Human versus Automatic Measurement of Biometric Sample Quality

Andy Adler, Tatyana Dembinsky
School of Information Technology, University of Ottawa, Canada

Introduction

- Measures of biometric quality are notoriously difficult
- Typically, we have considered (implicitly or explicitly) humans to be the correct judge of quality
- We wanted to understand the relationship between human quality measures and those from machines

Data sets

- 98 face images from the Mugshot Identification Database [1]
  - 29 different people
  - 3-4 samples

Fig. 1. Sample images from the Mugshot Identification Database.

- 84 iris images from the local database taken by an LG iris camera
  - 7 different people
  - 6 samples per eye

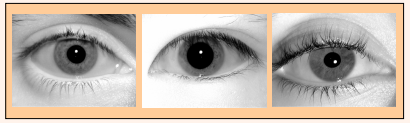

Fig. 2. Sample images from the local iris database.

Experiments

	Face images	Iris images
Human quality	8 subjects	8 subjects
Biometric quality	6 algorithms	1 algorithm [2]
Image quality measures	IQM [3]	IQM

Tab. 1. Experiments aimed at measuring image quality using different test methodologies.

Image quality evaluation by human participants

- Web based evaluation form was used
- Users were allowed to familiarize with the database
- Instructions: "Evaluate the quality of each image for a biometric identification application "

Fig. 3. Web based evaluation form for human evaluation of biometric image quality.

Quality from match scores

- *Model*: Match score (MS) from genuine comparisons is due to image qualities

$$MS_{i,j} = Q_i Q_j \qquad 0 < MS < 1,\ 0 < Q < 1$$

$$\log MS_{i,j} = \log Q_i + \log Q_j$$

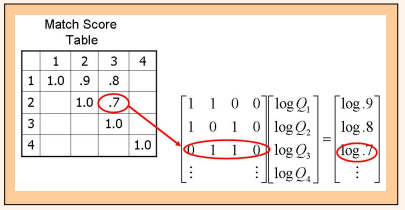

Fig. 4. An example of quality calculation for a set of match scores between 4 images of the same person.

Comparisons

- *Are humans consistent with each other?* **YES**
  - *Face*: average correlation coefficient $r = 0.613$, $p < 0.001$
  - *Iris*: average correlation coefficient $r = 0.723$, $p < 0.001$
- *Are algorithms consistent with each other?* **YES**
  - *Face*: average correlation coefficient $r = 0.534$, $p < 0.001$
  - *Iris*: impossible to analyze, only one algorithm was used
- *Are humans consistent with algorithms or other quality measures?* **NO**

	Mean algorithm	IQM
Mean human	0.234	0.159
Mean algorithm		0.003

Tab. 2. Correlation of biometric image quality measures for *face* images.

	Mean algorithm	IQM
Mean human	0.175	0.458
Mean algorithm		-0.036

Tab. 3. Correlation of biometric image quality measures for *iris* images.

- Human vs. IQM resulted in higher correlation due to preference for sharp images

Discussion

- In general, both algorithms and humans are consistent with others of the same group
- Correlations between different evaluator groups were not significant
- Naïve ideas about quality measures may not be relevant to algorithms
- Human evaluation of image quality may not be a good standard for biometric sample quality

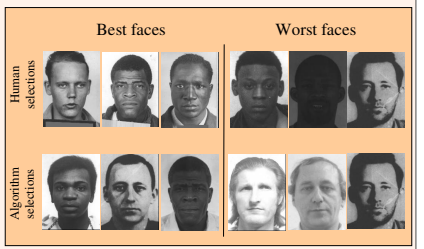

Fig. 5. The average highest and lowest quality selections of face images done by humans and algorithms.

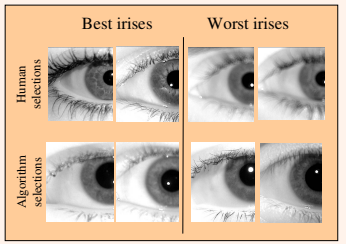

Fig. 6. The average highest and lowest quality selections of iris images done by humans and algorithms.

References

[1] NIST, *NIST Special Database 18: Mugshot Identification Database* (MID), http://www.nist.gov/srd/nistsd18.htm .
[2] L.Masek, "Recognition of Human Iris Patterns for Biometric Identification", BE Dissertation, The University of Western Australia, 2003, www.csse.uwa.edu.au/~pk/studentprojects/libor/index.htm .
[3] The MITRE Corporation, Image Quality Measure (IQM)© Software, http://www.mitre.org/tech/mtf/ .

For further information

Please contact *adler@site.uottawa.ca* or *tdembins@site.uottawa.ca* .
More information on this and related projects can be obtained at http://www.site.uottawa.ca/~adler/publications/publications.html .