sensorSift

Balancing Utility and Privacy in Sensor Data

Miro **Enev**
Liefeng **Bo**
Xiaofeng **Ren**
Jaeyeon **Jung**
Tadayoshi **Kohno**

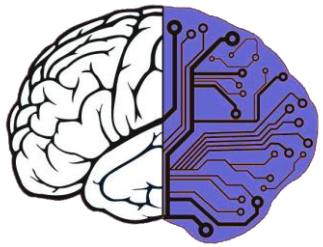
Rise of {Sensors + AI}

- People expect rich computational experiences to be available in every context



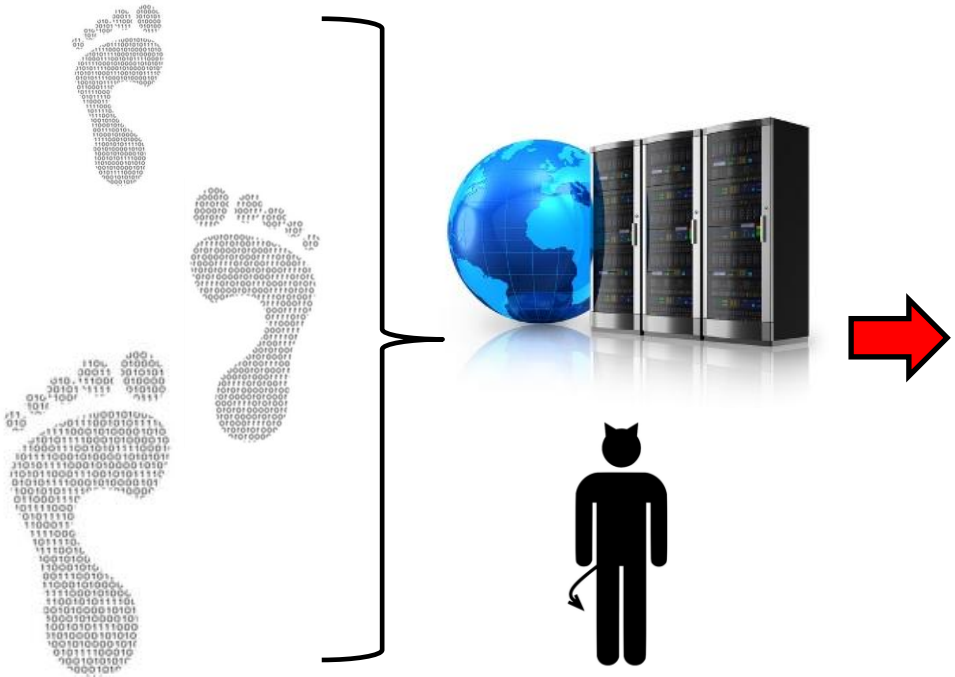

- As a result, our world is increasingly visible to intelligent computers
  - Minimal cost of sensors
  - Cheap computational power
  - Advances in machine reasoning

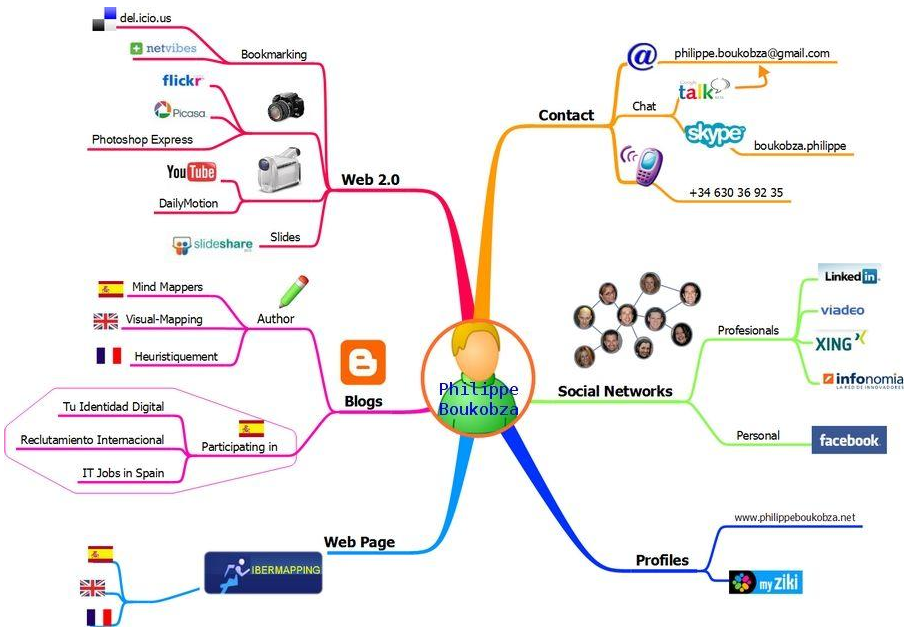




Lack of Balance

- There are many **benefits** of smart-sensor applications
  - **Increased Productivity, Connectivity, and Interactivity**
- However there are also potential **negative** effects
  - **Privacy Risks**

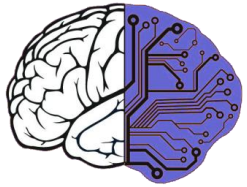




Goals

- Develop a quantitative framework for **balancing** privacy and utility in smart sensing applications.
  - Empower users with privacy guarantees
  - Applications retain functionality

- Evaluate the quality of our framework against state of the art machine inference



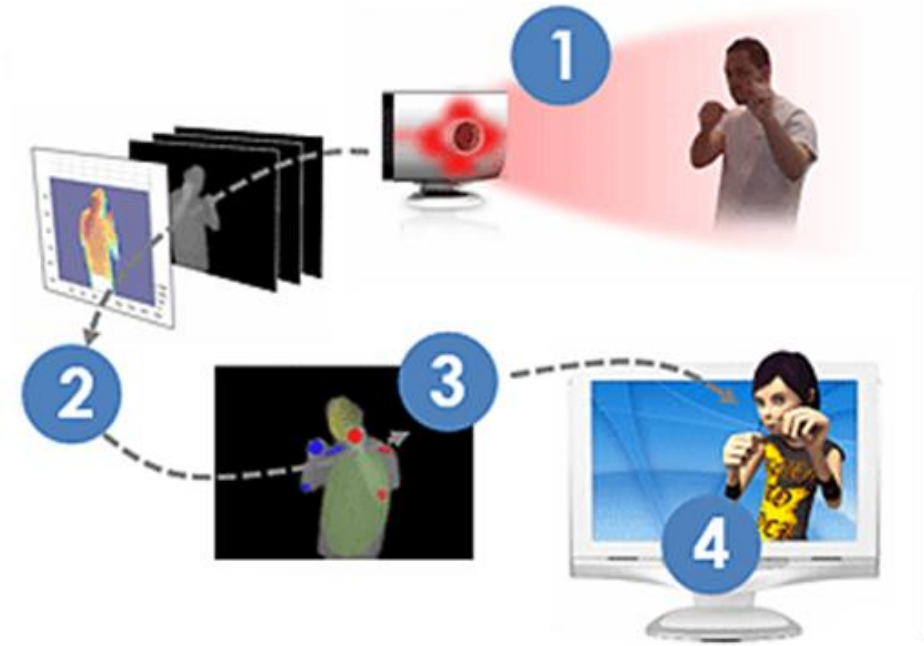
- Offer a flexible solution so that the future demands of users/applications can be supported

Usage Model 1

Sensor data releases to smart applications are often **risk carrying**

Common Practice: Sensor releases all of the raw data to an Application (e.g. MS Kinect)

Sensor :{ **1** sensor data } → **App** :{ **2** feature extract, **3** classify, **4** logic}

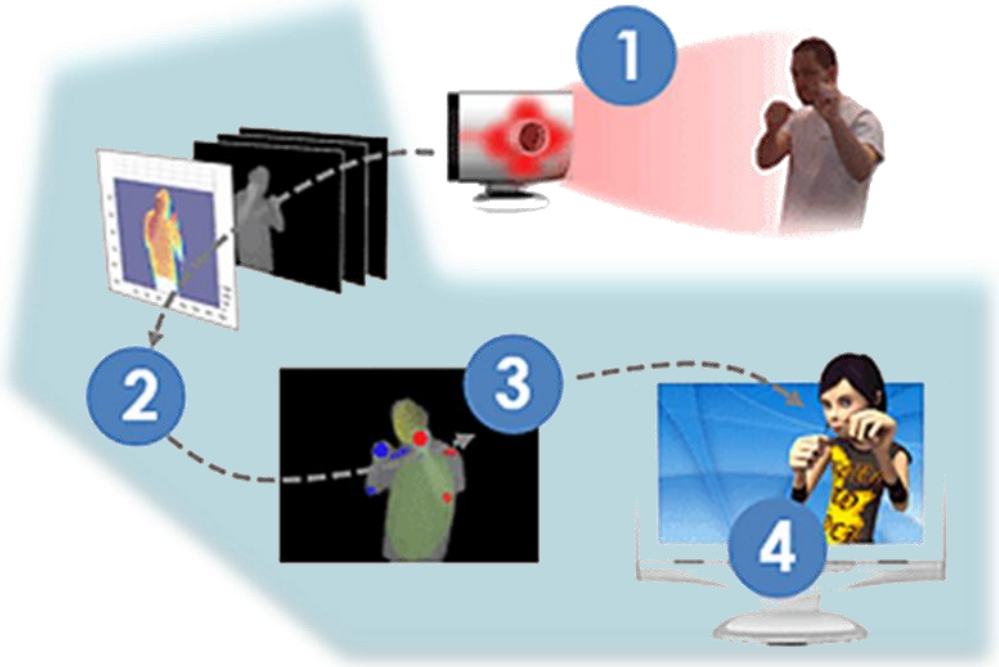


Usage Model 1

Sensor data releases to smart applications are often **risk carrying**

Common Practice: Sensor releases all of the raw data to an Application (e.g. MS Kinect)

Sensor :{ **1** sensor data } → **App** :{ **2** feature extract, **3** classify, **4** logic}



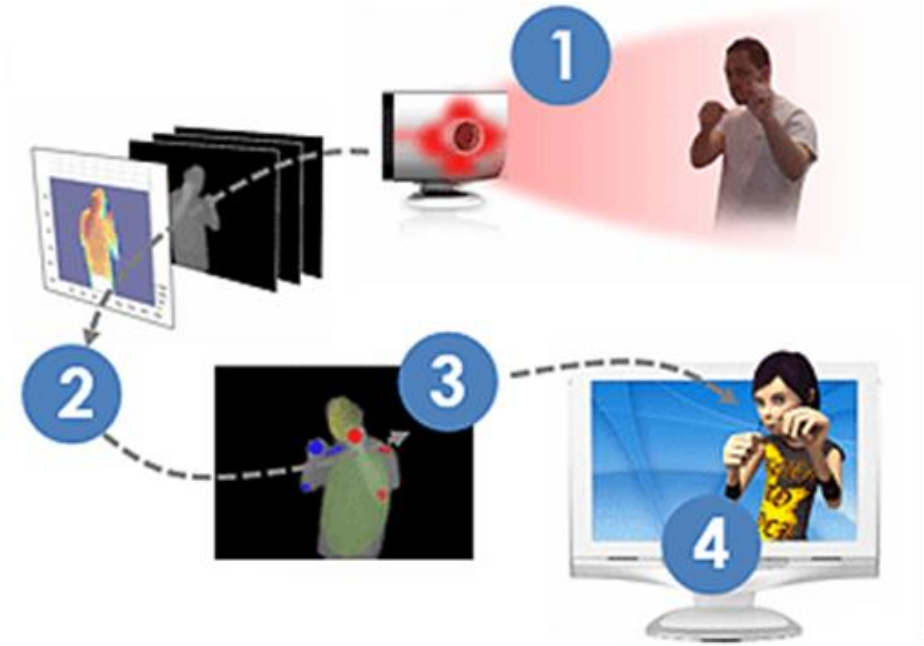
++ INNOVATION
– PRIVACY

Usage Model 2

Sensor data releases to smart applications are often **arbitrarily stifling**

Common Practice: Only a predefined set of features is available to an Application (e.g., iOS)

Platform :{ **1** sensor data , **2** feature extract, **3** classify } → **App** :{**4** logic}

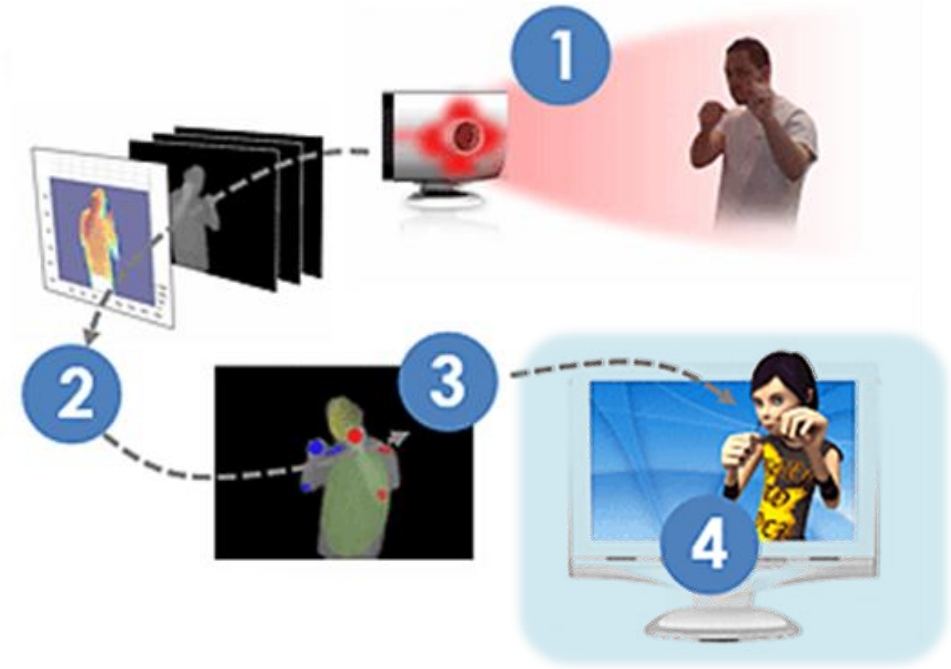


Usage Model 2

Sensor data releases to smart applications are often **arbitrarily stifling**

Common Practice: Only a predefined set of features is available to an Application (e.g., iOS)

Platform :{ **1** sensor data , **2** feature extract, **3** classify } → **App** :{**4** logic}



– **INNOVATION**
++ **PRIVACY**

Solution

- Users choose what attributes to keep **private**
- Applications can request non-private (**public**) attributes
  - Public attributes can be invented!

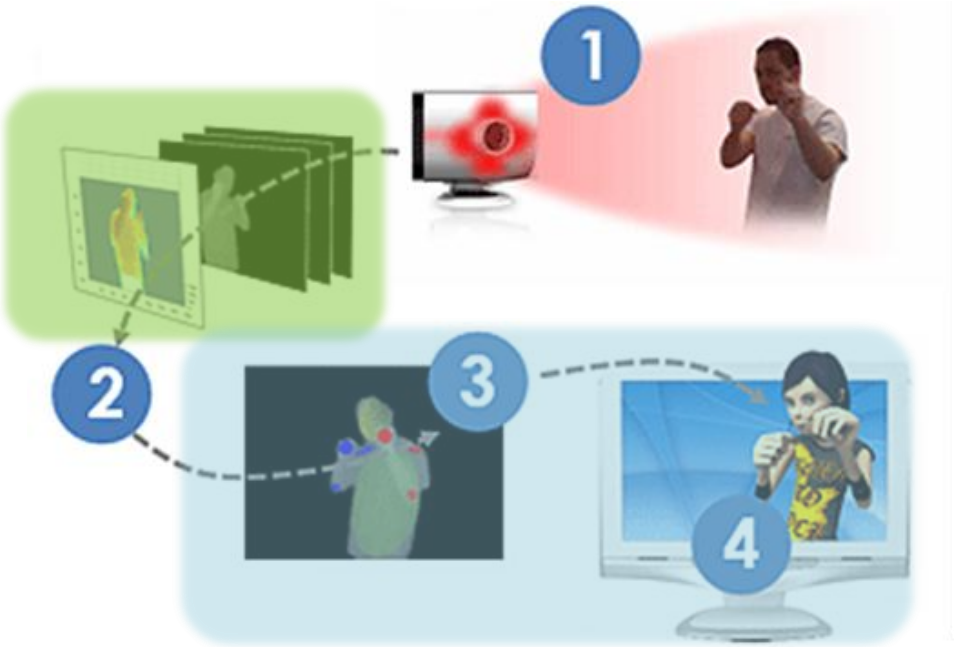
Solution

- Users choose what attributes to keep **private**
- Applications can request non-private (**public**) attributes
  - Public attributes can be invented!

POLICY

- We transform (sift) sensor data to reveal the **public** but hide the **private** attributes

Plat. : {**1** sensor data, **2** sift features } → **App** { **3** classify, **4** logic}



+ INNOVATION
+ PRIVACY

Evaluation Context

ATTRIBUTES: visually describable characteristics about a face

System Overview

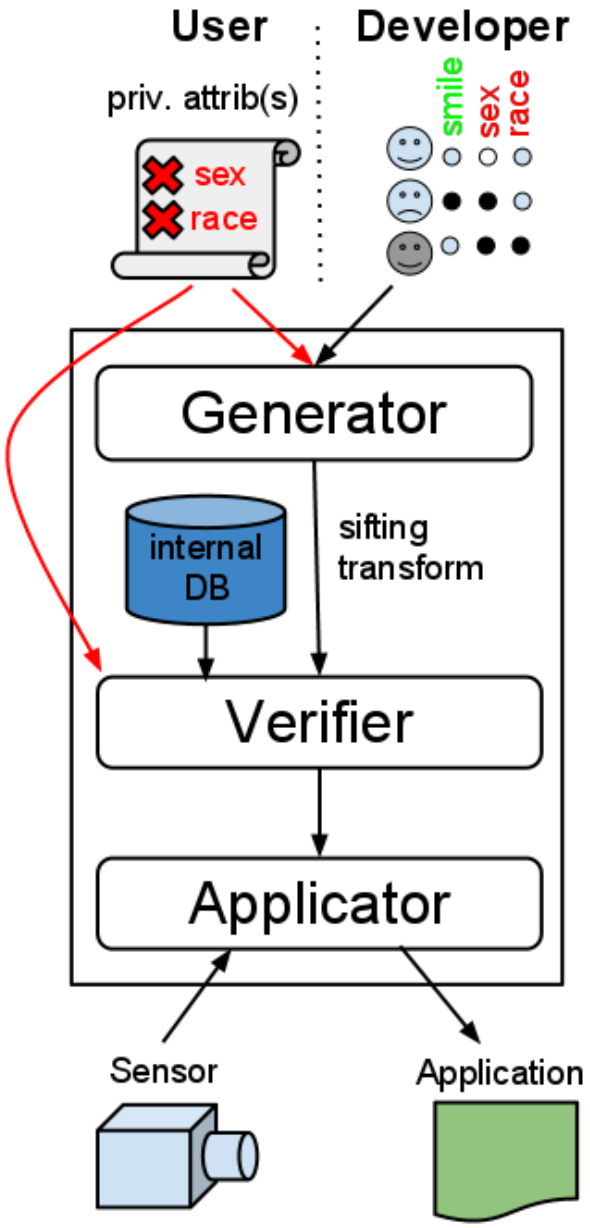
Scenario:

- **USER**: I don't want apps. to have knowledge about my **race** and **gender**
- **APPLICATION:** Is the user **smiling**?

> **POLICY:** **PRIVATE {race, gender}, PUBLIC {smiling}**

System:

1. Generates Sift
2. Verifies Sift
3. Applies Verified Sift



System Overview

Scenario:

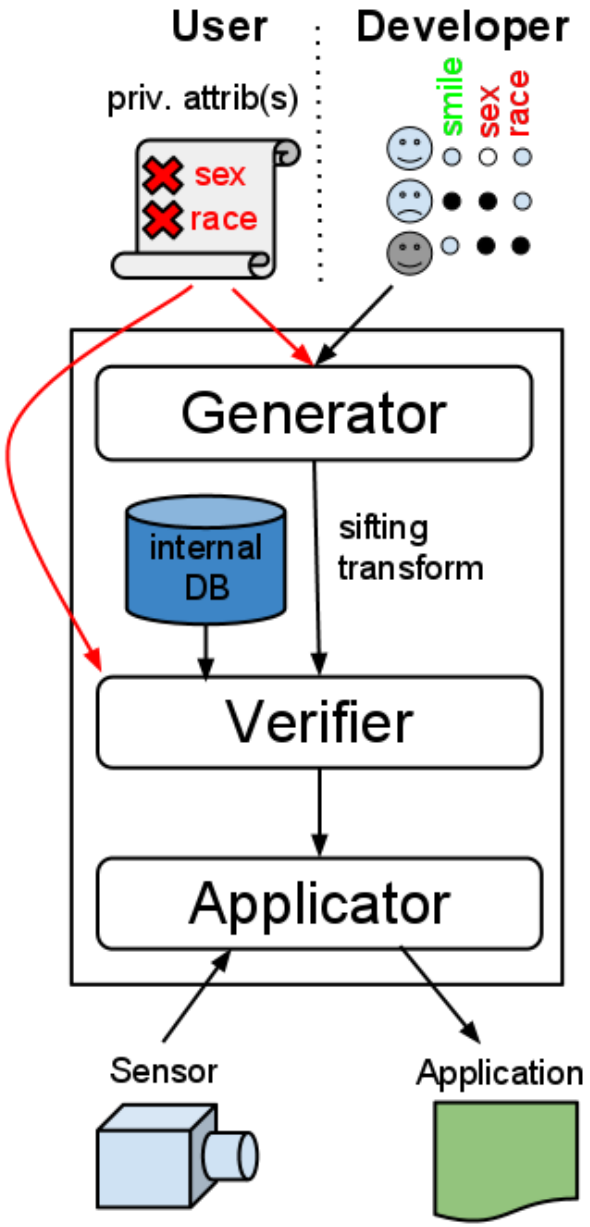
- **USER**: I don't want apps. to have knowledge about my **race** and **gender**
- **APPLICATION:** Is the user **smiling**?

> **POLICY:** **PRIVATE {race, gender}, PUBLIC {smiling}**

System:

1. Generates Sift
2. Verifies Sift
3. Applies Verified Sift

RUNTIME



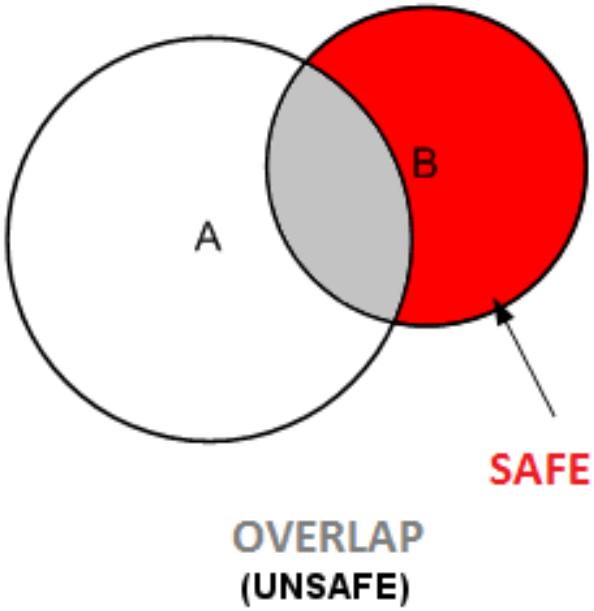
database

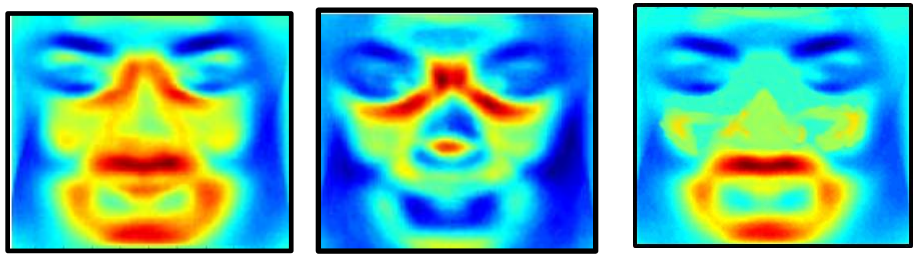
Generating Sifts

Intuitively, sifting finds the safe region(s) in feature space which are in the public feature set **B** but not in the private one **A**.

feature regions are based on a large database of sensor samples

A = eyewear (private)
B = gender (public)



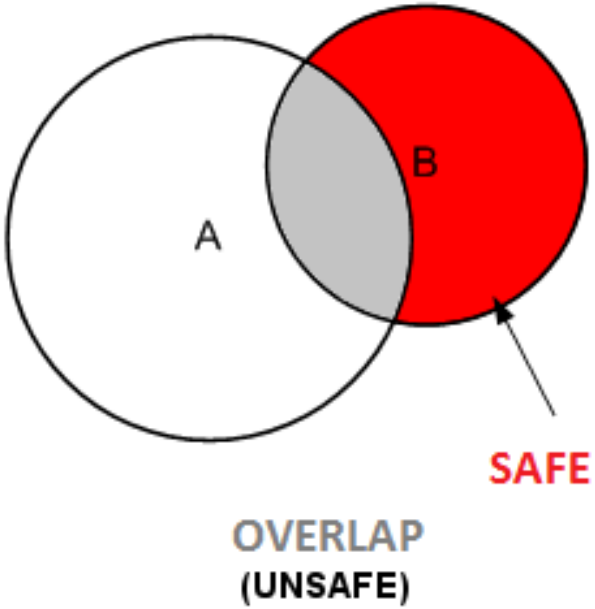


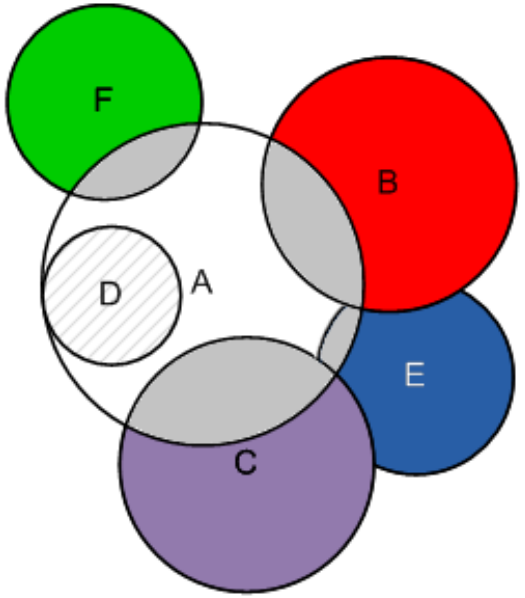
gender **eyewear** **safe region**

Generating Sifts

Intuitively, sifting finds the safe region(s) in feature space which are in the public feature set **B** but not in the private one **A**.

A = eyewear (private)
B = gender (public)





Safe region(s) may not always exist for certain attribute correlations.

Sifting Details

X = Raw Features
X′ = Sifted Features

$X_n, n > 100k$

sift

$X'_n, n \sim 5$

sifting in detail

PPLS

Algorithm 1 Privacy Partial Least Squares

1. Set $j = 0$ and cross-product $S_j = X^\top Y^+$
2. if $j > 0$, $S_j = S_{j-1} - P(P^\top P)^{-1} P^\top S_{j-1}$
3. Compute the largest eigenvector w_j: $\left[S_j^\top S_j - X^\top Y^- (Y^-)^\top X\right] w_j = \lambda w_j$
4. Compute $p_j = \frac{X^\top X w_j}{w_j^\top X^\top X w_j}$
5. If $j = k$, stop; otherwise let $P = [p_0, \cdots, p_j]$ and $j = j + 1$ and go back to step 2

$$find \quad \max_w \left[cov(Xw, Y^+)^2 - \lambda * cov(Xw, Y^-)^2\right]$$

Y+ = labels of public attribute(s)
Y- = labels of private attribute(s)

Performance Metrics

- A successful sift will have low scores on both **PubLoss** and **PrivLoss**

  - **PubLoss**: Decrease in sifted public attribute classification accuracy relative to the achievable accuracy using raw (unsifted) data.

  - **PrivLoss**: Gain in sifted private attribute classification accuracy relative to chance.

$$PubLoss = ML_m(X, Y^+) - ML_m(PM_{Y^+,Y^-}(X, K), Y^+)$$

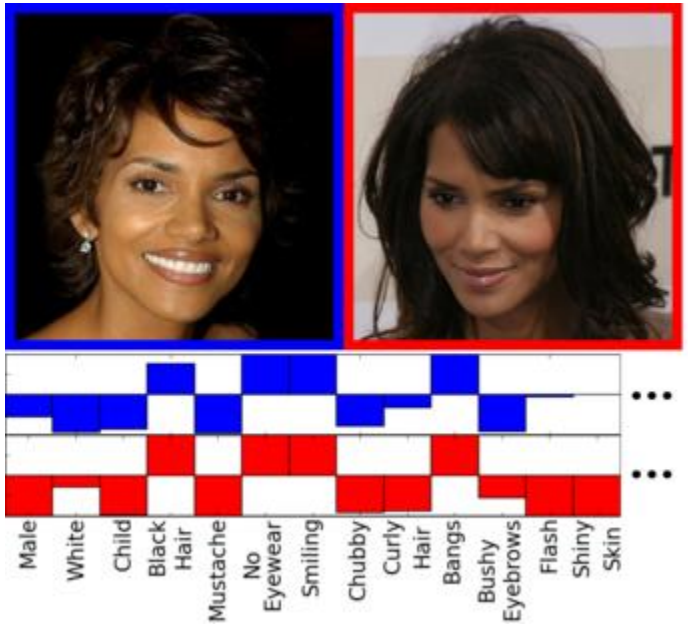
$$PrivLoss = ML_m(PM_{Y^+,Y^-}(X, K), Y^-) - .5$$

*Classifiers : Linear Support Vector Machine (SVM), Non-Linear SVM, Neural Network, Random Forest, kNearest Neighbors

Dataset & Attributes

PubFig Database ~**45,000** face images of **200** celebrities, **72** attributes

Attributes are [binary] labels for visually describable characteristics,

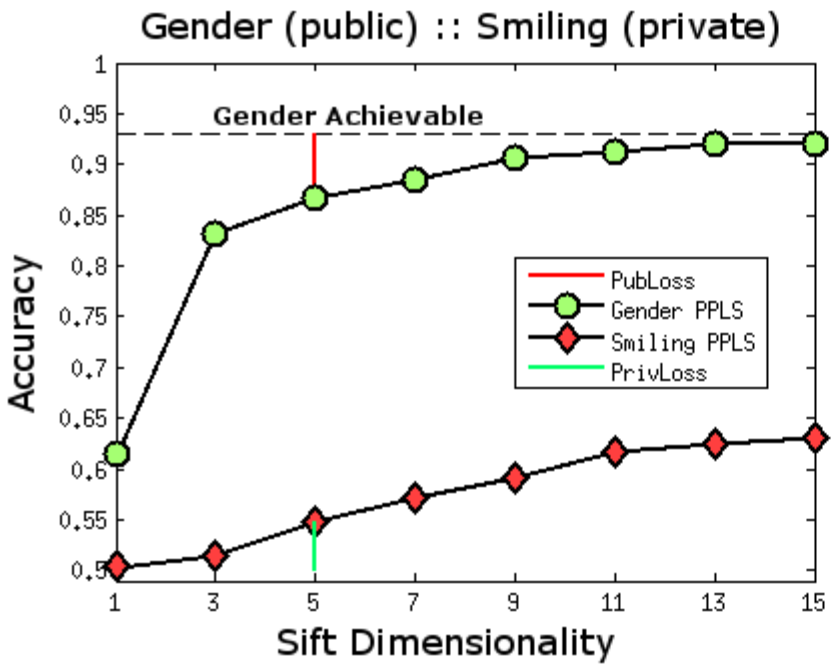


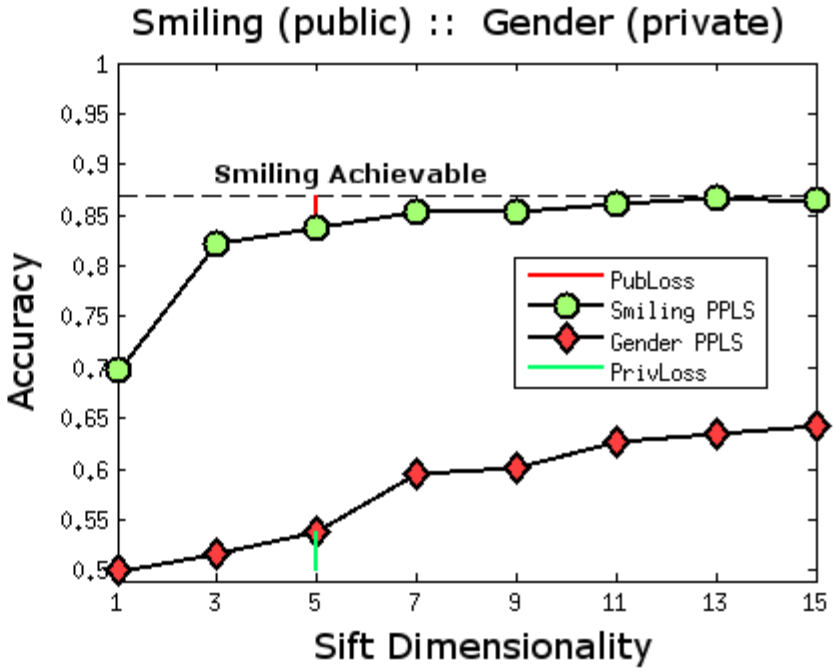
Attribute Clusters

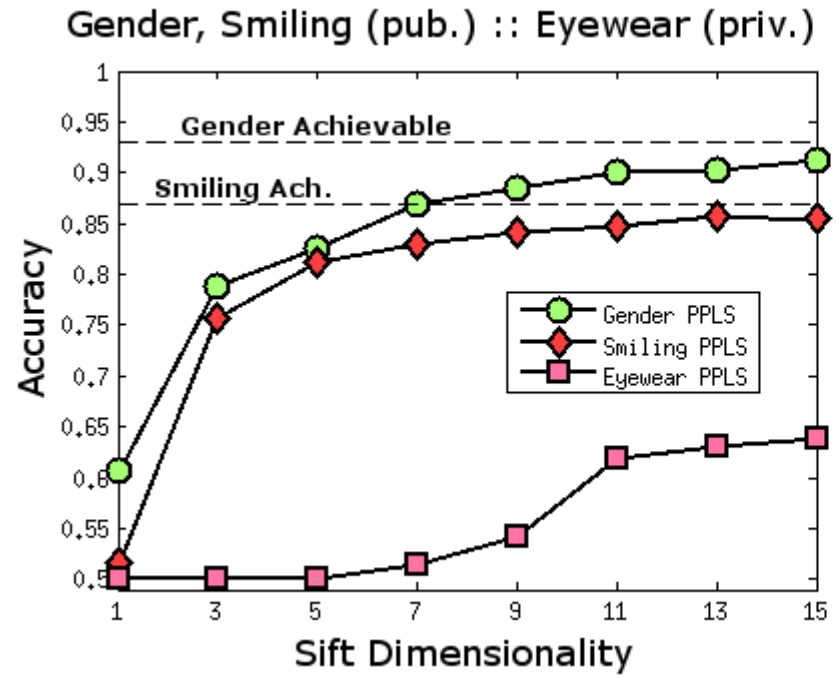
Wavy Hair
Arched Eyebrows
Wearing Lipstick
Blond Hair
Youth

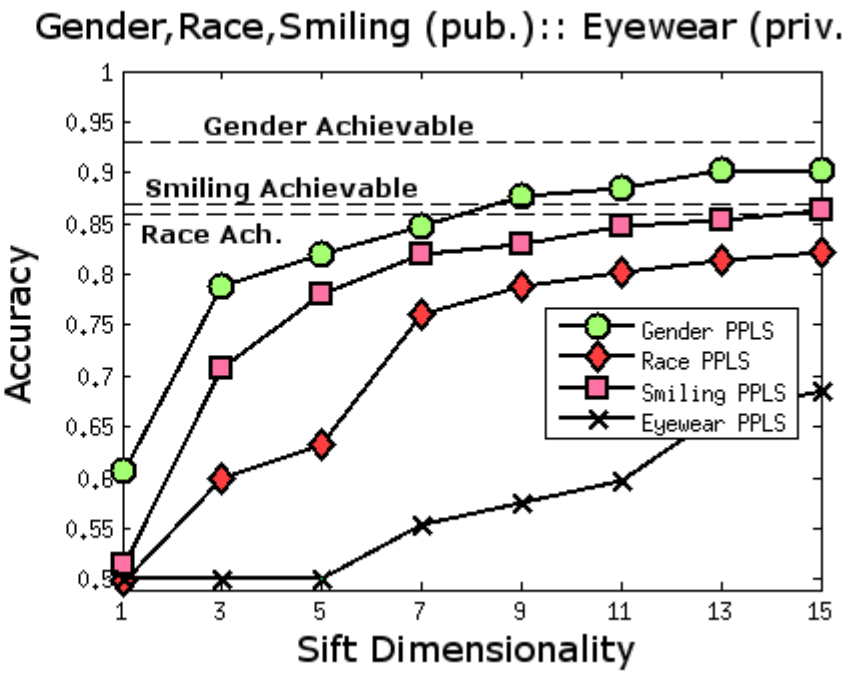
Attractive Female

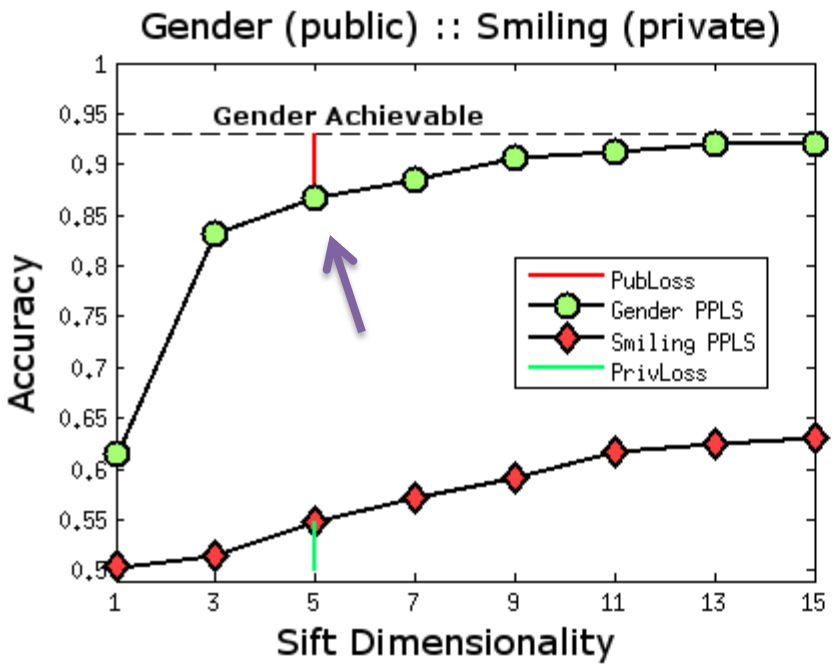
Male **- M,** Attractive Female **- AF,** White **- W,**
Youth **- Y,** Smiling **- S,** Frowning **- F,** No Eyewear **- nE,**
Obstructed Forehead **- OF,** No Beard **- nB,** and Outdoors **- O.**

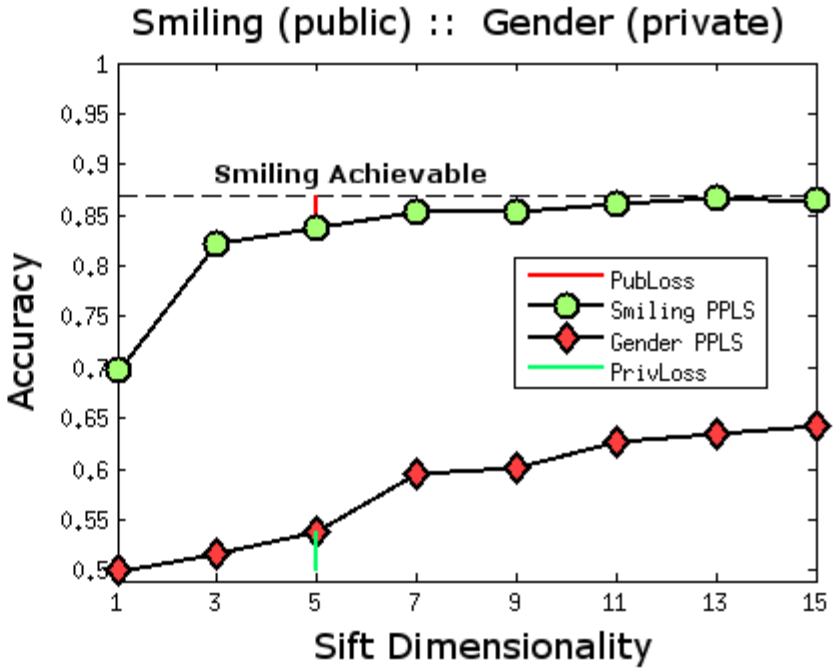

Gender (public) :: Smiling (private)
Gender Achievable
PubLoss
Gender PPLS
Smiling PPLS
PrivLoss
Accuracy
Sift Dimensionality

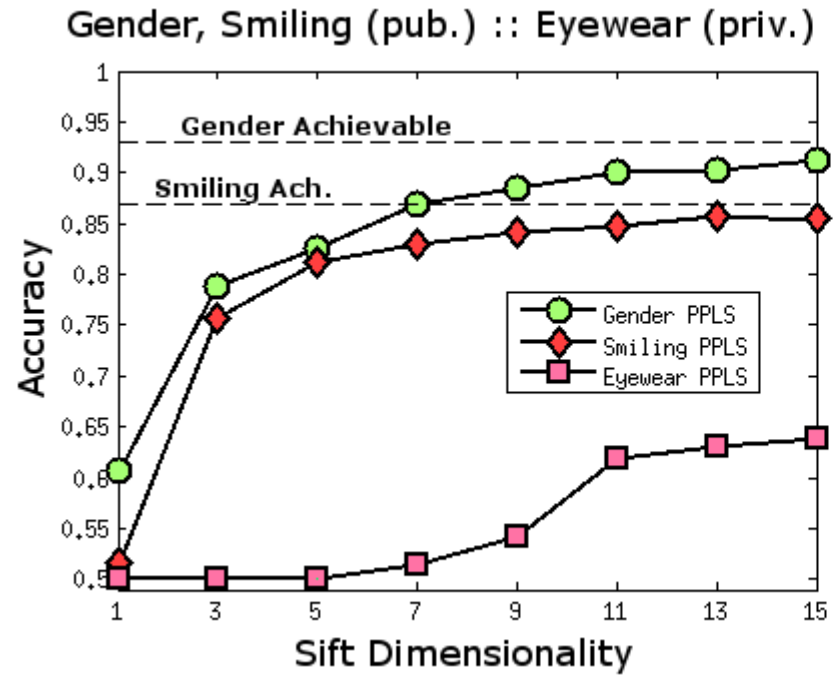

Smiling (public) :: Gender (private)
Smiling Achievable
PubLoss
Smiling PPLS
Gender PPLS
PrivLoss
Accuracy
Sift Dimensionality

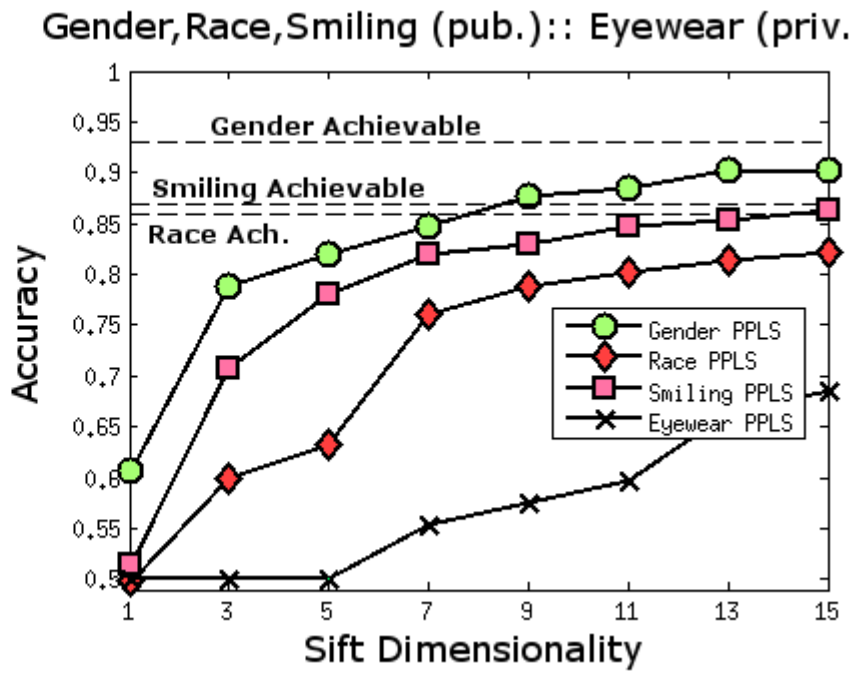

Gender, Smiling (pub.) :: Eyewear (priv.)
Gender Achievable
Smiling Ach.
Gender PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality

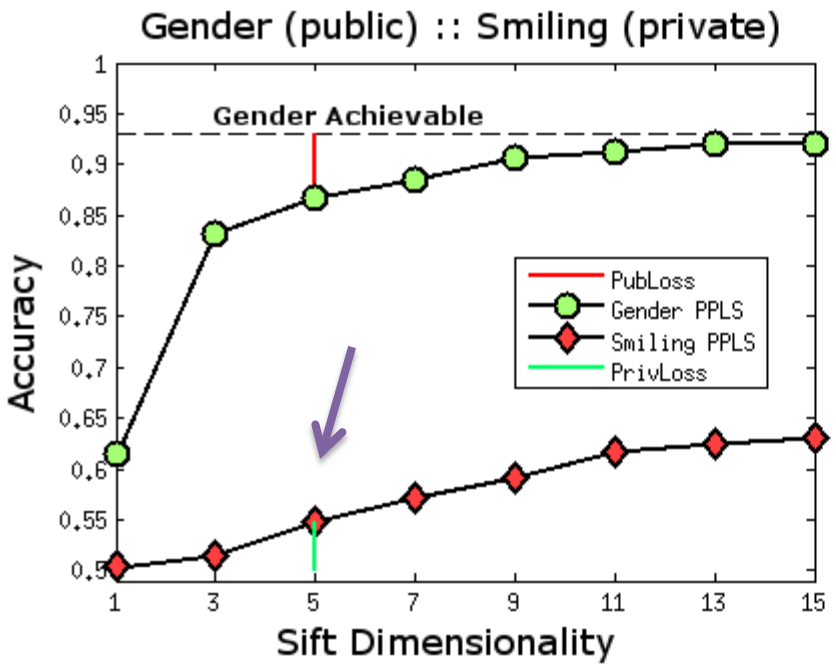

Gender,Race,Smiling (pub.):: Eyewear (priv.)
Gender Achievable
Smiling Achievable
Race Ach.
Gender PPLS
Race PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality

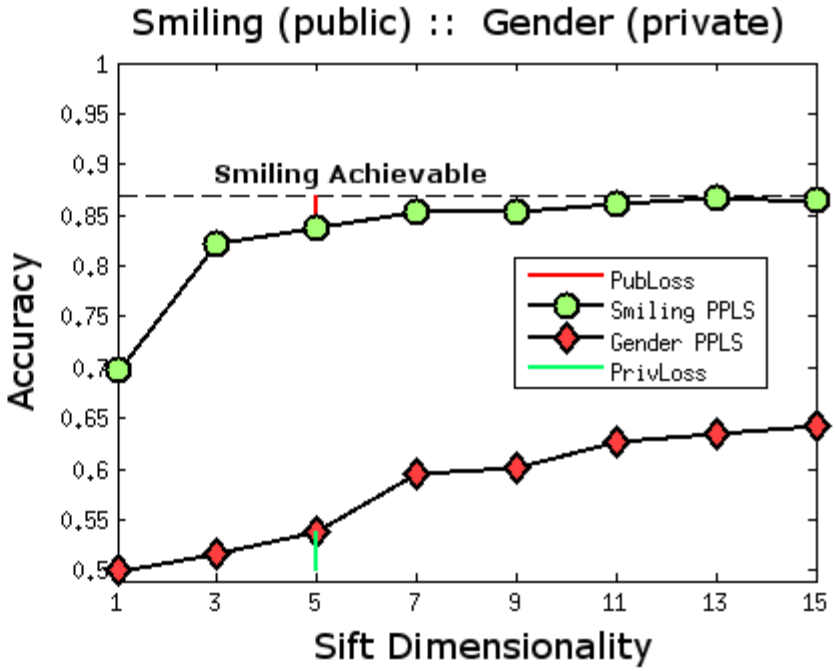

Gender (public) :: Smiling (private)
Gender Achievable
PubLoss
Gender PPLS
Smiling PPLS
PrivLoss
Accuracy
Sift Dimensionality

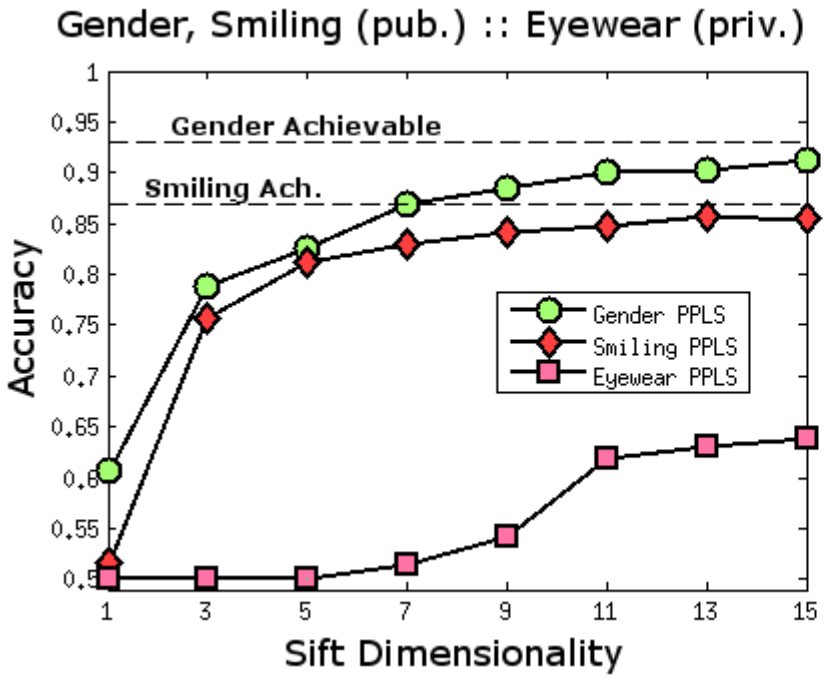

Smiling (public) :: Gender (private)
Smiling Achievable
PubLoss
Smiling PPLS
Gender PPLS
PrivLoss
Accuracy
Sift Dimensionality

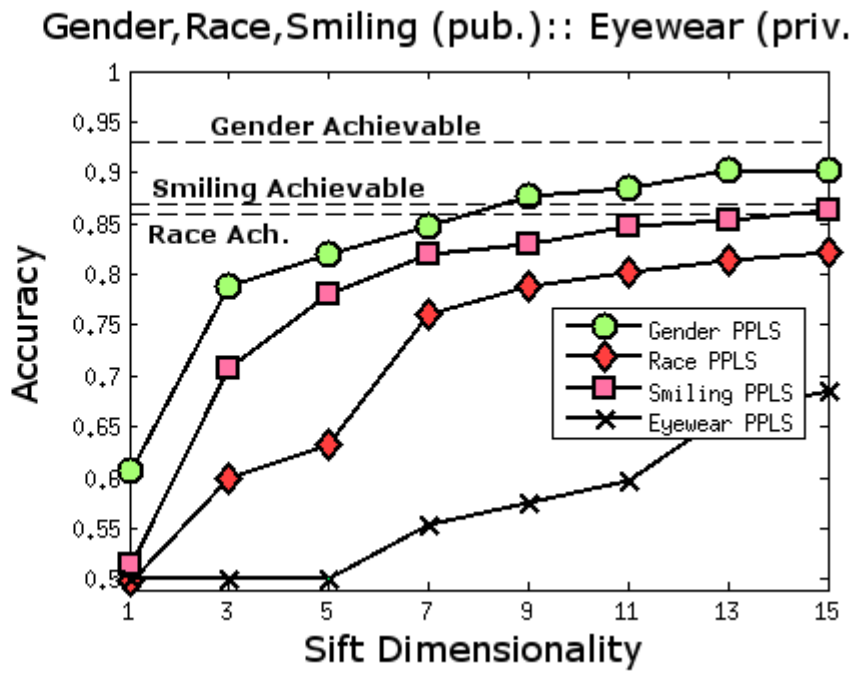

Gender, Smiling (pub.) :: Eyewear (priv.)
Gender Achievable
Smiling Ach.
Gender PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality

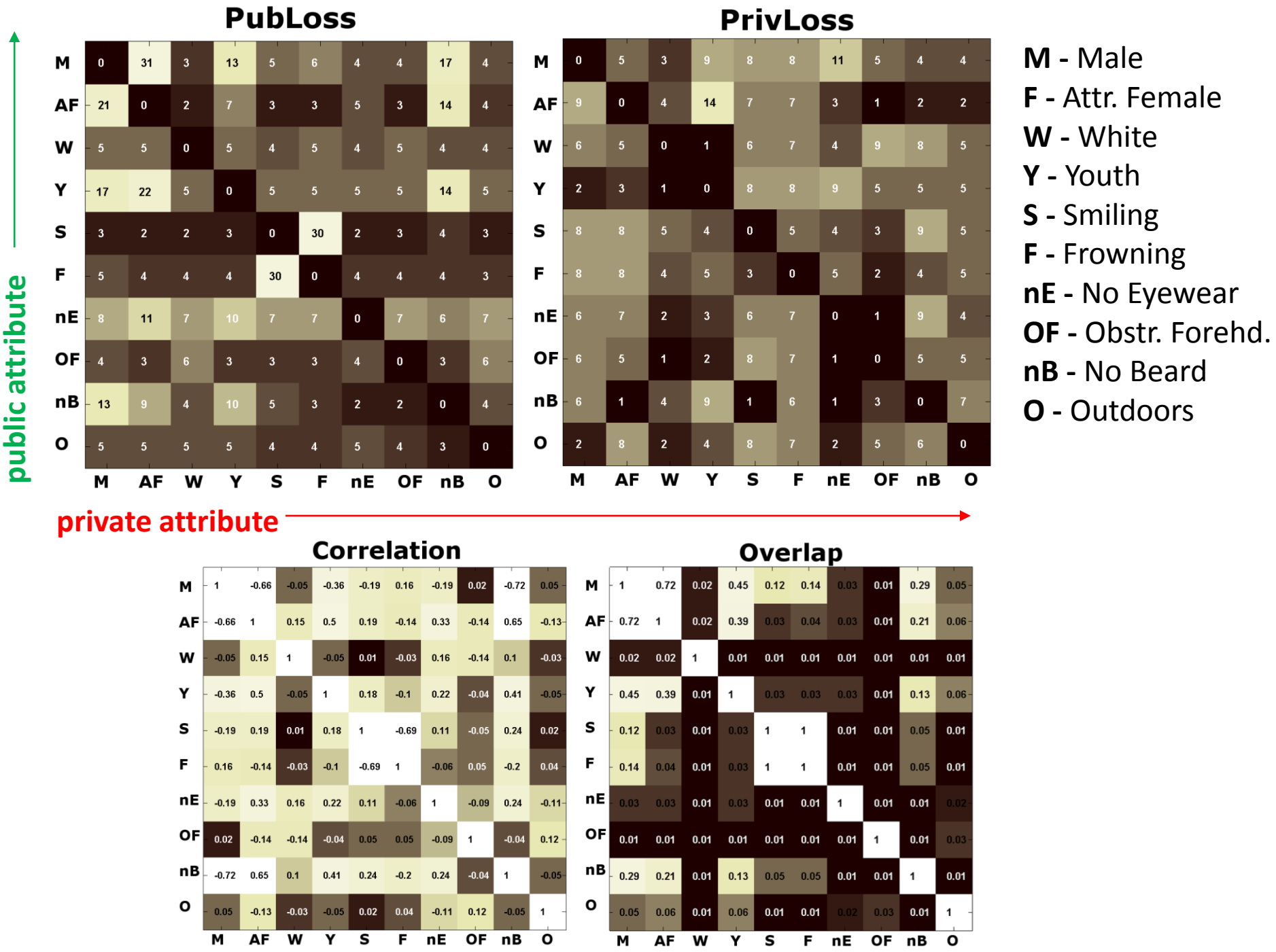

Gender,Race,Smiling (pub.):: Eyewear (priv.)
Gender Achievable
Smiling Achievable
Race Ach.
Gender PPLS
Race PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality


Gender (public) :: Smiling (private)
Gender Achievable
PubLoss
Gender PPLS
Smiling PPLS
PrivLoss
Accuracy
Sift Dimensionality


Smiling (public) :: Gender (private)
Smiling Achievable
PubLoss
Smiling PPLS
Gender PPLS
PrivLoss
Accuracy
Sift Dimensionality


Gender, Smiling (pub.) :: Eyewear (priv.)
Gender Achievable
Smiling Ach.
Gender PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality


Gender,Race,Smiling (pub.):: Eyewear (priv.)
Gender Achievable
Smiling Achievable
Race Ach.
Gender PPLS
Race PPLS
Smiling PPLS
Eyewear PPLS
Accuracy
Sift Dimensionality


PubLoss
PrivLoss
public attribute
private attribute
M AF W Y S F nE OF nB O
M - Male
F - Attr. Female
W - White
Y - Youth
S - Smiling
F - Frowning
nE - No Eyewear
OF - Obstr. Forehd.
nB - No Beard
O - Outdoors
Correlation
Overlap

Conclusions

- We proposed a theoretical framework for quantitative balance between utility and privacy though policy based control of sensor data exposure.

- In our analysis we found promising results when we evaluated the PPLS algorithm in the context of automated face understanding.

- The algorithm we introduce is general, as it exploits the statistical properties of the data; and in the future it would be exciting to evaluate SensorSift in other sensor contexts.

- Available as Open Source!

miro@cs.washington.edu

Thanks!

Liefeng

Xiaofeng

Jaeyeon

Yoshi

SecLab @ UW

ISTC PC

Intel Science and Technology Center for

Pervasive Computing

Microsoft Research

Questions?

http://homes.cs.washington.edu/~miro/**sensorSift**