An Application of Random and Hammersley Sampling Methods to Iris Recognition

Luis E. Garza Castañón[1], Saúl Montes de Oca[2], and Rubén Morales-Menéndez[1]

[1] Department of Mechatronics and Automation, ITESM Monterrey Campus
{legarza, rmm}@itesm.mx
[2] Automation Graduate Program Student, ITESM Monterrey Campus
saul_montesdeoca@yahoo.com.mx
Av. Eugenio Garza Sada Sur No. 2501
Monterrey, N.L. 64849 México

Abstract. We present a new approach for iris recognition based on a sampling scheme. Iris recognition is a method to identify persons, based on the analysis of the eye iris. A typical iris recognition system is composed of four phases: image acquisition and preprocessing, iris localization and extraction, iris features characterization, and comparison and matching. The main contribution in this work is in the step of characterization of iris features by using sampling methods and accumulated histograms. These histograms are built from data coming from sampled subimages of iris. In the comparison and matching step, a comparison is made between accumulated histograms of couples of iris samples, and a decision of accept/reject is taken based on samples differences and on a threshold calculated experimentally. We tested two sampling methods: random and Hammersley, and conduct experiments with UBIRIS database. Under certain conditions we found a rate of successful identifications in the order of 100 %.

1 Introduction

Iris recognition is related to the area of biometrics. The main intention of biometrics is to provide reliable automatic recognition of individuals based on the measuring of a physical characteristic or personal trait. Biometrics can be used for access control to restricted areas, such as airports or military installations, access to personal equipments such as laptops and cellular phones, and public applications, such as banking operations [11]. A wide variety of biometrics systems have been deployed and resulting systems include different human features such as: face, fingerprint, hand shape, palmprint, signature, voice and iris [7]. The last one may provide the best solution by offering a much more discriminating power than the others biometrics. Specific characteristics of iris such as a data-rich structure, genetic independence, stability over time and physical protection, makes the use of iris as biometric well recognized.

M. Ali and R. Dapoigny (Eds.): IEA/AIE 2006, LNAI 4031, pp. 520–529, 2006.
© Springer-Verlag Berlin Heidelberg 2006

In last years, there have been different implementations of iris recognition systems. Daugman's system [2] used multiscale quadrature wavelets (Gabor filters) to extract texture phase structure information of the iris to generate a 2,048-bit iris code and compared the difference between a pair of iris representations by their Hamming distance. In [10] iris features are extracted by applying a dyadic wavelet transform with null intersections. To characterize the texture of the iris, Boles and Boashash [1] calculated a one dimension wavelet transform at various resolution levels of a concentric circle on an iris image. In this case the iris matching step was based on two dissimilarity functions. Wildes [13] represented the iris texture with a Laplacian pyramid constructed with four different resolution levels and used the normalized correlation to determine whether the input image and the model image are from the same class. A Similar method to Daugman's is reported in [9], but using edge detection approach to localize the iris, and techniques to deal with illumination variations, such as histogram equalization and feature characterization by average absolute deviation. In [6] a new method is presented to remove noise in iris images, such as eyelashes, pupil, eyelids and reflections. The approach is based on the fusion of edge and region information. In [3] an iris recognition approach based on mutual information is developed. In that work, couples of iris samples are geometrically aligned by maximizing their mutual information and subsequently recognized.

Most of the previous work rely on some sort of filtering or transformation of iris data, and this step can be time consuming. We present an approach where direct information from selected areas of iris is used to build a set of features. In our work we apply standard techniques as integro-differential operators to locate the iris, and histogram equalization over extracted iris area to compensate for illumination variations. The main contribution of our work is in the feature extraction step, where we sample square subimages of iris, and build an acummulated histogram for each subimage. Every iris is represented by a set of features, and each feature is an accumulated histogram. The number and the size of square subimages were calculated experimentally. The comparison between iris sample and irises in database is done by computing the Euclidean distance between histograms, and a decision to accept or reject iris sample is taken based on a threshold calculated also experimentally.

We ran experiments with UBIRIS database [12] and tested two different sampling schemes: random and Hammersley. We have found under certain conditions a rate of succesful matching in the order of 100 %.

2 The Proposed Approach

The implementation of our approach relies on the use of colored eyes images from UBIRIS database. Eyes images include samples where iris is free from any occlusion, and others with moderate obstruction from eyelids and eyelashes (Fig. 1). We transform the images color representation to just grey level pixels, because this process is sufficient to reveal the relevant features of iris.

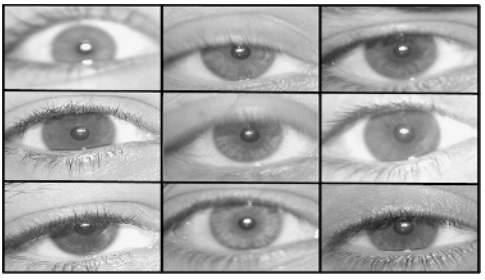


Fig. 1. Eyes samples with noise (moderate obstruction)

2.1 Iris Localization

The search of limbic and pupilar limits is achieved with the use of the integro-differential operator shown in eqn 1.

$$\max_{(r,x_0,y_0)} = \left| \frac{\partial}{\partial r} G(r) * \oint_{r,x_c,y_c} \frac{I(x,y)}{2\pi r} ds \right| \tag{1}$$

where $I(x, y)$ is an image containing an eye.

The operator behaves as an iterative circular edge detector, and searches over the image domain (x, y) for the maximum in the partial derivative with respect to an increasing radius r, of the normalized contour integral of $I(x, y)$ along a circular arc ds of radius r and center coordinates (x_0, y_0). The symbol $*$ denotes convolution and $G_\sigma(r)$ is a smoothimg function, tipically a Gaussian of scale σ.

This operator behaves well in most cases with moderate noise conditions, but requires some fine tuning of parameters, in order to deal with pupil reflections, obscure eyes and excess of illumination. Heavy occlusion of iris by eyelashes or eyelids needs to be handled by other methods. In our work, eye images with heavy occlusion were discarded.

The extracted iris image has to be normalized to compensate for pupil dilation and contraction under illumination variations. This process is achieved by a transformation from cartesian to polar coordinates, using equations 2 and 3.

$$x(r, \theta) = (1 - r)x_p(\theta) + rx_s(\theta) \tag{2}$$

$$y(r, \theta) = (1 - r)y_p(\theta) + ry_s(\theta) \tag{3}$$

where $x(r, \theta)$ and $y(r, \theta)$ are defined as a linear combination of pupil limits ($x_p(\theta)$, $y_p(\theta)$) and limbic limits ($x_s(\theta)$, $y_s(\theta)$). r is defined in the interval $[0, 1]$, and θ in the interval $[0, 2\pi]$.

2.2 Strip Processing

The iris image strip obtained in previous step, is processed by using an histogram equalization method, to compensate for differences in illumination conditions.

The main objective in this method is that all grey levels (ranging from 0 to 255) have the same number of pixels. Histogram equalization is obtained by working with the accumulated histogram, shown in eqn 4.

$$H(i) = \sum_{k=0}^{i} h(k) \tag{4}$$

where $h(k)$ is the histogram of the kth grey level.

A flat histogram, where every grey level has the same number of pixels, can be obtained by eqn 5.

$$G(i') = (i' + 1)\frac{N_r * M_c}{256} \tag{5}$$

Where N_r and M_c are the image dimensions and 256 is the number of grey levels.

2.3 Iris Sampling

Sampling strategies have been applied recently with certain degree of success in texture synthesis [4, 8]. The main idea in our work is to extract relevant features of iris, by sampling a set of subimages from the whole image. In our case we tested two sampling strategies: random and Hammersley. In the first strategy we randomly generate a set of coordinates (x, y) where the subimage is centered. In the second strategy we generate a more uniform sampling of coordinates by using Hammersley sequence sampling.

Hammersley sampling [5] is part of to the *quasi*-Monte Carlo methods, or low-discrepancy sampling family. The *quasi*-prefix refers to a sampling approach that employs a deterministic algorithm to generate samples in an n-dimensional space. These points are as close as possible to a uniform sampling. Discrepancy refers to a quantitative measure of how much the distribution of samples deviates from an ideal uniform distribution (i.e. low-discrepancy is a desired feature).

Quasi-Monte Carlo methods as Hammersley sequences show lower error bound in multidimensional problems such as integration. Error bounds for pseudo-Monte Carlo Methods are $\mathcal{O}(N^{-1/2})$, and for classical integration is $\mathcal{O}(N^{-2/n})$. However, Hammersley sequences has lower error bound with $\mathcal{O}(N^{-1}(log_{10}N)^{n-1})$ where N is the number of samples and n is the dimension of the design space. Usually, as n grows up Hammersley shows better results, note that pseudo-Monte Carlo error bound is a probabilistic bound.

Hammersley Algorithm. This algorithm generates N points using the radix-R notation of an integer. A specific integer, q, in radix-R notation can be represented as in eqn 6.

$$\begin{aligned} q &= q_m q_{m-1} \cdots q_2 q_1 q_0 \\ q &= q_0 + q_1 R + q_2 R^2 + \cdots + q_m R^m \end{aligned} \tag{6}$$

where $m = [log_R(q)] = [ln(q)/ln(R)]$, and the square brackets, [ ], denote the integer portion of the number inside the brackets. The inverse radix number

function constructs a unique number on the interval $[0, 1]$ by reversing the order of the digits of q around the decimal point. The inverse radix number function is:

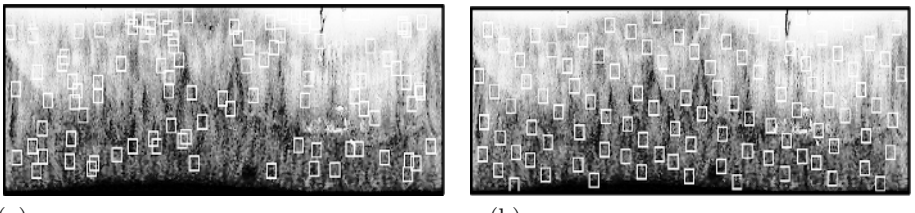
$$\phi_R(q) = q_0 q_1 q_2 ... q_m$$
$$\phi_R(q) = q_0 R^{-1} + q_1 R^{-2} + \cdots + q_m R^{-m-1} \tag{7}$$

The Hammersley sequence of $n-$dimensional points is generated as

$$x_n(q) = \left(\frac{q}{N}, \phi_{R_1}(q), \phi_{R_2}(q), \ldots, \phi_{R_{n-1}}(q)\right) \tag{8}$$

where $q = 0, 1, 2, \ldots, N-1$ and the values for $R_1, R_2, \ldots, R_{n-1}$ are the first $n-1$ prime numbers $(2, 3, 5, 7, 11, 13, 17, \ldots)$. This algorithm generates a set of N points in the $n-$dimensional design space $[0, 1]^N$.

Fig. 2 shows the samples generated by a random and Hammersley algorithms over iris strip. Uniformity properties can be appreciated in a qualitative fashion. Hammersley points have better uniformity properties because the algorithm exhibits an optimal design for placing n points on a k-dimensional hypercube.



(a) (b)

Fig. 2. Qualitative comparison of random (a) and Hammersley samplings (b), over an iris strip

2.4 Comparison and Matching

In our method, the iris features are represented by a set of accumulated histograms computed from sampled selected square subimages of iris strip. An accumulated histogram represents a feature and is built by using equation 4. The complete iris is represented by a set of accumulated histograms, one of them for every subimage. The optimal size of the number of features and subimage size, were determined empirically by experiments. A decision to accept or reject the iris sample is done according to the minimum Euclidean distance calculated from the comparison of iris sample and irises database, and also according to a threshold. Figure 3 shows the structure of this phase.

We can formalize the method as follows: Let I be an image, representing an iris strip, and $p \in I$ be a pixel and $\omega(p) \subset I$ be an square image patch of width S_f centered at p. We built iris features by forming a set of accumulated histograms with k bins, $P_m(i)$ $i = 1 \cdots k$, $m = 1 \cdots N_f$, from a set of N_f sampled patches

$$\{\Omega = \omega(p_1), \cdots, \omega_(p_{Nf})\}.$$

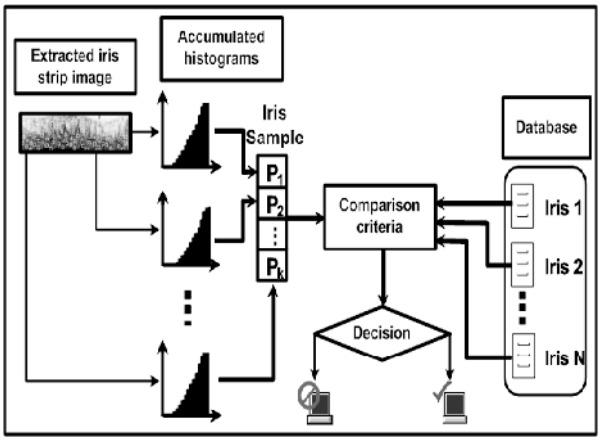


Fig. 3. The process of comparison and matching

The features of every iris in database are represented by a set of accumulated histograms $\{P_{DB_1}(i), P_{DB_2}(i), \cdots, P_{DB_{Nf}}(i)\}$.

An iris sample features set $\{P_{SMP_1}(i), P_{SMP_2}(i), \cdots, P_{SMP_{Nf}}(i)\}$ is compared against every iris features set in a database of size ℓ, according to the norm:

$$L = \min_n \sqrt{\sum_j \sum_i (P_{DB_j}(i) - P_{SMP_j}(i))^2} \tag{9}$$

with $n = 1 \cdots \ell$, $i = 1 \cdots k$, $j = 1 \cdots N_f$.

A decision to accept o reject the sample is taken based on the rule $L < \delta$, with δ being a threshold computed experimentally.

3 Experiments

Experiments were ran for UBIRIS database. Images with too much occlusion and noise were discarded, because the difficulty to locate the iris region with integro-differential operators. Then, our experimental database was built with 1013 samples coming from 173 users. With this database, we perform some experiments with 100 % of samples and others experiments where worst user samples were discarded. Table 1 shows the different databases used.

In table 2 the results of experiments from random sampling are shown. Every result is the average of 10 experiments performed. We can see that best results were obtained for cleaner databases. In table 3 the results of experiments from Hammersley sampling are shown. We can see that better results were obtained by using Hammersley sampling than for random sampling.

In Fig. 4 we can see the ROC curves for the different databases used in experiments with Hammersley sampling. databases with cleaner iris samples

Table 1. Databases used of experiments. First column refers to the percent of database used. For instance, 90 % means that 10 % of worst user samples were discarded. Second column refers to total number of iris samples, third column refers to the number of users, fourth column refers to the number of samples used to calculate the decision threshold, and fifth column refers to the total number of samples used for testing.

DB Size (%)	# of iris	# of users	# Threshold samples	# Test samples
100	1013	173	52	788
90	912	173	46	693
80	811	173	42	596
70	710	173	36	501
50	507	173	26	308

Table 2. Results of experiments (in % of accuracy) with random sampling, and different size (S_f) and number (N_f) of subimages

DB Size(%)	$N_f = 100$ $S_f = 5$	$N_f = 100$ $S_f = 15$	$N_f = 100$ $S_f = 25$	$N_f = 150$ $S_f = 5$	$N_f = 150$ $S_f = 15$	$N_f = 150$ $S_f = 25$	$N_f = 200$ $S_f = 5$	$N_f = 200$ $S_f = 15$	$N_f = 200$ $S_f = 25$
100	90.8	91.14	92.72	90.81	92.79	92.38	91.83	92.07	92.98
90	92.48	93.96	93.78	93.32	94.23	94.05	94.34	95.29	93.99
80	96.58	97.40	96.83	98.59	98.59	97.59	98.18	98.4	98.4
70	98.25	99.48	99.44	99.63	99.59	99.52	99.89	99.7	99.85
50	99.94	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3. Results of experiments (in % of accuracy) with Hammersley sampling, and different size (S_f) and number (N_f) of subimages. Highlighted results show the improvement over random sampling strategy.

DB Size(%)	$N_f = 100$ $S_f = 5$	$N_f = 100$ $S_f = 15$	$N_f = 100$ $S_f = 25$	$N_f = 150$ $S_f = 5$	$N_f = 150$ $S_f = 15$	$N_f = 150$ $S_f = 25$	$N_f = 200$ $S_f = 5$	$N_f = 200$ $S_f = 15$	$N_f = 200$ $S_f = 25$
100	90.47	92.12	93.14	90.48	93.0	93.64	92.12	92.76	93.64
90	92.56	93.85	94.42	94.29	94.64	94.42	94.21	95.07	94.37
80	97.87	98.72	98.34	98.84	99.0	98.5	99.09	99.13	98.31
70	99.48	99.85	99.52	99.85	99.74	99.59	99.96	100.0	99.85
50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

reflects better results. In Fig. 5 we can see the distribution curves of accumulated histograms distance for two databases used in experiments with Hammersley sampling. The Overlapping between distribution curves ing Fig. 5 (b) leads to worst results.

4 Comparison to Previous Work

Daugman's system [2] has been tested thoroughly with databases containing thousands of samples, and reports of 100 % of accuracy have been given. In [10],

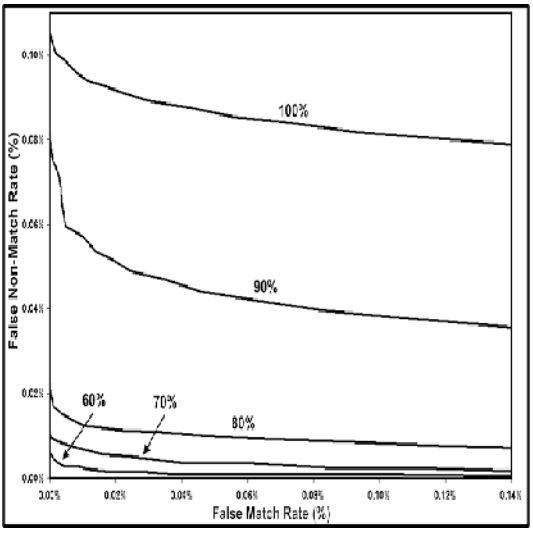
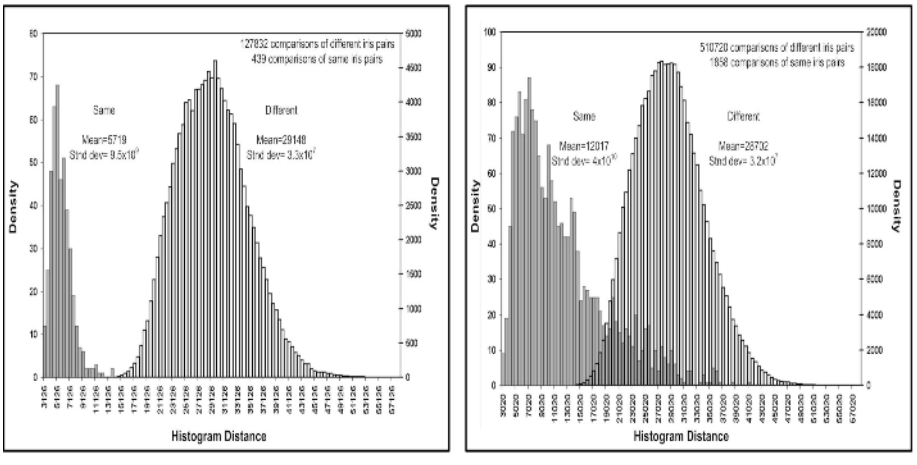


Fig. 4. Hammersley sampling ROC curve for different database sizes



(a) (b)

Fig. 5. Distribution of accumulated histograms distance (L), in Hammersley sampling, for 50 % database (a), and 100 % database (b)

the experimental results given are in the order of 97.9 %, by working with a database of 100 samples from 10 persons. Boles and Boashash [1] report best results in the order of 100 % but working with very small sets of images. Wildes [13] report results in the order of 100 % by working with a database of 600 samples coming from 40 individuals. In [9], a report is given about a performance

of 99.09 % in experiments with a database of 500 iris images from 25 individuals. In [6] the results are between 98 % and 99 % by working with CASIA database (2255 samples from 213 subjects). In [3] best results are in the order of 99.05 % with a database of 384 images from 64 persons. We are competitive with most of the methods mentiond above. In comparison our best results, are in the order of 100 % working with a database of 308 samples coming from 173 persons, and in the order of 99 % and higher with 501 samples.

5 Conclusions and Future Works

A new approach for iris recognition has been presented. The novel contribution relies on the feature characterization of iris by the use of two sampling techniques, random and Hammersley, and accumulated histograms. Although experimental results show better results for databases with cleaner eyes images, we claim that our method will conduct to an improved and faster approach where just a few samples of subimages taken from irises in database will be necessary to discard most of them in a first step, and then concentrate the effort of comparison and matching in a very reduced set of iris samples. This potential approach will lead us to test bigger databases.

References

1. W. Boles and B. Boashash, "Iris Recognition for Biometric Identification using dyadic wavelet transform zero-crossing", *IEEE Transactions on Signal Processing*, Vol. 46, No. 4, 1998, pp. 1185-1188.
2. J. Daugman, "How Iris Recognition Works", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1, 2004, pp. 21-30.
3. M. Dobes, L. Machala, P. Tichasvky, and J. Pospisil, "Human Eye Iris Recognition Using The Mutual Information", *Optik*, No. 9, 2004, pp. 399-404.
4. A. Efros, and T. Leung, "Texture Synthesis by Non-Parametric Sampling", in *Proceedings of the 7th IEEE International Conference on Computer Vision*, September 1999, Vol. 2, pp. 1033-1038.
5. J. Hammersley, "Monte Carlo Methods for Solving Multivariate Problems", *Annals of New York Academy of Science*, 1960, No. 86, pp. 844-874.
6. J. Huang, Y. Wang, T. Tan, and J. Cui, "A New Iris Segmentation Method for Iris Recognition System", In *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 554-557.
7. A. Jain, A. Ross, A. Prabhakar, "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 1, 2004, pp. 4-20.
8. L. Liang, C. Liu, Y. Xu, B. Guo, and H. Shum, "Real-time Texture Synthesis by Patch-based Sampling", *ACM Transactions on Graphics*, Vol. 20, No. 3, July 2001, pp. 127-150.
9. L. Ma, Y. Wang, T. Tan, and D. Zhang, "Personal Identification Based on Iris Texture Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, 2003, pp. 1519 - 1533.

10. D. de Martin-Roche, C. Sanchez-Avila, and R. Sanchez-Reillo, "Iris Recognition for Biometric Identification using dyadic wavelet transform zero-crossing", In *Proceedings of the IEEE 35th International Conference on Security Technology*, 2001, pp. 272-277.
11. M. Negin, Chmielewski T., Salganicoff M., Camus T., Cahn U., Venetianer P., and Zhang G. "An Iris Biometric System for Public and Personal Use ", *Computer*, Vol. 33, No. 2, 2000, pp. 70-75.
12. H. Proenca, and L. Alexandre,"UBIRIS: A Noisy Iris Image Database", in *Proceedings of the International Conference on Image Analysis and Processing 2005*, Vol. 1, pp. 970-977.
13. R. Wildes, "Iris Recognition: An Emerging Biometric Technology", *Proceedings of the IEEE*, Vol. 85, No. 9, 1997, pp. 1348-1363.