016 international conference on biometrics (ICB), pp 1–5, June 2016
2. Yanagawa T, Aoki S, Ohyama T (2007) Human finger vein images are diverse and its patterns are useful for personal identification. MHF Prepr Ser 12:1–7
3. Jeffers J, Arakala A, Horadam KJ (2010) Entropy of feature point-based retina templates. In: 2010 20th international conference on pattern recognition, pp 213–216, Aug 2010
4. Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30
5. Krivokuca V, Marcel S (2018) Towards quantifying the entropy of fingervein patterns across different feature extractors. In: 2018 IEEE 4th international conference on identity, security, and behavior analysis (ISBA), pp 1–8, Jan 2018
6. Takahashi K, Murakami T (2010) A metric of information gained through biometric systems. In: Proceedings of the 2010 20th international conference on pattern recognition. IEEE Computer Society, pp 1184–1187
7. Sutcu Y, Tabassi E, Sencar HT, Memon N (2013) What is biometric information and how to measure it? In: 2013 IEEE international conference on technologies for homeland security (HST), pp 67–72, Nov 2013

8. Adler A, Youmaran R, Loyka S (2006). Towards a measure of biometric information. In 2006 Canadian conference on electrical and computer engineering, pp 210–213, May 2006
9. Sutcu Y, Sencar HT, Memon N (2010) How to measure biometric information? In : 2010 20th international conference on pattern recognition, pp 1469–1472, Aug 2010
10. Wang Q, Kulkarni SR, Verdu S (2009) Divergence estimation for multidimensional densities via k-nearest-neighbor distances. IEEE Trans Inf Theory 55(5):2392–2405
11. Vanoni M, Tome P, El-Shafey L, Marcel S (2014) Cross-database evaluation using an open finger vein sensor. In: 2014 IEEE workshop on biometric measurements and systems for security and medical applications (BIOMS) proceedings, pp 30–35, Oct 2014
12. Ton BT, Veldhuis RNJ (2013) A high quality finger vascular pattern dataset collected using a custom designed capturing device. In: 2013 international conference on biometrics (ICB), pp 1–5, June 2013
13. Lee EC, Lee HC, Park KR (2009) Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction. Int J Imaging Syst Technol 19(3):179–186
14. Huang B, Dai Y, Li R, Tang D, Li W (2010) Finger-vein authentication based on wide line detector and pattern normalization. In: 2010 20th international conference on pattern recognition, pp 1269–1272, Aug 2010
15. Miura N, Nagasaka A, Miyatake T (2004) Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. Mach Vis Appl 15(4):194–203
16. Miura N, Nagasaka A, Miyatake T (2007) Extraction of finger-vein patterns using maximum curvature points in image profiles. IEICE Trans Inf Syst 90(8):1185–1194