Automated Facial Age Estimation

Mei Ngan + Patrick Grother

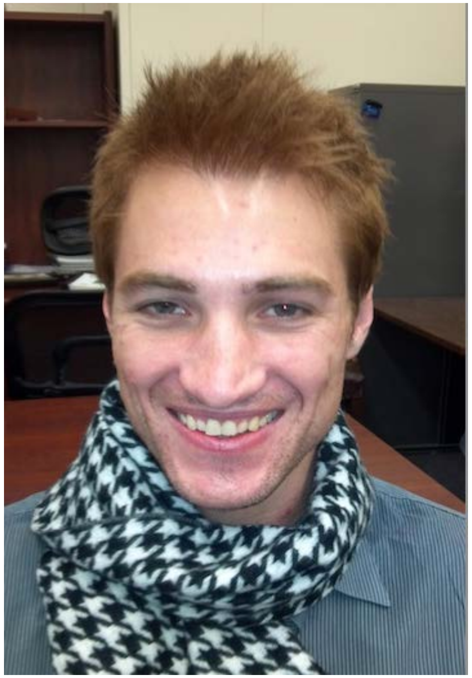
Information Technology Laboratory
National Institute of Standards and Technology (US),
United States Department of Commerce

NTIA
Thursday, November 6, 2014

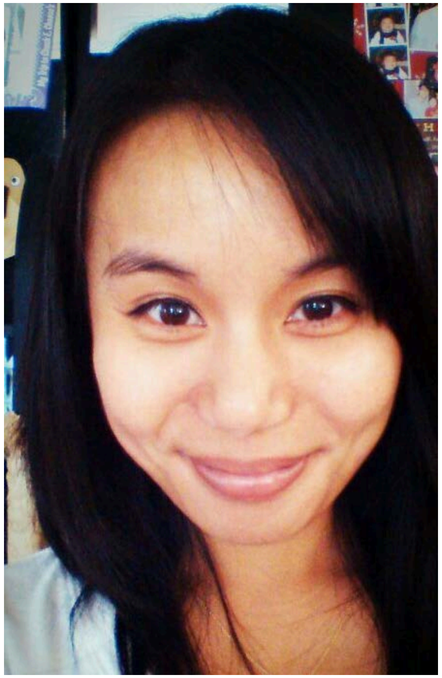
What is automated facial age estimation

How old are these people?

Estimated Age: 46
True Age: ??

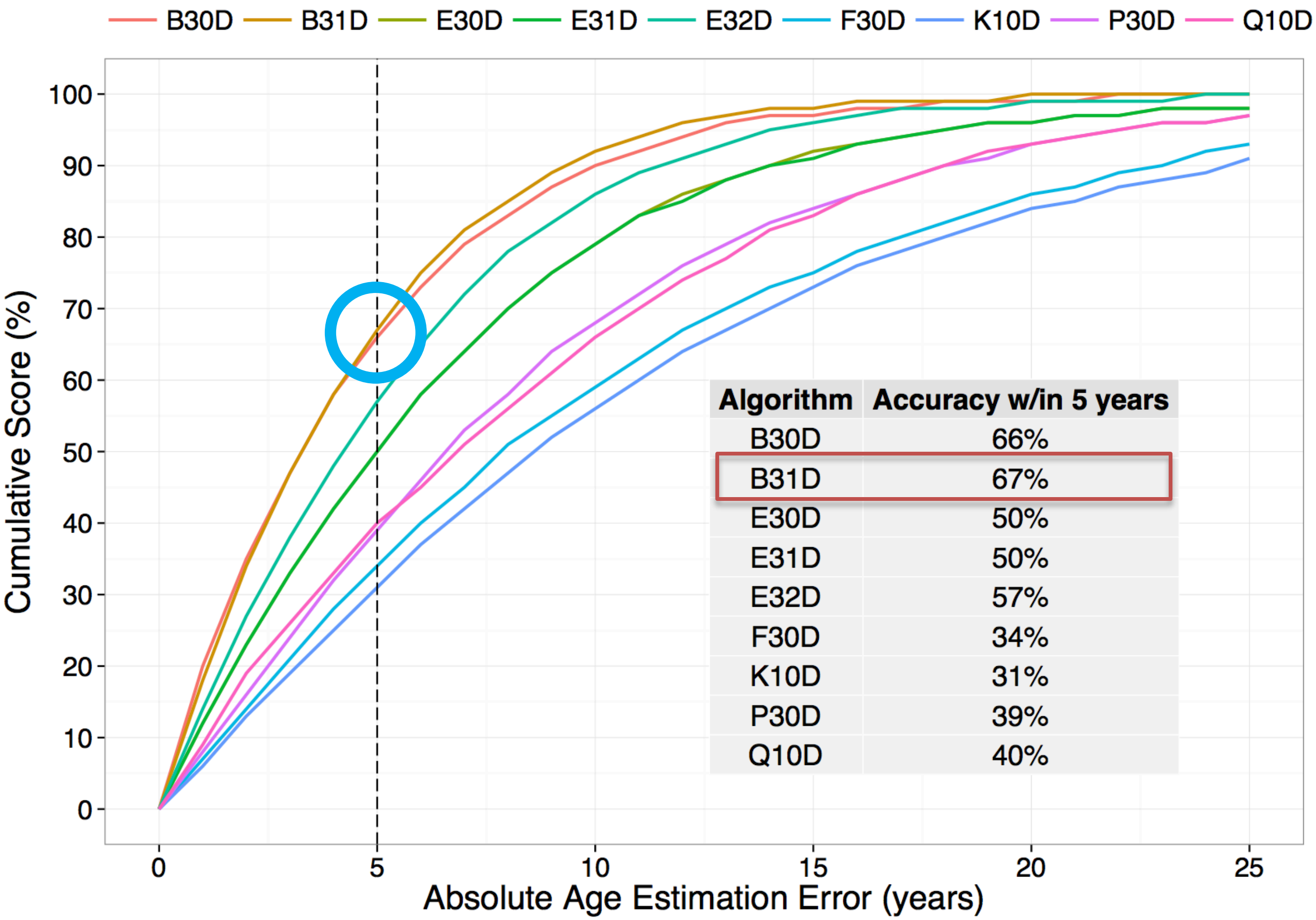


Estimated Age: 26
True Age: 32



Estimated Age: 16
True Age: 32

Age Estimation Accuracy & Error over Large Homogeneous Population of 6M



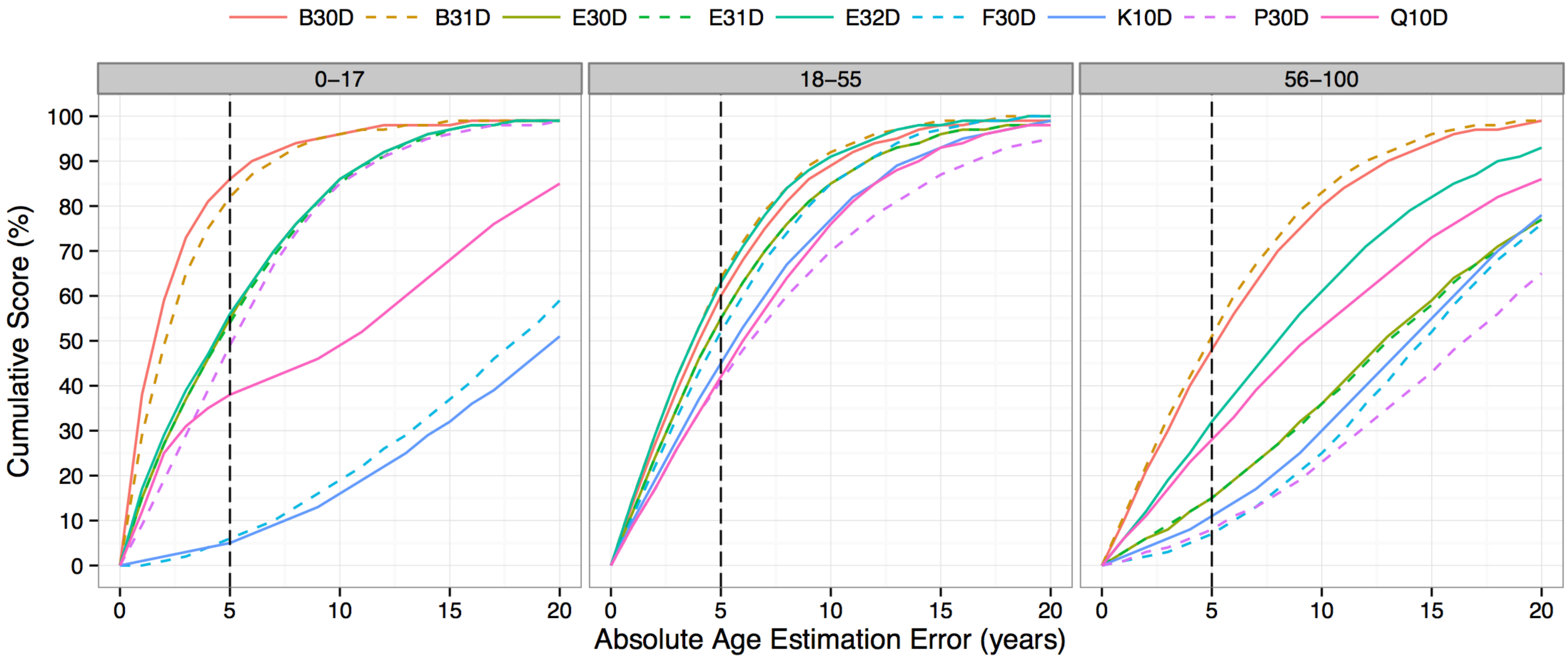
Algorithm	Accuracy w/in 5 years
B30D	66%
B31D	67%
E30D	50%
E31D	50%
E32D	57%
F30D	34%
K10D	31%
P30D	39%
Q10D	40%

Mean Absolute Error (years)

Algorithm	Mean
B30D	4.5
B31D	4.3
E30D	6.5
E31D	6.6
E32D	5.3
F30D	10.2
K10D	12.2
P30D	8.3
Q10D	8.5

For the most accurate algorithm, 67% of estimates are accurate within 5 years with a Mean Absolute Error (MAE) of 4.3 years.

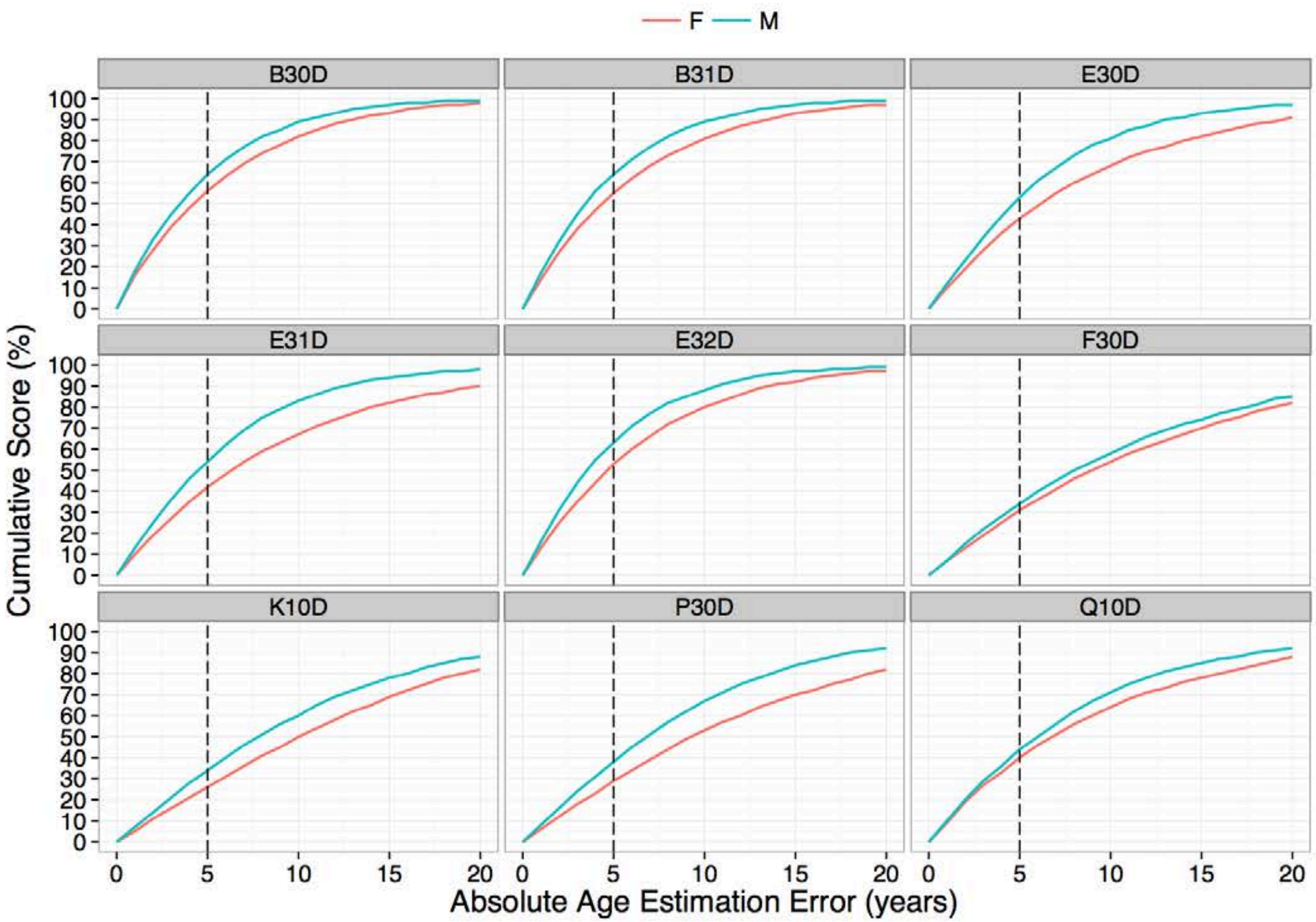
Age Estimation Accuracy & Error by Age Group



Age Group	Num Images	B30D	B31D	E30D	E31D	E32D	F30D	K10D	P30D	Q10D
0–17	1605807	2.6	3	5.3	5.4	5.3	18.6	21	6.1	10.9
18–55	3781607	4.9	4.5	5.5	5.5	4.6	5.6	6.6	7.6	7
56–100	785287	6.2	5.8	13.9	14	9	14.7	14	16.7	10.9

Mean Absolute Error (years)

Age Estimation Accuracy & Error by Gender



Mean Absolute Error (years)

Gender	Female	Male
Num Images	118108	124894
B30D	5.7	4.7
B31D	5.9	4.7
E30D	8.5	6.1
E31D	8.6	5.8
E32D	6.2	4.8
F30D	11.2	10.2
K10D	11.9	9.9
P30D	11.6	8.5
Q10D	9.5	8.2

Results:
All algorithms estimate age more accurately on males than females.

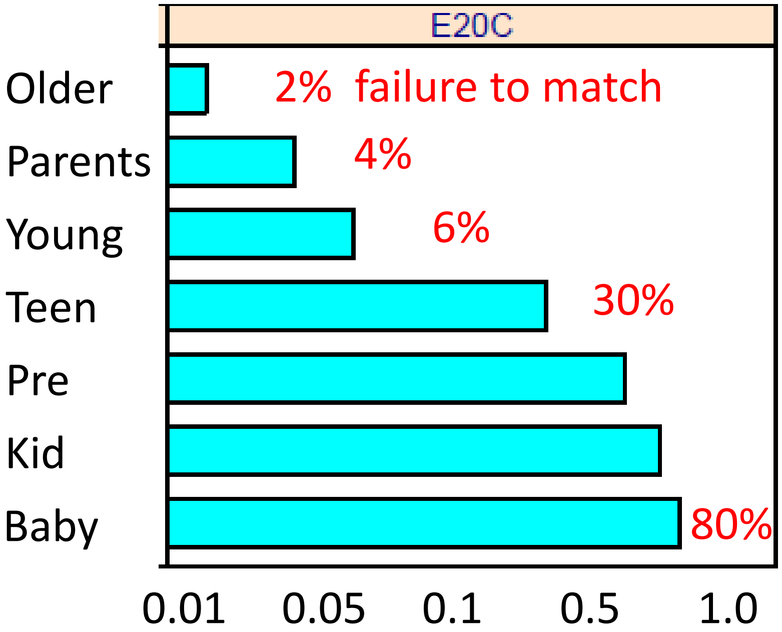
Face Recognition Accuracy By Age Group

Identification miss rates by age group

- » Older ← [56,120]
- » Parents ← [31,55]
- » Young ← [20,30]
- » Teen ←[14,19]
- » Pre ← [9,13]
- » Kid ← [4,8]
- » Baby ← [0,3]

Visa images:
Enrolled size, N = 19972

Mated searches = 19972
Non-mated searches = 203082



One-to-many "miss rate"
FNIR when threshold set to produce a false positive in only 1 in 100 non-mate searches (FPIR = 0.01)

Identification miss rates by age group, algorithm

Accuracy = F(Age, Ageing)

- Baby ← [0,3] Mean time lapse = 1.6
- Kid ← [4,8] Mean time lapse = 3.0
- Pre ← [9,13] Mean time lapse = 3.9
- Teen ←[14,19] Mean time lapse = 2.7
- Young ← [20,30] Mean time lapse = 2.0
- Parents ← [31,55] Mean time lapse = 2.1
- Older ← [56,120] Mean time lapse = 2.2

Accuracy by age group :: Summary

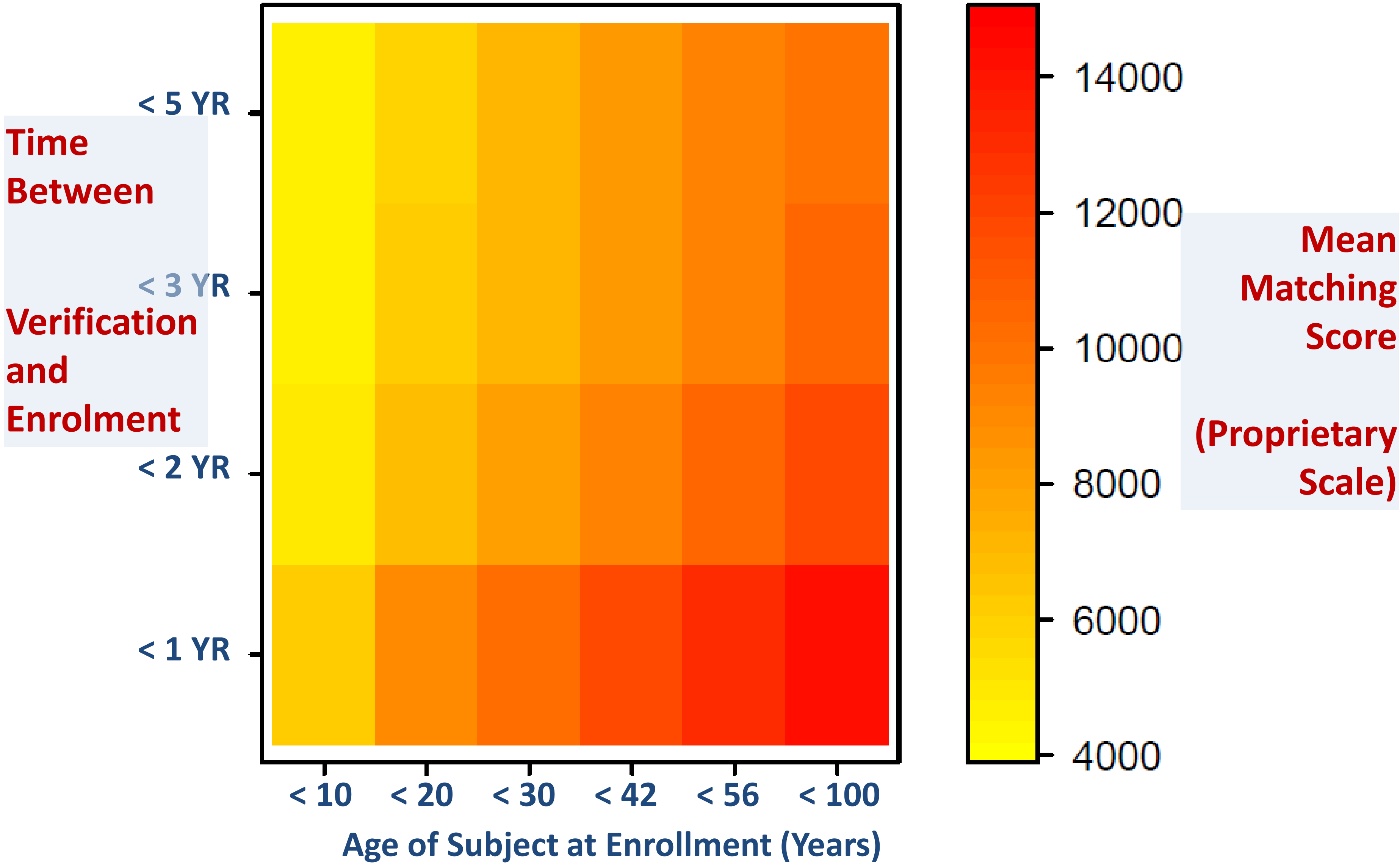
» Using visa photographs, younger people, especially but not limited to children, are more difficult to recognize.

» Lifelong trend to be more easily recognized. This is a big effect, larger than other drivers in face recognition.

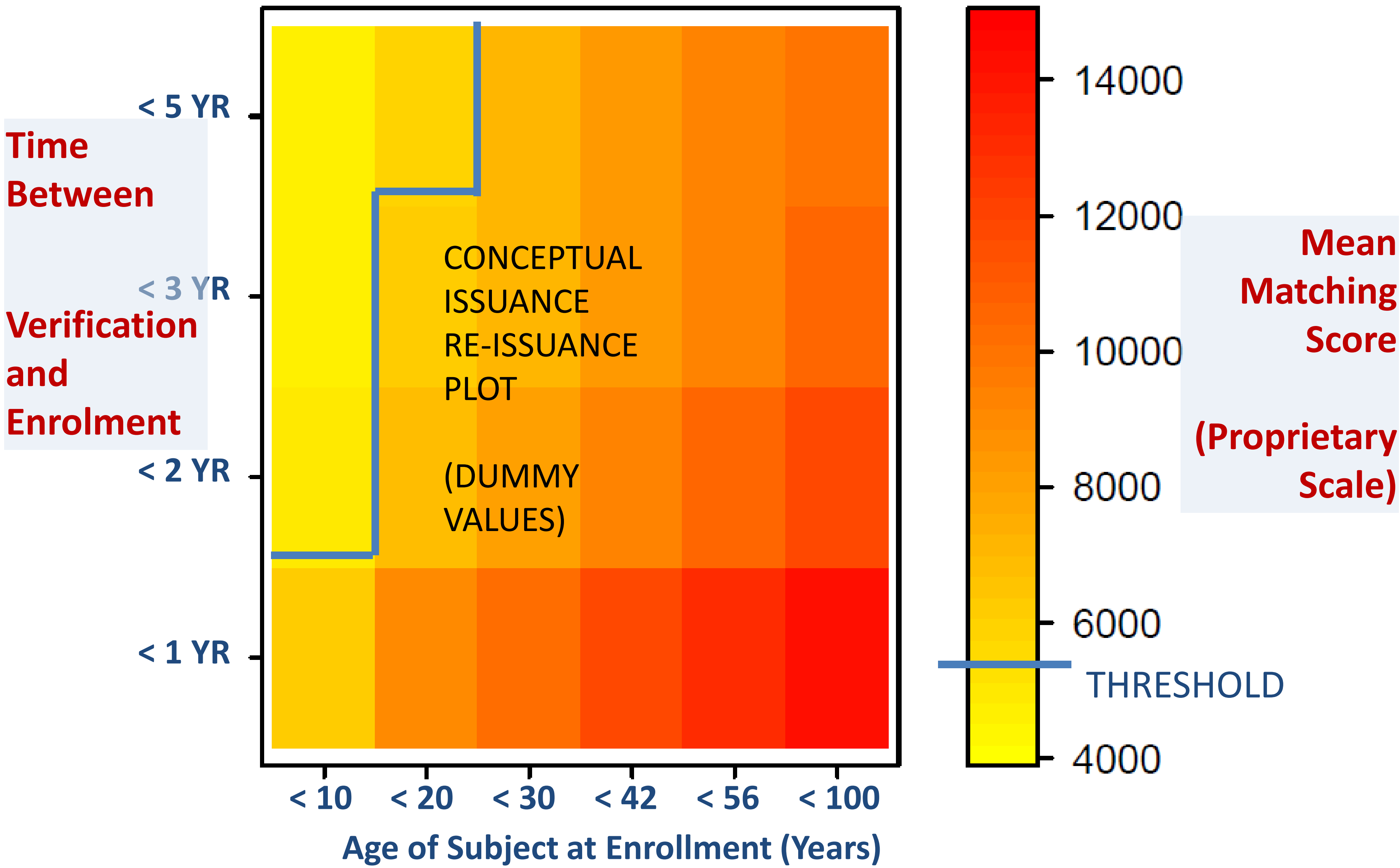
» Two effects:

- **Repeatability:** Older people more easily recognized as themselves.
- **Distinguishability:** Older people more easy to distinguish from others.

Face Visa Data :: Accuracy(Age, ΔT)



Face Visa Data :: Accuracy(Age, ΔT)



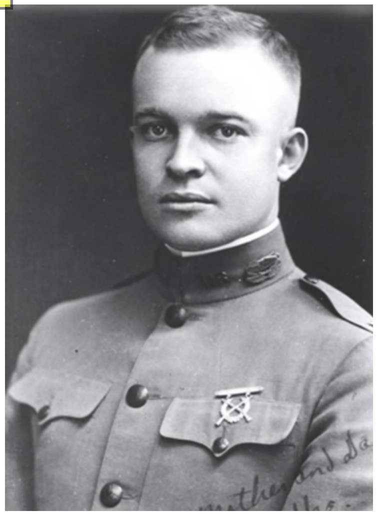
FACE RECOGNITION VENDOR TEST

NIST

Face Ageing
Quantification + Relevance

Patrick Grother + Mei Ngan
Information Access Division
National Institute of Standards and Technology

NTIA Meeting, Washington, DC
Thursday, November 6, 2014



Ageing: Permanent Appearance Change

Dwight D Eisenhower

ALGORITHM X	0.647	0.601	0.599	0.579
ALGORITHM Z	0.595	0.578	0.565	0.548

Green indicates successful 1:1 authentication at FMR = 0.001.

Red indicates failure.

FACE AGEING ⟶ DECREASED SIMILARITY.

IS THERE AN ANALOGOUS EFFECT FOR OTHER MODALITIES?

Photographs on exhibit at
Museum of Modern Art, NYC

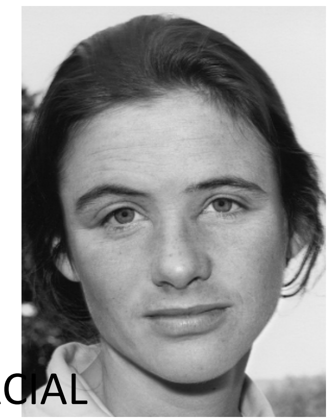
See Susan Minot's text in
NY Times Magazine
Sunday Oct 3 2014

The Brown Sisters

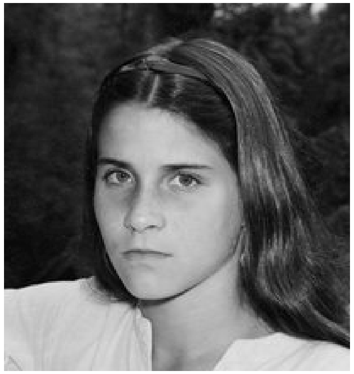
Photographed every year from
1975-2014

Ageing

Brown Sister #1

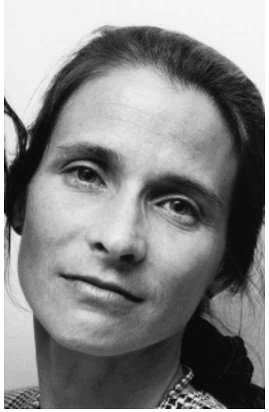
	T ~ 5	T ~ 10	T ~ 20	T ~ 30	T ~ 40 Years
					
X	0.632	0.608	0.584	0.602	0.576
Y	3004	2954	2755	2845	2781
Z	0.622	0.616	0.613	0.517	0.426

THREE LEADING COMMERCIAL FR ALGORITHMS



Ageing

Brown Sister #2

T ~ 5	T ~ 10	T ~ 20	T ~ 30	T ~ 40 Years
				

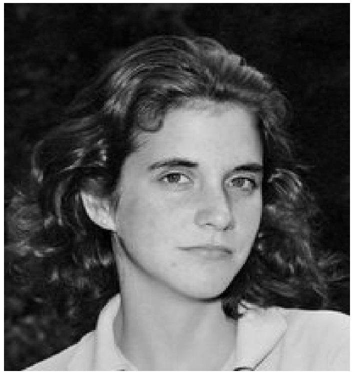
FR ALGORITHM	T ~ 5	T ~ 10	T ~ 20	T ~ 30	T ~ 40 Years
X	0.648	0.601	0.600	0.610	0.605
Y	2863	2821	2758	2752	2824
Z	0.617	0.593	0.506	0.531	0.533

Ageing

Brown Sister #3

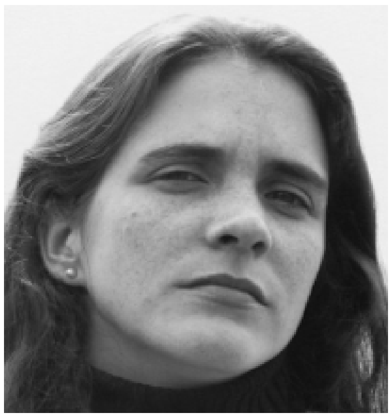
T ~ 5	T ~ 10	T ~ 20	T ~ 30	T ~ 40 Years
				

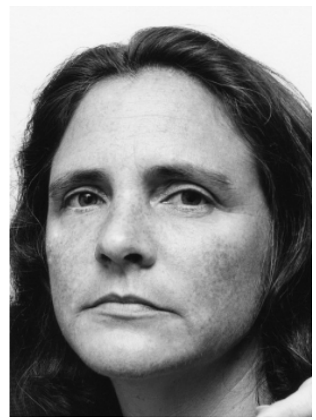
FR ALGORITHM					
X	0.673	0.635	0.627	0.607	0.586
Y	2847	2649	2687	2637	2630
Z	0.610	0.511	0.524	0.595	0.472



Ageing

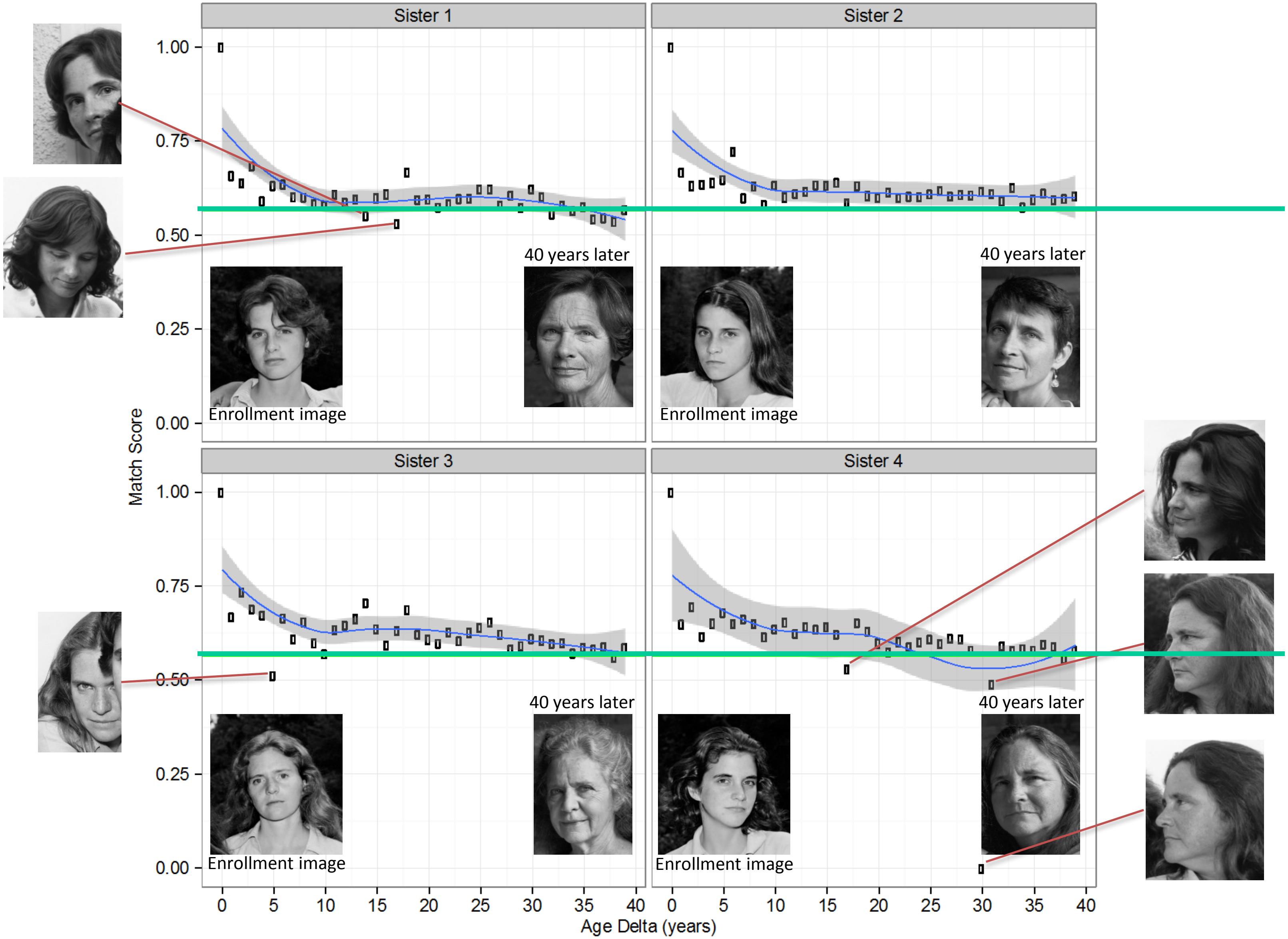
Brown Sister #4





FR ALGORITHM	T ~ 5	T ~ 10	T ~ 20	T ~ 30	T ~ 40 Years
X	0.652	0.654	0.603	0.591	0.578
Y	3055	2795	2743	2847	2607
Z	0.632	0.516	0.475	0.524	0.432

Verification over 40 years



Reasoning about Ageing

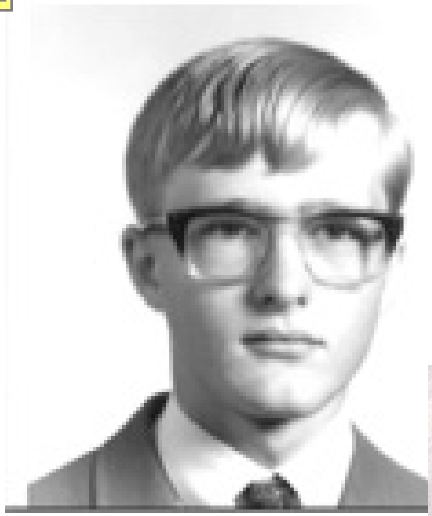
- The simplest conception of ageing is that:
  - Accuracy = F(Time-of-Enrollment – Time-of-Recognition) = F(ΔT)
- And we all ageing "steadily":
  - Accuracy = a - b ΔT *"if we're lucky, or simplistic, linear ageing"*
- Inexorable change: *"It's a one way street, and downhill at that"*
  - Accuracy = F(monotone(ΔT))
  - Modulo cosmetics(?), botox(?), surgery(?) and … photoshop
- But at least it's graceful:
  - Accuracy = F(slowly varying function(ΔT, n))
  - Absent injury, disease, abuse
- But … complications
  - Unsteady ageing: *"Five years at 30 is not five years at 40"*
    - Accuracy = F(Age-at-Enrollment; ΔT) or, simple Taylor expansion,
    - Accuracy = F(Age-at-Enrollment, Age-at-Recognition)
  - Person-specific ageing: *"Some age better than others"*
    - $\text{Accuracy}_i = F_i$(Age-at-Enrollment, Age-at-Recognition) subscript i

Longitudinal Analysis

Quantifying Permanence Using Data from a Large-Population Operational System

Ageing :: Longitudinal data

Brad Wing

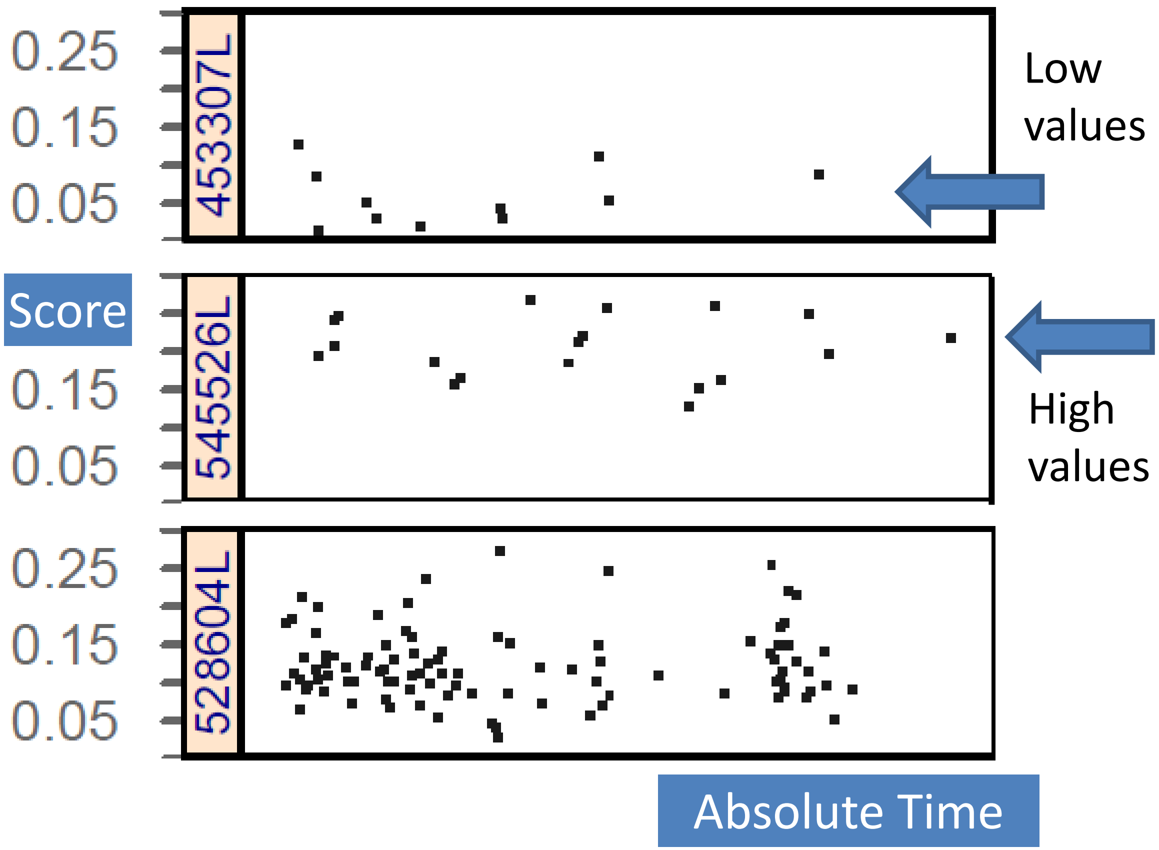


ALGORITHM E20A	0.617	0.578	0.532	0.541
ALGORITHM J20A	0.589	0.587	0.579	0.569

Green indicates successful 1:1 authentication at FMR = 0.001.
Red indicates failure.

LONGITUDINAL ANALYSIS APPLIED TO ALGORITHM SCORE DATA

Quantify ageing :: Individual recognition scores over time



» Often, visually flat

» Considerable variance within subject

» Considerable variance between subjects

» Irregular sampling

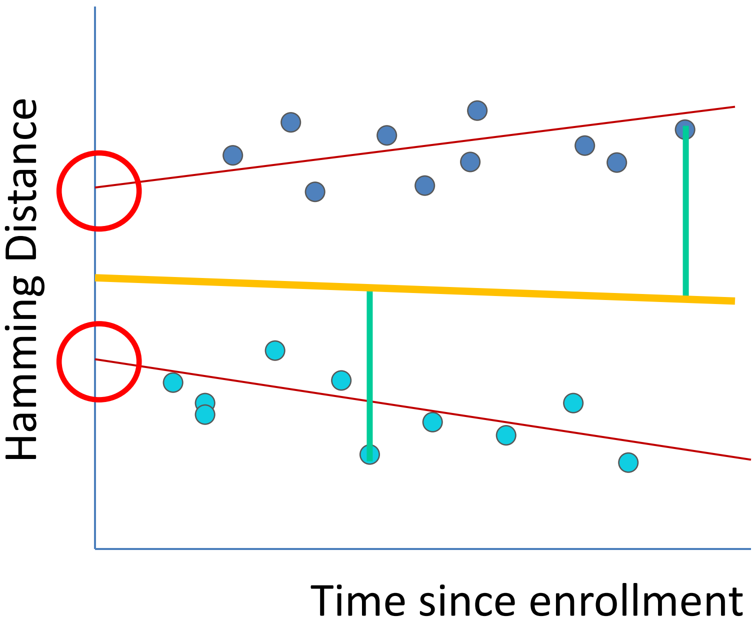
» Imbalanced sampling

» Mixed effects models

- Shared population part
- Individual part

TRAJECTORIES INDICATE HETEROGENEITY – INTERCEPTS (AND GRADIENTS) VARY WITH QUALITY OF THE ENROLLMENT IMAGE cf. DODDINGTON's ZOO

Quantifying permanence via mixed-effects regression



Model for the j-th score from the i-th eye

$$HD_{ij} = \pi_{0i} + \pi_{1i} T_{ij} + \epsilon_{ij}$$

Intercept is sum of population average term, the *fixed effect*, and an eye-specific *random effect*

$$\pi_{0i} = \gamma_{00} + \psi_{0i}$$

Slope is sum of population average term, the *fixed effect*, and an eye-specific *random effect*

$$\pi_{1i} = \underline{\gamma_{10}} + \psi_{1i}$$

Permanence stated by the population wide rate at which scores are decreasing.

Subject to assumptions:

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

$$\begin{bmatrix} \psi_{0i} \\ \psi_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01}^2 \\ \sigma_{10}^2 & \sigma_1^2 \end{bmatrix} \right)$$

MIXED EFFECTS MODEL RESPECT IDENTITY INFORMATION. SIMPLE LINEAR REGRESSION, IN YELLOW, DOES NOT AND HAS OTHER PROBLEMS

Conclusions

» Brown sisters: existence proof that 1:1 face authentication is possible over thirty years
  - But scores become weaker.
  - Successful 1:N identification demands stronger scores

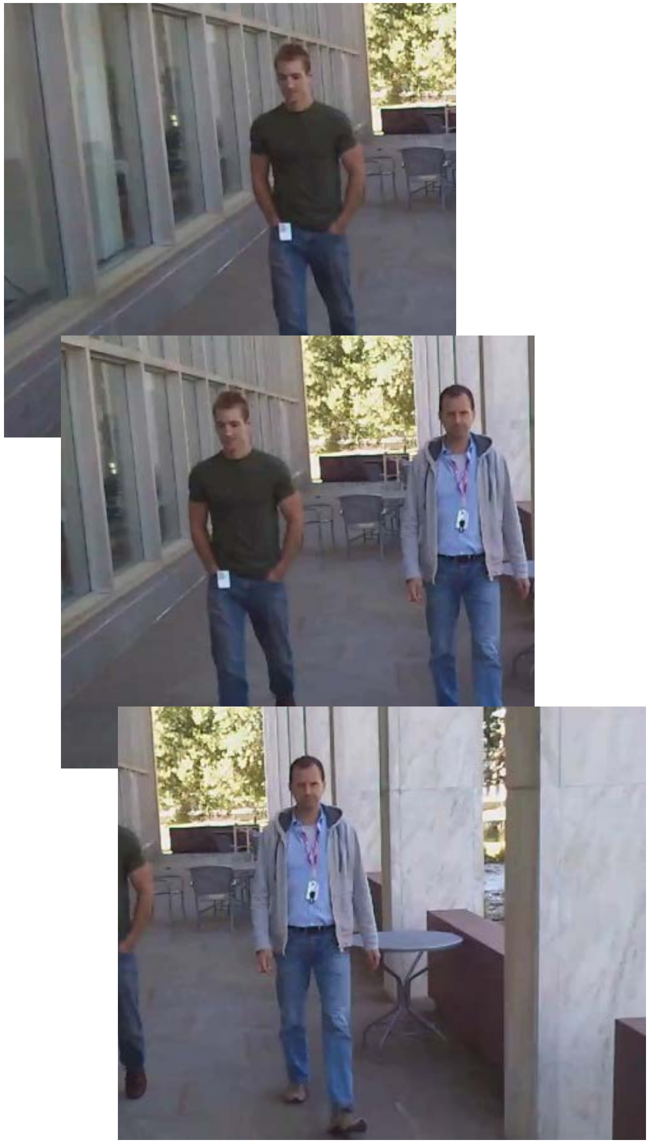
» No good long term face ageing studies. e-Passports and digital photography will change that… eventually.
  - And suitable longitudinal analysis methods are published (NIST, MSU)

» But, there’s a “so what” for some use cases:
  - Algorithms improve on a timescale shorter than ageing
  - Identity credentials are re-issued on a timescale shorter than ageing
    - But possibility to recycle old photos
  - Law enforcement + counter terrorism functions have no such luxury

F I V E

Face In Video Evaluation (FIVE)

FR in Video :: Scope



Face In Video Evaluation (FIVE)

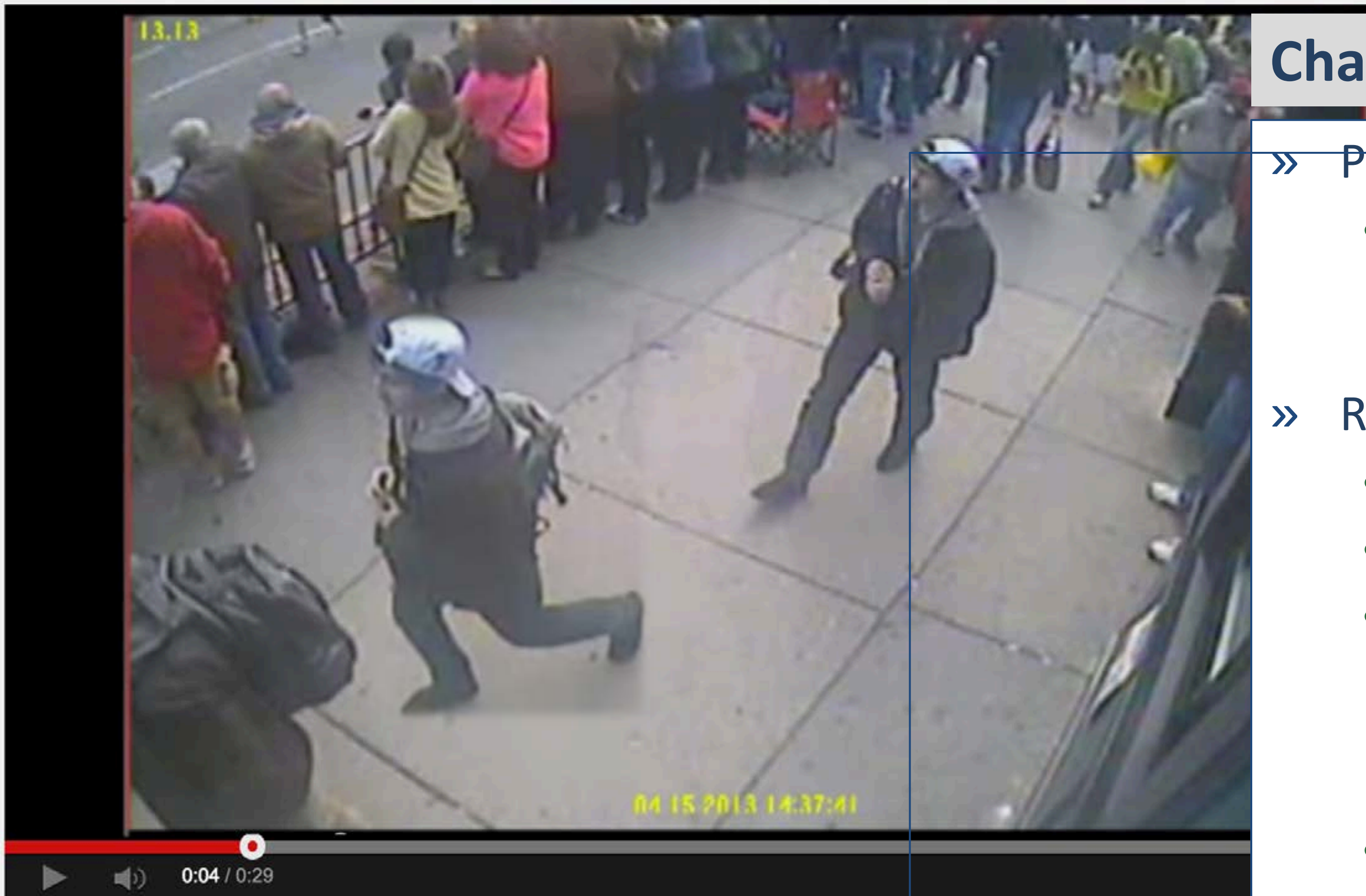
Goals

- » Comparative accuracy of algorithms
- » Absolute accuracy
- » Comparative computational cost
- » Iterative development with tech. providers
- » Threshold calibration
- » How to analyze + metrics → ISO/IEC 30137-2
- » Failure analysis → ISO/IEC 30137-1

Out-of-scope

- » Re-identification
- » Anomaly detection
- » Detection of un-coop, evasion
- » Other modalities + non-human

S2S – V2S – S2V – V2V :: Watchlist Surveillance



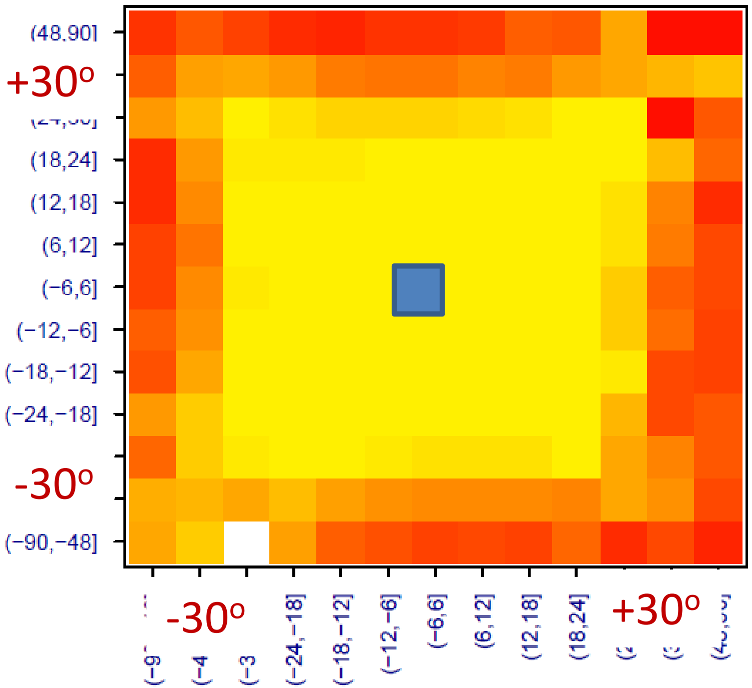
Surveillance Video Related to Boston Bombings

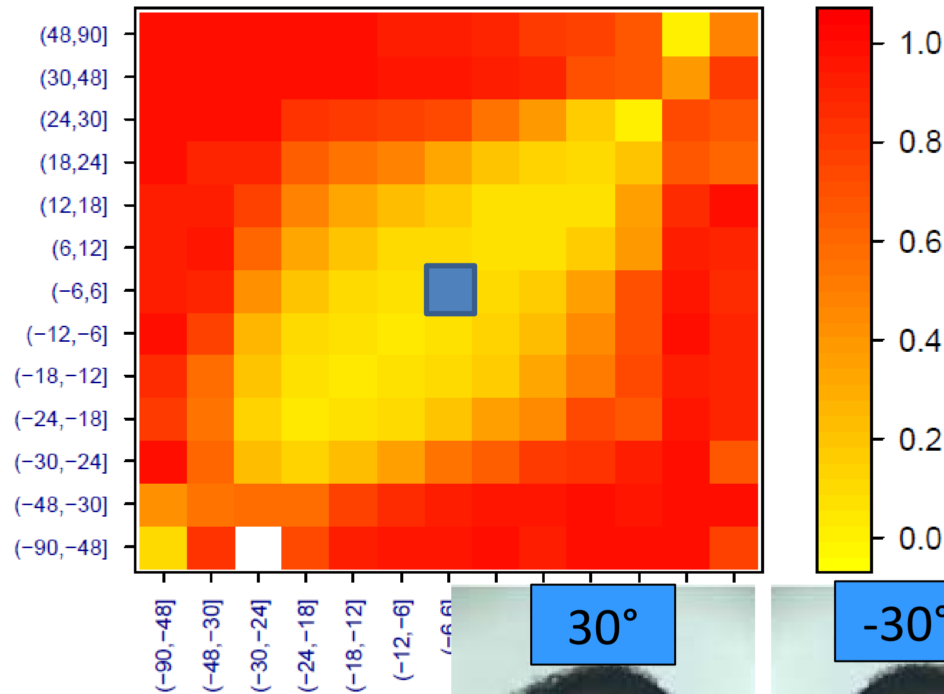
Challenges for FR

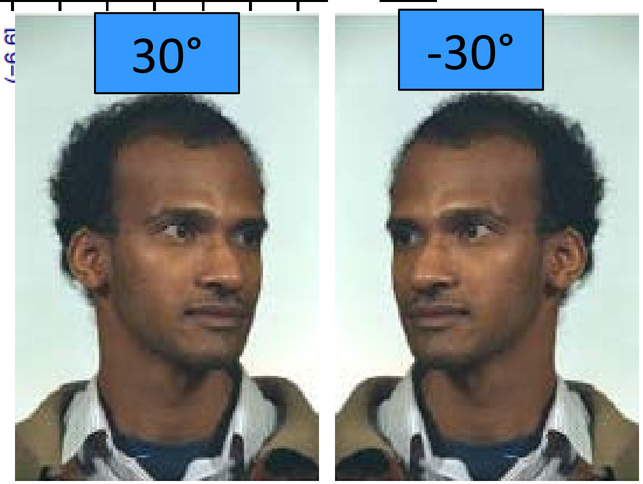
» Pose
  - Compound rotation of head to optical axis

» Resolution
  - Range to subject
  - Legacy camera
  - Adverse compression for storage or transmission
  - Motion blur

Off angle recognition: The problem for video



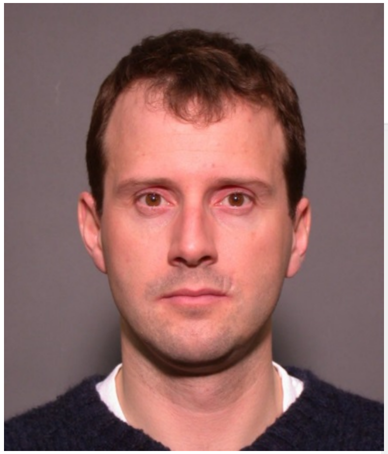




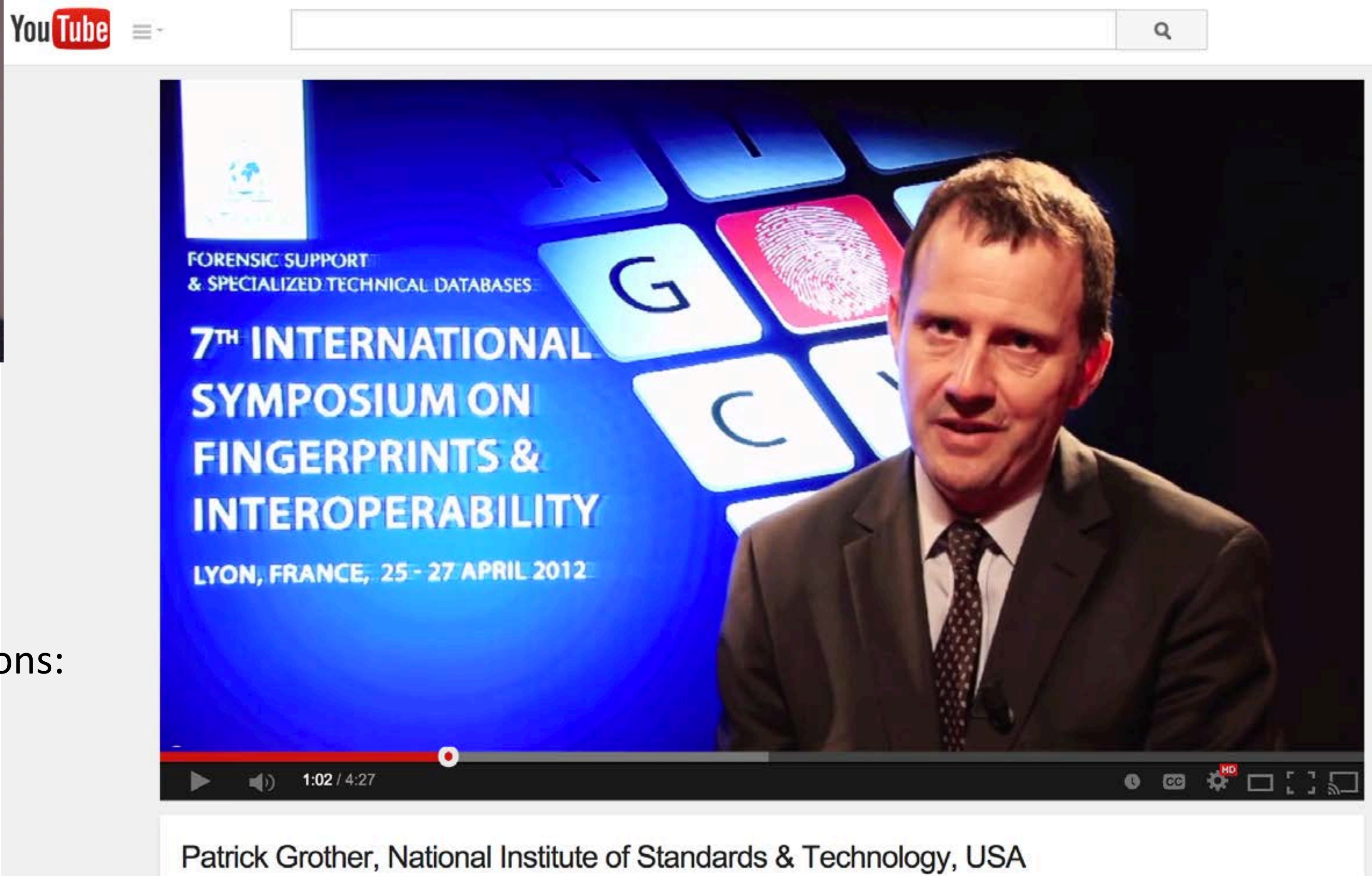
ISO standard tolerance for pristine imagery

S2S – V2S – S2V – V2V

Search = Mugshot



Enrolled = Video corpus, e.g. Youtube



Example applications:

1. Media search
2. Asylum re-identification

S2S – V2S – S2V – V2V

Search = Video corpus

Enrolled = Video corpus

Example applications:

1. Identity clustering
2. Re-identification

Thanks

patrick.grother@nist.gov

Time variation in three modalities

Iris

- » Healthy
  - Blink occlusion
  - Gaze direction
  - Dilation varies with mood, consumption, ambient light
- » Cosmetic
  - Contact lenses
  - Glasses
- » Ageing
  - Pupil constriction
  - Palpebral aperture
- » Disease

Fingerprint

- » Healthy
  - Facial expression
  - Mouth movement
  - Head motion, head orientation
  - Facial hair
- » Cosmetic
  - Moisturizers
- » Ageing
  - Arthritic fingers

Face

- » Healthy
  - Facial expression
  - Mouth movement
  - Head motion, head orientation
  - Facial hair
- » Cosmetic
  - Makeup
- » Ageing
  - Soft tissue folds
  - Stoop – pitch forward