

Mobile Iris CHallenge Evaluation II: results from the ICPR competition

Modesto Castrillón-Santana
Universidad de Las Palmas de Gran Canaria, Spain
modesto.castrillon@ulpgc.es

Maria De Marsico
Sapienza University of Rome, Italy
demarsico@di.uniroma1.it

Michele Nappi
University of Salerno, Italy
mnappi@unisa.it

Fabio Narducci
National Council of Research, Naples, Italy
fabio.narducci@icar.cnr.it

Hugo Proença
Universidade da Beira Interior, Portugal
hugocmp@ubi.pt

Abstract

The growing interest for mobile biometrics stems from the increasing need to secure personal data and services, which are often stored or accessed from there. Modern user mobile devices, with acquisition and computation resources to support related operations, are nowadays widely available. This makes this research topic very attracting and promising. Iris recognition plays a major role in this scenario. However, mobile biometrics still suffer from some hindering factors. The resolution of captured images and the computational power are not comparable to desktop systems yet. Furthermore, the acquisition setting is generally uncontrolled, with users who are not that expert to autonomously generate biometric samples of sufficient quality. Mobile Iris CHallenge Evaluation aims at providing a testbed to assess the progress of mobile iris recognition, and to evaluate the extent of its present limitations. This paper presents the results of the competition launched at the 2016 edition of the International Conference on Pattern Recognition (ICPR).

1. Introduction

Mobile biometric recognition by personal and/or wearable devices is the most advanced frontier for secure use of data and services. It provides a further application for user mobile equipment, which are ubiquitous nowadays. Moreover, it extends the functionality and capabilities of a traditional biometric identification systems, by allowing capture of biometric traits in any

place. Captured information can be compared with that stored either on the device itself, or even within RFID tags, smartcards or machine readable identification documents (IDs) for single user verification purposes, or on a remote server, for identification in a set of relevant subjects. Mobile devices used for biometric recognition must incorporate all necessary hardware equipment and software applications for capturing and processing data from one or more biometric traits. Moreover, they must be designed for intuitive operation, especially if it is not planned to assist users during sample capture. The captured data must be suitably converted by software into digital templates for storage and matching against other records. Feature extraction, storing and processing, require non negligible resources. Therefore, notwithstanding the continuous advances in technology and resources, transferring all the phases of biometric processing on a mobile device calls for faster and lighter procedures, and for more efficient storage.

Iris is a natural candidate for mobile biometric recognition for two main reasons: iris acquisition is little intrusive, and iris codes are among the less expensive templates from the storage point of view. As for other biometrics, even research results regarding related techniques underwent a quick progress, from the pioneering work by Daugman [1] and Wildes [10], mostly pertaining controlled settings, to the most advanced use of deep learning [5], through the recent challenges addressing iris recognition in less controlled and/or mobile settings [3] [8]. Most current iris recognition systems still require that subjects stand close to the capture device (about 1m or less) and look towards it for about 3s. The first iris biometric competitions have relied on

images acquired in these conditions. Among the most well-known, we can mention the Iris Challenge Evaluation (ICE) [6]. Proença and Alexandre [7] have rather tackled the problem of noisy iris recognition.

Even for this biometric trait, techniques targeted at mobile devices must be suitably adapted to their resources, and call for light processing possibly avoiding complex mathematical processing (e.g., see [2]).

The aim of the contest Mobile Iris CHallenge Evaluation II (MICHE-II), launched in conjunction with ICPR 2016, was to collect relevant contributions to the field of mobile iris recognition in both academy and industry. This paper presents the comparison of the seven best performing algorithms.

2. The challenge setup

The Noisy Iris Challenge Evaluation (NICE I) exploited images captured in unconstrained imaging environments, to evaluate how noise affects iris segmentation¹. To this aim, the proposed iris dataset UBIRIS.v2 [9] contains data captured in the visible wavelength, at-a-distance, namely between 4 and 8m, and on the move². The results achieved by participant methods confirm the major impact that uncontrolled conditions have on recognition performance. Recognition of visible wavelength (VW) iris images captured at-a-distance and on the move with less controlled protocols was the target of the further NICE II contest [7]. VW images usually contain much more features than the traditionally used near infrared (NIR) images, however they are also more seriously affected by many noisy artifacts, and their processing suffers from dark pigmentation.

MICHE-I challenge [3] moved to issues related to iris acquisition by mobile devices. In this new context, it is assumed that the subject to be recognized generally autonomously operate the capturing device. MICHE-I provided a dataset suitable to assess the performance of biometric applications related in this specific set up. Two opposite considerations hold. Capturing accuracy/quality may be enhanced due to the usually short distance (the length of a human arm) and the user quite naturally tends to assume a frontal pose. However, the quality of the captured image can suffer from possible lower resolution, motion blur and illumination distortions, caused by both the kind of device and by the lack of control on user capture operation. These possible issues call for robust detection/segmentation and encoding procedures. It is worth noticing that the accuracy of the latter is heavily affected by the quality of the former. The composition of the dataset used for MICHE-II

challenge is basically the same of MICHE-I, with the addition of new unpublished images to be used mostly in the competitors ranking process.

2.1. MICHE-II database

The aim of MICHE-I dataset, publicly available to the scientific community and representing the core of the still unpublished MICHE-II dataset, is to represent the starting core of a wider dataset to be collected thanks to a possible crowd-sourcing approach. This should better allow unbiased assessment of cross-demographic robustness, as well as of cross-device and cross-setting interoperability of recognition procedures. In particular, the dataset allows to measure both the ability to match samples of the same subject acquired with different devices, and in general the ability to handle samples acquired by devices with different characteristics without a significant performance degradation. Moreover performances in different illumination conditions are evaluated. We now detail how MICHE-I dataset differs from the most popular existing iris datasets. The Chinese Academy of Sciences collected the first group of publicly available datasets dealing with iris images, namely CASIA-Iris, that has been updated from CASIA-IrisV1 to CASIA-IrisV4 since 2002. Its images are collected under NIR illumination or synthesized. For these reasons, they cannot be reliably used for assessing methods entailing mobile acquisition, unless NIR sensors get more common in mobile devices too. Similar considerations hold for images used for ICE competitions. On the contrary, UBIRIS datasets, acquired and made available from SOCIA Lab at University of Beira Interior (Portugal), are captured in visible light and uncontrolled conditions. However, acquisition is carried out by cameras with a better resolution than average sensors built in mobile devices. MICHE-I is a dataset of iris images acquired in visible light by different mobile devices. The key features of the dataset are: (1) sufficiently large population of users; (2) the use of different mobile devices for the acquisition; (3) the realistic simulation of the acquisition process including different sources of noise; (4) and several acquisition sessions separated in time. A full metadata annotation completes the dataset. The subjects involved in data collection were asked to behave as they would do by using a real system, e.g., subjects wearing eyeglasses could either choose to remove or keep them. They had to take self-images of one of their irises, by holding the mobile device by themselves, with a minimum of four shots for each camera (two out of three devices were equipped with two cameras with different resolutions) and acquisition mode (indoor, outdoor). Indoor acquisition was

¹<http://nice1.di.ubi.pt>

²<http://nice2.di.ubi.pt/>

affected by various sources of artificial light, sometimes combined with natural light sources. Outdoor acquisition was carried out using natural light only. For each subject only one of the two irises was acquired. Three kinds of smartphones and tablets were used:

- Galaxy Samsung IV (GS4): Google Android Operating System, CMOS posterior camera with 13 Megapixel (72 dpi) and 2322×4128 resolution, CMOS anterior camera with 2 Megapixel (72 dpi) and 1080×1920 resolution;
- iPhone5 (IP5): Apple iOS Operating System, iSight posterior camera with 8 Megapixels (72 dpi) and 1536×2048 resolution, anterior Face-Time HD Camera with 1.2 Megapixels (72 dpi) and 960×1280 resolution;
- Galaxy Tablet II (GT2): Google Android Operating System, no posterior camera, 0.3 Megapixels anterior camera with 640×480 resolution.

As a consequence, the three groups of images have different levels of resolution, which is one of the factors that can negatively affect cross-device recognition. The sources of noise in the MICHE-II dataset include: (a) reflections caused by artificial/natural light sources, people or objects in the scene; (b) out of focus; (c) blur, due either to an involuntary movement of the hand holding the device, or to an involuntary movement of the head/eye; (d) occlusions, due to eyelids, eyeglasses, eyelashes, hair, or shadows; (e) device-specific artifacts, due to the low resolution and/or to the specific noise pattern of the device; (f) off-axis gaze; (g) variable illumination; and (h) different color dominants. The lack of precise localization and of fixed distance in the capture (both images containing well centred eyes and images containing half faces are present in dataset), result in variable sizes of the region useful for recognition. This is typical of mobile captures performed by the users, which are usually neither too close nor at arm-length. This introduces further difficulties, since eye localization must be performed in a pre-processing step. In some cases, the resulting size of the iris region is smaller. In other cases, it is possible to exploit the further possibilities offered by an extended periocular region. The dataset has been collected during several different data acquisition sessions separated in time. The time elapsed between the first and second acquisition of a subject varies from a minimum of two months to a maximum of nine. At present, MICHE-I contains images from 75 different subjects, with 1297 by GS4, 1262 images from IP5, and 632 images from GT2.

The XML annotations associated to each image consist of the following tags:

- *filename*: the name of the image to which the XML file refers; it is composed according to a conven-

tion allowing to quickly find the desired image;

- *img type*: the trait captured in the image, since face images will be included soon in the dataset;
- *iris*: indicates which iris was acquired (right, left or both when the image contains both irises);
- *distance from the device*: distance of the user from the acquisition camera, measured to provide a further assessment information;
- *session number*: the number of the acquisition session when the image was captured;
- *image number*: image ordinal number;
- *user*: identification number of the subject, together with age, gender and ethnicity;
- *device*: contains all information about the capture device: type, name, camera position (front or rear), resolution and dpi;
- *condition*: information about capture conditions: location, illumination;
- *author*: the XML file also contains the name of the laboratory/institution who made that acquisition.

The XML file structure allows a quick and reliable retrieval of any image as a function of any one of the above parameters.

MICHE-I is the dataset provided to participants to MICHE-II challenge. In addition, further sequestered images with similar characteristics were captured and used to evaluate the final ranking. The complete MICHE-II dataset will be soon available to the research community³.

2.2. The common segmentation algorithm

According to a policy established by NICE-I and NICE-II competitions the problem of iris recognition was tackled by two separate challenges: MICHE-I dealt with the problem of segmentation of iris images acquired by mobile devices, and the following MICHE-II started from the best segmentation algorithm as a fair preliminary processing step to feed the following phases of feature extraction and classification. This algorithm, by Haindl et al. [4], was provided to all competing groups registered for the challenge, in order to get an unbiased comparison. It is focused on the detection of the non-iris components inside the parametrised iris ring. The procedure starts by detecting reflections, and then applies form-fitting techniques that enable to find a parametrisation of the pupil. Next, data is converted into the polar domain, where texture analysis determines the regions of the normalized data that should not belong to the iris, according to a Bayesian paradigm. The MICHE-II competitor methods start from the segmentation produced.

³<http://biplab.unisa.it>

2.3. Performance evaluation

The competitors were left free to choose any distance measure, given that it was at least semi-metric. In more detail, the *dissimilarity score* chosen by each competitor is meant as the probability that two irises are from two different subjects. The higher is the dissimilarity, the higher is the probability that the two irises are not from the same person. Let I be set of images from MICHE-II database, and I_a and $I_b \in I$, the *dissimilarity function* D must be defined as:

$$D : I_a \times I_b \rightarrow [0, 1] \subset \mathbb{R} \quad (1)$$

and satisfy the following properties:

1. $D(I_a, I_a) = 0$
2. $D(I_a, I_b) = 0 \rightarrow I_a = I_b$
3. $D(I_a, I_b) = D(I_b, I_a)$

Each submitted algorithm has to fill a dissimilarity matrix among input probe and gallery sets. A possible bias implied by embedding special processing into the algorithms to improve performance is avoided by the competition procedure itself, since new images were added and new distance matrices were computed in order to create the final rank. Distance matrices produced by each methods were used to compute the usual Figures Of Merit (FoM) to rank them, namely Recognition Rate (RR) for identification, and Receiver Operating Characteristic (ROC) curves, in particular the Area Under Curve (AUC), for verification.

3. Summary of Competitor Algorithms

Rank 1 - Nasir Uddin Ahmed, Slobodan Cvetkovic, Md. Erfanul Hoque Siddiqi, Andrey Nikiforov and Iliia Nikiforov participated the challenge with the algorithm with label `tiger_miche`. Iris biometric matching is performed using a combination of a popular iris code approach and a periocular biometric based on the Multi-Block Transitional Local Binary Patterns (LBP). The authentication scores are calculated separately, and the results are combined to improve the system performance. The proposed algorithm uses a score-level fusion. The scores from periocular matching and iris code Hamming distance are combined together to produce a final score. The matchers produce outputs in different ranges, with very different score distributions. Therefore, z-score normalization is used.

Rank 2 - The algorithm with label `Bata` was submitted by Zhou Shujuan. The pre-processing phase entails local histogram equalization to enhance iris texture. As a first step, a 1D feature $F = (f_1, f_2, \dots, f_{64})$ is computed by using the normalized and equalized

512×64 pixel gray scale image, which is obtained by summing up row values to obtain single values. Then the algorithm takes 64×8 fixed sample points in the normalized iris image, and calculates 2D Gabor wavelet coefficients on 5 frequencies and 8 directions for each sample point, therefore obtaining 40 complex numbers, stored in a binary array of 80 bits for each sample point. Eye image is extracted using the iris mask and is normalized to 100×150 pixel size. Then, LBP features are calculated in this eye image. The final matching of two iris images is obtained by combining the similarities measures obtained, namely Hamming distance for both binary arrays and LBP, and cosine distance for 1D features.

Rank 3 - Karan Ahuja, Rahul Islam, Ferdous Barbhuiya and Kuntal Dey presented the algorithm with label `karanahujax`. They proposed a baseline model, namely Root Scale-Invariant Feature Transform (SIFT), and two stacked convolution-based deep learning models (Convolutional Neural Networks - CNNs), for identifying a given individual from a periocular image. This was obtained by training the CNNs on a given set of periocular images as part of the learning phase, and verifying a pair of images during the testing phase. The two convolution-based models for verifying a pair of periocular images containing the iris are compared amongst each other as well as with the baseline model. In the first approach, deep learning is implemented in an unsupervised manner. The method uses a stacked convolutional architecture, using external models learned a-priori on external facial and periocular data, on top of the baseline model applied on the provided data. Afterwards different score fusion models are applied. In the second approach, the authors again use a stacked convolution architecture, but the feature vector is learned in a supervised manner.

Rank 4 - The algorithm with label `irisom` was submitted by Fabio Narducci, Silvio Barra, Luigi Gallo and Andrea Abate. It combines simple image processing techniques, like contrast enhancement and histogram adjustment, with unsupervised learning by Self Organizing Maps (SOM). The algorithm first composes the original image in polar coordinates with the segmenting mask, to discard all non-significant pixels in the surrounding of the iris. A SOM network is then configured and trained with pixels of the pre-processed image thus building the feature matrix that clusters the iris pixels. The SOM network is fed with RGB triples together with local statistical descriptors. These are kurtosis and skewness, which are computed at pixel level in a neighborhood window of 3×3 size. The output of the network is a feature map representing the activation status of the neurons for each pixel. Such a map represents, in other

terms, the cluster decomposition of the image which projects the problem of iris recognition onto a lower dimensional space. On the obtained feature maps, the algorithm computes the Histogram of Gradients (HOG), which is finally used as a feature vector representing the subject iris. To verify the subject identity, the Pearson coefficient in $[0,1]$ real interval is used, to measure the correlation between the two images. The Pearson correlation is used as the probability that the two irises are from the same subject.

Rank 5 - The algorithm proposed by Chiara Galdi and Jean-Luc Dugelay has label `FICO_matcher`. Its key features are: (i) the use of a combination of classifiers exploiting the iris colour and texture information; (ii) its limited computational time, particularly suitable for fast identity checking on mobile devices; (iii) the high parallelism of the code, making this approach also appropriate for identity verification on large database.

Rank 6 - The algorithm with label `otsedom` was submitted by Naiara Aginako, José María Martínez-Otzeta, Basilio Sierra, Modesto Castrillón-Santana and Javier Lorenzo-Navarro. It exploits both Machine Learning paradigms, and Computer Vision techniques. Descriptors are obtained by well known approaches, such as LBP, LPQ, and WLD. They are used individually in order to construct a classifier, and then some of them are combined to outperform the obtained accuracy. The final algorithm combines the best five descriptors to obtain a dissimilarity measure of two given iris images. Machine Learning classifiers have been used to perform the classification, and hence to obtain the a-posteriori probability distribution for each of the two iris images. Histogram distance between the two distributions is used to compute the dissimilarity. To perform the final classifier combination, five different classifiers are used, each of one giving a different a-posteriori distribution for each image. The mode of each a posteriori probability for each class value is used to combine the five classifiers, and the distance of the two mode histograms (one for each iris) is used as dissimilarity measure.

Rank 7 - Naiara Aginako, José María Maqrtínez-Otzeta, Igor Rodríguez, Elena Lazkano and Basilio Sierra submitted the algorithm with label `ccpsiarb`. Even in this case, the proposed approach is a combination of image transformations and classification, using techniques both from Machine Learning and Computer Vision. First, an image classification process is carried out in order to be able to classify the images as belonging to one of those which are defined in a given set of classes. This step involves both Machine Learning paradigms, in order to perform the classification itself, and image transformations from the Com-

puter Vision area in an intent to improve the accuracies of the obtained models. As classifiers the authors use some well known ML supervised classification algorithms, with completely different approaches to learning, and a long tradition in different classification tasks: IB1, NaiveBayes, Random Forest and C4.5. After several experiments with a combination of different image transformations and Machine Learning algorithms, the Edge transformation followed by IB1 classification has shown the best results. As a novelty, the dissimilarity computation between two images has been computed as an a-posteriori histogram difference of the classes distribution returned by the Machine Learning algorithm.

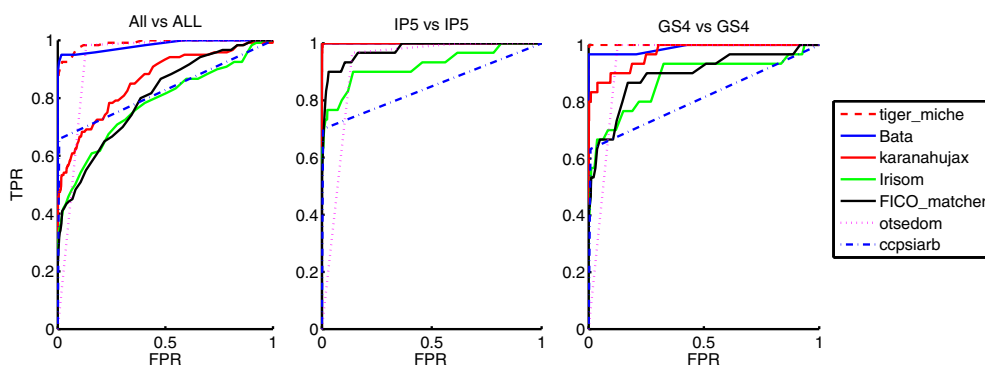
4. Summary of competition results

This section summarizes the main results obtained from the comparison of participating methods. Table 1 reports the final rank list, including the best performing version among the ones submitted for each author. The ranking has been obtained by averaging the Recognition Rate (RR) and the Area Under Curve (AUC) achieved.

Results were achieved with Probe and Gallery containing all sequestered images from MICHE-II with the addition of random images from the public set. The test set contains a total of 60 images for each set (IP5 and GS4). Both sets only included images from front cameras, with IP5 images captured indoor and GS4 images captured outdoor. The table accounts for both the total of 3600 matching operations, in the column ALLvsALL, and for the results achieved by homogeneous sets of images, in the columns IP5vsIP5 and GS4vsGS4. Actually, ALLvsALL presents quite degraded performance due to the double cross-matching carried out by some operations, that involve images captured with both different devices and in different settings (indoor vs. outdoor). It is interesting to notice that, notwithstanding the lower resolution of front cameras with respect to posterior ones, the performance degradation is less dramatic than expected. It is also interesting that, notwithstanding the lower resolution of IP5 front camera, the results obtained in IP5vsIP5, with images captured indoor, are much better (except for `otsedom`) than those obtained in GS4vsGS4, with images captured outdoor. This confirms that resolution and ambient conditions are complementary factors affecting iris recognition. This especially holds when images are captured by mobile devices, since conditions are less controlled and unpredictable. Figure 1 shows the ROC curves obtained in the three settings, and visually confirms the observed considerations.

Table 1. Final ranking of ICPR-MICHE-II competition

Rank	Algorithm	ALLvsALL	GS4vsGS4	Ip5vsIP5	Final Rank
1	tiger_miche	0.99	1.00	1.00	1.00
2	Bata	0.98	0.98	1.00	0.99
3	karanahujax	0.89	0.89	0.96	0.91
4	irisom	0.79	0.82	0.88	0.83
5	FICO_matcher	0.77	0.78	0.92	0.82
6	otseudom	0.78	0.80	0.78	0.79
7	ccpsiarb	0.75	0.72	0.77	0.75

**Figure 1. ROC curves obtained by the competitor methods in the three settings.**

5. Conclusions

MICHE-II challenge at ICPR aimed at assessing the present status of iris recognition on mobile devices in Visible Light conditions. Recent smart devices allow to capture images in NIR spectrum, that in general allows to achieve better performance. In fact, illumination conditions play a less critical role and dark irises are better processed. However, problems that are typical of mobile uncontrolled image capture, e.g., significant off-axis or bad eye framing, as well as cross-condition issues, would hold even with NIR approaches. Clearly, mobile iris recognition is still not able to achieve the same results of desktop based one, especially if the whole processing must be carried out on the device itself. However, achieved results are encouraging, and suggest further research.

References

- [1] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE transactions on pattern analysis and machine intelligence*, 15(11):1148–1161, 1993.
- [2] M. De Marsico, M. Nappi, and D. Riccio. Noisy iris recognition integrated scheme. *Pattern Recognition Letters*, 33(8):1006–1011, 2012.
- [3] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols. *Pattern Recognition Letters*, 57:17–23, 2015.
- [4] M. Haindl and M. Krupička. Unsupervised detection of non-iris occlusions. *Pattern Recognition Letters*, 57:60–65, 2015.
- [5] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan. DeepIris: Learning pairwise filter bank for heterogeneous iris verification. *Pattern Recognition Letters*, 2015.
- [6] P. J. Phillips, W. T. Scruggs, A. J. O’Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):831–846, 2010.
- [7] H. Proença and L. A. Alexandre. The NICE.I: Noisy Iris Challenge Evaluation-part I. In *Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–4. IEEE, 2007.
- [8] H. Proença and L. A. Alexandre. Toward covert iris biometric recognition: Experimental results from the NICE contests. *IEEE Transactions on Information Forensics and Security*, 7(2):798–808, 2012.
- [9] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1529–1535, 2010.
- [10] R. P. Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.