RELIABILITY INDEX FOR IRIS CODE MATCHING IN IRIS RECOGNITION SYSTEMS

B. H. Shekara, Sharada S Bhatb, Archil Maysuradze

a Mangalore University, Mangalore, Karnataka India - bhshekar@gmail.com
b Government First Grade College, Ankola, Karnataka, India - sharadasbhat@gmail.com
c Lomonosov Moscow State University, Moscow, Russia - archil@mail.ru

Commission II, WG II/10

KEY WORDS: Iris code matching, Reliability Index, Iris recognition

ABSTRACT:

Iris code matching is an important stage of iris biometric systems which compares the input iris code with stored patterns of enrolled iris codes and classifies the code into one of classes so that, the claim is accepted or rejected. Several classifier based approaches are proposed by the researchers to improve the recognition accuracy. In this paper, we discuss the factors affecting an iris classifier’s performance and we propose a reliability index for iris matching techniques to quantitatively measure the extent of system reliability, based on false acceptance rate and false rejection rates using Monte Carlo Simulation. Experiments are carried out on benchmark databases such as, IITD, MMU v-2, CASIA v-4 Distance and UBIRIS v.2.

1. INTRODUCTION

Iris code matching and classifying is a crucial stage in any iris recognition system. During this stage, the query iris code is compared with an enrolled iris code to generate a match score. Match score is a similarity or a dissimilarity measure which is computed using a distance function. If it is authentication (verification) system, the query iris code is compared with the iris patterns of the claimer so that, the database of enrolled iris codes to be searched is smaller in size and hence, the number of match scores generated are less. Where as, in case of identification systems, query iris code is compared with all the enrolled iris patterns so that the database size is quite larger and hence, large number of match scores are generated. However, in both the systems, a classifier (iris code matcher) has to be adopted at this stage. In authentication systems the iris matcher has to accept or reject the claim (binary classification) and in identification systems, the classifier has to classify the query iris code into one of the N + 1 classes where N is the number of subjects enrolled in the database (multi-class classifier). In password-based recognition systems a perfect match between the two password strings is necessary. Where as, in biometric recognition systems, identity of a person is determined based on a close match between query and enrolled iris codes. Closeness of the match score is decided by a threshold value. (Jain et al., 2011).

Several approaches have been proposed in the literature for iris code matching (Bowyer et al., 2008, Bowyer and Burge, 2016). The well known Daugman’s approach (Daugman, 1993) uses Hamming distance with exclusive OR operator (XOR operator) to generate a dissimilarity score between the two iris codes and statistical decision theory for classification. Two iris codes with extracted features are not all statistically independent of each other. However, given the Hamming distance distributions for a large number of genuine match scores (scores between the two irises of the same subject-mate irises) and impostor match scores (scores between the two irises of different subjects-nommate irises), the distributions can be fit with a binomial curve to find the effective number of degrees of freedom which calculates a confidence limit for a match of two iris codes. To measure the goodness-of-match between the feature vectors of query iris and enrolled irises Wildes et.al (Wildes, 1997) have used normalised correlation of the feature vectors and have utilised Fisher’s linear discriminant function for decision making. Le Ma et.al. (Ma et al., 2004) have devised a matching scheme on the expanded feature vectors based on XOR operator. For comparing the two iris codes, a nearest neighbor approach is used by Monro et.al. (Monro et al., 2007) and to compute the distance they have proposed product-of-sum of Hamming distance of individual sub features extracted from each iris. These early methods perform well with the constrained imaging systems. However, recent researchers are working with the noisy iris images taken at-a-distance and in the relaxed imaging scenario wherein, the images suffer from occlusion, bad illuminations and reflections (Nguyen et al., 2017) and the research is referred to as unconstrained iris recognition.

In order to undermine the effect of noise in unconstrained iris recognition, the localised iris region is divided into N subregions (called as patches) using m tracks and n sectors. Proença and Alexandre (Proença and Alexandre, 2007) have extracted the features of individual patch, computed the match score between the corresponding patches and then proposed a classification rule to combine the scores. Hollingsworth et.al. (Hollingsworth et al., 2011) have proposed a fragile bit distance, the match score obtained by which, is then combined with the hamming distance. Wang et.al. (Wang et al., 2012) have adopted AdaBoost learning to determine the iris patches with most discriminative features. Pillai et.al. (Pillai et al., 2011) have divided the unwrapped iris image into four sectors, extracted the features from each sector and used a sparse based recognition algorithm for classification. Different distance measures and classifiers are existing in the literature are Euclidean distance (Proença and Briceño, 2013, Tan and Kumar, 2013), $\chi^2$ distance, (Bharadwaj et al., 2010), normalised cosine distance, (Juefei-Xu and Savvides, 2012), supervised Mahalanobis distance(Nie et al., 2014) and weighted mean sum.

The key question is: Is there any reason to prefer one classifier?
over the other? As stated by No-free-lunch theorem (Wolpert and Macready, 1997), any learning algorithm which performs better for a particular class of pattern recognition (PR) problem is not superior in general for all PR problems. The ugly-duckling theorem (Kamishima et al., 2012) states that there is no feature (no pattern representation) that will yield better classification performance, unless we make some assumptions about the data. In (Kryszkuch et al., 2005), authors have presented a framework to predicate and correct the classification errors on modality reliability measures in multi-modal biometric system, with the case study of face and speech recognition.

Though, the existing performance measures of biometric recognition systems measure the efficiency of the systems, they don’t answer the question, how to select a better classifier for an iris recognition system. In this paper, we propose a reliability index (RI) to measure the performance of iris classifiers. We define RI based on false acceptance rate (FAR) and false reject rate (FRR) of the classifier’s output. We propose Monte Carlo simulation to compute RI thus defined. We also discuss the various factors that affect the reliability of iris code matching techniques.

Rest of the paper is organised as follows. Section 2 briefs the existing methods of measurements of performance of iris recognition systems. In Section 3 we explain the proposed reliability index and how to compute this using Monte Carlo simulation method. The experimental illustration of the proposed RI is discussed in Section 4 and Section 5 concludes this paper.

2. EXISTING MEASURES

Based on Daugman’s statistical decision theory (Daugman, 1993), in case of iris authentication systems, failure of iris classifiers is measured by the two error rates, false acceptance rate (FAR), accepting a non-mate iris as mate iris and false reject rate (FRR), rejecting a mate iris as non-mate iris. The goal of any iris classifier is to minimise these two error rates. Let \( \omega_g \) and \( \omega_i \) denote genuine and impostor classes respectively and \( P(x \mid \omega_g) \) and \( P(x \mid \omega_i) \) be their probability distributions where \( x \) is a variable representing an arbitrary dissimilarity score. Figure 1 presents an example of \( P(x \mid \omega_g) \) and \( P(x \mid \omega_i) \) computed with respect to MMU v.2 iris dataset using the methods discussed in (Masek et al., 2003).

![Figure 1. Probability distributions of genuine and impostor scores](https://example.com/image1.png)

FAR is the probability of an impostor score \( x \) to be accepted by the classifier, where as, FRR is the probability of a genuine score to be rejected. Since these probabilities change with respect to the system threshold \( \tau \), they are defined as functions of \( \tau \), as follows (Jain et al., 2011).

\[
FAR(\tau) = P(x \leq \tau \mid \omega_i) = \int_{-\infty}^{\tau} P(x \mid \omega_i)dx
\]

\[
FRR(\tau) = P(x > \tau \mid \omega_g) = \int_{\tau}^{\infty} P(x \mid \omega_g)dx
\]

Further, by computing FAR and FRR at various threshold values two smooth curves, detection error tradeoff (DET) and receiver operating characteristic (ROC), can be obtained which are the graphical tools to measure the performance of iris authentication systems. DET curve is obtained by plotting FRR against FAR at different threshold values on a normal deviate scale and ROC curve plots true acceptance rate (1-FRR) against FAR using linear logarithmic scale. Iris authentication systems can also be evaluated by single valued measures like, equal error rate (EER) which is the point on DET curve where FAR equals FRR and \( d_{prime} \) value which measures the separation between the probability distributions \( P(x \mid \omega_g) \) and \( P(x \mid \omega_i) \) and is defined as follows.

\[
d_{prime} = \frac{\mu_g - \mu_i}{\sqrt{\frac{1}{2} (\sigma_g^2 + \sigma_i^2)}}
\]

where, \( \mu_g \) and \( \sigma_g \) are the mean and standard deviation of impostor distribution and \( \mu_i \) and \( \sigma_i \) are those of genuine distribution. Researchers have been using these measuring tools to prove the efficiency of their algorithms with respect to the best performance on a dataset. However, these tools do not provide any clue about how reliable an algorithm is, with respect to a random scenario. We propose an index which measures the reliability of a matching algorithm when the algorithm is applied on different datasets. Similar to \( d_{prime} \) or EER, proposed RI is also single valued measure which can tell whether a classifier is reliable against the other classifiers in a given scenario.

3. PROPOSED RELIABILITY INDEX

Reliability of a system means the ability of the system to fulfill what is required of it. Quantitatively it is measured as the probability that a system operates successfully (Crowder, 2017). Statistical meaning of simple reliability index for a system with random variables which are described by their mean values, variances and covariances (Ditlevsen and Madsen, 1996) is explained in the next section.

3.1 Statistical meaning of RI

Assume that the input variables \( \{x_1, x_2, ..., x_n\} \) with some uncertainties are modeled as random variables \( \{X_1, X_2, ..., X_n\} \) with the given mean values,

\[
E[X_1], E[x_2], ..., E[X_n]
\]

and covariances are defined by,

\[
Cov[X_i, X_j] = E[X_iX_j] - E[X_i]E[X_j]
\]

Let \( R \) be a new random variable given by,

\[
R = a_1X_1 + a_2X_2 + ... + a_nX_n + b = a^T X + b
\]

where \( a = [a_1, a_2, ..., a_n] \), \( X = [X_1, X_2, ..., X_n] \) and \( a_1, a_2, ..., a_n \) and \( b \) are given constants. The mean and variance of the variable
Hence, the reliability index is defined as,

\[
\beta = \frac{E[R]}{\sqrt{Var[R]}} = \frac{a^T E[X] + b}{\sqrt{a^T Cov[X, X^T]a}}
\]

If the random variable \( R \) defined in equation (6) is normalised with zero mean and unit standard deviation then, geometrically simple reliability index is the perpendicular distance from the origin to the hyperplane represented by normalised \( R \) (see Figure 2).

### 3.2 Reliability index for iris classifiers

Apart from the statistical definition, in general, \( RI \) of any system is defined based on the probability of the failure of the system.

\[
RI = 1 - \text{Probability(failure)}
\]

We define \( RI \) in terms of probability of failure using equation (10). Since our definition is based on FAR and FRR, \( RI \) is also a function of system threshold \( \tau \).

#### Definition 3.1
Let \( \omega^\text{FAR}_r \) and \( \omega^\text{FRR}_r \) represent the classes of dissimilarity scores falling under false acceptance and false reject regions respectively for a particular value of \( \tau \), so that, both are subsets of \( \omega_i \cap \omega_j \), then, the probability of failure is given by,

\[
P(x \mid \omega^\text{FAR}_r \cup \omega^\text{FRR}_r) = 1 - P(x \leq \tau \mid \omega^\text{FAR}_r) - P(x > \tau \mid \omega^\text{FRR}_r)
\]

Hence, the reliability index \( RI \) is given by,

\[
RI(\tau) = 1 - P(x \mid \omega^\text{FAR}_r \cup \omega^\text{FRR}_r)
\]

Since, \( \omega^\text{FAR}_r \) and \( \omega^\text{FRR}_r \) are disjoint classes, it can be further simplified as,

\[
RI(\tau) = 1 - \left( P(x \mid \omega^\text{FAR}_r) + P(x \mid \omega^\text{FRR}_r) \right)
\]

Using equations (1) and (2) we can write \( RI \) as,

\[
RI(\tau) = 1 - (FAR(\tau) + FRR(\tau))
\]

RI given in equation (14) is trivial and straightforward. However, the reliability of an iris classifier is affected by several other factors also. Intermingling all the effects together, we propose the following expression for \( RI \) using a normalising factor \( \eta \).

\[
RI(\tau) = 1 - \frac{1}{\eta} (FAR(\tau) + FRR(\tau))
\]

In the following section we discuss factors that affect a classifier’s performance in iris authentication systems and how the expression \( \eta \) can be computed.

### 3.3 Factors affecting the iris classifier’s performance

The most influential factor that affects an iris classifier’s performance is iris training dataset. Publicly available near infrared (NIR) and visual wavelength (VW) iris datasets possess variety of degraded images for training the classifiers. These degraded images, unless handled properly, reduce the recognition accuracy, thereby making the system less reliable. In a recent paper (Kabra et al., 2015) authors have proposed a distance measure which helps to tune the most influential training dataset that causes classifier’s errors. Size of the dataset and the training to test ratio are the other factors that affect the performance. In the applications like, bank teller machines, search for the best match is against a smaller and specific set of iris templates and such authentication systems are expected to possess zero percentage of false acceptance rate (FAR) with highest possible reliability index. i.e. such applications require the systems with \( RI = 1 \).

In (Liu et al., 2018), authors have proposed inner reliability of a classifier, represented by a matrix \( R_{i.e.c} \) ( \( \omega \) being the number of classes in the data set), to characterize the conditional probability of the object belonging to class number \( i \) ( \( i = 1, 2, ..., c \) ) when it is classified into class number \( j \) ( \( j = 1, 2, ..., c \) ) by the given classifier. Authors have proposed a method to evaluate the reliability matrix using training information and they claim that this matrix provides much more refined reliability knowledge than the training accuracy value.

Thus, considering the effect of training datasets, we take \( \eta \) to be proportional to the total size of the classes \( \omega_i \) and \( \omega_j \).

\[
\eta \propto |\omega_i \cup \omega_j|
\]

In addition, \( RI \) is also affected by the segmentation errors and the filtering errors of the feature extraction algorithms and hence, \( \eta \) also depends on the percentages of segmentation and filtering error.

### 3.4 Application of Monte Carlo Simulation

To compute the reliability index, several statistical methods are discussed in (Crowder, 2017) and (Ditlevsen and Madsen, 1996). Terry and Mae (Koo and Li, 2016) have used intra-class correlation coefficient for reliability analysis in clinical research. Monte Carlo simulation (MCS) has been a powerful tool for reliability assessment of distribution systems in various fields (Liang and Goel, 1997). MCS is based on the probability concept that, while behavior of average samples of random variables is mathematically established, these behavioral tendency is stabilized as the sample size increases largely. The necessary condition to apply MCS is the randomness of the input variables which is well satisfied in case of probability distribution of iris dissimilarity scores.

To explain the application of MCS for evaluation of RI in case of iris classifiers, an arbitrary dissimilarity score \( x \) may fall in any area of the distributions \( P(\omega_i) \) and \( P(\omega_j) \) (Refer Figure 1). When experimented with a small sample of dissimilarity scores,
Table 1. $RI$ obtained by the experiments conducted using different filters with respect to the four databases. Distance measure is the Hamming distance in all the cases.

<table>
<thead>
<tr>
<th>Method</th>
<th>IITD v-2</th>
<th>MMU v-2</th>
<th>CASIA v-4 Dist</th>
<th>UBIRIS v-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCT</td>
<td>0.9560</td>
<td>0.9355</td>
<td>0.9470</td>
<td>0.9332</td>
</tr>
<tr>
<td>LOG-Gabor</td>
<td>0.9638</td>
<td>0.9500</td>
<td>0.9543</td>
<td>0.9500</td>
</tr>
<tr>
<td>Riesz</td>
<td>0.9773</td>
<td>0.9673</td>
<td>0.9643</td>
<td>0.9401</td>
</tr>
<tr>
<td>TSE</td>
<td>0.9793</td>
<td>0.9681</td>
<td>0.9649</td>
<td>0.9365</td>
</tr>
</tbody>
</table>

Table 2. $RI$ obtained by using Riesz filter with two different distance measures HD and WMHD, on all the databases.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>HD</th>
<th>WMHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IITD v-2</td>
<td>0.9773</td>
<td>0.9800</td>
</tr>
<tr>
<td>MMU v-2</td>
<td>0.9673</td>
<td>0.9770</td>
</tr>
<tr>
<td>CASIA v-4 Distance</td>
<td>0.9634</td>
<td>0.9765</td>
</tr>
<tr>
<td>UBIRIS v-2</td>
<td>0.9401</td>
<td>0.9760</td>
</tr>
</tbody>
</table>

Figure 3. $RI$ showing that, with respect to all databases, WHMD is more reliable than normal Hamming distance.

Reader can refer (Shekar and Bhat, 2017) for detailed explanation. Table 2 and Figure 3 present the results of second set of experiments. Figure 3 shows that with respect to all databases WHMD is more reliable than normal Hamming distance.

During the third set of experiments, we have computed the $d_{prime}$ using the equation (3) and $RI$ values on the same dataset (IITD) with respect to different filters. In these experiments $RI$ is based only on the filtering errors. Results presented in Table 3 illustrate that both $RI$ and $d_{prime}$ are positively correlated. Though, $d_{prime}$ is one of the popular performance measures of biometric systems, proposed $RI$ checks the extent to which the algorithm is reliable.

5. CONCLUSION

We have proposed a quantitative measure for reliability ($RI$) for iris code matching techniques. Proposed $RI$ is tested by conducting rigorous experiments on different data sets using different feature extraction methodologies and matching techniques. We also have discussed the various factors that affect the performance of an iris code matcher. It has been observed that a technique which is more reliable on one particular system may not be the superior in context of another system.

We have proposed $RI$ based on the two error rates FAR and FRR which are based on the identification system errors. When we consider the errors in identification systems, we get two more error

Table 3. $RI$ and $d_{prime}$ values obtained by different filters and two distance measures HD and WMHD, on IITD database.

<table>
<thead>
<tr>
<th>Method</th>
<th>HD d_{prime}</th>
<th>WMHD d_{prime}</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCT</td>
<td>0.9500</td>
<td>2.7604</td>
</tr>
<tr>
<td>LOG-Gabor</td>
<td>0.9638</td>
<td>5.4817</td>
</tr>
<tr>
<td>Riesz</td>
<td>0.9773</td>
<td>5.6306</td>
</tr>
<tr>
<td>TSE</td>
<td>0.9793</td>
<td>5.8728</td>
</tr>
</tbody>
</table>
rates, false positive identification rate (FPIR) and false negative identification rate (FNIR). By making minor modifications in the expression given in equation (15), RI can also be computed for iris identification systems.

ACKNOWLEDGEMENT
This work is supported jointly by the Department of Science & Technology, Govt. of India and Russian Foundation for Basic Research, Russian Federation under the grant No. INT/RUS / RFBR /P-248.

REFERENCES
