

Human vs. Automatic face recognition performance

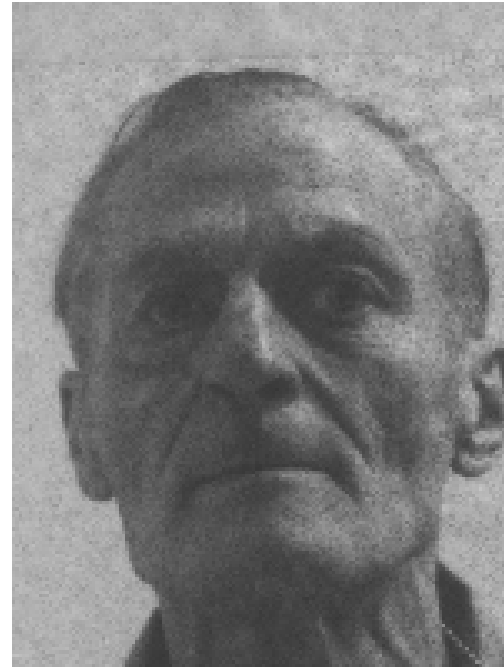
Andy Adler

School of Information Technology and Engineering
University of Ottawa, Canada

James Maclean

3M Security Systems Division,
1545 Carling ave, Ottawa, Canada

Are these the same person?



Same person? **Yes**

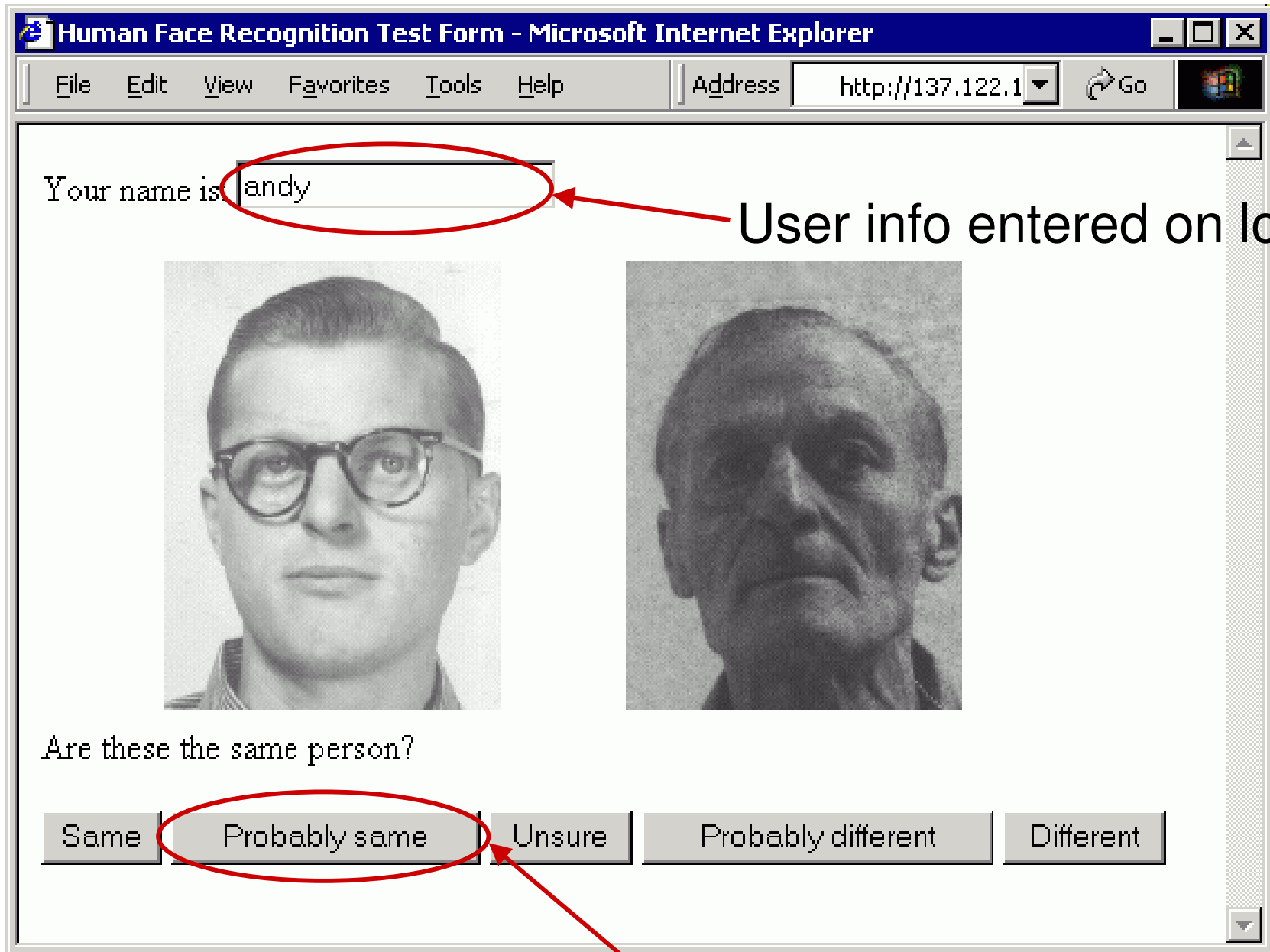
- I have just demonstrated a massively parallel face recognition computer
- Of all biometric modalities, automatic face recognition is most often compared to human performance
- Surprisingly little work has been done to quantify these levels of performance

Talk Outline

- Test design
- Results
- Discussion
 - issues and unknowns

Test Design

Participants	Employees of 3M Security Systems Division (then AiT) in Ottawa, Canada
Participation	Voluntary – announcement at company weekly meeting
Participant demographics	16 Male, 5 Female, ages 20-40, predominantly Caucasian
Test format	Web based: subject participated from their office
Instructions	Focus on accurate results



User info entered on login page

Select choice: no time limit

Choice of images

- *Goldilocks* problem:
 - Too easy test -> all score 100%
 - Too hard test -> all score 0%
- Database used: *NIST Mugshot*
 - Large age changes between captures
 - Population that tends to change appearance

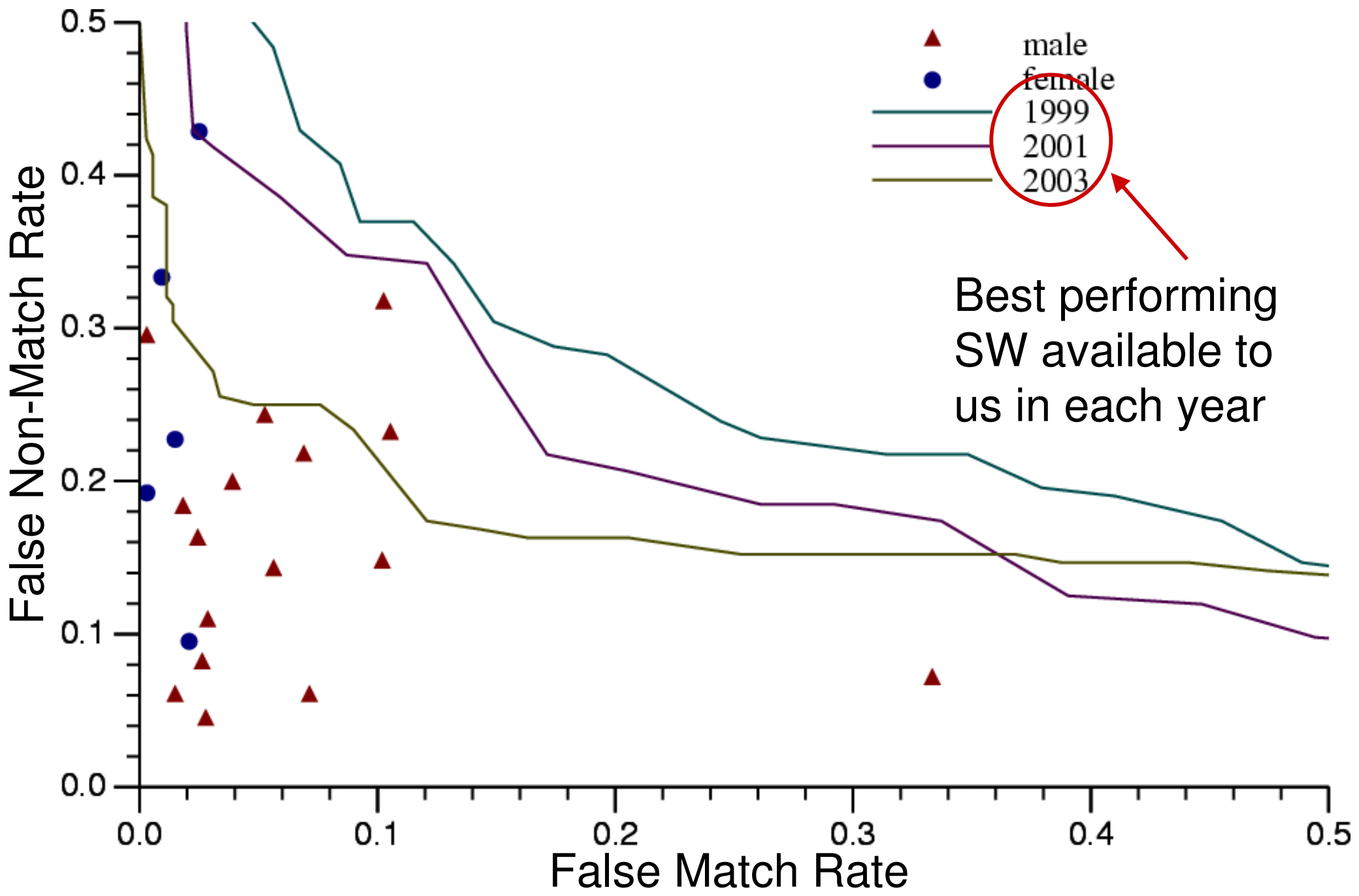
Analysis

■ Human results

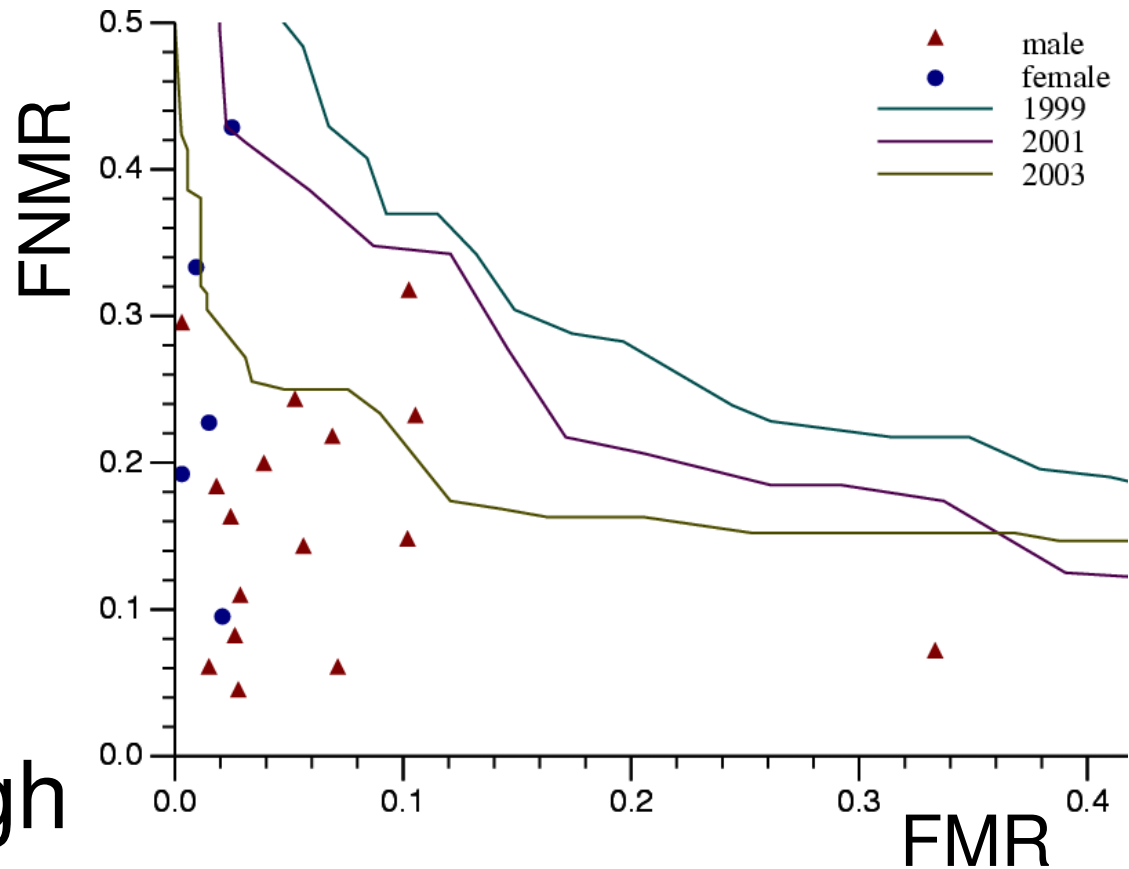
- Post-processed to choose optimal “threshold” for them
- An operating point FMR/FNMR calculated

■ Software results

- Same images presented to FR software (worked with 13 packages and versions)
- ROC curve calculated



Results



- Error rates are high
- Significant improvement in SW 1999-2003
- Most motivated, attentive humans can outperform face recognition software
- No significant difference male/female

Conclusion

- Currently, most people are able to significantly outperform FR software on difficult data sets
 - Unlimited time
 - Motivated staff

Discussion

This is a limited study

- Our participants

- Untrained,
- Motivated
- Attentive

- Unlimited time given (average was about 15s)

- Challenging database

- 1/3 matching, 2/3 non-matching pairs presented
- Keeps interest level
- ... but is it representative?

Previous work

- Significant body of work in psychological literature about human face rec.
- Much is not of much interest for biometrics
 - Eg. Recognize familiar faces, famous faces

Other studies

- Kemp et al. (1997) analysed the ability of supermarket cashiers to identify shoppers from photos on credit cards
 - Results show poor performance.
- Chang Hong et al. (2003) analyzed the ability of people to match poor-quality video footage against high-quality photographs
 - Results show high performance.
- Burton et al. (1998,2001) compared PCA based and graph-matching algorithms against human ratings
 - Primarily to elucidate aspects of human memory not to evaluate algorithms

Unanswered questions

1. How do humans perform as familiarity increases?
2. What is the effect of motivation, routine and boredom?
3. Do experts outperform untrained recognizers?
4. What distinguishes good recognizers from poor ones?
5. What if a live subject is available?

References

www.site.uOttawa.ca/~adler

- Burton A M, Miller P, Bruce V, Hancock P J B, Henderson Z (2001) Human and automatic face recognition: a comparison across image formats *Vision Research* **41** 3185-3195
- Chang Hong L, Seetzen H, Burton A M, Chaudhuri A (2003) Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied*. **9** 33-41.
- Hancock P J B, Bruce V, Burton M A (1998) A comparison of two computer-based face identification systems with human perceptions of faces *Vision Research* **38** 2277-2288
- Kemp R, Towell N, Pike G (1997) When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*. **11** 211-222