MICHIGAN STATE UNIVERSITY

CVPR 2017

FACE RECOGNITION PERFORMANCE UNDER AGING

Debayan Deb, **Lacey Best-Rowden**, Anil K. Jain
Michigan State University, PRIP Lab
{debdebay, bestrow1, jain}@msu.edu

July 21, 2017

Motivation

- Unlike PIE facial variations, aging is intrinsic and cannot be controlled by image capture or subject cooperation
- Critical to determine state-of-the-art robustness to facial aging
  - Large time lapse can result in false reject errors
  - Expiration periods for ID credentials (5-10 years)
    - Should be informed by FR performance
- Longitudinal datasets available for research are lacking
  - FG-NET: 82 subjects total, relatively poor image quality
  - MORPH: only ~300 subjects with at least 5 images over at least 5 years


19 years


25 years


30 years


35 years

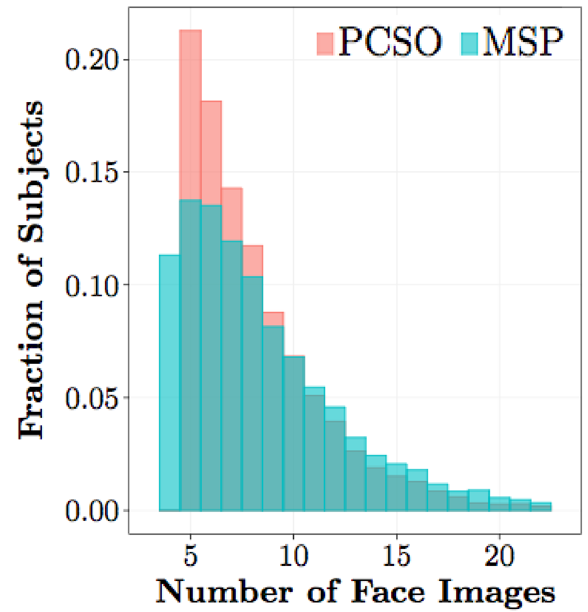

25 years

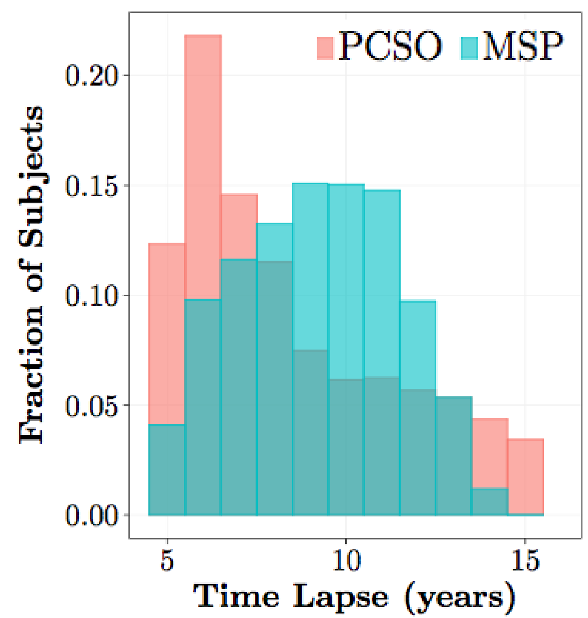

32 years

Longitudinal Face Datasets

- Subsets of larger mugshot databases obtained from law enforcement agencies
  - Pinellas County Sheriff's Office (PCSO)
  - Michigan State Police (MSP)

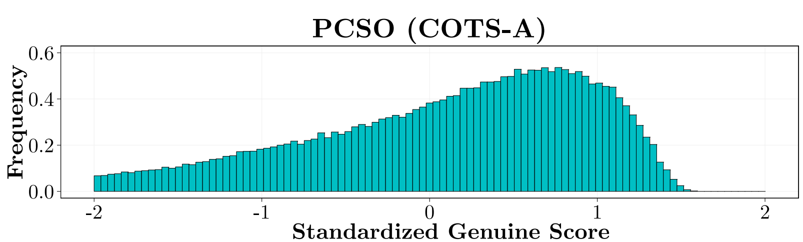
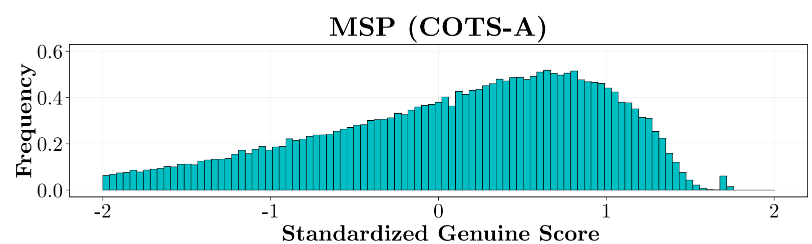
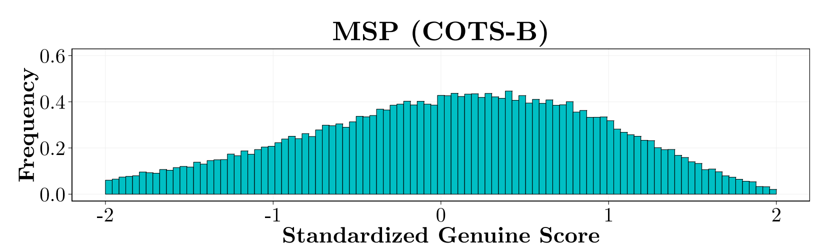
	PCSO	MSP
No. of images	147,784	82,450
No. of subjects	18,007	9,572
Avg. no. of images per subj.	8	9
Avg. time lapse (yrs)	8.5	9.0
Min - Max time lapse (yrs)	5 - 16	4 - 14
Age Range (yrs)	18 - 83	18 - 78
Male / Female (%)	83 / 17	88 / 12
Black / White (%)	61 / 39	52 / 48





Face Matchers

- Two state-of-the-art face matchers
  - COTS-A: one of the top-3 performers in NIST FRVT 2014
  - COTS-B: based on deep convolutional network
- *At what rate do genuine similarity scores change over time due to time lapse (and other covariates)?*

	PCSO	MSP
COTS-A		
COTS-B	*COTS-B is not evaluated on PCSO because, according to the vendor, COTS-B is trained on PCSO.*	

Approach: Mixed-effects Statistical Models

- Level-1 Model (within-subject variability)

$$y_{ij} = \varphi_{0i} + \varphi_{1i} \boxed{x_{ij}} + \varepsilon_{ij} \qquad \varepsilon_{ij} \sim N\left(0, \sigma_{\varepsilon}^{2}\right)$$

Genuine similarity score between the *enrollment and jth* images of subject *i*

Covariate: elapsed time, age, face quality, etc.

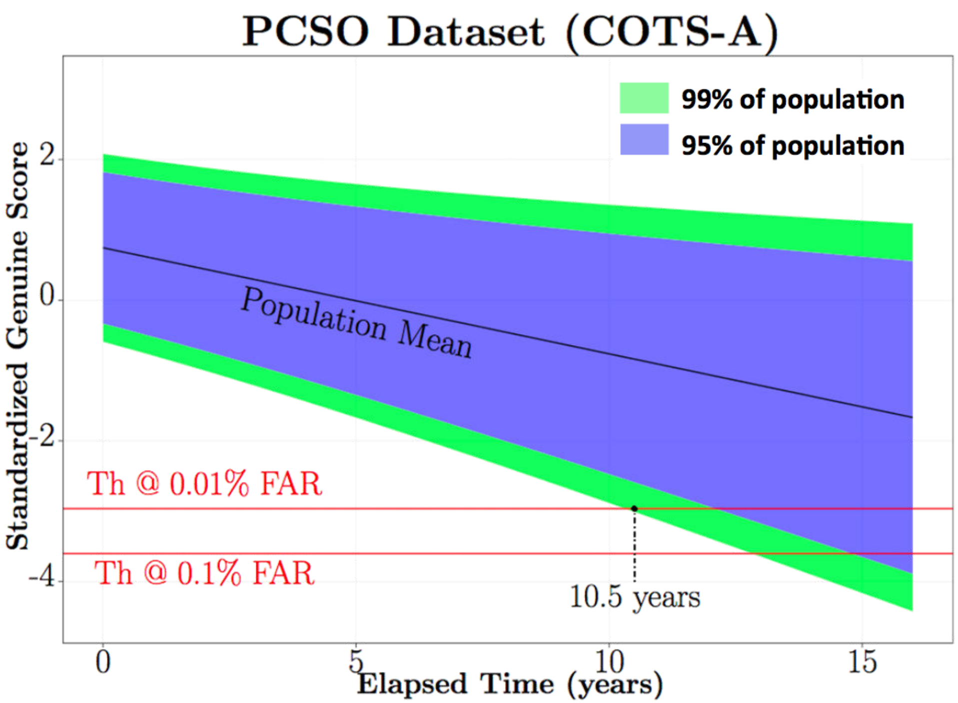
- Level-2 Model (between-subject variability)

$$\begin{aligned} \varphi_{0i} &= \boxed{\beta_{00}} + \boxed{b_{0i}} \\ \varphi_{1i} &= \boxed{\beta_{10}} + \boxed{b_{1i}} \end{aligned} \qquad \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

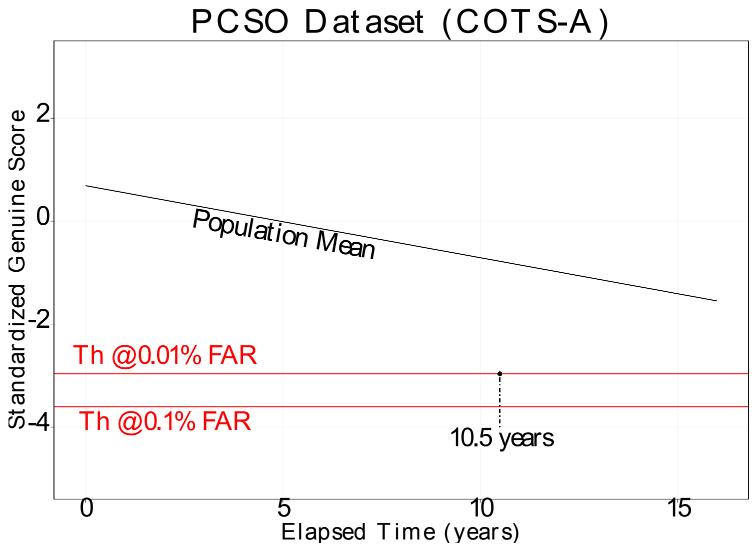
Fixed effects: Population-mean trend

Random effects: Deviations from population-mean trend

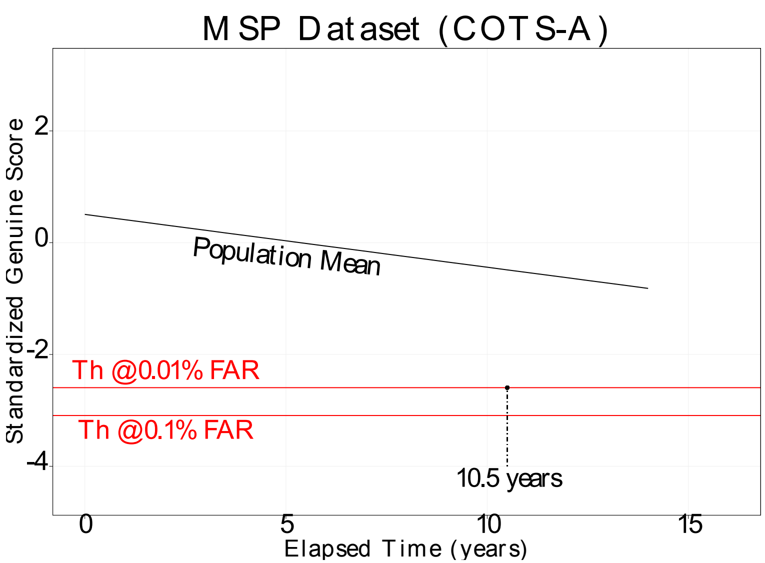
Results: COTS-A on PCSO

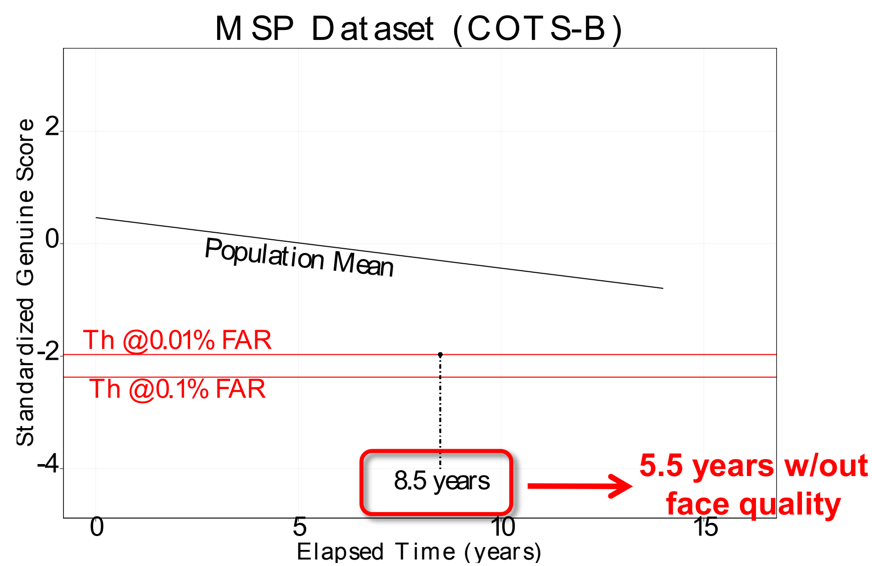


Results: Time Lapse & Face Quality



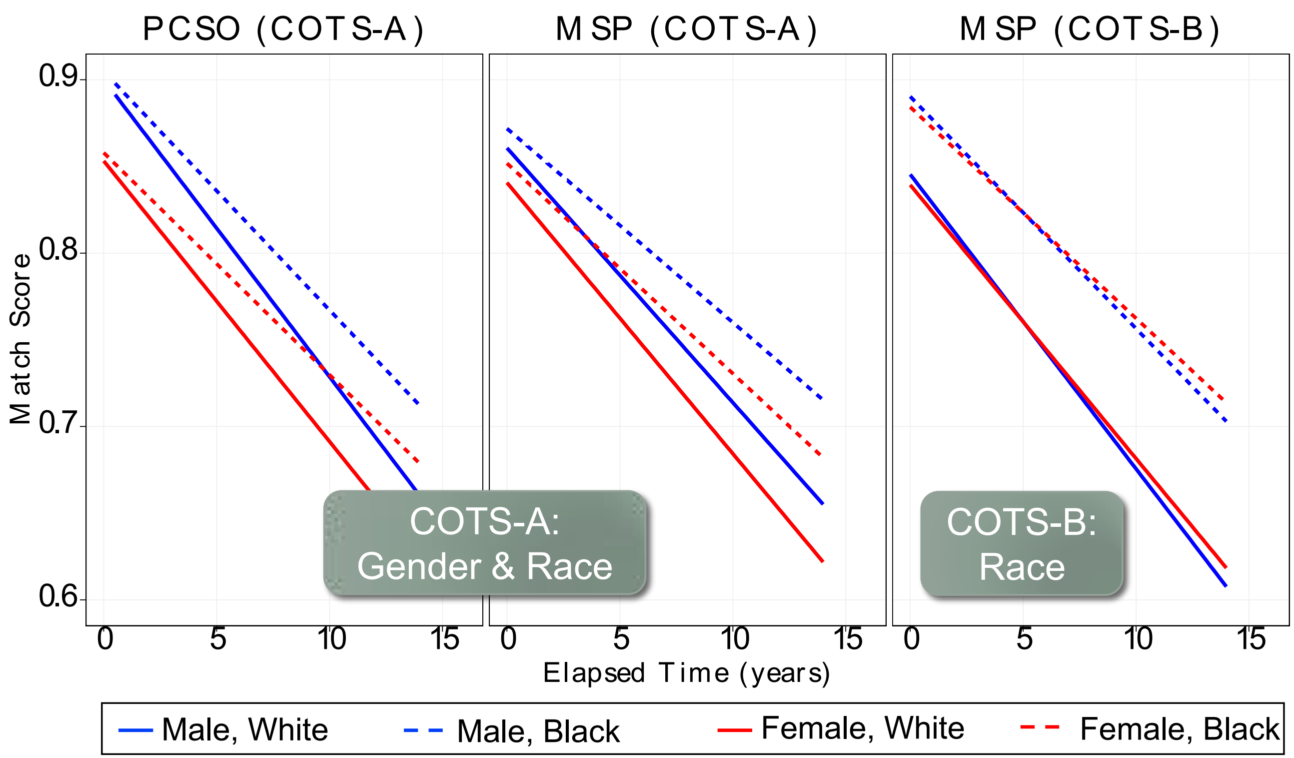
- Face quality measure from COTS-B SDK
- Figures shown are plotted for average quality mugshot images





Results: Gender and Race

Is subject-specific variability explained by demographics?

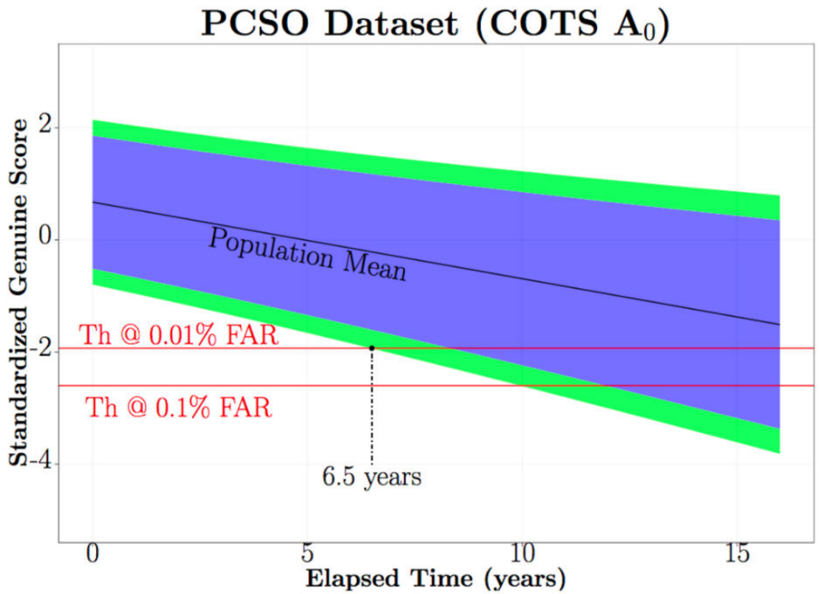


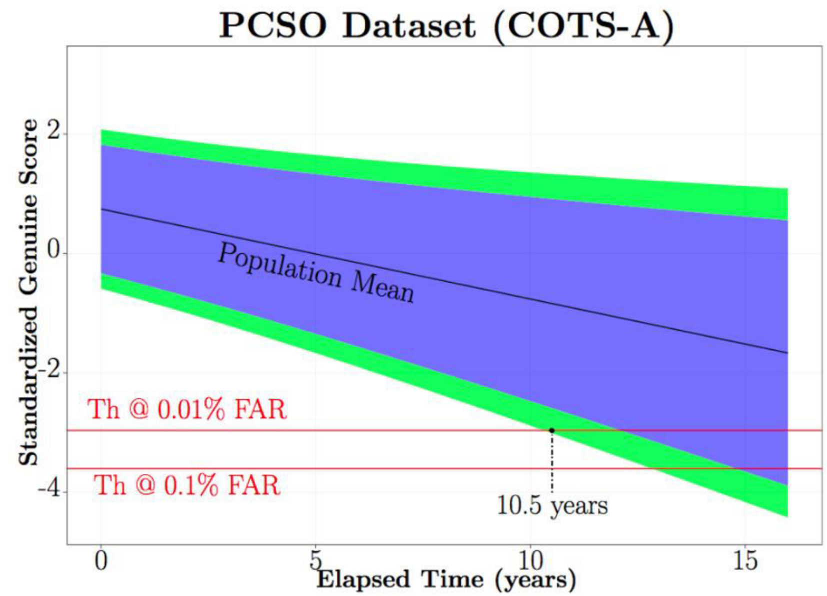
- Consistent for COTS-A on both PCSO and MSP
- Matcher-dependent for MSP database
  - Differences between COTS-A and COTS-B likely due to training set

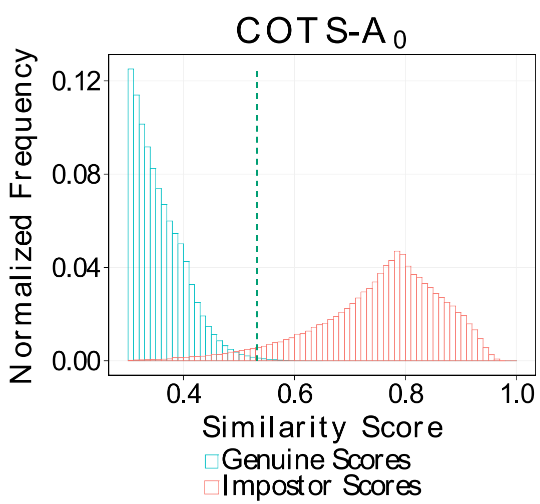
Conclusions

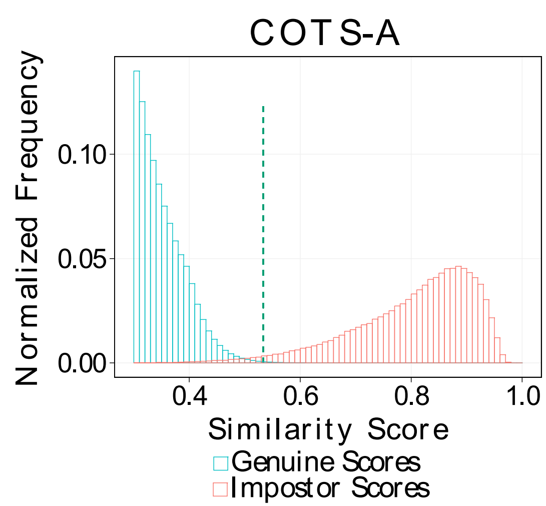
- Elapsed times tolerated by COTS face matchers
  - At 0.01% FAR, COTS-A can verify 99% of population up to 10.5 years; accuracy drops to 95% of the population after an additional 2-3 years
- COTS-B is overall weaker matcher than COTS-A, but *longitudinal performance* is comparable
  - After accounting for face quality
- Demographic effects on genuine scores
  - Database-independent for COTS-A
  - Matcher-dependent on MSP datasets
- Methodology should be periodically conducted to reassess state-of-the-art robustness to facial aging

Results: COTS-A vs. COTS-A_0









Results: Time Lapse & Face Quality

	Max time lapse (years) at which 99% of population can be verified	
	0.01% FAR	
	Without FQ	At average mugshot FQ
PCSO (COTS-A)	10.5	10.5
MSP (COTS-A)	9.5	10.5
MSP (COTS-B)	5.5	8.5