Controllable Face Privacy

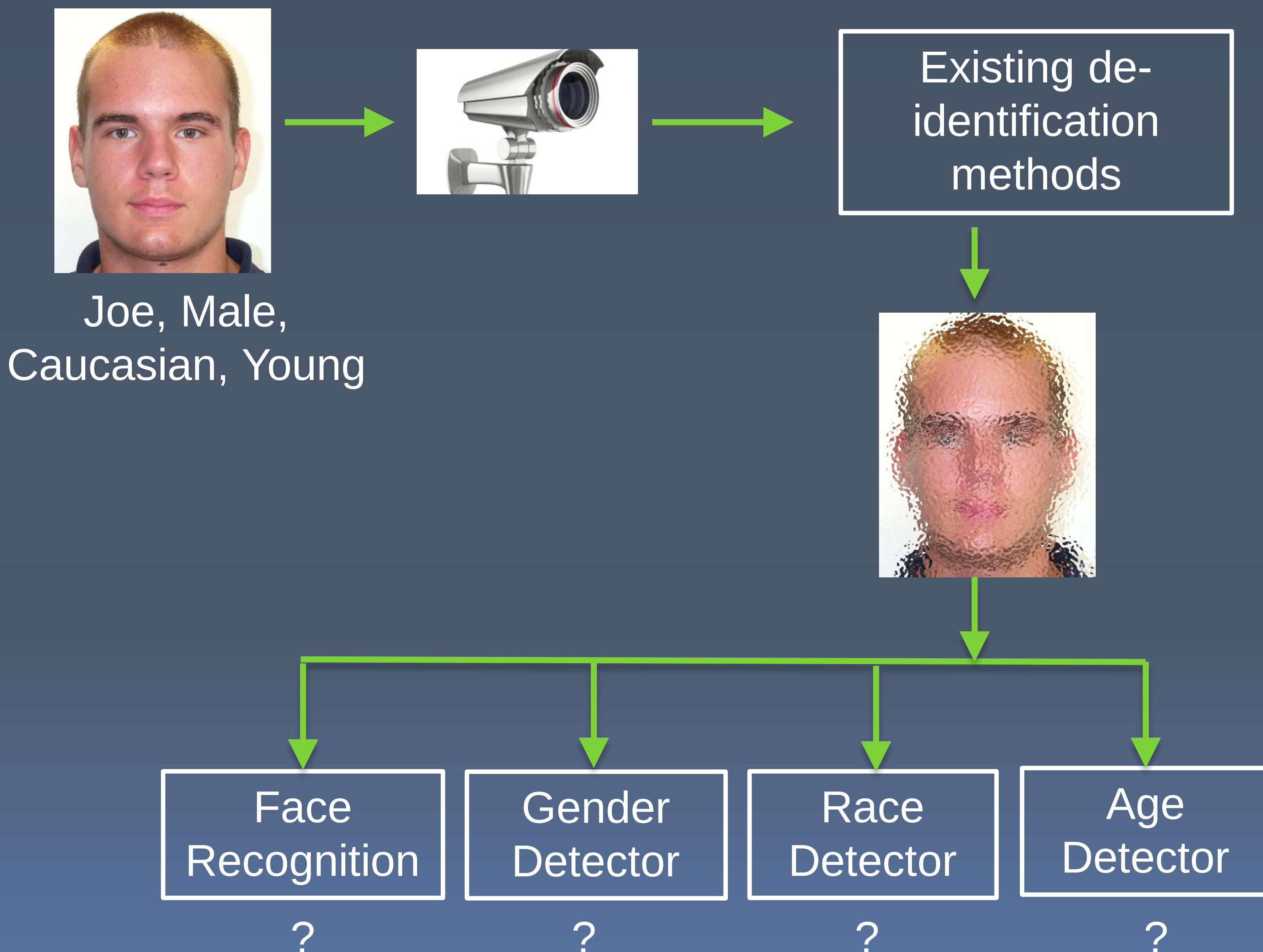
Terence Sim, Li Zhang

NUS National University of Singapore | School of Computing

Leading The World With Asia's Best

Joe, Male, Caucasian, Young

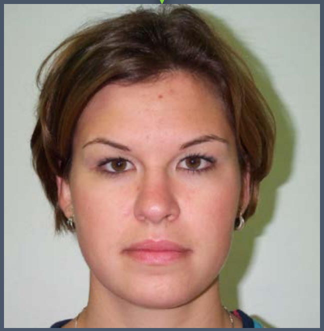
Existing de-identification methods

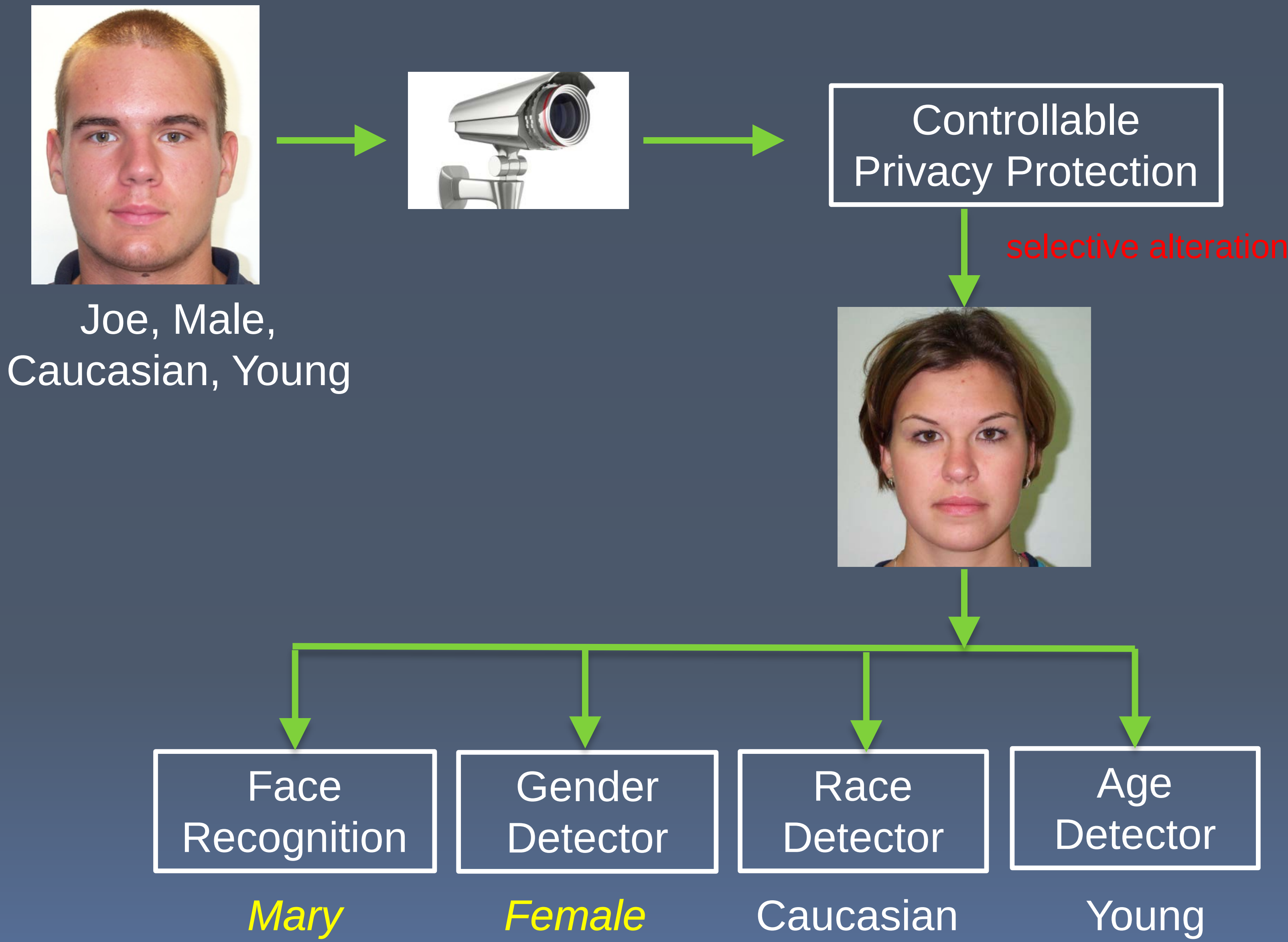

Joe, Male,
Caucasian, Young
Existing de-identification methods
Face Recognition
?
Gender Detector
?
Race Detector
?
Age Detector
?

Joe, Male,
Caucasian, Young

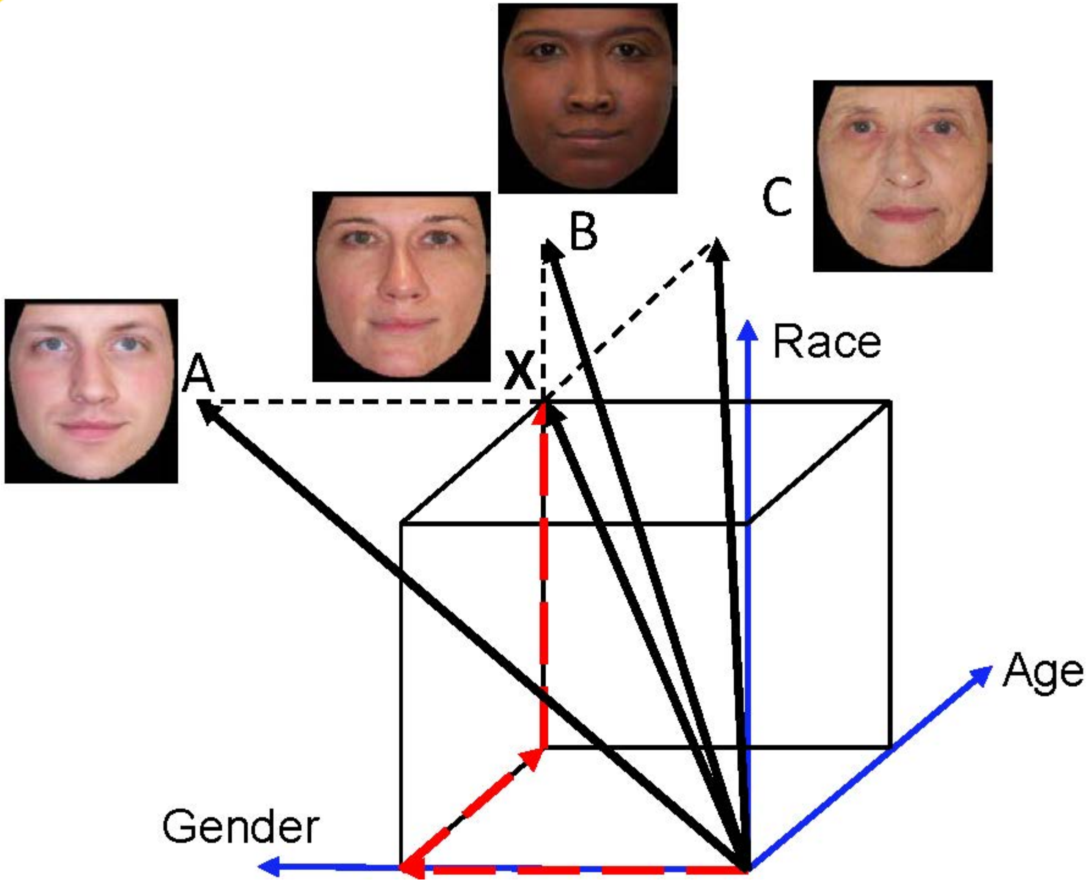
Controllable
Privacy Protection

selective alteration




Joe, Male,
Caucasian, Young
Controllable
Privacy Protection
selective alteration
Face
Recognition
Gender
Detector
Race
Detector
Age
Detector
Mary
Female
Caucasian
Young

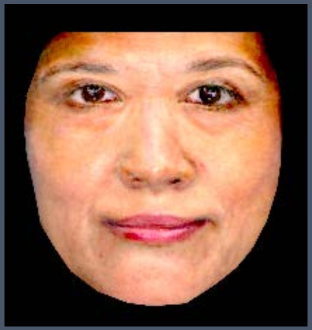
Key Idea



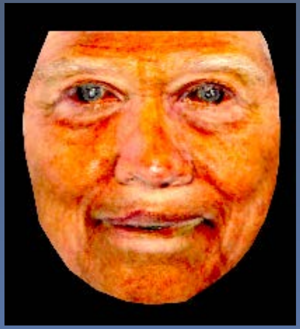
Orthogonal subspace decomposition

Semantic Faces

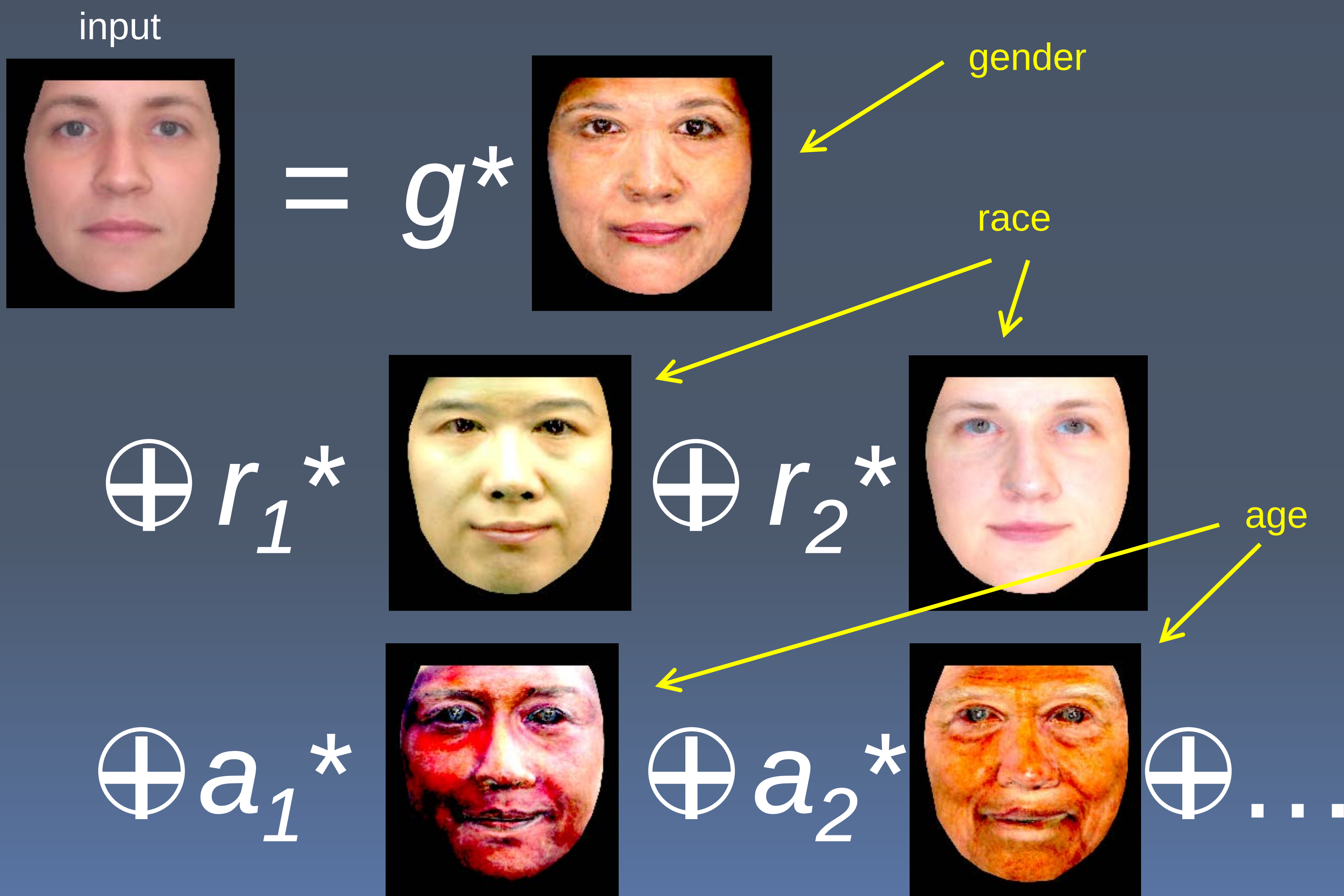
input

$= g*$ 

$\oplus\, r_1*$  $\oplus\, r_2*$ 

$\oplus\, a_1*$  $\oplus\, a_2*$  $\oplus\, \ldots$

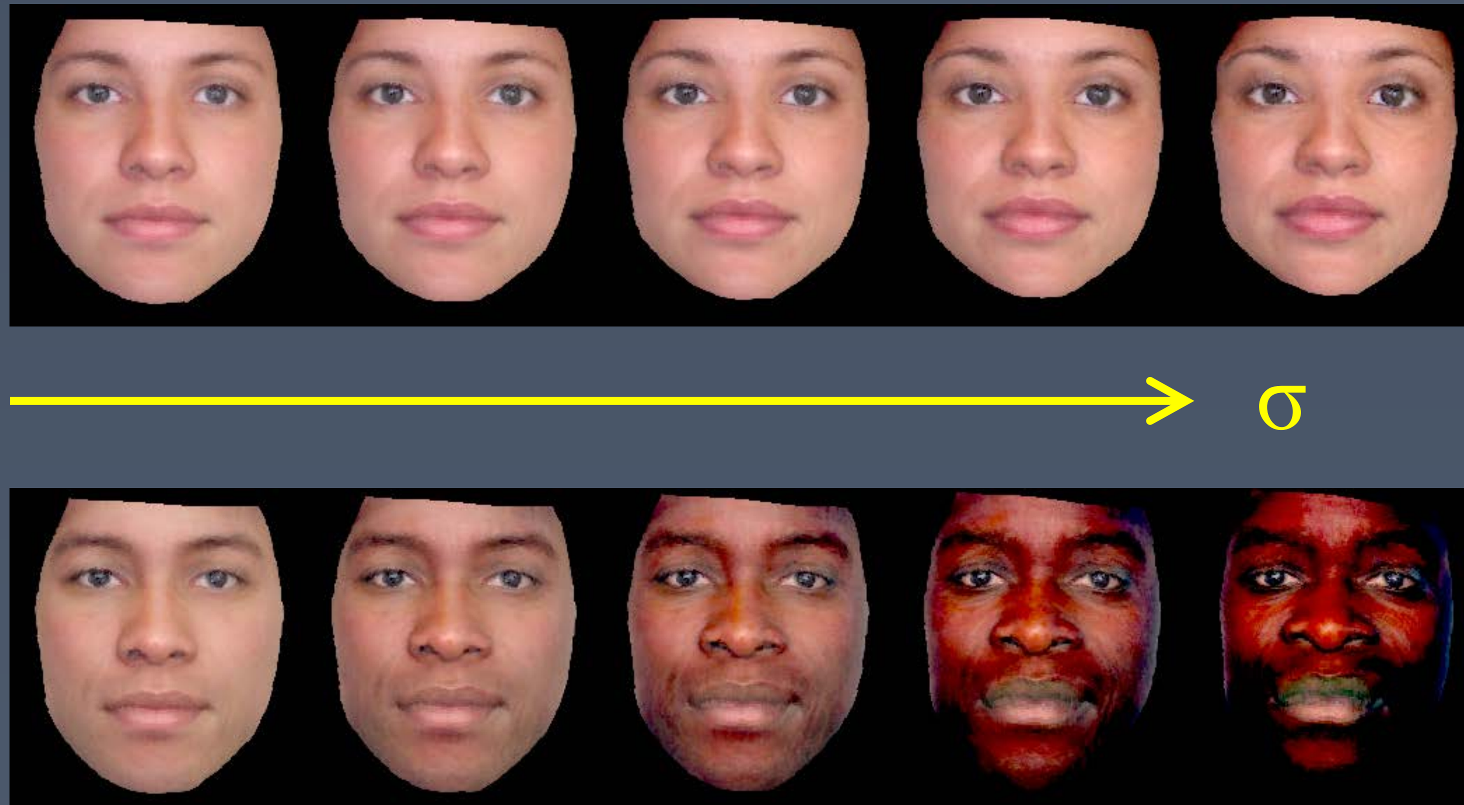
Semantic Faces



Increasing
Femininity

Increasing Femininity

Intensifying “African”-ness



Increasing
Femininity

σ

Intensifying
“African”-ness

Growing
older

How to learn Semantic Faces?

[Sim et al., ICCV 2009]

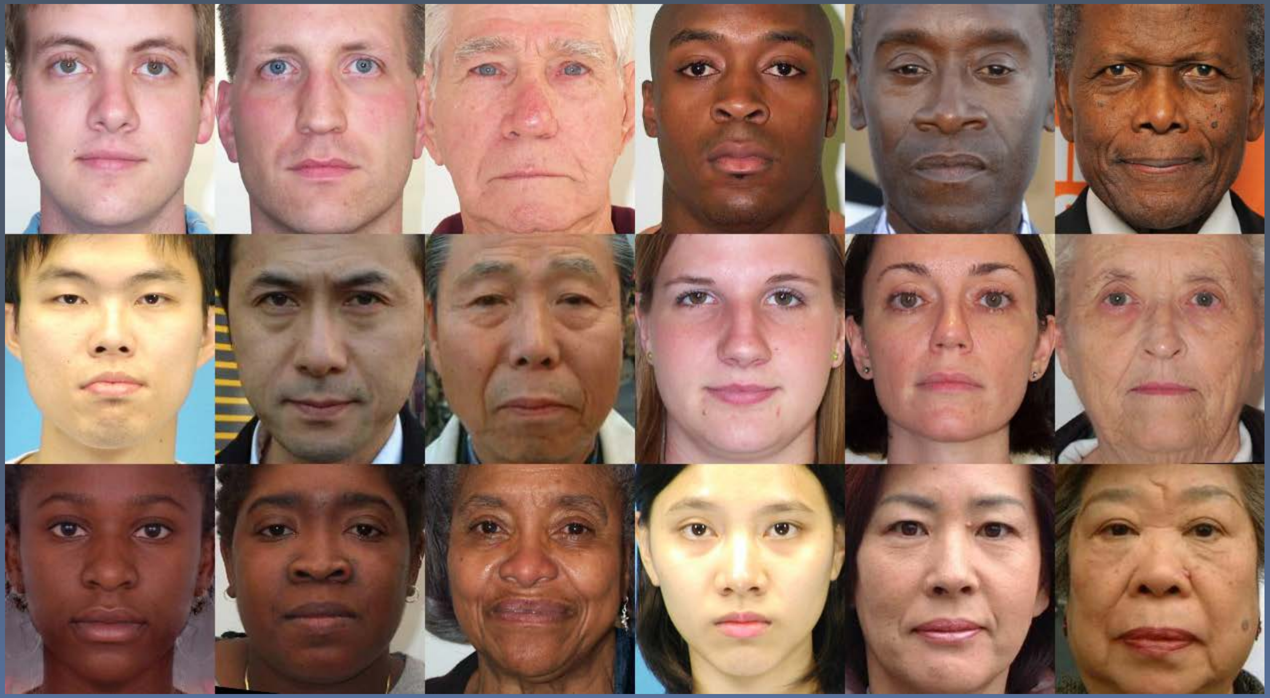
Multimodal Discriminant Analysis

Key Idea

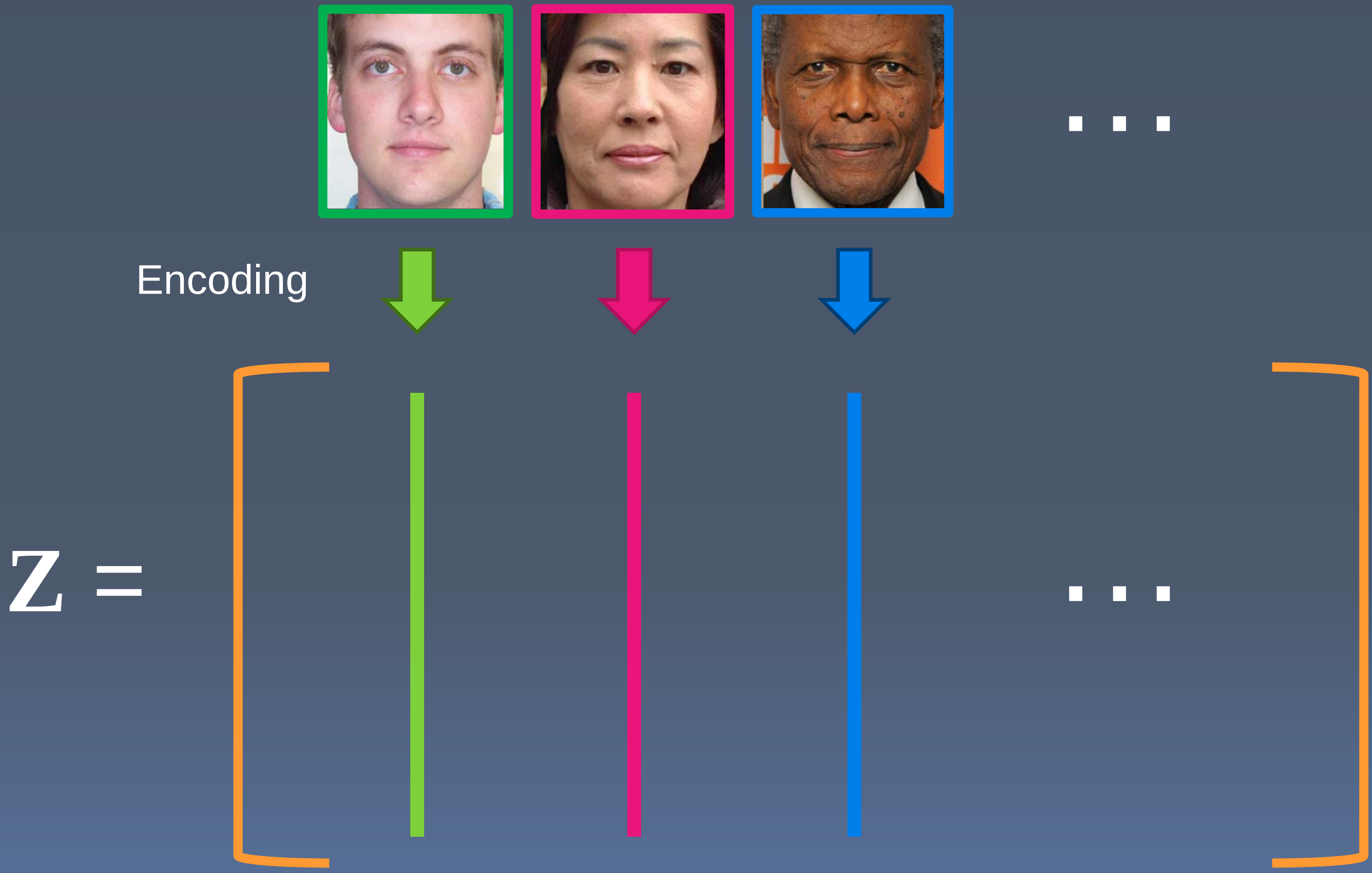
repeated Fisher Linear Discriminant analysis

N=18 Labeled Training images

Gender:	Male, Female
Race:	Caucasian, African, Oriental
Age:	Young, Middle-aged, Old



Data matrix



1. Center to mean

$$\mathbf{X} = \mathbf{Z} - \mathbf{M}$$

2. Whiten

$$\tilde{\mathbf{X}} = \mathbf{P}^\top \mathbf{X}$$

where

$$\mathbf{X}\mathbf{X}^\top = \mathbf{U}\mathbf{D}\mathbf{U}^\top$$

$$\mathbf{P} = \mathbf{U}\mathbf{D}^{-1/2}$$

3. For each mode *i* = {gender, race, age}, compute between-class and within-class scatter matrices on : $\tilde{\mathbf{X}}$

$$\mathbf{S}_b^i, \ \mathbf{S}_w^i$$

4. Maximize Fisher Criterion

$$\mathbf{J}_F(\mathbf{V}^i) = \frac{(\mathbf{V}^i)^\top \mathbf{S}_b^i (\mathbf{V}^i)}{(\mathbf{V}^i)^\top \mathbf{S}_w^i (\mathbf{V}^i)}$$

5. Form matrix

$$\mathbf{V} = [\,\mathbf{V}^{gender}\,\mathbf{V}^{race}\,\mathbf{V}^{age}\,\mathbf{V}^{0}\,]$$

5. Form matrix

$$\mathbf{V} = [\,\mathbf{V}^{gender}\ \mathbf{V}^{race}\ \mathbf{V}^{age}\ \mathbf{V}^{0}\,]$$

Residual Space

- $\mathbf{V}$ is an (N-1) X (N-1) orthogonal matrix, whose columns are the bases for the gender (1-dim), race (2-dim) and age (2-dim) subspaces.
- These subspaces capture the variations of gender, race and age, respectively.
- Residual space is 12-dim, and captures facial identity.

6. Decomposition

$$\mathbf{y} = \mathbf{V}^\top \mathbf{P}^\top \mathbf{x}$$

$$\mathbf{y}^\top = [\, g \quad r_1 \quad r_2 \quad a_1 \quad a_2 \quad \mathbf{s}^\top \,]$$

gender

race

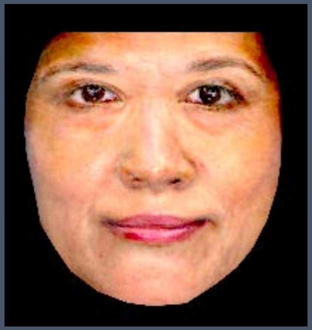
age

residual

Semantic Faces

input

$= g^*$



$\oplus r_1^*$

$\oplus r_2^*$

$\oplus a_1^*$

$\oplus a_2^*$

$\oplus \ldots$

7. Synthesis

$$\mathbf{x}' = \mathbf{P}_r \mathbf{y}' = \mathbf{Q}\mathbf{y}'$$

This equation allows the synthesis of new faces.

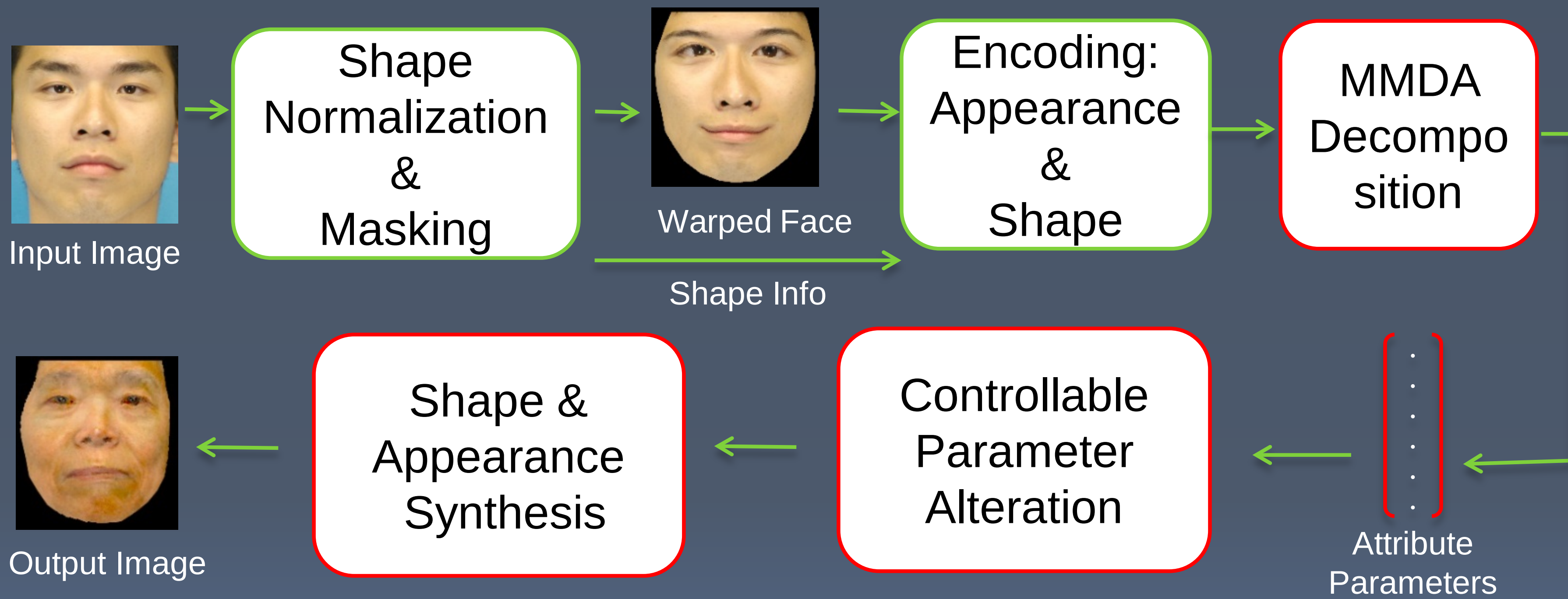
$$\mathbf{P}_r = \mathbf{U}\mathbf{D}^{1/2}$$

Undoes the whitening operation

Visualizing Q

Taking appropriate linear combinations of these Semantic Faces will alter gender, race, age, and identity.

Algorithm overview



Face normalization & encoding

Roughly align

- Affine transform

Locate landmarks

- Stasm or Face++

Warp to reference face

- Thin-plate Spline
- Masking

Encode as vector

- XYZ color space
- (x,y) landmark coords before warping

Results

Altering 2 attributes

Original 1.0 2.0

Gender + Age

Race + Gender

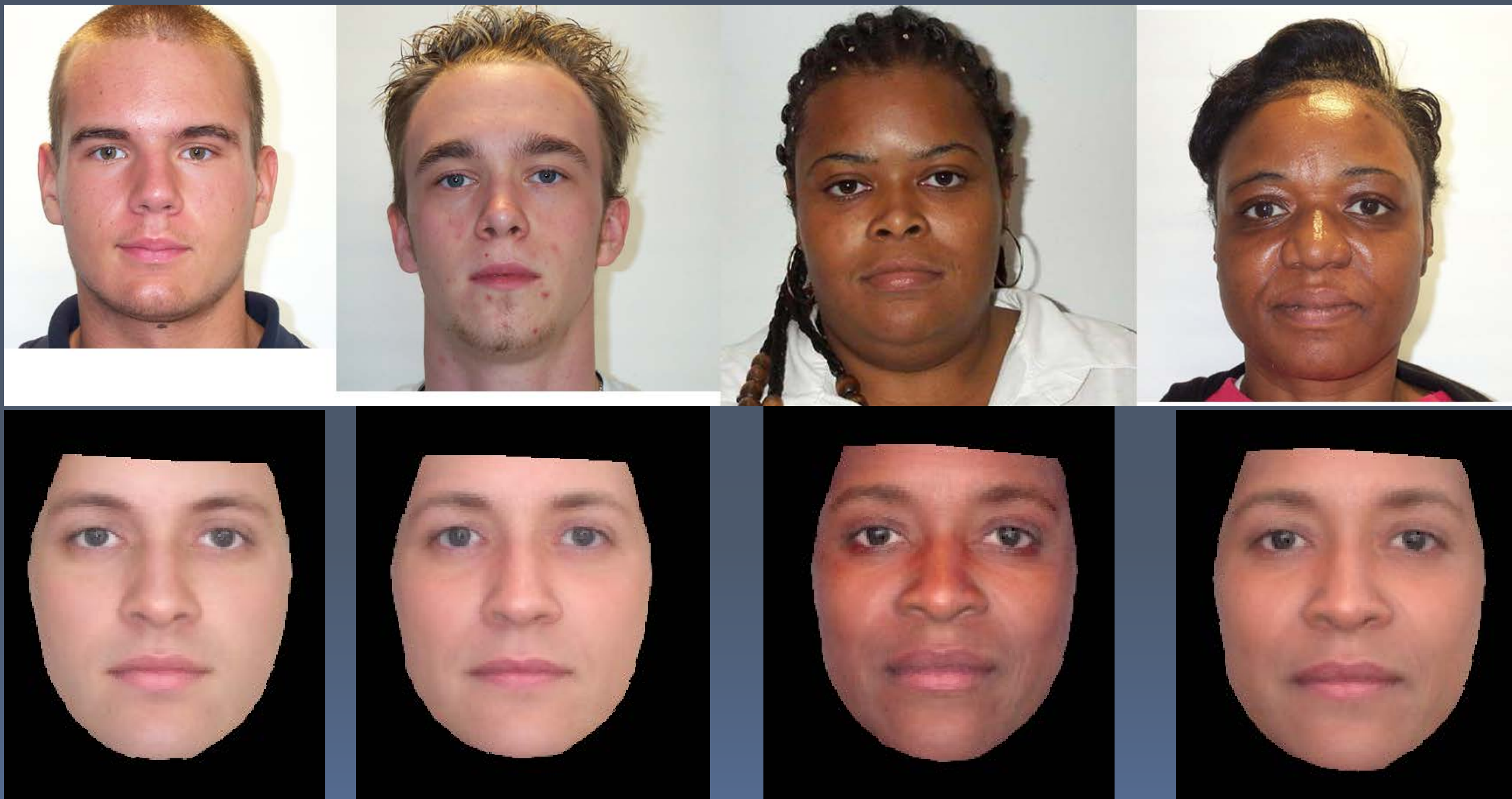
Race + Age

Altering 3 attributes

Original 1.0 2.0

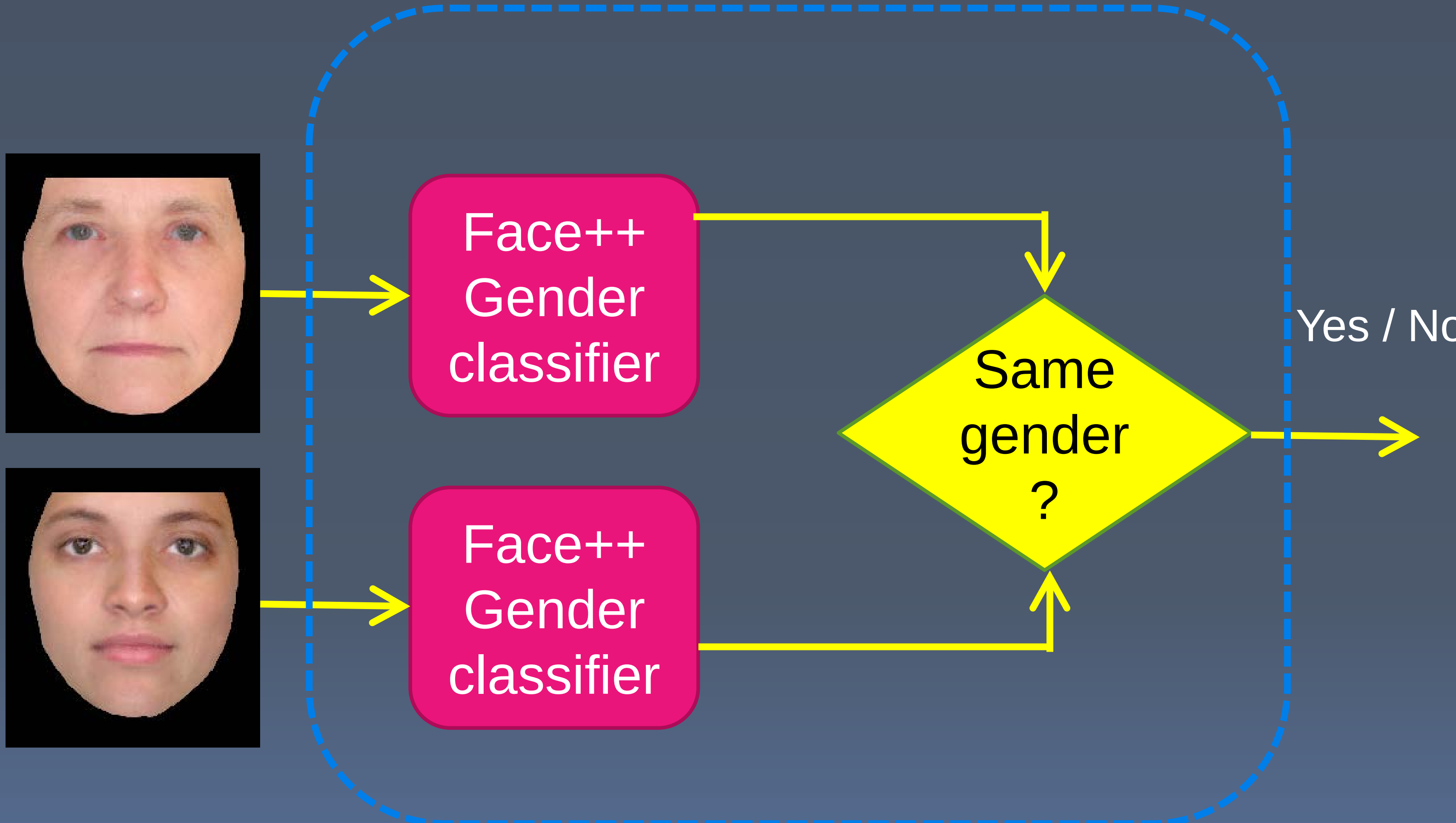
Gender + Race + Age

Altering Identity irreversibly



But can we fool vision algorithms?

Change Detector



Gender Change Detector

Experiments: Single-attribute change

"Changed rates" of 3 Change Detectors

	Intensity σ	0.5	1.0	1.5	2.0	2.5
Gender Change	Gender CD	**0.50**	**0.63**	**0.75**	**0.88**	**1.00**
	Race CD	0.31	0.31	0.31	0.31	0.31
	Age CD	0.56	0.19	0.38	0.38	0.19
Race Change	Gender CD	0.38	0.31	0.44	0.38	0.31
	Race CD	**0.57**	**0.64**	**0.70**	**0.76**	**0.89**
	Age CD	0.19	0.19	0.47	0.66	0.28
Age Change	Gender CD	0.31	0.38	0.31	0.38	0.19
	Race CD	0.25	0.25	0.12	0.12	0.25
	Age CD	**0.66**	**0.84**	**1.00**	**1.00**	**1.00**

Altered attributes are detected as changed, while unaltered attributes are unchanged

Experiments: Multi-attribute change

"Changed rates" of 3 Change Detectors

	Gender+Race			Gender+Age			Race+Age			All 3 attributes		
Intensity σ	Gender	Race	Age	Gender	Race	Age	Gender	Race	Age	Gender	Race	Age
1.0	0.45	0.52	0.23	0.55	0.26	0.71	0.39	0.54	0.77	0.36	0.47	0.72
2.0	0.60	0.60	*0.68*	0.53	0.25	0.82	0.33	0.72	0.88	0.50	0.71	0.82

Higher intensity needed to fool detectors.
Age attribute affected.

Experiments: Identity change

Confidence values returned by Face++

Intensity σ	0.5	1.0	1.5	2.0
Average Confidence	0.9137	0.943	0.951	0.966

Fun results

Fun results

Conclusion

Controllable Face Privacy

- Nuanced alteration of facial attributes

Orthogonal subspace decomposition

- Separate the parameters that control gender, race and age variations

Privacy guarantees

- K-anonymity, L-diversity, t-closeness may be incorporated

Experimental validation

- Extensive experiments with commercial detector confirm the effectiveness of our method

The End: but come see our demo!

Additional slides

$$\beta = \frac{\alpha - f_p}{t_p - f_p}$$

α : observed “changed rate”

β : actual “changed rate”

f_p : false positive rate

t_p : true positive rate