Pattern Recognition Letters 000 (2017) 1–8

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Results from MICHE II – Mobile Iris CHallenge Evaluation II

Maria De Marsico[a,*], Michele Nappi[b], Hugo Proença[c]

[a] Sapienza University of Rome, Italy
[b] University of Salerno, Italy
[c] IT: Instituto de Telecomunicações, University of Beira Interior, Portugal

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Mobile iris recognition
MICHE-II competition
Biometric authentication

ABSTRACT

Mobile biometrics represent the new frontier of authentication. The most appealing feature of mobile devices is the wide availability and the presence of more and more reliable sensors for capturing biometric traits, e.g., cameras and accelerometers. Moreover, they more and more often store personal and sensitive data, that need to be protected. Doing this on the same device using biometrics to enforce security seems a natural solution. This makes this research topic attracting and generally promising. However, the growing interest for related applications is counterbalanced by still present limitations, especially for some traits. Acquisition and computation resources are nowadays widely available, but they are not always sufficient to allow a reliable recognition result. Most of all, the way capture is expected to be carried out, i.e., by the user him/herself in uncontrolled conditions and without an expert assistance, can heavily affect the quality of samples and, as a consequence, the accuracy of recognition. Among the biometric traits raising the interest of researchers, iris plays an important role. Mobile Iris CHallenge Evaluation II (MICHE II) competition provided a testbed to assess the progress of mobile iris recognition, as well as its limitations still to overcome. This paper presents the results of the competition and the analysis of achieved performance, that takes into account both proposals submitted for the competition section launched at the 2016 edition of the International Conference on Pattern Recognition (ICPR), as well as proposals submitted for this special issue.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Mobile equipment is ubiquitous nowadays. Smartphones are inexorably substituting "old" cellular phones, that in turn had replaced traditional landlines at a significant extent. As a matter of fact, it is not very frequent to find a public telephone, that was in the past a characteristic element in the urban scenario. The possibility to communicate almost wherever and whenever represents a significant sociological phenomenon. Notwithstanding this, the "whatever" option offered by the new communication devices and protocols can be seen as the actual revolution. The use of smart mobile devices for mere communication, in its traditional meaning of "speaking with somebody" has become quite marginal with respect to more advanced communication modalities. These include sending in real time almost any kind of multimedia information, therefore dramatically decreasing the time needed for dissemination of ideas and event accounts. One of the aspects related to this new technological scenario is the possibility to store personal, and often sensitive data directly on the mobile devices, and remotely connecting to either social platforms or to personal services. The latter often entail the exchange of sensitive information, and requires a twofold approach to address an increased security need: from one side, it is necessary to reliably identify the owner *before* the use of the device, while from the other side it is necessary to reliably identify the user of a remote service *at the moment* the device is used to connect to it. It is well accepted that biometrics can both enforce and make authentication simpler. In simple cases, the presentation of a *robust* biometric trait for recognition can substitute complex passwords or cards. The next step on which research is focusing is to move biometrics on mobile. Mobile biometric recognition is the new advanced frontier for secure use of data and services, either locally or remotely. It provides a further application for user mobile equipment, which are ubiquitous nowadays. These include both by personal devices, e.g., smartphones and tablets, but also the incoming wearable devices, e.g., smart watches, given that they are equipped with suitable sensors. Capture of biometric traits in any place is both the strength of this new approach. Captured information (sample) can be compared with that stored either on the device itself, or even within RFID tags, smartcards or machine readable identification documents (IDs), to verify owner's identity by a 1:1 match operation. As an alternative, the possibly

* Corresponding author.
E-mail address: demarsico@di.uniroma1.it (M. De Marsico).

http://dx.doi.org/10.1016/j.patrec.2016.12.013
0167-8655/© 2016 Elsevier B.V. All rights reserved.

pre-processed sample can be sent to a remote server, for identification in a set of relevant subjects by a 1: *N* matching operation. Mobile devices used for biometric recognition must therefore embed suitable hardware equipment and software applications, allowing to capture and possibly process data from one or more biometric traits. However, the *any-time, any-place* option is also one of the weak points of this new authentication paradigm.

Applications must be designed for intuitive operation, especially if it is not planned to assist users during sample capture. The environmental conditions might be adverse and cause a number of artifacts, e.g., shadows, reflections, or acoustic noise, hindering an accurate recognition. Moreover, the user might not be able to evaluate the sufficient quality of the obtained capture, able to allow a reliable processing. The captured data must be suitably converted by software into digital templates for storage and matching against other records. Feature extraction, storing and processing, might 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 (e.g., see De Marsico et al. [12]), and for more efficient storage, that might not be universally available yet. Techniques targeted at mobile devices must still be suitably adapted. It is interesting therefore to assess the level of performance that we can realistically expect from a mobile biometric application.

Among possible biometric traits that can be processed in a mobile setting, iris plays a relevant role. Iris acquisition is little intrusive, since it can be carried out at a reasonable distance from the eye to avoid user discomfort. In addition, iris codes are among the lighter templates to store. Related research achieved a quick performance increase. The pioneering works by Daugman [11] and Wildes [25], as it happened for other biometric traits, mostly pertain controlled settings. More recent challenges address iris recognition in less controlled and/or mobile settings [13,22]. Among the most advanced approaches, it is worth mentioning the use of deep learning [18]. At present, iris recognition systems operating in controlled conditions still require a distance of about 1m or less between the subject and the capture device. Moreover, the user has to look towards the device 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) [20]. Proença and Alexandre [21] have rather tackled the problem of noisy iris recognition. A further difference is that they aimed at assessing performance over images captured in visible light (VL) instead of near-infrared (NIR). This point is still debated in the research community. VL images usually contain more features and, in some cases, more details than the NIR images. However, they are also more seriously affected by problems due to uneven illumination and may present many noisy artifacts, especially reflections of light sources and/or objects present in the environment. Moreover, their processing suffers from dark pigmentation [17]. This raises an apparent contradiction: more detail does not necessarily mean more information. The work by Hollingsworth et al. [17] demonstrated that humans can better recognize images of the periocular region, including skin patches and other elements like eyebrows, if acquired in visible light. As a matter of fact, these images show melanin-related differences that do not appear in NIR images. However, the situation is often reversed, when the recognition operation regards the iris alone. IR images contain cleaner and more easily distinguishable features, and this makes iris recognition in IR generally more feasible than in visible light. In addition, dark pigmentation represents a hard condition in VL, since it may hide details that are otherwise well detectable in lighter colors irises. According to a balance between pros and cons of the different approaches, commercial iris recognition uses NIR acquisition. Since mobile biometrics are becoming quite popular, the market also started to offer NIR attachments made for mobile devices. However, in everyday devices, both NIR sensors and NIR attachments are still quite rare, and it is difficult to anticipate if they will reach a sufficient diffusion, if no largely shared user requirement arises. The aim of Mobile Iris CHallenge Evaluation II (MICHE-II) competition, was to assess which level of performance can be achieved without special equipment. The aim of the contest, was to collect relevant contributions to the field of mobile iris recognition in both academy and industry. A section of the contest was launched in conjunction with ICPR 2016, while further proposals were submitted for this Special Issue. This paper presents the comparison of the eight best performing algorithms. In the following, Section 2 will present the benchmark used for the competition. Section 3 describes the competition setup, by presenting the common segmentation algorithm provided to competitors and the performance measures used for the evaluation. Section 4 briefly presents the best eight algorithms participating in MICHE-II challenge. Section 5 presents and discusses the competition results. Finally, Section 6 draws some conclusions.

2. MICHE-II database

In biometrics research, the advancements in recognition techniques, and in particular the progressive loosening of constraints on acquisition conditions, usually calls for increasingly challenging benchmarks. This happened for iris too. The Chinese Academy of Sciences collected and made available to the scientific community the first group of publicly available datasets of significant size dealing with iris images, namely CASIA-Iris. since 2002, the original dataset has been updated from CASIA-IrisV1 to CASIA-IrisV4. In all dataset versions, images are collected under NIR illumination or contain synthesized parts (CASIA-IrisV1). For these reasons, they cannot be used for a general assessment of methods entailing mobile acquisition, where NIR sensors are not sufficiently widespread in mobile devices. The mentioned ICE competitions used images acquired with the same class of sensors. UBIRIS datasets were collected in order to rather assess proposals of iris recognition in VL. Their images were acquired and made available from SOCIA Lab at University of Beira Interior (Portugal), that organized the two NICE challenges that will be discussed in the next section. The images were captured in visible light and uncontrolled conditions. Such conditions are similar to those that MICHE-II competition wanted to address. However, acquisition exploited cameras with a better resolution than average ones embedded in mobile devices.

The aim of MICHE-I dataset, which is publicly available to the scientific community and represents the core of the still unpublished MICHE-II dataset, is to represent the starting core of a wider dataset to be collected using different mobile devices, in uncontrolled conditions, with different illumination, and without the assistance of an operator. This should better allow unbiased assessment of cross-device and cross-setting interoperability of recognition procedures. In fact, the dataset allows a twofold assessment. It allows to measure the ability to match samples of the same subject from different devices, and in different environments. In addition, more in general, it allows to measure the ability to handle samples acquired by devices with different characteristics and in different settings, without a significant degradation of recognition performance. MICHE-I, the core of MICHE-II, is a dataset of iris images acquired in visible light by different mobile devices. Its key features are: (1) a sufficient population of users; (2) the use of different mobile devices embedding sensors with different characteristics; (3) the acquisition process including different sources of noise; and (4) more acquisition sessions in different times. The images in the dataset are further fully annotated, as detailed in the following. Three kinds of devices were used for iris image capture:

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013

- Galaxy Samsung IV (GS4) smartphone: Google Android Operating System, two cameras: CMOS posterior camera with 13 Megapixel (72 dpi) and 2322 × 4128 resolution, and CMOS anterior camera with 2 Megapixel (72 dpi) and 1080 × 1920 resolution;
- iPhone5 (IP5) smartphone: Apple iOS Operating System, two cameras: iSight posterior camera with 8 Megapixels (72 dpi) and 1536 × 2048 resolution, and anterior FaceTime 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.

In order to have images with the same characteristics of those expected in a mobile operation, the subjects participating in data collection were instructed to behave as they would do in a normal situation, e.g., subjects wearing eyeglasses could either choose to remove or keep them, and take a different choice in different sessions. They captured self-images of one of their irises only, by holding the mobile device at the distance they considered to be suitable, with a minimum of four shots for each camera and acquisition mode (indoor, outdoor). Indoor acquisition set up included various sources of artificial light, sometimes combined with natural light sources. Outdoor acquisition was carried out using natural light only. As a consequence of the different sensor characteristics, the different groups of images, for each device and for each camera, have different levels of resolution, which is one of the factors that possibly negatively affect cross-device recognition. MICHE-I dataset images are affected by different sources of noise: (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, eyelashes, hair, eyeglasses, 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 framing and of fixed distance in the capture, result in variable sizes of the region useful for recognition. As a matter of fact, both images containing well centered eyes and images containing half faces are present in dataset. This can be considered as typical of mobile captures performed by the users, which usually hold the device neither too close nor at arm-length. As a consequence, eye localization, that must be performed in a pre-processing step, may result more difficult. In some cases, it is possible to exploit the extended periocular region for recognition, if this is included in the captured image, but this cannot be taken for granted. The dataset has been collected during different acquisition sessions, of course separated in time. The time elapsed between the first and second acquisition of a same subject varies from two to nine months. At present, MICHE-I contains images from 75 different subjects, with 1297 by GS4, 1262 images from IP5, and 632 images from GT2.

As already mentioned, MICHE-I images are annotated to maintain a number of useful information. The XML annotations include the following tags:

- *filename*: the name of the image to which the XML file refers; it is composed according to a convention allowing a quick retrieval of the desired image(s);
- *img type*: the trait captured in the image, since face images will be included soon in the dataset;
- *iris*: the iris that 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 image acquisition session;
- *image number*: image ordinal number;
- *user*: identification number of the subject, together with age, gender and ethnicity;
- *device*: 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.

MICHE-I was the dataset provided to participants to MICHE-II challenge for training. Further sequestered images with similar characteristics were captured and used to evaluate the final ranking. These will be added to MICHE-I to obtain the complete MICHE-II dataset, that will be soon available to the research community.[1]

Figs. 1 and 2 in Section 3.2 below show some examples of images in MICHE-I.

3. MICHE-II competition setup

This section presents the setup that was chosen for MICHE-II competition, starting from the elements that distinguish it with respect to preceding challenges, and then presenting the details of the common framework provided to participants.

3.1. The goal of MICHE-II competition

The previous section has already mentioned the first wide challenges in iris recognition, namely ICE. Since those competitions used NIR images of the iris, acquired in controlled conditions, it is clear how they differ from MICHE-II. A more similar context in entailed bu NICE contests. The Noisy Iris Challenge Evaluation (NICE I),[2] organized by SOCIA Lab at University of Beira Interior (Portugal), exploited images captured in unconstrained imaging environments to evaluate at which extent noise can affects iris segmentation. To this aim, the proposed iris dataset UBIRIS.v2 [23] contains data captured in the visible light (VL), at-a-distance, namely between 4 and 8m, and on the move. The results achieved by participant methods confirm the major impact that uncontrolled conditions have on recognition performance. Recognition of visible light (VL) iris images captured at-a-distance and on the move with less controlled protocols was the target of the further NICE II contest[3] [21]. While NICE I contest was focused on iris segmentation, NICE II rather addressed iris recognition. The segmentation masks obtained by the best algorithm from NICE I were provided to participants, that used them to select the relevant region for feature extraction and matching. The pair of competitions MICHE-I and MICHE-II has been conceived along a similar line.

MICHE-I challenge [13] addressed issues related to iris acquisition and segmentation by mobile devices. In this new context, the basic assumption is that the subject to be recognized autonomously operates the capturing device. MICHE-I provided to participants the dataset with the same name, to be used as a common benchmark, and suitable to assess the performance of biometric applications related to this specific setup. Two opposite considerations hold. The usually short distance (the length of a human arm at maximum) can increase capturing accuracy/ quality. A further element to consider is the natural tendency of the user tends to assume a frontal pose in selfies. The reverse of the medal is that the quality of the captured image can suffer from both possible lower resolution, due to the capture device, and from out of

[1] http://biplab.unisa.it.
[2] http://nice1.di.ubi.pt.
[3] http://nice2.di.ubi.pt/.

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013

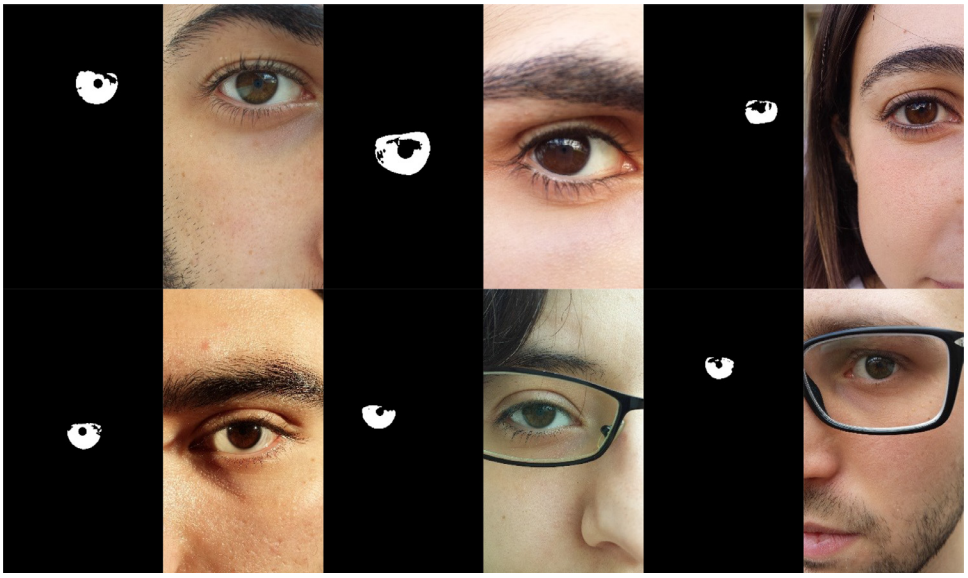

Fig. 1. Examples of segmentation of "good" images with the algorithm proposed by Haindl and Krupička for MICHE-I.

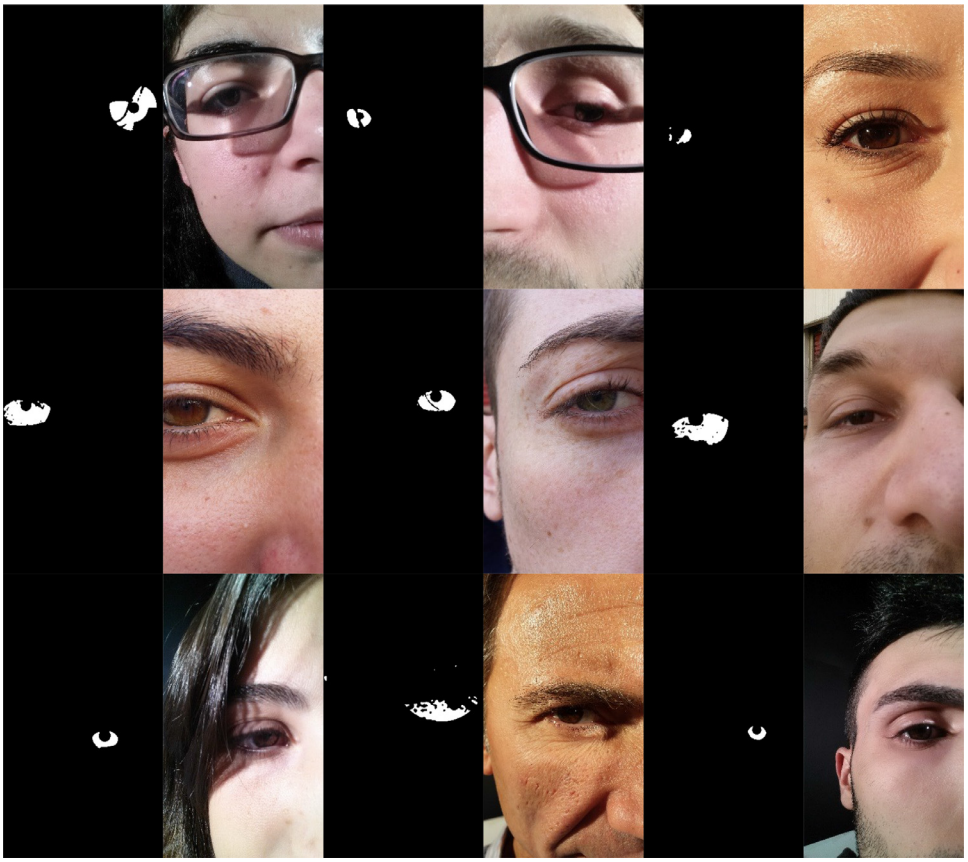

Fig. 2. Examples of segmentation of "problematic" images with the algorithm proposed by Haindl and Krupička for MICHE-I.

focus/motion blur, incorrect framing, and illumination distortions, caused by both the kind of device and by the lack of experience of the user in controlling the capture operation. Due to these distortion factors, the robustness of detection/segmentation and encoding procedures must be suitably improved in order to achieve good results in this context. Obviously the accuracy of the encoding, i.e., the ability to correctly extract relevant and discriminative features, and the following recognition, are heavily affected by the quality of the segmentation. As mentioned above, 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 in the competitors ranking process.

3.2. The common segmentation algorithm

According to the policy established by NICE-I and NICE-II competitions, the problem of iris recognition was tackled by the two separate challenges: MICHE-I addressing segmentation, and the

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013

following MICHE-II addressing recognition. In order to ensure a fair and unbiased evaluation focused on recognition only, all competing groups registered for the challenge had to use a common segmentation algorithm, namely the best segmentation algorithm from MICHE-I. This algorithm, proposed by Haindl and Krupička [16], is focused on the detection of the non-iris components inside the parametrized iris ring. The segmentation procedure starts from a reflection detection step. After this, form-fitting techniques produce a parametrization of the pupil. Next, as it often happens in iris-related methods, data is converted into the polar domain, where texture analysis identifies the regions of the normalized data according to a Bayesian paradigm separating iris from non-iris. All the MICHE-II competitor methods start from the segmentation produced. Fig. 1 shows some examples of "ideal" images captured both indoor and outdoor, with the corresponding segmentation mask obtained by the algorithm by Haindl and Krupička.

It is interesting to notice that, though reflections are more evident outdoor, a more diffused and uniform illumination creates favorable conditions for localization and segmentation.

Fig. 2 shows some examples of problematic images. They summarize a number of problems that can be found in MICHE-I dataset, that can be generically classified as "occlusions". For instance, they can depend from the presence of glasses, that can create sharp shadows in the region of interest, especially indoor, and in general to a scarcely visible iris region, so that the number of pixels that can be used for recognition is reduced.

It is possible to underline that in some cases the eye is framed quite well in the captured image, while in other cases a half face is captured.

3.3. Performance evaluation

Different feature extraction procedures can produce different kinds of templates, that may require specific approaches to similarity/distance evaluation. For this reason, the competitors were left free to choose any suitable distance measure for the produced iris templates, with the only constraint to be a semi-metric at least. The *dissimilarity score* chosen by each competitor was meant to represent 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. According to this, given I the set of images from MICHE-II database, and I_a and $I_b \in I$, the *dissimilarity function* D had to be defined as:

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

and had to satisfy the following properties:

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

The final result returned by each algorithm had to be a full dissimilarity matrix among probe and gallery sets provided in. The competition procedure avoided a possible bias implied by embedding special processing into the algorithms to improve performance, 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.

4. Competitor algorithms at a glance

This section briefly presents the best eight algorithms submitted for MICHE-II challenge. Algorithms are listed according to the final ranking achieved. The given information is sufficient to give a self-contained account for results. More details can be found in the papers in this Special Issue. Each algorithm is labeled in order to identify it in the tables in the section presenting the experiments.

The best recognition results were achieved by the algorithm denoted by the label `tiger_miche` and described in [6] (extended in [7]), with the algorithm that in the following will be . This algorithms carries out iris matching by a combination of a popular iris coding approaches and a periocular biometric processing based on the Multi-Block Transitional Local Binary Patterns (LBP). The algorithm computes iris code Hamming distance and periocular matching scores separately, and then combines the results by a score-level fusion to improve the system accuracy. Since the matchers produce values in different ranges and with very different score distributions, z-score normalization is used.

The following algorithm is labeled `karanahujax` and is described in [8] (extended in [9]). The paper first presents a baseline model, based on Root Scale-Invariant Feature Transform (SIFT). The segmented iris image is overlaid with the binary mask to get the iris image rid of occlusions. Then Dense color Root SIFT descriptors [10] are computed, giving keypoints with identical size and orientation. Then two stacked convolution-based deep learning models (Convolutional Neural Networks – CNNs) are designed to identify a given subject from a periocular image. The CNNs are trained on a set of periocular images as part of the learning 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 (Model 1). The method uses a stacked convolutional architecture, after learning external models 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 tested to get the final result. In the second approach, the authors still use a stacked convolution architecture, but the discriminative feature representations is learned in a supervised manner (Model 2).

The algorithm with label `Raja` is due to Raja et al. [24]. They propose multi-patch deep features using deep sparse filters, in order to obtain robust features for iris recognition. The patch level representation is obtained by dividing the iris image into a number of patches, that are mapped onto a collaborative subspace to perform classification via maximized likelihood. The approach is designed to work also with a single-sample gallery. Of course, the advantage entailed by the patch based approach is that the noise from external illumination and eyelids can be considered as localized within individual blocks, so that other blocks are relatively noise free and can be reliably used for extracting robust features. In order to avoid missing relevant global information, the features are extracted from the holistic image which is jointly represented with the patches. The information corresponding to red, green and blue color channels separately. The images obtained for each channel from segmentation and normalization are divided into a number of blocks. Along with the number of blocks, the whole image is also processed to obtain deep sparse histograms using the set of deep sparse filters. The set of histograms obtained from different channels and blocks are concatenated to form the final feature vector. In order to in order to demonstrate the robust feature representation provided by the proposed scheme, the authors employ simple distance measures using χ^2 and *Cosine* distance.

The next algorithm is labeled `irisom` and details can be found in [1] (extended in [19]). Simple image processing techniques, like contrast enhancement and histogram adjustment, are used together with unsupervised learning by Self Organizing Maps (SOM). The algorithm first matches the original image in polar coordinates with the segmentation mask, in order 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

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013

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 $3\times$ 3 size. The output of the network is a feature map with the activation status of the neurons for each pixel. In practice, such a map represents 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) over the obtained feature maps, and the result is used as a feature vector representing the iris. To verify the subject identity, the Pearson coefficient in the [0,1] real interval is used, to measure the correlation between two images. The Pearson correlation is used as the probability that the two irises are from the same subject. Results reported here refer to 5×5 and 10×10 SOM.

The algorithm proposed by Galdi and Dugelay has label `FICO_matcher` [14] (extended in [15]). It uses of a combination of classifiers exploiting the iris color and texture information. Moreover cluster information is coded, where "clusters" are the small color spots that often characterize the human iris. For each cluster, centroid coordinates, orientation, and eccentricity are concatenated in a feature vector. A multi-layer approach is proposed, where layers are derived for each color channel of the color space at hand. The best performances are obtained using the a^* and b^* channels of the CIE 1976 $L^*a^*b^*$ color space. Color values are first normalized between 0 and 255, then dividing the resulting grey values into 8 intervals of size $32 (32 * 8 = 256)$, i.e. 8 layers (images) are obtained from each color channel. Two versions of the algorithm were proposed for the competition, differing in the fusion formula for the DIST(ances) computed over the different descriptors (color, texture and cluster). As for Version 1 (V1), $DIST_V1final = DIST_color * 0.2 + DIST_texture * 0.4 + DIST_cluster * 0.4$; as for Version 2 (V2), $DIST_V2final = DIST_color * 0.2 + DIST_cluster * 0.8$ (texture descriptor is not used). An interesting characteristic is represented by its limited computational time, particularly suitable for fast identity checking on mobile devices. In addition, the code is highly parallel, so that this approach is also appropriate for identity verification on large database.

Aginako-Bengoa et al. [4] (extended in [5]) propose the algorithm labeled `otsedom`. It exploits both Machine Learning paradigms, and Computer Vision techniques. Well known descriptors are computed, such as LBP, LPQ, and WLD. They are used individually in order to construct a classifier, and then subsets of them are combined to outperform the obtained accuracy. The final algorithm combines the best five descriptors to obtain a robust dissimilarity measure of two given iris images. Machine Learning classifiers are used to perform the individual classifications, 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 one giving a different a-posteriori distribution for each image. The mode of each a-posteriory 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.

Finally, a different group of participants submitted the algorithm with label `ccpsiarb` [2] (extended in [3]). Even in this case, the proposed approach is a combination of techniques from Machine Learning and Computer Vision. First, an image classification process is carried out to classify the images as belonging to one of a given set of classes. This step involves both Machine Learning paradigms, to perform the classification itself, and image transformations to improve the accuracies of the obtained models. Regarding the latter, different modifications over the original pictures provide different views of the images. Examples of the transformations considered are Equalization, Gaussian, Median, etc. The main goal of this phase is to have variability in the aspect the picture offers, so to obtain different values for the same pixel positions. The classifiers derive from 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. Experiments took into account the 19 image collections obtained by applying single transformations and the four different classifiers, giving a total 76 experiments. After testing these combinations of different image transformations and Machine Learning algorithms, the Edge transformation followed by IB1 classification (identified as combination `ccpsiarb_17`) is identified as the combination providing the best results. Following this combination, very close results were achieved by `ccpsiarb_2` and `ccpsiarb_42`, obtained by Equalize + IB1, and by Gaussian filter + IB1 respectively. 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.

5. Competition results and discussion

In order to ensure a fair comparison, all algorithms were run from scratch at BipLab - University of Salerno, over an extended set of images (see Section 3.1) after segmenting them with the segmentation algorithm provided for the competition (see Section 3.2). The final rank list in Table 1 reports the best performing version among the ones submitted for each author (label). The rank was obtained by averaging the Recognition Rate (RR) and the Area Under Curve (AUC) achieved, and considering images captured by the two smartphones. We remind that authors could propose their own distance measure, given that it was a metric (see Section 3.3).

Table 1 shows that the better the ranking achieved, the more stable the method with respect to the test setting. The "hardest"one is of course ALLvsALL, entailing that gallery and probe images may come from different devices. All methods provide consistently lower performances in this condition. On the average, the images over which the best results are achieved in homogeneous settings (gallery and probes from the same device) come from IP5 (mean value = 0,89), and the achieved results further present a lower standard deviation (0,086). This might be unexpected, due to the lower resolution of the camera, and seems to suggest that higher resolution may also increase the way the noise typical of iris images can affect recognition. A complementary observation regards the way the different methods behave with respect to the different devices. Excluding `tiger_miche` , all the others present better results either with one camera or the other with significant differences, except of `otsedom`. In particular, four of them work better with IP5 (lower resolution images).

In order to evaluate the stability of the methods submitted in more versions, the extended ranking shown in Table 2 includes the different versions in the list of results.

We can notice that some methods are more stable across their variants. For instance, `karanahujax_Model1` and `karanahujax_Model1` achieve the same final score, but Model2 was preferred in the final ranking due to the better behavior in ALLvsALL. A similar constant behavior is observed for `ccpsiarb`, while `FICO_matcher_V2` achieves dramatically worse results than `FICO_matcher_V1`.

Finally, Table 3 shows in detail the results in terms of RR and AUC achieved by the competing method, dividing them according to the inter-device or intra-device setting, and also reports the computing time per matching operation in seconds.

Due to the possible change in ranking among methods in different settings, no specific ordering is provided for the values in

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013

Table 1
The final ranking of the best methods submitted to MICHE-II competition.

Rank	Algorithm	ALLvsALL	GS4vsGS4	Ip5vsIP5	Final Rank
1	tiger_miche	0,99	1,00	1,00	1,00
2	karanahujax_Model2	0,89	0,89	0,96	0,91
3	Raja	0,82	0,95	0,83	0,86
4	irisom_10_10	0,79	0,82	0,88	0,83
5	FICO_matcher_V1	0,77	0,78	0,92	0,82
6	otsedom	0,78	0,80	0,78	0,79
7	ccpsiarb_17	0,75	0,72	0,77	0,75

Table 2
Results from different versions of the competing methods.

Rank	Algorithm	ALLvsALL	GS4vsGS4	Ip5vsIP5	Final Rank
1	tiger_miche	0,99	1,00	1,00	1,00
2	karanahujax_Model2	0,89	0,89	0,96	0,91
3	karanahujax_Model1	0,82	0,90	1,00	0,91
4	Raja	0,82	0,95	0,83	0,86
5	irisom_10_10	0,79	0,82	0,88	0,83
6	FICO_matcher_V1	0,77	0,78	0,92	0,82
7	irisom_5_5	0,77	0,75	0,90	0,81
8	otsedom	0,78	0,80	0,78	0,79
9	ccpsiarb_17	0,75	0,72	0,77	0,75
10	ccpsiarb_2	0,74	0,72	0,75	0,74
11	ccpsiarb_42	0,73	0,72	0,72	0,73
12	FICO_matcher_V2	0,61	0,65	0,75	0,67

Table 3
Detailed results in terms of RR, AUC and methods computing time per matching operation.

Algorithm	All vs ALL				GS4 vs GS4				IP5 vs IP5			
	RR	AUC	Global	Time (s)	RR	AUC	Global	Time (s)	RR	AUC	Global	Time (s)
ccpsiarb_17	0,68	0,83	0,75	57,64	0,63	0,81	0,72	61,43	0,70	0,85	0,77	54,30
ccpsiarb_2	0,65	0,82	0,74	25927	0,63	0,81	0,72	26538	0,63	0,86	0,75	25284
ccpsiarb_42	0,65	0,81	0,73	28620	0,63	0,81	0,72	28927	0,63	0,81	0,72	28339
FICO_matcher_V1	0,73	0,80	0,77	1,00	0,67	0,89	0,78	1,00	0,87	0,98	0,92	1,01
FICO_matcher_V2	0,48	0,73	0,61	0,57	0,50	0,79	0,65	0,57	0,57	0,93	0,75	0,57
irisom_10_10	0,80	0,78	0,79	3,48	0,77	0,88	0,82	3,45	0,83	0,92	0,88	3,51
Irisom_5_5	0,75	0,79	0,77	3,03	0,63	0,88	0,75	3,00	0,87	0,93	0,90	3,05
karanahujax_Model1	0,88	0,76	0,82	5,68	0,83	0,97	0,90	5,72	1,00	1,00	1,00	5,63
karanahujax_Model2	0,92	0,86	0,89	4,65	0,83	0,95	0,89	4,62	0,93	0,98	0,96	4,69
otsedom	0,63	0,93	0,78	41,74	0,67	0,94	0,80	42,37	0,63	0,92	0,78	41,10
tiger_miche	1,00	0,99	0,99	1,71	1,00	1,00	1,00	1,72	1,00	1,00	1,00	1,71
Raja	0,82	0,81	0,82	12,91	0,93	0,96	0,95	12,78	0,77	0,89	0,83	12,80

Table 3. No dramatic differences appear among intra- and inter-device settings regarding the matching time, therefore we will consider the latter for the following discussion. It is interesting to observe how the best method `tiger_miche` also achieves the best result in terms of time required by the single matching operation. Only `FICO_matcher_V1` does better, and even more `FICO_matcher_V2`. However, the latter one provides much lower recognition accuracy. On the other extreme we find the methods `otsedom` and the group of `ccpsiarb` variants, with a maximum of about 286 seconds, i.e., about 5 minutes. This seems to testify that, overall, the presented ML techniques are not suited for a real time operational setting, even if this parameter was not evaluated for the competition.

6. Conclusions

Mobile devices are going to play a relant role in biometric recognition too. In this paper we discussed in detail the results of MICHE-II contest, the follow-up of MICHE-II. They were the first international contests aiming at demonstrating the feasibility of iris/ocular recognition using data acquired from different types of mobile devices. After briefly summarizing the competition setting and the participating algorithms, the paper compared their performance in the different experiments carried out on MICHE-II dataset. The level of performance achieved by the methods participating in MICHE-I were not comparable to those achieved by other investigation performed so far on iris recognition. However, the best methods participating in MICHE-II, focusing on feature extraction and recognition only, achieve extremely promising results. The most interesting aspect is that images were acquired in uncontrolled conditions and in visible light, that are widely recognized as extremely adverse conditions. On the other hand, most mobile devices are equipped with high-resolution RGB cameras, and not with Near Infrared (NIR) sensors that would allow better results. MICHE-II Evaluation Challenge aimed at testing the feasibility of using state-of-the-art methods for iris recognition on mobile devices. The dataset provided for the competition contains images from indoor/outdoor, frontal/rear camera and acquired by the participants on their own by multiple devices and without any supervision. This created the opportunities to try the submitted methods on very challenging images, in order to answer the question: “is it feasible to recognize human irises from mobile with sufficient level of accuracy?”. The results achieved are encouraging and help identifying interesting research lines to follow. As a matter of fact, in the best cases, they are comparable to state of the art in iris recognition in more controlled conditions. The aim of the work was to assess the present state of the art in this specific field and highlight existing limitations. However, the same results suggest that it

is worth further exploring possible improvements of the available techniques.

Acknowledgments

The third author acknowledges the support given by FCT project UID/EEA/50008/2013.

References

[1] A. Abate, S. Barra, L. Gallo, F. Narducci, SKIPSOM: skewness & kurtosis of iris pixels in self organizing maps for iris recognition on mobile devices, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–5.
[2] N. Aginako-Bengoa, J. Martínez-Otzerta, I. Rodriguez, E. Lazkano, B. Sierra, Machine learning approach to dissimilarity computation: Iris matching, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–5.
[3] N. Aginako-Bengoa, J. Martínez-Otzerta, I. Rodriguez, E. Lazkano, B. Sierra, Iris matching by means of machine learning paradigms: a new approach to dissimilarity computation, Pattern Recognit. Lett. Same Volume, TBA.
[4] N. Aginako-Bengoa, J. Martínez-Otzerta, B. Sierra, M. Castrillón-Santana, J. Lorenzo-Navarro, Local descriptors fusion for mobile iris verification, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–5.
[5] N. Aginako-Bengoa, J. Martínez-Otzerta, B. Sierra, M. Castrillón-Santana, J. Lorenzo-Navarro, Periocular and iris local descriptors for identity verification in mobile applications, Pattern Recognit. Lett. Same Volume, TBA.
[6] N. Ahmed, S. Cvetkovic, E. Siddiqi, A. Nikiforov, I. Nikiforov, Using fusion of iris code and periocular biometric for matching visible spectrum iris images captured by smart phone cameras, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–5.
[7] N. Ahmed, S. Cvetkovic, E. Siddiqi, A. Nikiforov, I. Nikiforov, Combining iris and periocular biometric for matching visible spectrum eye images, Pattern Recognit. Lett. Same Volume, TBA.
[8] K. Ahuja, R. Islam, F. Barbhuiya, K. Dey, A preliminary study of cnns for iris and periocular verification in the visible spectrum, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–6.
[9] K. Ahuja, R. Islam, F. Barbhuiya, K. Dey, Cnns for ocular smartphone-based biometrics, Pattern Recognit. Lett. Same Volume TBA.
[10] R. Arandjelovic, A. Zisserman, Three things everyone should know to improve object retrieval, in: Conference on Computer Vision and Pattern Recognition, CVPR 2012, IEEE, 2012, pp. 2911–2918.
[11] J.G. Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Trans. Pattern Anal. Mach. Intell. 15 (11) (1993) 1148–1161.
[12] M. De Marsico, M. Nappi, D. Riccio, Noisy iris recognition integrated scheme, Pattern Recognit. Lett. 33 (8) (2012) 1006–1011.
[13] M. De Marsico, M. Nappi, D. Riccio, H. Wechsler, Mobile Iris Challenge Evaluation (MICHE)-I, biometric iris dataset and protocols, Pattern Recognit. Lett. 57 (2015) 17–23.
[14] C. Galdi, J.-L. Dugelay, Fusing iris colour and texture information for fast iris recognition on mobile devices, in: International Conference on Pattern Recognition, ICPR 2016, 2016, pp. 1–5.
[15] C. Galdi, J.-L. Dugelay, Fire: Fast iris recognition on mobile phones by combining colour and texture features, Pattern Recognit. Lett. Same Volume TBA.
[16] M. Haindl, M. Krupička, Unsupervised detection of non-iris occlusions, Pattern Recognit. Lett. 57 (2015) 60–65.
[17] K.P. Hollingsworth, S.S. Darnell, P.E. Miller, D.L. Woodard, B.K. W., P.J. Flynn, Human and machine performance on periocular biometrics under near-infrared light and visible light., IEEE Trans. Inf. Forensics Secur. 7 (2) (2012) 588–601.
[18] N. Liu, M. Zhang, H. Li, Z. Sun, T. Tan, Deepiris: learning pairwise filter bank for heterogeneous iris verification, Pattern Recognit. Lett. (2015).
[19] F. Narducci, A. Abate, S. Barra, L. Gallo, Kurtosis and skewness at pixel level as input for som networks to iris recognition on mobile, Pattern Recogn. Lett. Same Volume TBA.
[20] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe, FRVT 2006 and ICE 2006 large-scale experimental results, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 831–846.
[21] H. Proença, 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, IEEE, 2007, pp. 1–4.
[22] H. Proença, L.A. Alexandre, Toward covert iris biometric recognition: experimental results from the NICE contests, IEEE Trans. Inf. Forensics Secur. 7 (2) (2012) 798–808.
[23] H. Proença, S. Filipe, R. Santos, J. Oliveira, L.A. Alexandre, The UBIRIS.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance, IEEE Trans. Pattern Anal. Mach. Intell. 32 (8) (2010) 1529–1535.
[24] K. Raja, R. Raghavendra, S. Venkatesh, C. Busch, Multi-patch deep sparse features for iris recognition in visible spectrum using collaborative subspace for robust verification, Pattern Recognit. Lett. Same Volume TBA.
[25] R.P. Wildes, Iris recognition: an emerging biometric technology, Proc. IEEE 85 (9) (1997) 1348–1363.

Please cite this article as: M. De Marsico et al., Results from MICHE II – Mobile Iris CHallenge Evaluation II, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2016.12.013