Face Recognition Performance
Role of Demographic Information

B. Klare[1] J .Klontz[1] M. Burge[1] R. Vorder Bruegge[2]
Anil K. Jain[3]

[1]The MITRE Corporation
A Federally Funded Research & Development Center

[2]Science and Technology Branch
Federal Bureau of Investigation

[3]Michigan State University

IBPC 2012

This is revision 2012-03-06:72b4bd4 built at 11:05 on Tuesday 6th March, 2012.

Explore impact of demographics on face recognition performance

- Commercial matchers
  - COTS-A (Asia)
  - COTS-B (North America)
  - COTS-C (Europe)
- Non-trainable matchers
  - Local Binary Patterns (LBP) *Tan and Triggs 2010*
  - Gabor Wavelets
- Trainable matcher
  - Spectrally Sampled Structural Subspace Features (4SF) *Klare 2011*
- Eight cohorts across three demographics isolated:
  - Race (White, Black, Hispanic)
  - Gender (male and female)
  - Age (18 to 30 years old, 30 to 50 y.o., 50 to 70 y.o.)

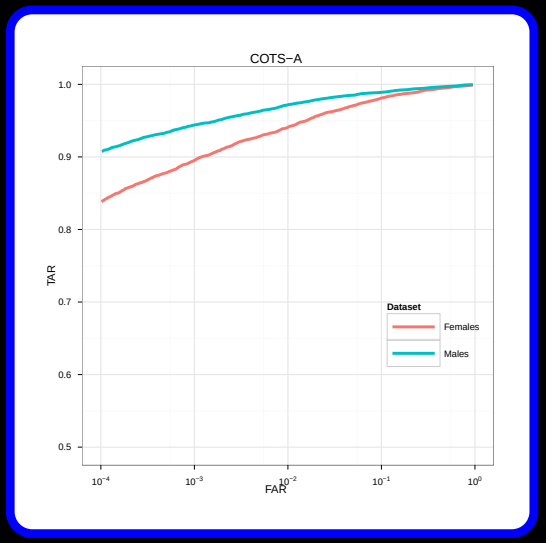
- Confirm previous biases reported in COTS algorithms
  NIST: FRVT 2002, MBE 2010
- Understand if performance on different cohorts is a function of
  - Unbalanced training
  - Inherent difficulty of the demographic cohort
- Determine if performance on a cohort can be improved by training exclusively on that cohort

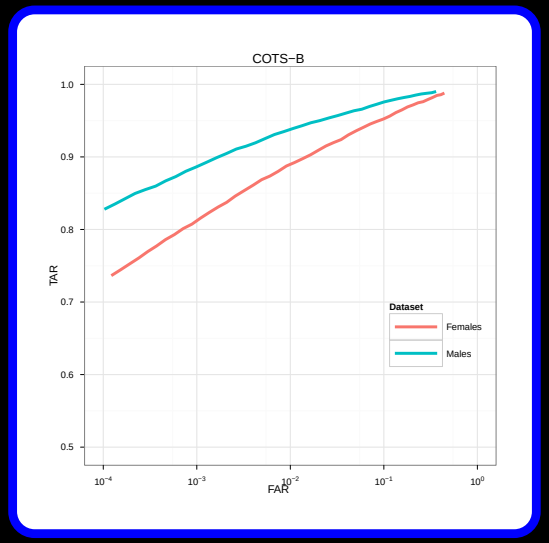
- Experiment 1:
  - Measure COTS FRS performance within each demographic cohort
  - Goal: Confirm previous biases (MBE 2010) on a different dataset
- Experiment 2:
  - Measure performance of non-trainable face recognition algorithms
  - Goal: Understand inherent (or a priori) difficulty of each cohort
- Experiment 3:
  - Train several versions of the 4SF algorithm (one for each cohort)
  - Apply each version of 4SF to test sets from each cohort
  - Goal: Investigate influence of a training set on recognition performance
  - Goal: Leverage demographics to improve FR performance through demographic-based training

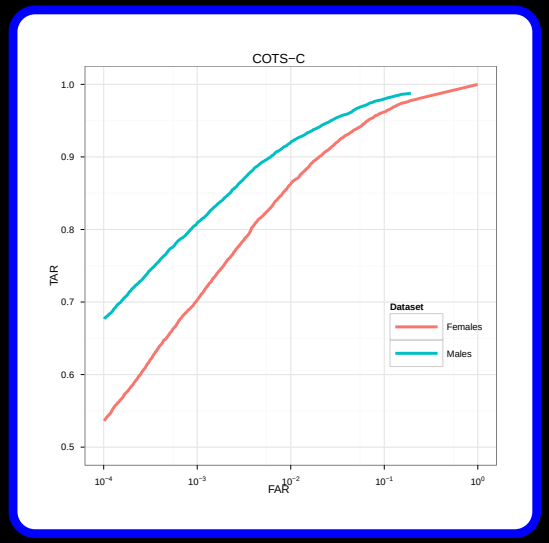
Demographic	Cohort	# Training	# Testing
Gender	Female	7995	7996
	Male	7996	7998
Race	African	7993	7992
	Caucasian	7997	8000
	Hispanic	1384	1425
Age	18 to 30	7998	7999
	30 to 50	7995	7997
	50 to 70	2801	2853

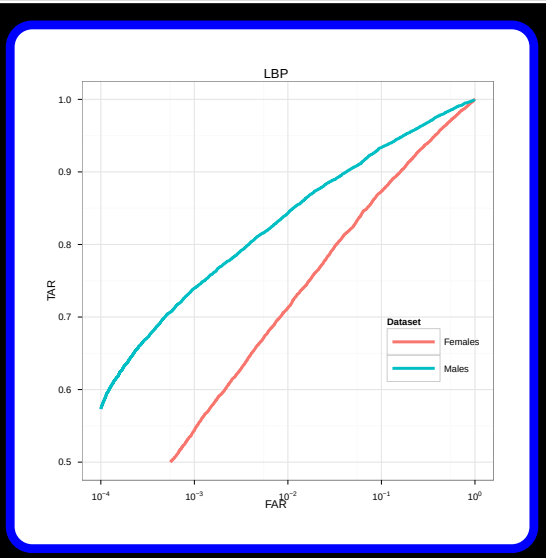
Table: Distribution of subjects by demographic category. Two images per subject. Disjoint Training and Testing sets. Order of 200,000 face images.

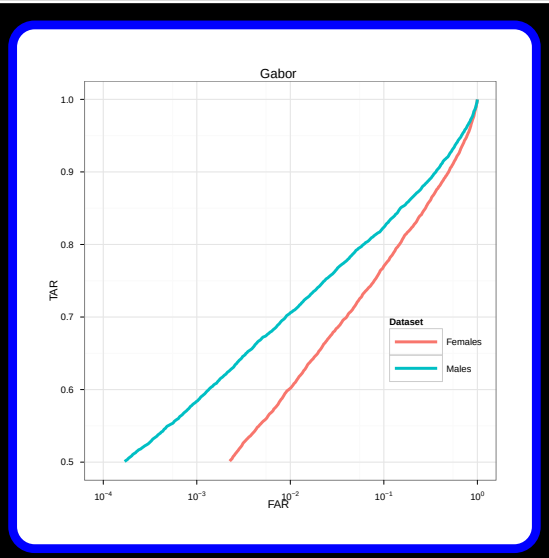
Demographic	Cohort	# Training	# Testing
Gender	Female	7995	7996
Gender	Male	7996	7998

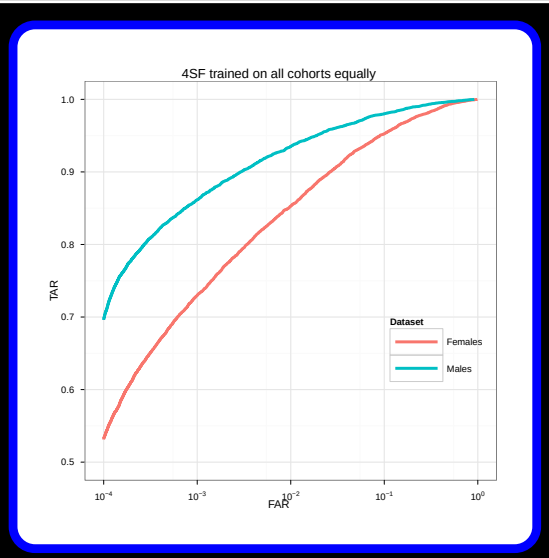

COTS−A
TAR
FAR
Dataset
Females
Males

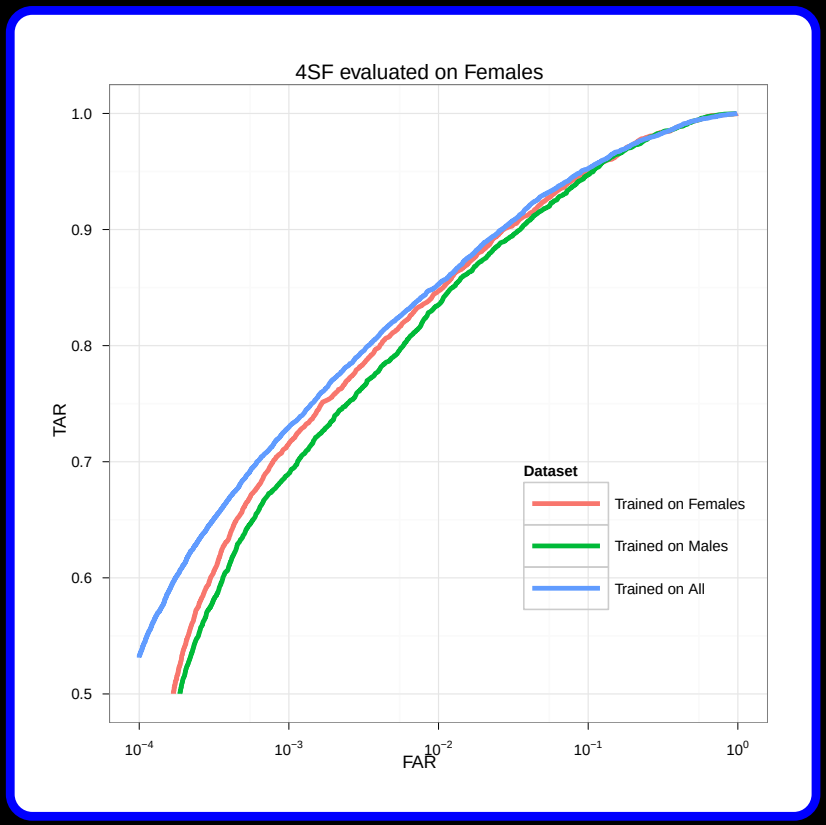

COTS−B
TAR
FAR
Dataset
Females
Males

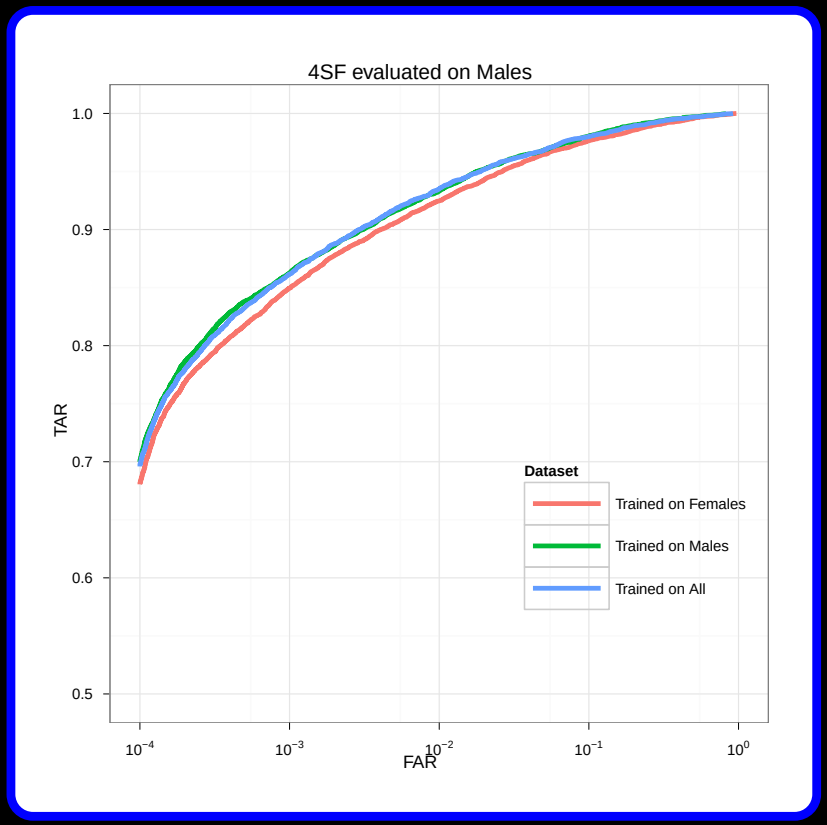

COTS−C
TAR
FAR
Dataset
Females
Males


LBP
TAR
FAR
Dataset
Females
Males


Gabor
TAR
FAR
Dataset
Females
Males

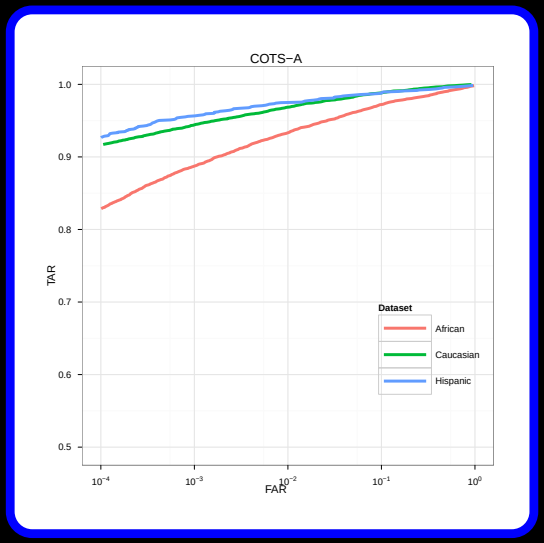

4SF trained on all cohorts equally
TAR
FAR
Dataset
Females
Males

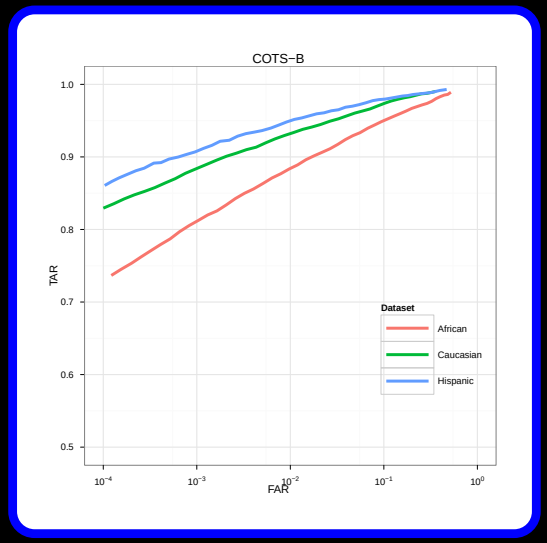

4SF evaluated on Females
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on Females
Trained on Males
Trained on All

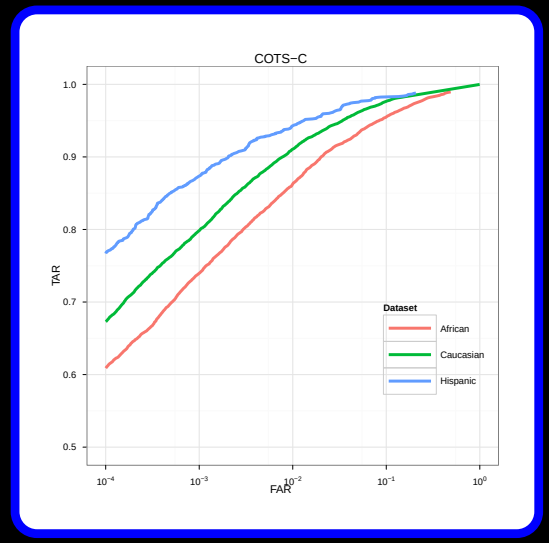

4SF evaluated on Males
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on Females
Trained on Males
Trained on All

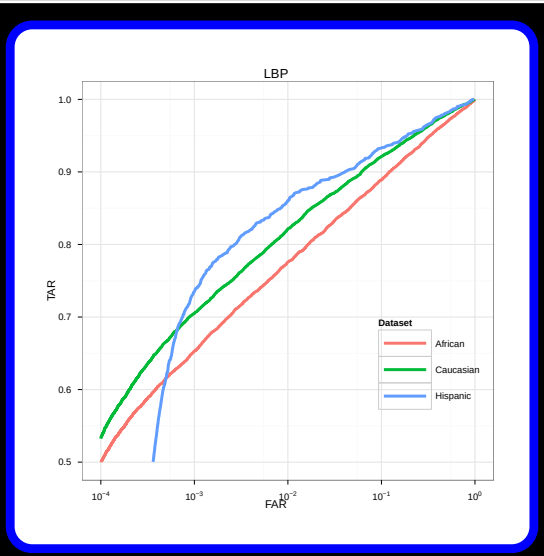
- All commercial matchers had lower performance on the female cohort.
- All non-trainable matchers had lower performance on the female cohort.
- Training on gender cohorts did not improve performance.
- Females appear to be intrinsically more difficult to match.

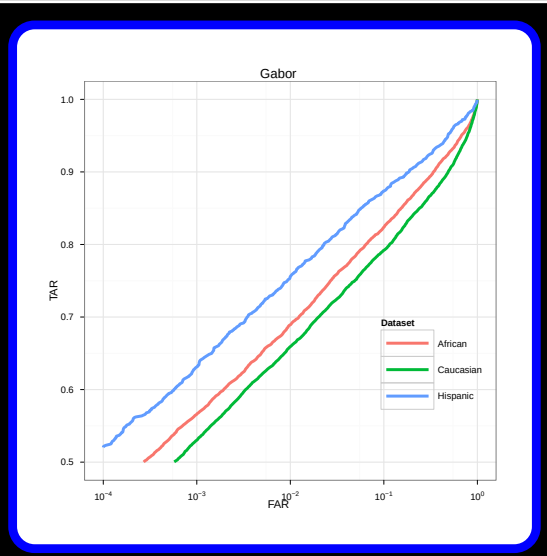
Demographic	Cohort	# Training	# Testing
Race	African	7993	7992
Race	Caucasian	7997	8000
Race	Hispanic	1384	1425

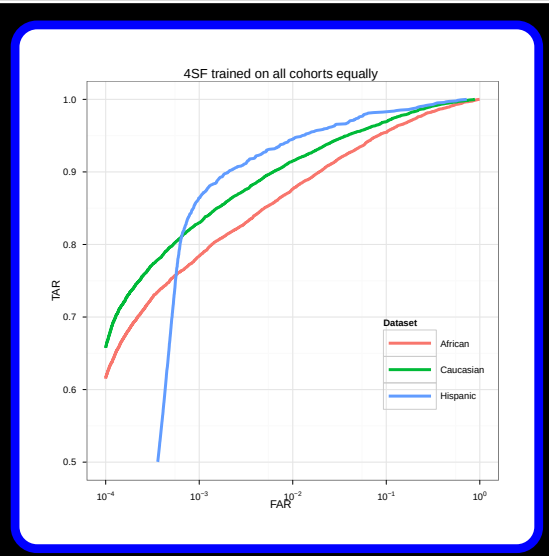

COTS−A
TAR
FAR
Dataset
African
Caucasian
Hispanic

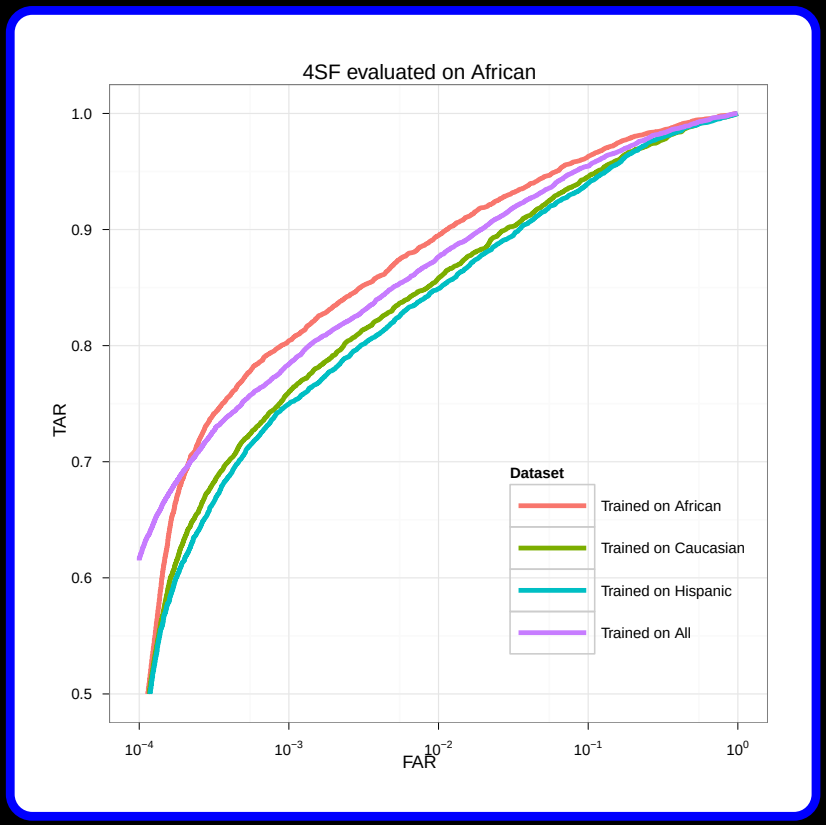

COTS−B
TAR
FAR
Dataset
African
Caucasian
Hispanic

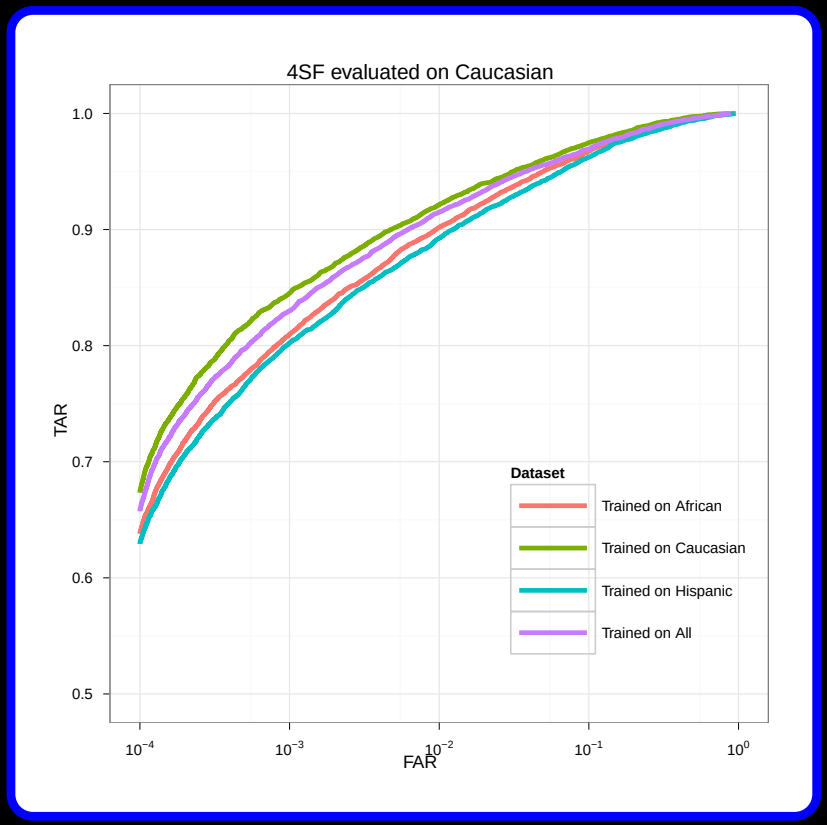

COTS−C
TAR
FAR
Dataset
African
Caucasian
Hispanic


LBP
TAR
FAR
Dataset
African
Caucasian
Hispanic


Gabor
TAR
FAR
Dataset
African
Caucasian
Hispanic

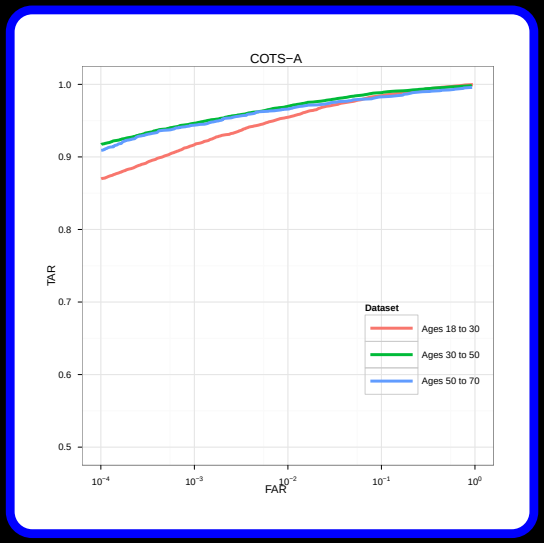

4SF trained on all cohorts equally
TAR
FAR
Dataset
African
Caucasian
Hispanic

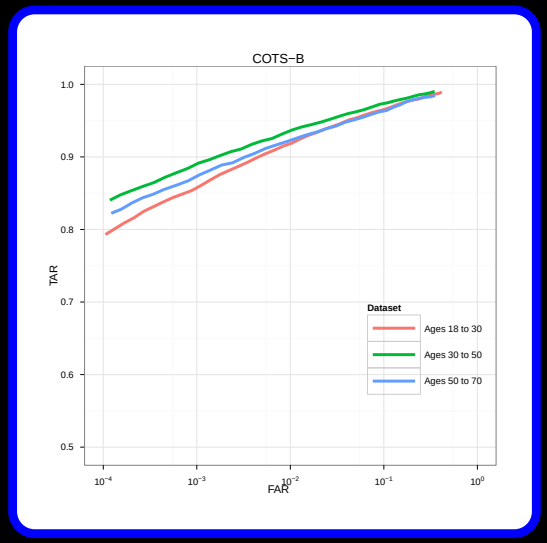

4SF evaluated on African
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on African
Trained on Caucasian
Trained on Hispanic
Trained on All

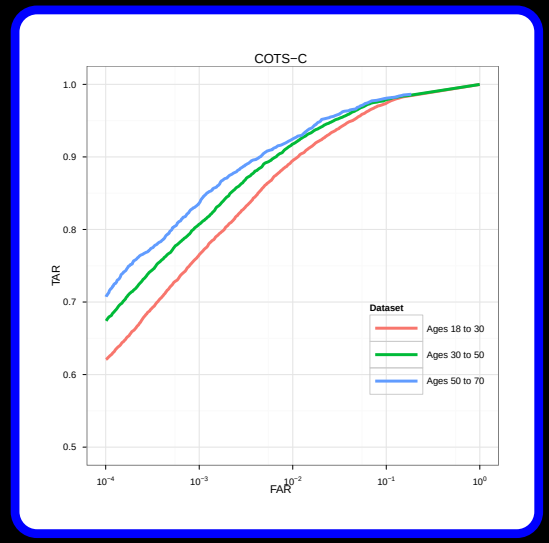

4SF evaluated on Caucasian
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on African
Trained on Caucasian
Trained on Hispanic
Trained on All

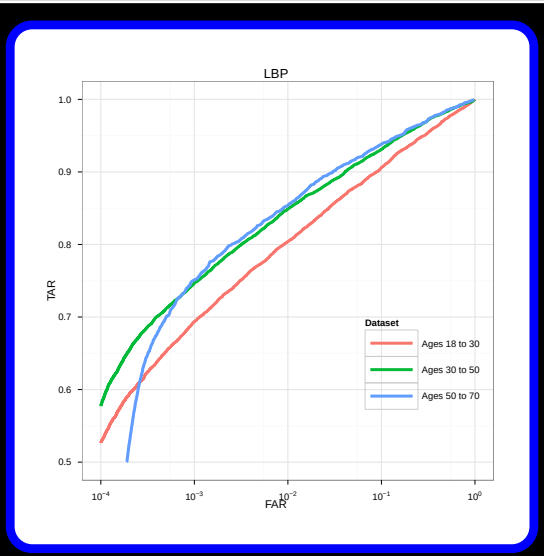
- All commercial matchers had lower performance on Black cohort.
- All non-trainable matchers had lower performance on the Black cohort.
- Training on a race cohort improved performance on that cohort at a slight detriment to other race cohorts.
- Races performance appears to be benefit from different encodings.

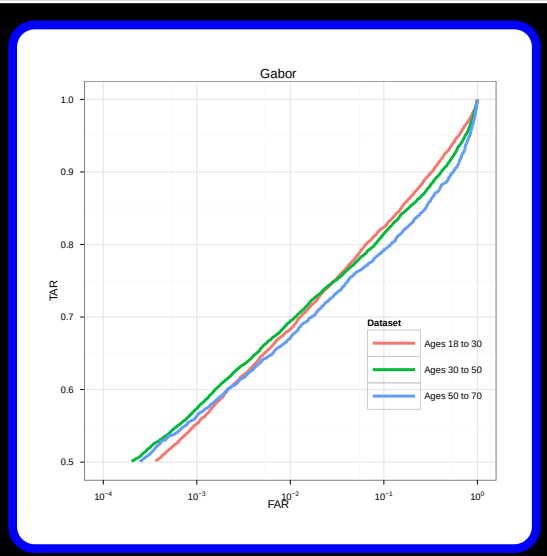
Demographic	Cohort	# Training	# Testing
Age	18 to 30	7998	7999
Age	30 to 50	7995	7997
Age	50 to 70	2801	2853

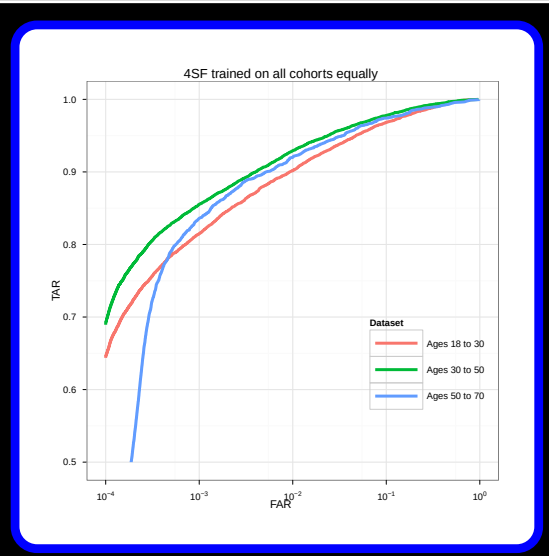

COTS−A
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70

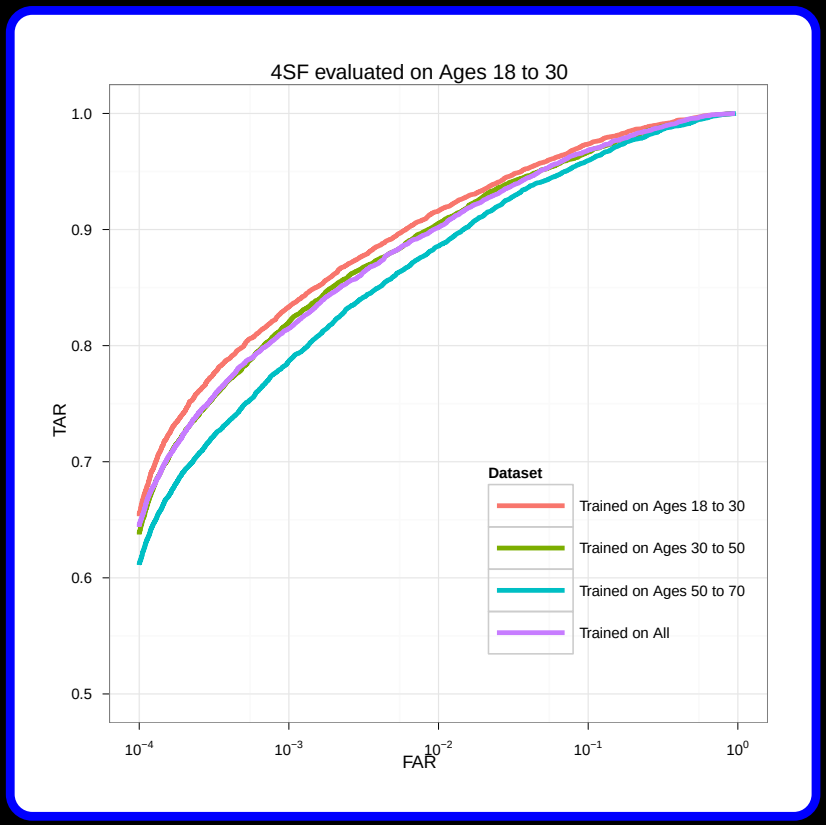

COTS−B
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70

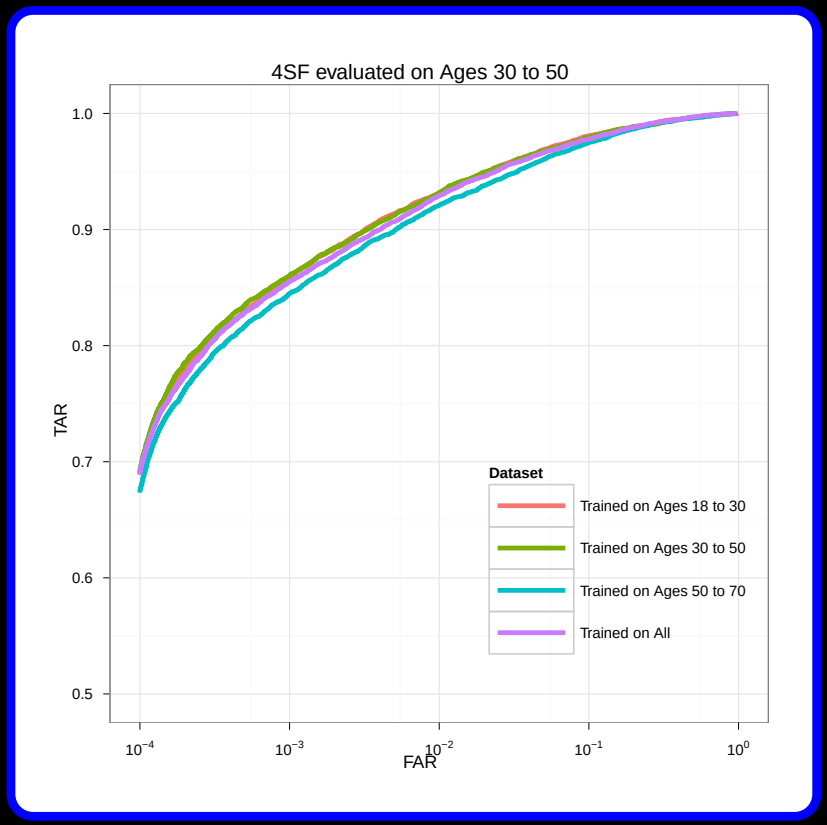

COTS−C
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70


LBP
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70


Gabor
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70


4SF trained on all cohorts equally
TAR
FAR
Dataset
Ages 18 to 30
Ages 30 to 50
Ages 50 to 70


4SF evaluated on Ages 18 to 30
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on Ages 18 to 30
Trained on Ages 30 to 50
Trained on Ages 50 to 70
Trained on All


4SF evaluated on Ages 30 to 50
TAR
FAR
1.0
0.9
0.8
0.7
0.6
0.5
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Dataset
Trained on Ages 18 to 30
Trained on Ages 30 to 50
Trained on Ages 50 to 70
Trained on All

- All matchers had lower performance on the 18-30 cohort.
- Training on an age cohort improved performance on that cohort at a slight detriment to other cohorts.

- The Female, Black, and 18-30 cohorts were difficult for all matchers.
- Performance on race and age cohorts improves when training exclusively on a similar cohort.
- Performance on gender cohorts does not improve when training exclusively on a similar cohort.
- Training FR algorithms on datatsets well distributed across all demographics is critical.

Face Recognition Performance
Role of Demographic Information

B. Klare[1] J .Klontz[1] M. Burge[1] R. Vorder Bruegge[2]
Anil K. Jain[3]

[1]The MITRE Corporation
A Federally Funded Research & Development Center

[2]Science and Technology Branch
Federal Bureau of Investigation

[3]Michigan State University

IBPC 2012

This is revision 2012-03-06:72b4bd4 built at 11:05 on Tuesday 6th March, 2012.

References I

- B. Klare. "Spectrally Sampled Structural Subspace Features (4SF)". In: *Michigan State University Technical Report, MSU-CSE-11-16.* 2011.
- Xiaoyang Tan and B. Triggs. "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions". In: *IEEE Trans. on Image Processing* 19.6 (2010), pp. 1635 –1650.