Iris-Biometric Comparators: Minimizing Trade-Offs Costs between Computational Performance and Recognition Accuracy

Christian Rathgeb, Andreas Uhl and Peter Wild

Multimedia Signal Processing and Security Lab (WaveLab)
University of Salzburg, Austria

Abstract

The intricate structure of the iris constitutes a powerful biometric utilized by iris recognition algorithms to extract discriminative biometric templates. In order to provide a rapid comparison of biometric templates the vast majority of feature extraction methods are designed to generate binary biometric templates, applying the Hamming distance as (dis-)similarity metric. Based on this concept several feature extraction techniques have been proposed in literature, while potential improvements in comparison procedures are commonly neglected. In this paper trade-off costs between the computational performance and recognition accuracy of iris-biometric comparators are investigated. Different comparison techniques of binary biometric templates, and a composition of these, are proposed, where emphasis is put on the trade-off between computational cost and improvement of recognition accuracy, i.e. recognition accuracy is improved at minimal additional computational cost. Experimental results confirm the soundness of the proposed approaches.

Keywords: Biometrics, iris recognition, iris biometric comparators, bit-reliability, score-level fusion.

1 Introduction

Iris recognition is gaining popularity as a robust and reliable biometric technology. The iris’s complex texture and its apparent stability hold tremendous promise for applying iris recognition in diverse application scenarios, such as border control, forensic investigations, as well as cryptosystems. Several existing approaches to iris recognition achieve auspicious performance, reporting recognition rates above 99% and equal error rates of less than 1% on diverse data sets [1]. Generic iris recognition systems comprise four key components: image acquisition, pre-processing, feature extraction, and template comparison. In the acquisition step the image of a subject’s eye is captured (e.g. using a near-infrared camera). At pre-processing the iris is detected and prepared for subsequent feature extraction, which commonly involves an un-wrapping of the iris to a rectangular image, as well as contrast enhancement. Based on the resulting iris texture, feature extraction is applied in order to generate a biometric template. The majority of iris recognition algorithms extract binary templates, i.e. iris-codes, applying the Hamming distance to calculate (dis-)similarity scores, providing (1) a rapid authentication (even in identification mode) and (2) a compact storage of biometric templates. Alignment of biometric templates is achieved by a circular bit-shift of iris-codes (to some degree), where the minimum obtained Hamming distance corresponds to an optimal alignment. Figure 1 illustrates the common processing chain of an iris recognition algorithm. While most approaches to iris recognition algorithms focus on extracting highly discriminative iris-codes, potential improvements within comparators are frequently neglected.

The contribution of this work is the proposal of different improved iris-biometric comparators. In order to maintain a fast comparison and compact storage of biometric templates, emphasis is put on trade-off costs between computational performance, storage cost, and recognition accuracy. The aim is to gain performance with respect to recognition accuracy at negligible cost of computational performance and template storage. By introducing two different comparison techniques, and a composition of these, the accuracy of different iris recognition systems is increased on diverse databases, confirming the soundness of the proposed approaches.

This paper is organized as follows: related work regarding iris-biometric template comparison is briefly summarized (Section 2). Subsequently, different iris-biometric comparators are proposed and described in detail (Section 3). A comprehensive experimental evaluation of both methods and a composition of these is presented (Section 4). Finally, a conclusion is given (Section 5).

2 Template Comparison in Iris Recognition

Focusing on iris recognition, a binary representation of biometric features offers two major advantages:

1. Rapid authentication (even in identification mode).
2. Compact storage of biometric templates.

Comparisons between binary biometric feature vectors are commonly implemented by the simple Boolean exclusive-OR operator (XOR) applied to a pair of binary biometric feature vectors, masked (AND’ed) by both of their corresponding mask templates to prevent occlusions caused by eyelids or eyelashes from influencing comparisons. The XOR operator ⊕ detects disagreement between any corresponding pair of bits,
Approach

HD

1

"Best Bits"

high

\gg

low

\gg

1

Levenshtein Distance

Weighted

high

high

Context-based

\gg

high

Enroll. Sam.

moderate

high

low

Accuracy

Comp. Cost

Majority Decoding

medium

medium

medium

Table 1. Proposed iris-biometric comparators.

recognition rates as well. Based on the idea that large connected matching parts of iris-codes indicate genuine samples and non-genuine samples tend to cause more randomized distortions according context-based match scores are extracted. Uhl and Wild [13] have proposed the use of a constrained version of the Levenshtein distance to tolerate e.g. segmentation inaccuracies or non-linear deformations by employing inexact matching. The technique was reported to be 4-5 times slower than the Hamming distance, but lied in same complexity class and increased recognition accuracy. Typically, minor improvements do not lead to significant performance gain with respect to accuracy. On the other hand, more complex comparison techniques do not provide a rapid comparison of biometric templates, yielding a trade-off between computational effort and recognition accuracy.

3 Reliable Bits and Shifting Variation

The proposed approaches build upon and extend previous works [11, 12]. In order to outperform traditional comparison techniques based on the fractional Hamming distance as well as approaches proposed in literature comparators have to fulfill three major requirements:

1. Minimize additional computational and storage cost.
2. Maximize performance gain with respect to accuracy.

Two different iris-biometric comparators which are referred to as bit reliability-driven template comparison (BRD) and shifting score fusion (SSF) are proposed:

3.1 Bit-Reliability Driven Template Comparison

At the time of enrollment an iris-code $I_K$ of length $N$ is obtained from subject $K$ and stored as biometric template, $I_K \in \{0,1\}^N$. Additionally, a user-specific reliability mask, denoted by $W_K$, of length $N$ of weights, which should indicate the reliability at each bit position of $I_K$ is stored. Initially, each value of $W_K$ is set to 1, i.e. all bits of $I_K$ reveal the same reliability. The fractional Hamming distance $HD_{KL}$ between two iris-codes $I_K$ and $I_L$ is calculated by,

$$HD(I_K, I_L) = \frac{\sum_{i=1}^{N} I_{Ki} \oplus I_{Li}}{N}$$

Figure 1. Iris recognition: (a) image of eye (b) detection of pupil and iris (c) unrolled iris texture (d) preprocessed iris texture (e) sample iris-code.

while the AND operator $\cap$ ensures that the compared bits are both deemed to have been uncorrupted by noise. The norms $\|\cdot\|$ of the resulting bit vector and of the AND’ed mask template are then measured in order to compute a fractional Hamming distance (HD) as a measure of the (dis-)similarity between pairs of binary feature vectors \{codeA, codeB\} and the according mask bit vectors \{maskA, maskB\} [2]:

$$HD = \frac{||\text{codeA} \oplus \text{codeB}) \cap \text{maskA} \cap \text{maskB}||}{||\text{maskA} \cap \text{maskB}||}. \quad (1)$$

Apart from the fractional Hamming distance several other techniques of how to compare iris-codes have been proposed. Table 1 summarizes proposed iris-biometric comparator with respect to computational cost, obtained accuracy and the number of required enrollment samples. To obtain a representative user-specific iris template during enrollment Davida et al. [3] and Ziauddin and Dailey [15] analyze several iris-codes. Davida et al. propose a majority decoding where the majority of bits is assigned to according bit positions in order to reduce Hamming distances between genuine iris-codes. Experimental results are omitted. Ziauddin and Dailey suggest to assign weights to each bit position, defining the stability of bits at according positions. Hollingsworth et al. [6] examined the consistency of bits in iris-codes resulting from different parts of the iris texture. The authors suggest to mask out so-called “fragile” bits for each user, where these bits are detected from several iris-code samples. In experimental results the authors achieve a significant performance gain. Obviously, applying more than one enrollment sample yields better recognition performance [4], however, commercial applications usually require single sample enrollment. Rathgeb and Uhl [9] have demonstrated that a context-based comparison of binary iris-codes increases
and if $HD(I_K, I_L)$ is below a predefined threshold $t$, $HD(I_K, I_L) < t$, successful authentication is yielded. Additionally, bit-masking can be applied to mask out bits of iris-codes which originate from parts of the iris which are occluded by eye lids or eye lashes (bit-masks need to be extracted at each feature extraction). In order to assign a degree of reliability to each bit position a weighted Hamming distance $WHD(I_K, I_L, W_K)$ is estimated. The mask $W_K$ of the claimed identity $K$ is applied to weight the comparison process by summing up weights at bit positions of mis-matching bits where the result is divided by the sum of all weights,

$$WHD(I_K, I_L, W_K) = \frac{\sum_{i=1}^{N}(I_{K_i} \oplus I_{L_i}) \cdot W_{K_i}}{\sum_{i=1}^{N} W_{K_i}}.$$  \hspace{1cm} (3)

Initially $||W_K||$ is $N$ since each bit of $W_K$ is 1, i.e. $WHD(I_K, I_L, W_K)$ is equivalent to $HD(I_K, I_L)$. In case of successful authentication ($WHD(I_K, I_L, W_K) < t$), for any $I_{K_i}$, the $i$th bit of the stored mask $W_K$ of user $K$ is updated,

$$W_{K_i} = \begin{cases} W_{K_i} + (I_{K_i} \oplus I_{L_i}), & \text{if } WHD(I_K, I_L, W_K) < t, \\ W_{K_i}, & \text{otherwise}. \end{cases} \hspace{1cm} (4)$$

Upon each successful authentication weights of $W_K$ are incremented at each position $i$ where $I_{K_i}$ is equal to $I_{L_i}$, resulting in the updated mask $W_K’$. An example for the proposed comparison procedure is illustrated in Figure 2. The predefined decision threshold $t$ has to be set up according to the applied iris recognition algorithm (in generic iris recognition algorithms this threshold is settled around $HD = 0.4$). The threshold $t$ remains unaltered for all subjects independent of the number of authentications. Since the weighted Hamming distance is calculated in relation to stored reliability masks inter-class distances are not expected to decrease, in contrast intra-class distances are expected to decrease due to the fact that unreliable bits are weighted less.

Extracted weights are user-specific and after several successful authentications the mask of each subject adapts to the iris-code extracted during enrollment. Less reliable parts of the iris-code, which result from parts of the iris texture which suffer from occlusions, tend to reveal low weights while others exhibit high weights. If a user-specific weighted comparison is performed recognition rates are expected to improve [15].

3.2 Shifting Score Fusion

In traditional iris comparison [1], in order to obtain a comparison score indicating the (dis-)similarity between two iris-codes, the minimum fractional Hamming distance over different bit shifts is calculated. The main reason for shifting one of the two paired iris-codes is to obtain a perfect alignment, i.e. to tolerate a certain amount of relative rotation between the two iris textures. It is a very natural approach to preserve the best match only, i.e. the minimum Hamming distance value over different shifts, because this value most likely corresponds to the best alignment of two codes. The impact of bit shifts on inter-class comparisons has been shown to just skew the distribution to the left and reduce its mean [2]. However, there is no evidence, that the other computed Hamming distance scores of less perfect alignments can not contribute to an improved recognition accuracy. Since calculating Hamming distances at different shifting positions is obligatory when comparing a single pair of iris-codes, it is interesting to look at the shifting variation, i.e. difference between maximum and minimum obtained Hamming distance scores. Let $s(l, m)$ denote an iris-code $l$ shifted by $m \in M_n = \{z \in Z : |z| \leq n\}$ bits and $HD(I_K, I_L)$ be the Hamming distance of two iris-codes, then the shifting variation (SV) score for two iris-codes $I_K, I_L$ is defined as:

$$SV(I_K, I_L) = \max_{m \in M_n}\left(HD(I_K, s(I_L, m))\right) - \min_{m \in M_n}\left(HD(I_K, s(I_L, m))\right).$$  \hspace{1cm} (5)

Since multiplication and addition with constant values does not alter the receiver operation characteristic (ROC) behavior of $SV$ scores, we perform some trivial operations to illustrate an interesting connection between $SV$ and sum rule fusion:

$$SSF(I_K, I_L) = \frac{1}{2} \left( 1 - \max_{m \in M_n}\left(HD(I_K, s(I_L, m))\right) \right) + \min_{m \in M_n}\left(HD(I_K, s(I_L, m))\right).$$  \hspace{1cm} (6)
### Table 2. Proposed iris-biometric comparators.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BRD Comp.</td>
<td>$WHD$</td>
<td>Mask $W$</td>
<td>1</td>
</tr>
<tr>
<td>SSF Comp.</td>
<td>$\max(\text{HD})$</td>
<td>$WHD + \max(\text{HD})$</td>
<td>Mask $W$</td>
</tr>
</tbody>
</table>

That is, shifting variation corresponds to a score level fusion of the minimum (i.e. best) Hamming distance and one minus the maximum (i.e. worst) Hamming distance using the sum rule [5]. By combining “best” and “worst” observed Hamming distance scores the variation between these scores is tracked, which represents a good indicator for genuine and imposter classes. The basic operation mode of the proposed technique is illustrated in Figure 3.

### 3.3 Composition of BRD and SSF Comparator

A composition of the bit reliability-driven comparator and the shifting score fusion comparator can be applied by fusing the minimum and one minus the maximum obtained weighted Hamming distance according to user-specific reliability masks. The composition of both comparators is defined as:

$$BSF(I_K, I_L, W_K) = \frac{1}{2} \left( \left(1 - \max_{m \in M_n} \left( WHD(I_K, s(I_L, m), W_K) \right) \right) + \min_{m \in M_n} \left( WHD(I_K, s(I_L, m), W_K) \right) \right) \right).$$

(7)

In case the estimated score is below a predefined threshold $t$ the according reliability mask is updated where matching bits are obtained from the shifting position of the minimum obtained Hamming distance.

### 3.4 Discussion

Focusing on the bit reliability-driven comparator additional computational cost is caused calculating the weighted Hamming distance according to a distinct reliability mask. However, the weighted Hamming distance can be efficiently estimated by successively adding values of reliability masks after bit-wise XORing iris-codes, i.e. the ones of the vector resulting from XORing a pair of iris-codes point at the positions of reliability mask values which need to be summed up. Furthermore, additional storage is required for reliability masks. In case bit-masks are applied it is suggested to integrate these to reliability masks by setting weights of stored reliability masks to zero, if bits of a bit-mask indicate that iris-code bits result from parts of the iris texture where some kind of distortions were detected. Thus, the comparison procedure does not need to be modified further and a bit-mask could be stored within a reliability mask.

Applying the shifting score fusion comparator improvement with respect to recognition accuracy comes at almost no additional computational cost, since a calculation of the minimum Hamming distance already involves a calculation of all Hamming distances in a specified range. The only required additional operation is a tracking of the maximum observed Hamming Distance (besides the minimum Hamming Distance) and an application of the fusion rule outlined before. Furthermore, this comparator does not require the storage of any additional information. Table 2 summarizes additional computational costs and additional storage costs of the proposed comparators. Both techniques, as well as a fusion of these, can be applied in a single sample enrollment scenario.

### 4 Experimental Evaluation

#### 4.1 Experimental Setup

Experiments are carried out on the CASIAv3-Interval iris database\(^1\) and on the IIT Delhi Iris Database v1\(^2\), two public available iris datasets. Both databases consist of good quality NIR illuminated indoor images, sample images of both databases are shown in Figure 4. These datasets are fused in order to obtain one comprehensive test set. The resulting test set consists of over 800 classes as shown in Table 3, allowing a comprehensive evaluation of the proposed systems.

At pre-processing, the pupil and the iris of a given sample are detected by applying Canny edge detection and Hough circle detection. Having localized the pupil and iris boundaries, the area between them is transformed to a normalized rectangular texture of $512 \times 64$ pixel, according to the “rubbersheet” approach by Daugman. In a final step, lighting across the texture is normalized using block-wise brightness estimation, see Figure 1. In the feature extraction stage we employ custom implementations of two different algorithms used to extract binary iris-codes, which have been evaluated in previous work [10]. The first feature extraction method follows an implementation by Masek [8] in which filters obtained from a Log-Gabor function are applied. The second one was proposed by Ma et al. [7]. Within this algorithm a dyadic wavelet transform is performed on signals obtained from texture stripes, and local

---

\(^1\)CASIA Iris Image Database, URL: http://www.idealtest.org

\(^2\)IIT Delhi Iris Database version 1.0, URL: http://www4.comp.polyu.edu.hk/˜csajaykr/IITDDatabase_Iris.htm
minima and maxima of two fixed subbands are encoded. For both feature extraction methods the final code is 10240 bit.

4.2 Performance Evaluation

Recognition accuracy is evaluated in terms of relations between genuine acceptance rate (GAR) and false acceptance rate (FAR), i.e. GARs at certain FARs. The GAR defines the "proportion of verification transactions with truthful claims of identity that are correctly confirmed", and the FAR defines the "proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed" (ISO/IEC FDIS 19795-1). As score distributions overlap the equal error rate (EER) of the system is defined. At each authentication 8 circular bit-shifts are performed in each direction for both feature extraction methods and proposed comparators.

The intra-class and inter-class score distributions for the original algorithm of Masek and Ma et al. are plotted in Figure 5. At a FAR of 0.01% the algorithm of Masek results in a GAR of 96.35% and the algorithm of Ma et al. results in a GAR of 98.52% applying the minimum obtained Hamming distance as (dis-)similarity metric. The corresponding EERs are 1.04% and 0.69%, respectively. Receiver operation characteristic (ROC) curves are plotted in Figure 8.

Obtained score distributions for the bit reliability-driven comparator are shown in Figure 6. In order to provide a meaningful performance evaluation inter-class comparisons are performed between each authentication sequence of genuine samples, i.e. the impact of reliability masks to inter-class comparisons is estimated. While the comparator does not significantly affect the inter-class score distribution, intra-class distances are further reduced achieving GARs at 0.01% FARs of 97.46% and 98.76% for both algorithms. Compared to the original score distributions of Figure 5 probability densities of intra-class distributions increase only slightly at high comparison scores (approximately 40%) and, on the other hand, increase at low scores (approximately 5%). Figure 7 illustrates, how both systems reveal improved performance rates of 0.91% and 0.62% EER after several successful authentication sequences. It is observed, that convergence is hit relatively fast, i.e. updating reliability masks may be stopped at a certain bit depth, reducing memory consumption (e.g. 2 or 3 bits per position).

For the proposed shifting score fusion comparator, score distributions are plotted in Figure 9. Accumulations of intra-class distributions at rather low scores result from the fact that the IITDv1 database is acquired under very favorable conditions (see Figure 4). Thus, small Hamming distance values (∼15-25%) between pairs of genuine subjects are further reduced through high variations of the minimum and maximum obtained scores, for both feature extraction methods. At a FAR of 0.01% GARs of 97.34% and 98.89% are obtained for the feature extraction of Masek and Ma et al., while the one minus maximum Hamming distance scores yield unpractical recognition rates for both algorithms. The ROC curves for the minimum Hamming distance, one minus the maximum Hamming distance and the proposed fusion of both metrics are plotted in Figure 8 where the shifting score fusion comparator obtains EERs of 0.93% and 0.59%, respectively.

Recognition rates can be further improved, in case both comparators are composed. Figure 10 shows the score distribution for intra-class and inter-class comparisons in the composition scenario. GARs of 97.77% and 98.91% are obtained

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>GAR (FAR=0.01) (%)</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>min(HD)</td>
<td>96.35</td>
<td>1.04</td>
</tr>
<tr>
<td>1 - max(HD)</td>
<td>18.16</td>
<td>16.91</td>
</tr>
<tr>
<td>Bit Reliability-driven</td>
<td>97.46</td>
<td>0.91</td>
</tr>
<tr>
<td>Sifting Score Fusion</td>
<td>97.46</td>
<td>0.93</td>
</tr>
<tr>
<td>BSF</td>
<td>97.77</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Table 4. Experimental results for the algorithm of Masek.
at 0.01% FAR. The according ROC curves are plotted in Figure 11 achieving EERs of 0.86% and 0.58%. Obtained performance rates for the entire experimental evaluation for the feature extraction method of Masek and Ma et al. are summarized in Tables 4 and 5. The bit reliability-driven comparator and the shifting score fusion comparator significantly improve the recognition accuracy of both feature extraction methods. Compared to the minimum Hamming distance which exhibits EERs of 1.04% and 0.69% a composition of both comparators improved the recognition accuracy to EERs of 0.86% and 0.58%.

5 Conclusion

In this work two different iris-biometric comparison techniques, which are referred to as bit reliability-driven (BRD) and shifting score fusion (SSF) comparator, are proposed. In experiments, which are carried out on a comprehensive data set, both approaches (and a composition of these) significantly improve the recognition accuracy of diverse iris recognition systems based on binary iris-codes. In contrast to enhanced comparison techniques proposed in literature emphasis is put systems based on binary iris-codes. In contrast to enhanced comparison techniques proposed in literature emphasis is put

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>GAR (FAR=0.01) (%)</th>
<th>EER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>min($HD$)</td>
<td>98.52</td>
<td>0.69</td>
</tr>
<tr>
<td>1 − max($HD$)</td>
<td>94.31</td>
<td>2.25</td>
</tr>
<tr>
<td>Bit Reliability-driven Sifting Score Fusion</td>
<td>98.76</td>
<td>0.62</td>
</tr>
<tr>
<td>BSF</td>
<td>98.91</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table 5. Experimental results for the algorithm of Ma et al.

References


