

A Contrario Detection of False Matches in Iris Recognition

Marcelo Mottalli, Mariano Tepper and Marta Mejail

Departamento de Computación, Universidad de Buenos Aires, Argentina

Abstract. The pattern of the human iris contains rich information which provides one of the most accurate methods for recognition of individuals. Identification through iris recognition is achieved by matching a biometric template generated from the texture of the iris against an existing database of templates. This relies on the assumption that the probability of two different iris generating similar templates is very low. This assumption opens a question: how can one be sure that two iris templates are similar because they were generated from the same iris and not because of some other random factor?

In this paper we introduce a novel technique for iris matching based on the *a contrario* framework, where two iris templates are decided to belong to the same iris according to the unlikelyness of the similarity between them. This method provides an intuitive detection thresholding technique, based on the probability of occurrence of the distance between two templates. We perform tests on different iris databases captured in heterogeneous environments and we show that the proposed identification method is more robust than the standard method based on the Hamming distance.

1 Introduction

The human iris is regarded as one of the richest biometric features. It is an external and visible part of the eye and contains enough information for the identification of a person with very low error rates. This fact allows to build an automatic system for identification of individuals via iris recognition.

The fundamentals of iris recognition were proposed by Daugman [4]. Starting from a picture of the human eye with enough quality, the picture is segmented to locate the iris and isolate it from the rest of the image. Once the iris is isolated, a *biometric template* is generated from its texture and the recognition takes place by comparing this template against a set of templates previously stored in a database.

The identification itself is done by global thresholding on the distance between two iris templates, for an appropriate distance function (usually the Hamming distance). Both templates are assumed to be generated by the same iris if and only if the distance between them is lower than a given threshold. This distance-based identification step provides good results, but three issues remain open:

1. Which value for the decision threshold provides an acceptable level of identification errors while not producing many rejections?
2. Once this threshold is defined, how can we be certain that two templates that are considered similar were obtained from the same iris and not by chance (for example, if both templates were generated from two low-quality images of different iris)?
3. Does a change in the database affect the identification results? For example, after a new iris template is added to the database or if the capture conditions change, is the same threshold still suitable?

One of the central problems in any automated recognition system is to have a tool to make a sound judgement about the accuracy of its output, however, the vast majority of related works concentrate on other parts of the recognition process. For example, for iris recognition, most works focus on segmentation or template generation algorithms, while template matching is often overlooked and is done by simple template distance thresholding [2].

In this work we analyse and improve iris template matching. For this, we propose a novel method based on an *a contrario* [3] detection of meaningful matches between templates (sections 2 and 3) and we analyse its performance (section 4).

2 The *A Contrario* Framework

The *a contrario* framework [3] provides a way to address the aforementioned issues. The framework is based on the Helmholtz principle that, in this case, states that a match between two iris templates is *meaningful* when it is not likely to occur in a context where noise overwhelms the information. In other words, *we detect by modeling what we do not want to detect*.

Formally, assume we have one query iris template Q and an iris template database $\mathcal{T} = \{T_i : 1 \leq i \leq N\}$ composed of N templates. Given an appropriate distance function, the distances between Q and each T_i , $d(Q, T_i)$ can be seen as observations of a random variable D that follows some unknown random process.

Under these assumptions, we can perform an hypothesis test for each pair (Q, T_i) , where we have two hypothesis:

- \mathcal{H}_0 (*null hypothesis*): $d(Q, T_i)$ is observed by chance, i.e. because the database is large.
- \mathcal{H}_1 (*alternate hypothesis*): $d(Q, T_i)$ is observed because of some causality, i.e. because both templates were generated from the same iris.

On one hand, $P(D \mid \mathcal{H}_0)$ (the probability distribution of the distances between iris templates) can be modeled with relative ease, even if the model is not perfectly realistic. On the other hand, it is not possible to model $P(D \mid \mathcal{H}_1)$ (the probability distribution of the distance when two templates belong to the same iris) because we assume no other information than the observed templates. Hence, the full hypothesis test cannot be done: we cannot control false negatives.

However, controlling false positives, or *false alarms*, under \mathcal{H}_0 is enough to decide whether a given iris template matches another template in the database. In other words, two templates are considered as belonging to the same iris *a contrario* if the distance between them has a very low probability of occurring by chance.

We formally state the *a contrario* hypothesis as:

\mathcal{H}_0 : $d(Q, T_i)$ are observations of the random variable D that follows some stochastic process.

3 Iris template matching

Now, given a pair of iris templates (Q, T_i) and the distance between them $\delta_i = d(Q, T_i)$ we want to know what is the probability of occurrence of their distance under \mathcal{H}_0 , i. e. the probability of false alarms $p_i = P(D \leq \delta_i \mid \mathcal{H}_0)$. If it is small enough, we can claim that Q and T_i are not similar just by chance [3].

Since we do not make any assumptions on the distribution of D , the probability p_i cannot be calculated directly. However, p_i can be estimated empirically over the database as the cumulative histogram of the distances between Q and each T_j , $1 \leq j \leq N$.

The expected *number of false alarms* (NFA) between Q and T_i under \mathcal{H}_0 on a database of size N is defined as:

$$\text{NFA}(Q, T_i) = N \cdot P(D \leq \delta_i \mid \mathcal{H}_0) \quad (1)$$

If $\text{NFA}(Q, T_i) \leq \varepsilon$ for a given ε , then the pair (Q, T_i) is said to be an ε -*meaningful match*.

We claim that Q and T_i are not similar just by chance if the match (Q, T_i) is ε -meaningful, i. e. they must be generated by the same iris. This provides a simple rule to decide whether a single pair of templates (Q, T_i) belongs to the same individual or not. Moreover, the expected number of ε -meaningful matches in a set of random matches can be proven to be smaller than ε [3]. Thus the threshold ε has a clear and intuitive meaning: it represents a bound on expected number of false matches (or *false alarms*) we're willing to accept when a template is compared against another template stored in a database if those two templates are considered to be generated by the same iris.

Basing the decision on the NFA we obtain a robust method, since the threshold ε is taken on the probability of false alarms instead of the distances themselves. Also, the same threshold ε is suitable for different database configurations because the probabilities are computed with respect to the entire database.

This provides us with a straightforward way of comparing the query template Q against a template T_i in the database: if $\text{NFA}(Q, T_i) \leq \varepsilon$, that is, if (Q, T_i) is an ε -meaningful match, it is assumed that Q and T_i are not similar just by chance, and it is assumed that they were generated by the same iris.

3.1 Partitioning the iris template

Estimating $P(D \leq \delta_i \mid \mathcal{H}_0)$ using the cumulative histogram of the distances between Q and every T_j poses a problem: since any bin of the histogram is at least $1/N$ (one occurrence over the database), this necessarily means that $P(D \leq \delta_i \mid \mathcal{H}_0) \geq 1/N$. Then:

$$\text{NFA}(Q, T_i) = N \cdot P(D \leq \delta_i \mid \mathcal{H}_0) \geq N \cdot \frac{1}{N} = 1 \quad (2)$$

That is, we would never be able to achieve a NFA lower than 1. Since having in average more than one false alarm for every query is not desirable, this is not an acceptable bound.

Following [9], we solve this problem by *partitioning* the iris template into C independent and non-overlapping parts, not necessarily of the same size. We refer to the k -th partition of the template T using the notation $T^{(k)}$, where $1 \leq k \leq C$.

Under these assumptions, given the query template Q and a template T_i in the database, we redefine

$$\delta_i = \max_{1 \leq k \leq C} \delta_i^{(k)}, \quad \text{where} \quad \delta_i^{(k)} = d(Q^{(k)}, T_i^{(k)}) \quad (3)$$

and $d(\cdot, \cdot)$ is a properly chosen distance.

As before, we formulate the *a contrario* hypothesis:

\mathcal{H}_0 : δ_i and $\delta_i^{(k)}$ are observations of identically distributed variables D and $D^{(k)}$, respectively, that follow some stochastic process.

Then, the probability of false alarms is:

$$\begin{aligned} P(D \leq \delta_i \mid \mathcal{H}_0) &= P\left(\max_{1 \leq k \leq C} D^{(k)} \leq \delta_i \mid \mathcal{H}_0\right) \\ &= \prod_{k=1}^C P\left(D^{(k)} \leq \delta_i \mid \mathcal{H}_0\right) \end{aligned} \quad (4)$$

and the NFA is:

$$\text{NFA}(Q, T_i) = N \cdot \prod_{k=1}^C P\left(D^{(k)} \leq \delta_i \mid \mathcal{H}_0\right) \quad (5)$$

Now, since we have C different histograms to compute the NFA, it turns out that $\text{NFA}(Q, T_i) \geq N \cdot \left(\frac{1}{N}\right)^C = \frac{1}{N^{C-1}}$. That is, we can achieve lower values for the NFA by partitioning the iris template into $C > 1$ independent parts.

It is important to note that, while most bits in the iris template present local spatial correlation [6], it is expected that this correlation will only be observable in one of the parts and it will not spread through all the parts, thus we can still claim that all the parts are independent between themselves.

3.2 Other considerations

In addition to partitioning the template, two other issues need to be addressed when matching the templates: the noise mask and rotation invariance.

The noise mask marks, for each template, which bits should not be matched. This can happen for example when those bits were calculated in regions obscured by eyelids or eyelashes, or affected by external illumination. When performing the *a contrario* matching, the masks will also need to be partitioned the same way as the templates. Since the distances between the parts are usually normalized by the number of available bits, equation 5 still holds.

Regarding rotation invariance, the most common solution is shifting the iris template on its x axis in both directions and keeping the lowest distance. The same method is used to calculate the distance between parts: one of the templates is rotated, partitioned, and the minimum distance is selected for each part.

4 Performance evaluation

We tested the proposed matching method using the publicly available Bath iris database [1] composed of 1000 images of 25 persons, with 40 iris images per person (20 for each eye). Iris templates are generated using Daugman's algorithm introduced in [4]: first, the iris images are segmented using an algorithm based on flexible contours [8]. Then, the iris texture is isolated from the image and normalized using the *ruber sheet* model. The texture is then filtered using a set of 2D Gabor filters, and resulting texture is then quantized to create a 2D binary iris template of 256×8 bits. Additionally, a *noise mask* is also associated to the template. This mask marks the bits of the template that were influenced by noise (eyelashes, eyelids and light reflections) and should not be considered on the matching stage.

As mentioned in section 3.1, the iris template must be partitioned to achieve a lower number of false alarms. The best results were achieved when the template is partitioned in four equal parts in the angular direction, exploiting the fact that the iris texture presents more variation in the angular direction than on the radial direction [5]. Thus, an iris template T is partitioned in four parts $T^{(k)}$, $k = 1 \dots 4$.

Recognition is performed as follows: given a query template Q and an iris template database $\mathcal{T} = \{T_i : 1 \leq i \leq N\}$, the distance $\delta_i^{(k)}$ between the k -th partition of Q and the k -th partition of T_i is calculated using the normalized Hamming distance. As mentioned before, one of the templates is shifted in the radial direction to compensate for rotations in the eye, and the minimum distance for each independent partition is used.

Then, a histogram of the values of $\delta_i^{(k)}$ is computed for each partition k , giving a total of four different histograms, as seen in Fig. 1. Once the histograms are obtained, the NFA between Q and each T_i is computed using Eq. 5.

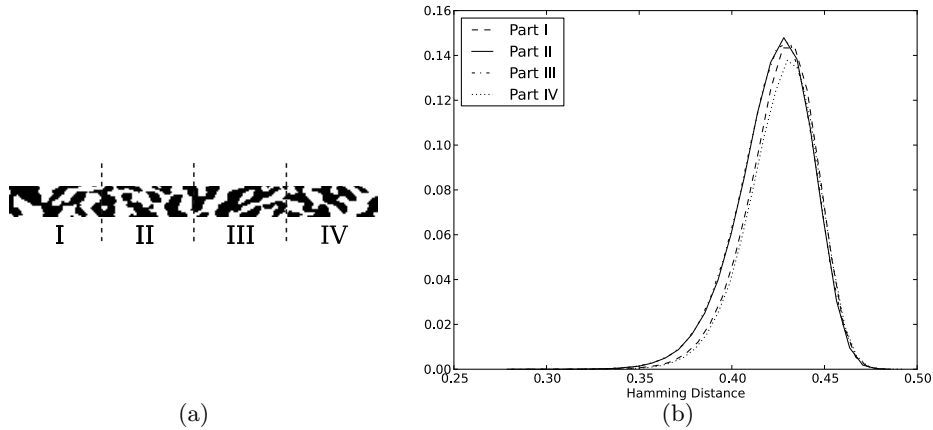


Fig. 1: Partitioning the templates. (a) The iris template is splitted into four partitions. (b) Four histograms obtained while comparing each partition of a given iris template against an entire database of partitioned templates (one for each partition).

4.1 Results

Following this approach we calculate the NFA between every pair of templates stored in the database using a leave-one-out cross-validation approach. We calculate the False Match Rate (FMR) for different values of ε and the corresponding False Nonmatch Rate (FNMR), where a pair of templates (Q, T_i) is considered a match if $\text{NFA}(Q, T_i) \leq \varepsilon$.

The Table 1 shows the FMR and FNMR for different values of ε . It can be observed that the FMR is bounded by ε . This was expected for the reasons mentioned in section 3, and it implies that by setting the value of ε beforehand we can effectively control the number of false matches of the system.

ε	FMR	FNMR
10^{-4}	0	9.07×10^{-2}
10^{-3}	1.34×10^{-6}	2.34×10^{-2}
10^{-2}	5.49×10^{-5}	5.55×10^{-3}
10^{-1}	2.24×10^{-3}	1.49×10^{-3}

Table 1: FMR and FNMR for different values of ε

Additionally, the performance of the system is analysed and compared in Table 2 by using the Equal Error Rate (EER) and the decidability d' , which is a dimensionless measure of the distance between the distributions of the inter-class

and intra-class comparisons. It is also compared against recent results obtained by Hollingsworth et al [7].

Method	EER	d'
Hamming distance	9.7×10^{-3}	3.49
Hollingsworth et al [7]	6.99×10^{-3}	6.06
NFA matching	1.5×10^{-3}	7.72

Table 2: Performance results using different methods

It can be observed that the EER has decreased and at the same time the decidability has been more than doubled with respect to using the Hamming distance alone. This means that the separation of the distribution between intra-class and inter-class comparisons is greatly increased using the NFA rather than the Hamming distance, as can be seen on Fig. 2.

It should be noted that the *a contrario* method only allows us to set a limit on the number of false matches, and it has no control over the false nonmatches. However, the results indicate that this method has considerably better performance than using the Hamming distance alone.

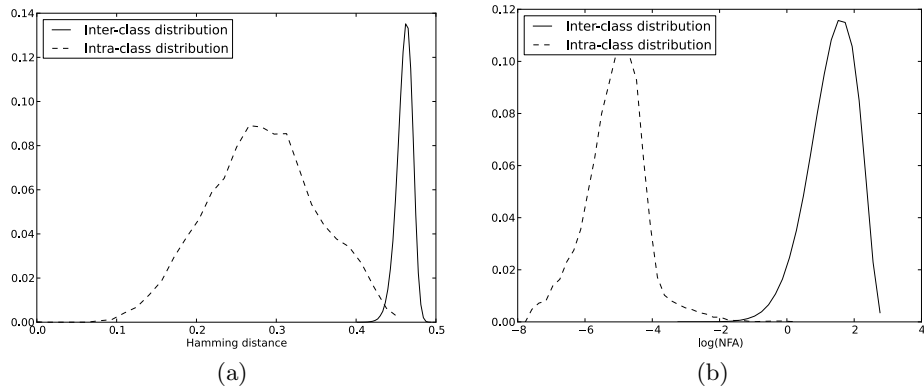


Fig. 2: Comparison of inter-class and intra-class distances (a) using the Hamming distance and (b) using the NFA (the NFA is plotted in logarithmic scale for visualization purposes). A greater separation between classes can be seen using the NFA.

5 Conclusions

The novel contribution in this work is to apply the *a contrario* detection framework for template based iris recognition. The *a contrario* matching has the important advantage of not requiring to compute by hand an “optimum” threshold. It is possible to directly define the acceptable number of false matches for the system with *no* a priori information on the database.

The proposed method proved to give results consistent with the theory presented in section 2. The number of false matches observed in the tests were consistent with the expected number of false alarms defined beforehand with the parameter ε .

Since the *a contrario* method does not make assumptions about the method used for encoding the iris texture or for measuring the distance between templates, it can be adapted for different scenarios. This is an important feature, since most works deal with increasing the separation between the classes by improving the segmentation or codification algorithms. We show that it is possible to achieve this just by changing the matching algorithm, which is often overlooked in the bibliography.

Also, since the NFA is calculated against all the templates in the database, it is a robust measure that is not affected by changes in the capture conditions. For example, if the database is composed by templates generated from blurry or low quality images, then a query template generated in similar conditions will be more likely to be rejected by our method.

References

1. Smart sensors iris database. <http://www.irisbase.com/>
2. Bowyer, K., Hollingsworth, H., Flynn, P.: Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding* 110, 281–307 (October 2007)
3. Cao, F., Lisani, J.L., Morel, J.M., Musé, P., Sur, F.: *A Theory of Shape Identification*, vol. 1948. Springer (2008)
4. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(11), 1148–1161 (November 1993)
5. Gentile, J.E., Ratha, N., Connell, J.: Slic: Short-length iris codes. In: *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*. pp. 1–5 (Sept 2009)
6. Hollingsworth, K., Bowyer, K., P., F.: The best bits in an iris code. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1–10 (2009)
7. Hollingsworth, K., Peters, T., Bowyer, K., Flynn, P.: Iris recognition using signal-level fusion of frames from video. *IEEE Transactions on Information Forensics and Security* 4(4), 837–848 (December 2009)
8. Mottalli, M., Mejail, M., Jacobo-Berles, J.: Flexible image segmentation and quality assessment for real-time iris recognition. In: *16th IEEE International Conference on Image Processing (ICIP)*. pp. 1941–1944 (November 2009)
9. Proena, H., Alexandre, L.: Toward non-cooperative iris recognition: A classification approach using multiple signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4), 607–612 (April 2007)