Human vs. Automatic Measurement of Biometric Sample Quality

Andy Adler, Tanya Dembinsky
University of Ottawa

Background

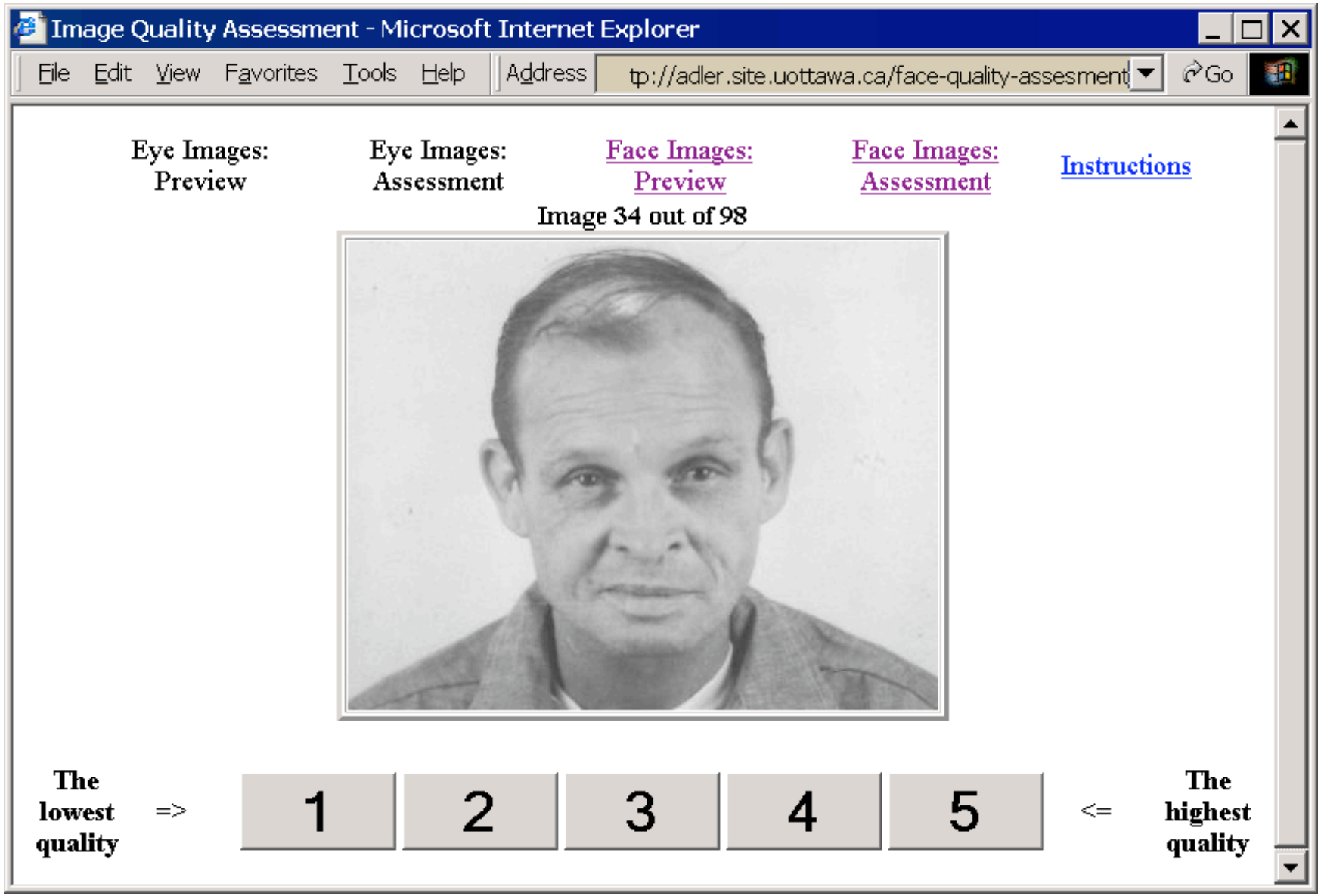
- Measures of biometric quality are notoriously difficult
- Typically, we have considered (implicitly or explicitly) humans to be the correct judge of quality
- We wanted to understand the relationship between human quality measures and those from machines

Experiments

	Face Mugshot DB	**Iris** Our DB
Human Quality	8 subjects	8 subjects
Biometric Quality	6 algorithms	1 algorithm
Image Quality Measures	IQM[1]	IQM[1]

[1] www.mitre.org/tech/mtf/

Human Quality Evaluation



Issues in Human Evaluations

- Scale differences
  - Analysis cannot compare raw values
- Training Effect
  - Users were allowed to familiarize with database
- What is evaluated?
  - Instructions were: "assess biometric image quality"

Quality from Match scores

Model: *MS from genuine comparisons is due to image qualities*

Except:

- Identical comparisons
- Different pose / age / etc.

$$MS_{i,j} = Q_i Q_j$$

$0 < MS < 1$

$0 < Q < 1$

Quality from Match Scores

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

Match Score Table

	1	2	3	4
1	1.0	.9	.8	
2		1.0	.7	
3			1.0	
4				1.0

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ \vdots & & & \vdots \end{bmatrix} \begin{bmatrix} \log Q_1 \\ \log Q_2 \\ \log Q_3 \\ \log Q_4 \end{bmatrix} = \begin{bmatrix} \log .9 \\ \log .8 \\ \log .7 \\ \vdots \end{bmatrix}$$

Comparisons

- Are humans consistent with each other?
- Are algorithms consistent with each other?
- Are humans consistent with algorithms, or other quality measures?

Are humans consistent?

Face

- Yes ($p<.001$)
- Average correlation coefficient *r=.613*

Iris

- Yes ($p<.001$)
- Average correlation coefficient *r=.723*

Are algorithms consistent?

Face

- Yes ($p<.001$)
- Average correlation coefficient *r=.534*
- Highest correlations not between different versions of same vendors SW

Iris

- Could not analyse (only one alg.)

Humans vs. algorithms

Face

Iris

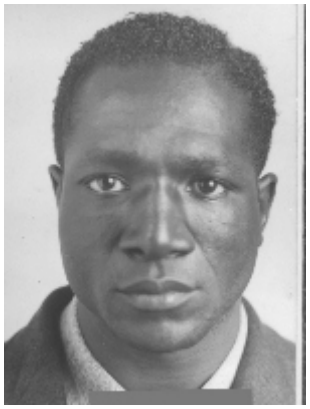
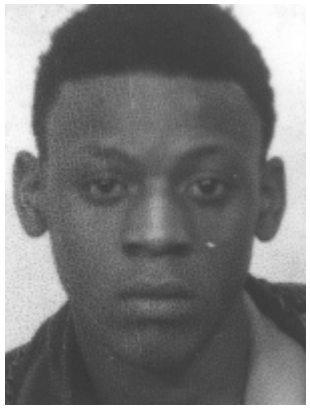
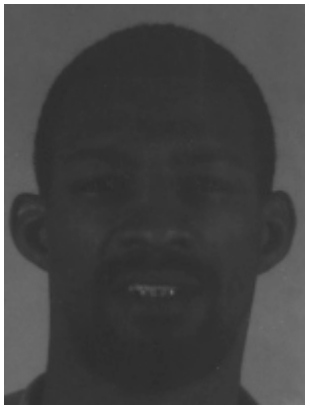
	Mean Human	Mean FR Alg	IQM
Mean Human		.234	.159
Mean Alg	.175		.003
IQM	.458	-0.036	

← Best Faces

Worst Faces →

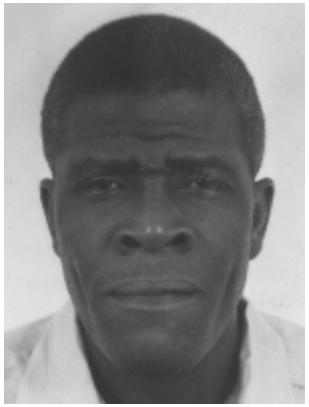
Human Selections



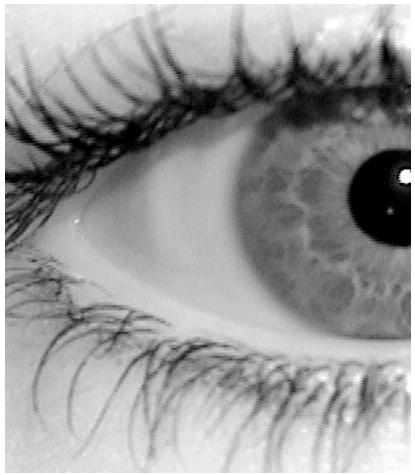
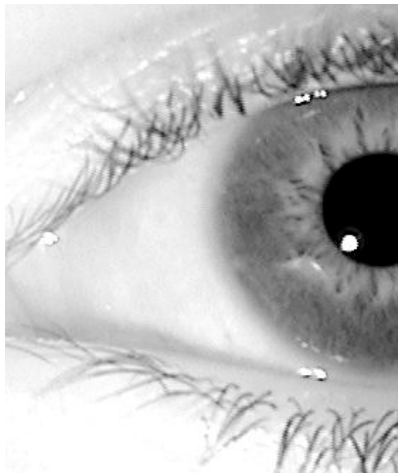

Algorithm Selections



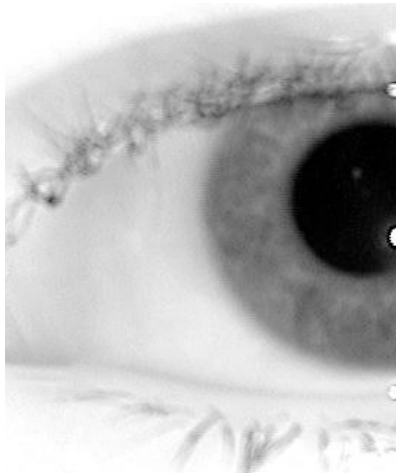



← Best Irises

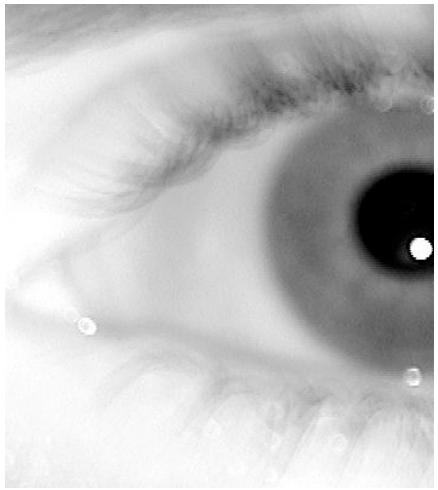
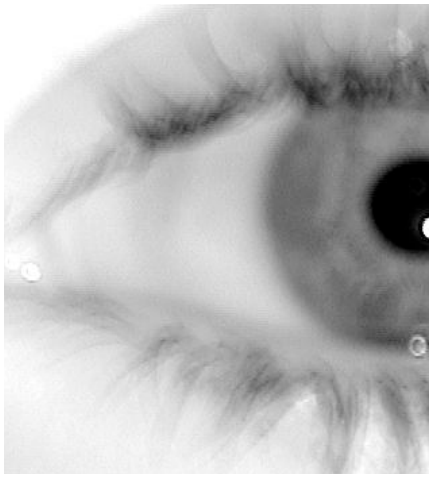
Human Selections

Algorithm Selections




Worst Irises →

Discussion

- Work done on Face / Iris.
  - Fingerprints are different because there are fingerprint experts
- Humans are consistent
- Algorithms are consistent

- ***But,*** humans are not consistent with algorithms

What does this mean?

- Naïve ideas about quality measures may not be relevant to algorithms

- Some countries are vetting submitted passport photos for Face Rec
  - How useful is this really?

Comment: *Quality*

- *Quality* is a value laden term
- Can we tell users this?

- Maybe we need another term: *Clarity?*