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

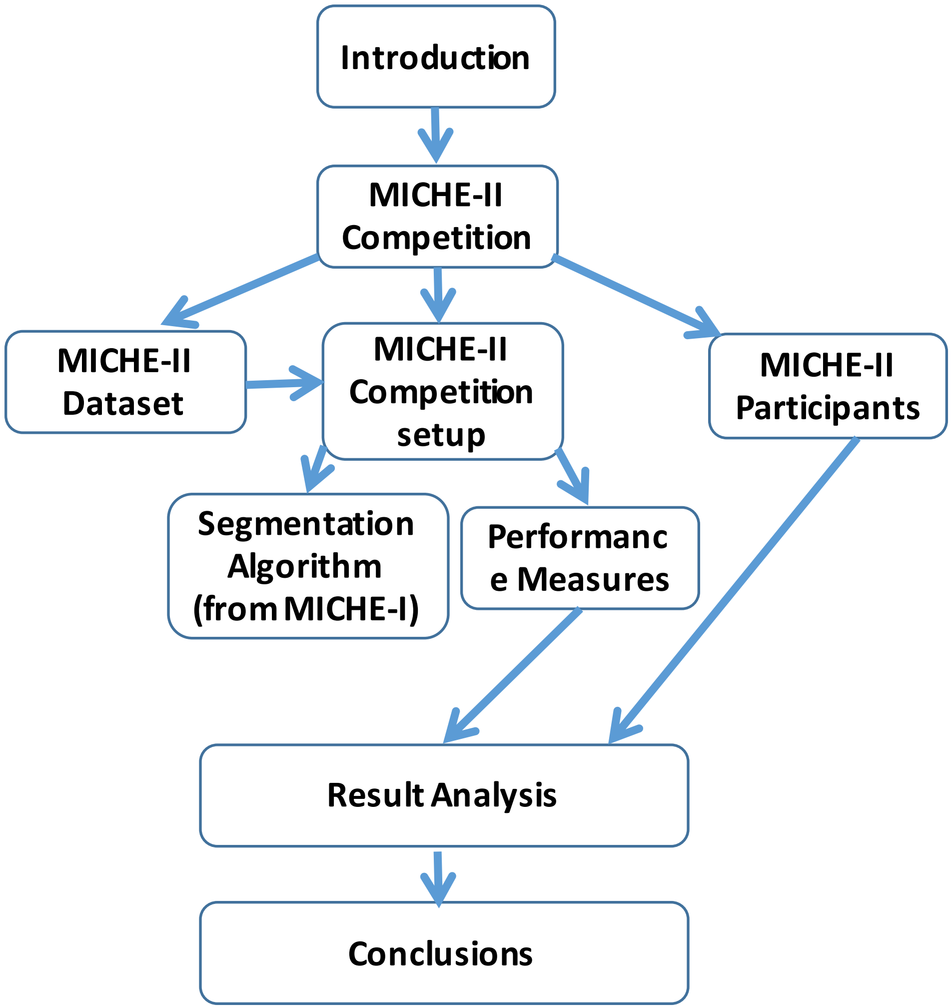
Maria De Marsico
Sapienza University of Rome, Italy
demarsico@di.uniroma1.it

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

Fabio Narducci
University of Salerno, Italy
fnarducci@unisa.it

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

OUTLINE

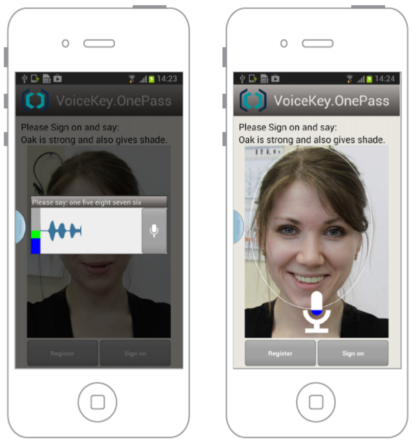


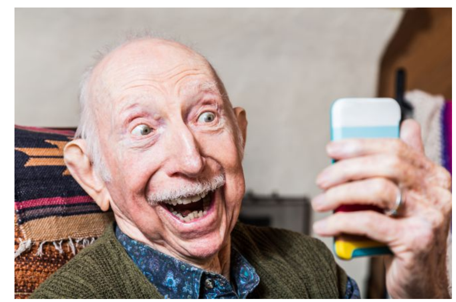
Introduction
MICHE-II Competition
MICHE-II Dataset
MICHE-II Competition setup
MICHE-II Participants
Segmentation Algorithm (from MICHE-I)
Performanc e Measures
Result Analysis
Conclusions

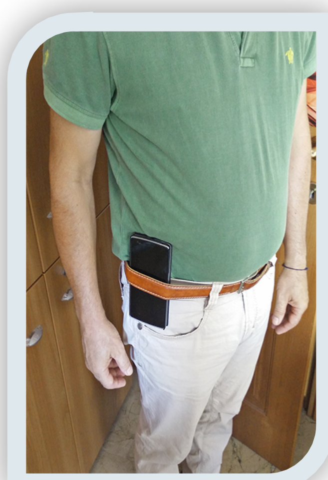
MOTIVATIONS OF RELATED RESEARCH

Increasing popularity of mobile biometrics

Uncontrolled conditions
Non-technical users
Lower computational resources

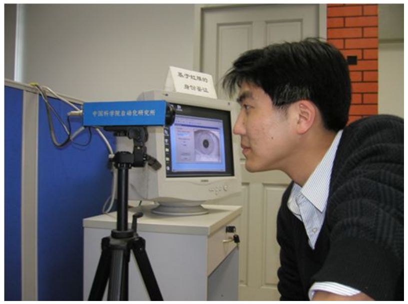


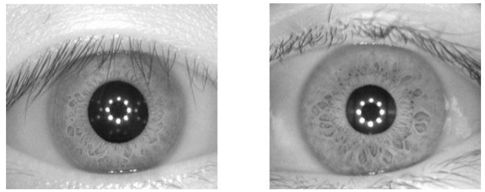




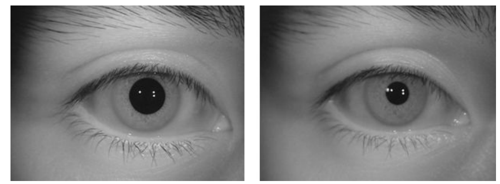
WHY MICHE-II

CASIA DATASET (LATEST VERSION)



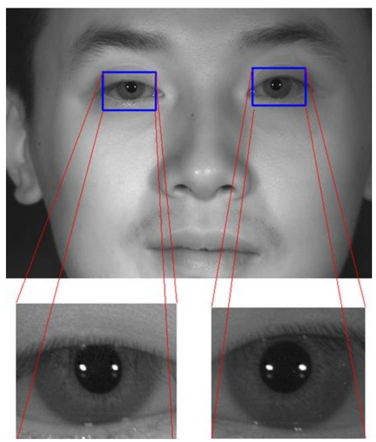


CASIA_INTERVAL



CASIA_LAMP

CASIA_DISTANCE

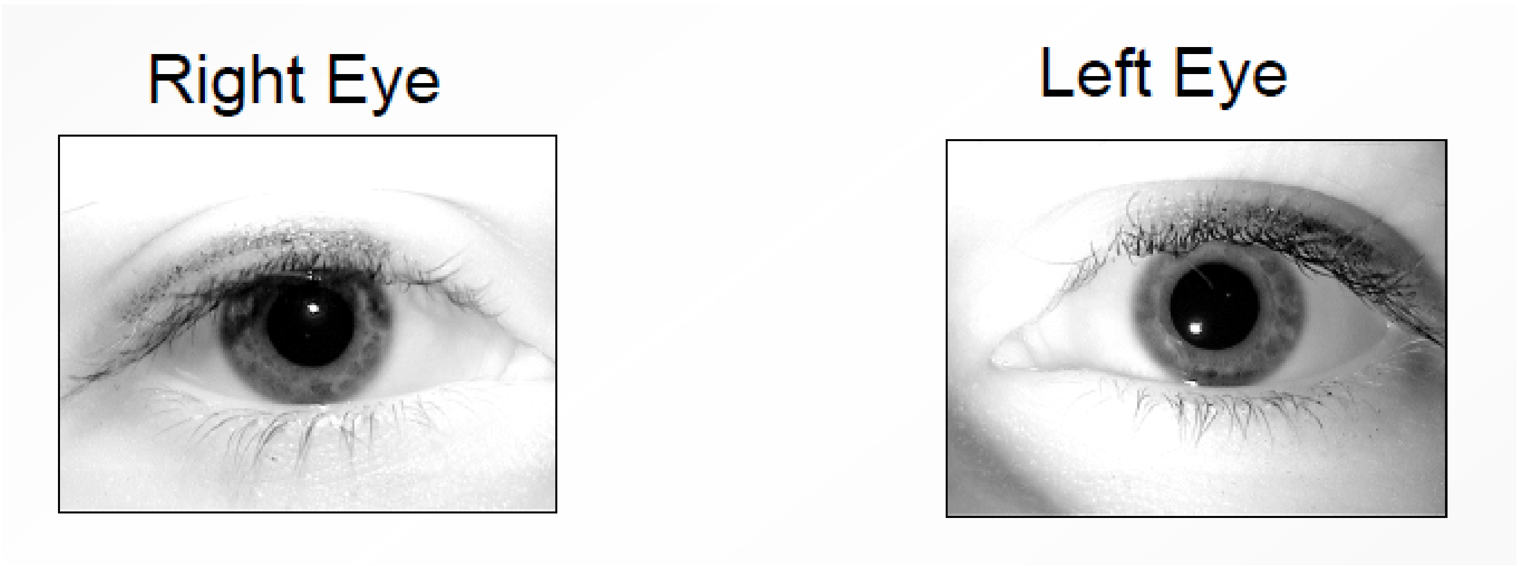


CASIA_IRISTHOUSANDs

Why not suited for mobile testing?
A few mobile devices at present are equipped with (cheap and easy usable) NIR sensors

WHY MICHE-II

ICE (Iris Challenge Evaluation) COMPETITIONS by NIST

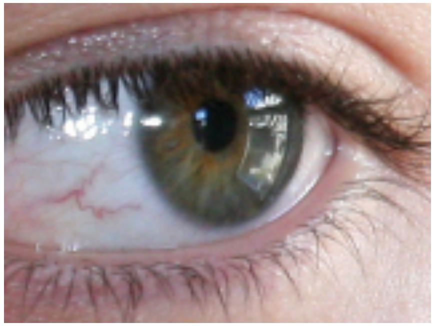


Why not suited for mobile testing?
A few mobile devices at present are equipped with (cheap and easy usable) NIR sensors

WHY MICHE-II

NICE (Noisy Iris Challenge Evaluation) COMPETITIONS by SOCIA LAB AT BEIRA INTERIOR

DA COMPLETARE



First crucial difference: images in visible light
Second crucial difference: uncontrolled conditions, normal equipment

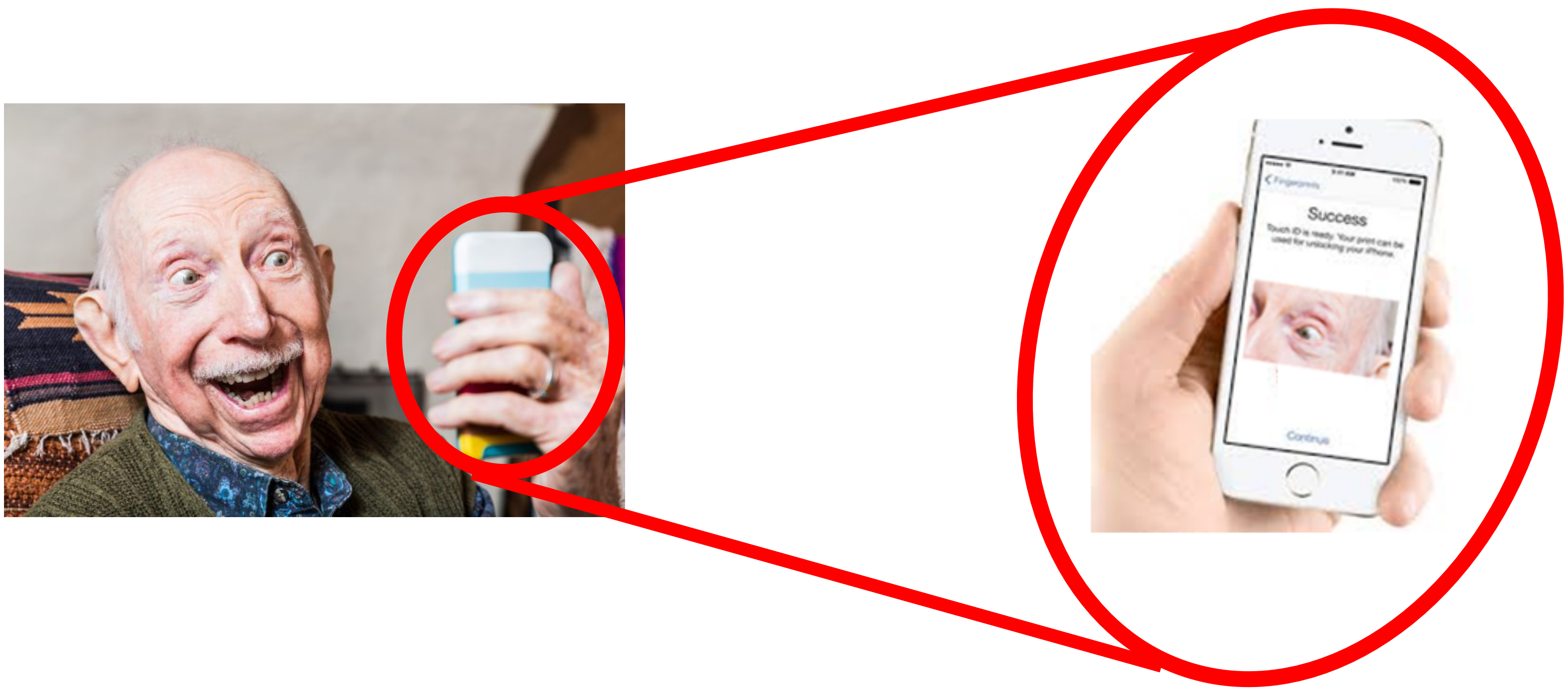
Why not much suited for mobile testing?
Very high image resolution

MICHE DATASET

- Images captured
  - in visible light
  - by «normal» user-level mobile devices

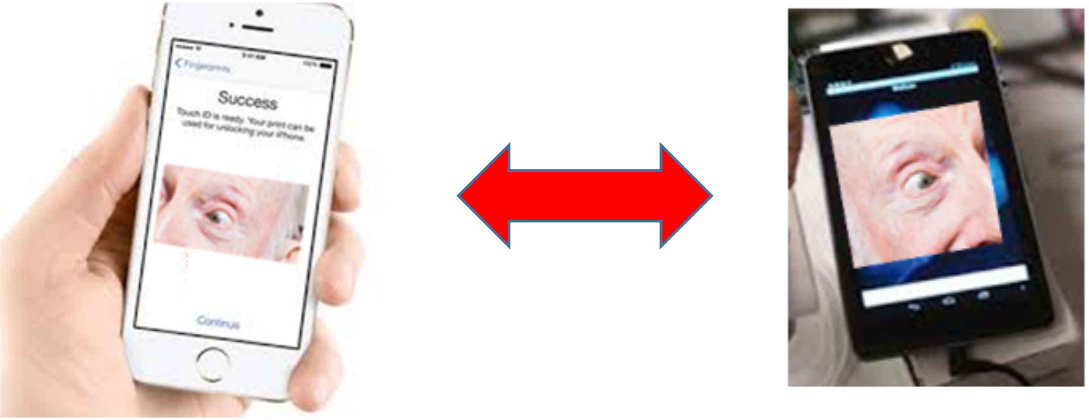
MICHE DATASET

- Images captured by the user in uncontrolled/unattended conditions



MICHE DATASET

- Images captured by different devices for cross-device matching



MICHE II COMPETITION SETUP – COMMON SEGMENTATION

- All participants had to exploit the results of segmentation provided by the best algorithm in MICHE I competition*

- For all participants, tests were repeated at BIPlab according to the common protocol

*Haindl, M., Krupiˇcka, M., 2015. Unsupervised detection of non-iris occlusions. Pattern Recognition Letters 57, 60–65.

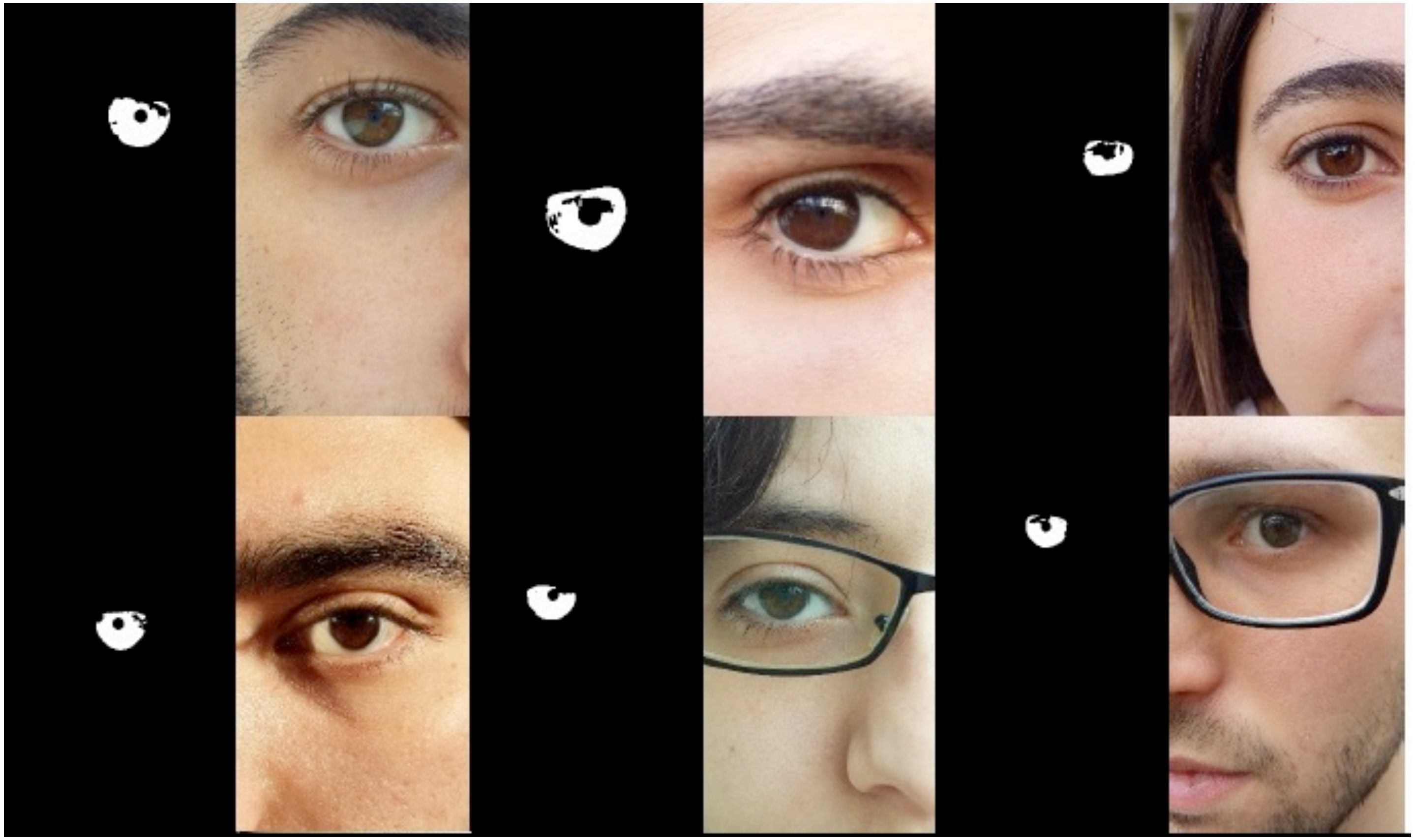
MICHE II COMPETITION SETUP – DISTANCE MEASURE

- Participants were to provide any distance measure they deemed suitable for their approach, given that it was a (semi)metric.

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

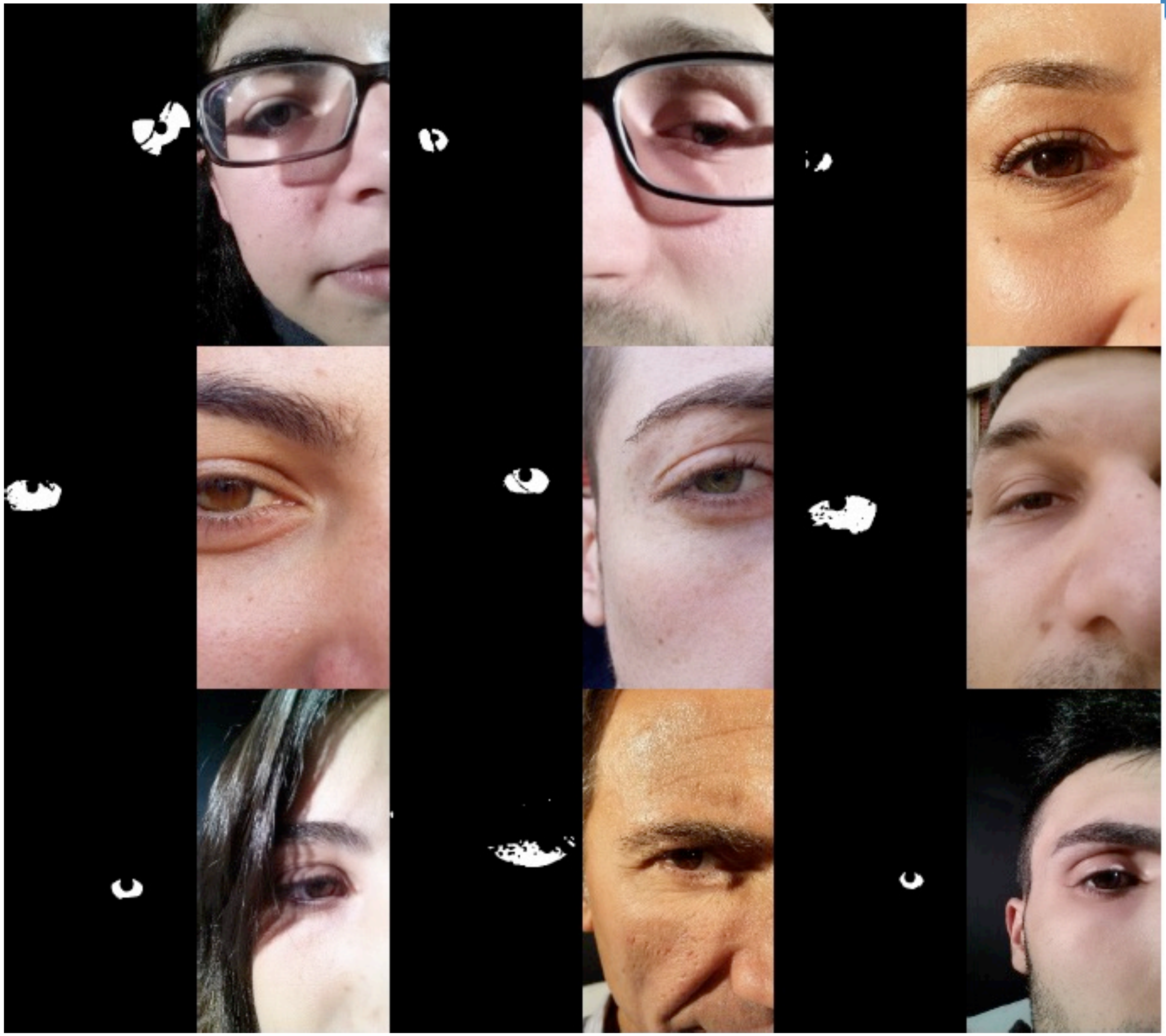
1. $D(I_a, I_a) = 0$
2. $D(I_a, I_b) = 0 \rightarrow I_a = I_b$
3. $D(I_a, I_b) = D(I_b, I_a)$

MICHE II COMPETITION SETUP – «GOOD» CASES



- Good segmentation
- Sufficient (and connected) iris region for matching

MICHE II COMPETITION SETUP – «BAD» CASES



- Poor segmentation (possible fragments)
- Possibly insufficient iris region for matching

MICHE II COMPETITION SETUP – PERFORMANCE MEASURES

- Identification (1:N) = Recognition Rate (RR)
- Verification (1:1) = Area Under (ROC) Curve (AUC)

MICHE II COMPETITION – PARTICIPANTS

- Algorithms presented in this session
- We adopted a suitable labeling in reporting results:

  - `tiger_miche` described in Ahmed et al.
  - `karanahujax` described in Ahuja et al.
  - `irisom` described in Abate et al.
  - `FICO_matcher` described in Galdi and Dugelay
  - `otsedom` described in Aginako-Bengoa et al.
  - `ccpsiarb` presented in Aginako-Bengoa et al.

  - The authors of a further proposal labeled `Bata` did not provide the details requested for publication, therefore we will not include the corresponding results in the presentation

MICHE II COMPETITION – PARTICIPANTS

- Participants with more proposals
  - `karanahujax`
    - Model1: Deep Learning in unsupervised mode
    - Model2: Deep Learning in supervised mode
  - `irisom`
    - 5×5 and 10×10 refer to the size of the exploited Self Organizing Map (SOM)
  - `FICO_matcher`
    - V1: distance measure considers color, texture and cluster descriptors
    - V2: distance measure only considers color and cluster descriptors
  - `ccpsiarb` (different combinations of computer vision transformations and machine learning-based classifiers)
    - `ccpsiarb_17`: Edge transformation + IB1 classifier
    - `ccpsiarb_2`: Equalize transformation + IB1 classifier
    - `ccpsiarb_42`: Gaussian transformation + IB1 classifier

MICHE II COMPETITION – RESULTS FOR SAME-DEVICE MATCHING

Algorithm	GS4 vs GS4			
	RR	AUC	Global Score	Time(seconds)
tiger_miche	1,00	1,00	1,00	1,72
karanahujax_Model1	0,83	0,97	0,90	5,72
karanahujax_Model2	0,83	0,95	0,89	4,62
irisom_10_10	0,77	0,88	0,82	3,45
otsedom	0,67	0,94	0,80	42,37
FICO_matcher_V1	0,67	0,89	0,78	1,00
Irisom_5_5	0,63	0,88	0,75	3,00
ccpsiarb_42	0,63	0,81	0,72	289,27
ccpsiarb_17	0,63	0,81	0,72	61,43
ccpsiarb_2	0,63	0,81	0,72	265,38
FICO_matcher_V2	0,50	0,79	0,65	0,57

`karanahujax` versions achieve comparable results, as well as `ccpsiarb`

`irisom` is less stable

the most significant difference in performance is achieved by `FICO_matcher` versions

time was not considered for ranking, but we observe that best methoda also achieve acceptable processing times (single match)

MICHE II COMPETITION – RESULTS FOR SAME-DEVICE MATCHING

Algorithm	IP5 vs IP5			
	RR	*AUC*	*Global Score*	*Time(seconds)*
tiger_miche	1,00	1,00	1,00	1,71
karanahujax_Model1	1,00	1,00	1,00	5,63
karanahujax_Model2	0,93	0,98	0,96	4,69
FICO_matcher_V1	0,87	0,98	0,92	1,01
Irisom_5_5	0,87	0,93	0,90	3,05
irisom_10_10	0,83	0,92	0,88	3,51
otsedom	0,63	0,92	0,78	41,10
ccpsiarb_17	0,70	0,85	0,77	54,30
FICO_matcher_V2	0,57	0,93	0,75	0,57
ccpsiarb_2	0,63	0,86	0,75	252,84
ccpsiarb_42	0,63	0,81	0,72	283,39

`karanahujax` versions are still consistent, as well as different versions of `ccpsiarb`

`irisom` is stable this time (lower resolution = less noise?)

again, the most significant difference in performance is achieved by `FICO_matcher` versions

as for time, the same observation regarding the best methods hold, but it is worth observing the speed of `FICO_matcher_V2`

MICHE II COMPETITION – RESULTS FOR CROSS-DEVICE MATCHING (HARDEST)

Algorithm	**All vs ALL**			
	RR	***AUC***	***Global Score***	***Time(seconds)***
tiger_miche	1,00	0,99	0,99	1,71
karanahujax_Model2	0,92	0,86	0,89	4,65
karanahujax_Model1	0,88	0,76	0,82	5,68
irisom_10_10	0,80	0,78	0,79	3,48
otsedom	0,63	0,93	0,78	41,74
Irisom_5_5	0,75	0,79	0,77	3,03
FICO_matcher_V1	0,73	0,80	0,77	1,00
ccpsiarb_17	0,68	0,83	0,75	57,64
ccpsiarb_2	0,65	0,82	0,74	259,27
ccpsiarb_42	0,65	0,81	0,73	286,20
FICO_matcher_V2	0,48	0,73	0,61	0,57

`karanahujax` versions are still consistent, as well as different versions of `ccpsiarb`

`irisom` is quite stable

the most significant difference in performance is achieved by `FICO_matcher` versions

as for time, the same observation regarding the best methods hold, but it is worth observing the speed of `FICO_matcher_V2`

MICHE II COMPETITION – SUMMARY RESULTS (ALL VERSIONS)

Rank	Algorithm	ALLvsALL	GS4vsGS4	Ip5vsIP5	Final Score
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

most methods achieve better results with IP5vsIP5 (lower resolution = less noise?)

MICHE II COMPETITION – FINAL RESULTS

Rank	Algorithm	ALLvsALL	GS4vsGS4	Ip5vsIP5	Final Score
1	tiger_miche	0,99	1,00	1,00	1,00
3	karanahujax_Model2	0,89	0,89	0,96	0,91
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

we choose `karanahujax_Mode2` instead of `karanahujax_Model` (they got the same final score) since it has a more stable behaviour across operational settings, and also achieves better results in cross-device matching

FINAL OBSERVATIONS

- In general, methods presented in different versions are quite stable across test coditions, except for FICO_matcher
- The best metods also generally require less computational time per single match
- The better the ranking achieved, the more stable the method with respect to the test setting
- All methods provide consistently lower performances in ALLvsALL condition.
- On the average, the images over which the best results are achieved in same-device settings come from IP5 (mean value= 0,89), and the achieved results further present a lower standard deviation (0,086). Lower resolution = less noise?

	ALLvsALL	GS4vsGS4	Ip5vsIP5
Average Global Score achieved by proposed algorithms	0,85	0,87	0,89
Variance of Global Score achieved by proposed algorithms	0,090596	0,0943836	0,0865148

Conclusions

- The performance achieved by the methods participating in MICHE-I were not comparable to 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 accuracy.
- Results suggest that it is worth further exploring possible improvements of the availabletechniques.

THANK YOU!