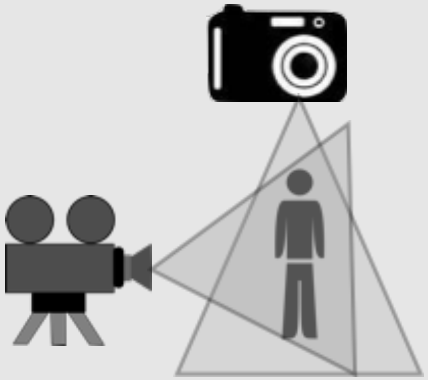


TagSense: A Smartphone-based Approach to Automatic Image Tagging

Chuan Qin, Xuan Bao, Romit Roy Choudhury, Srihari Nelakuditi
University of South Carolina, Duke University

Image retrieval

- Digital pictures are undergoing an explosion

- Image retrieval becomes crucial

  - Image searching

  - Personal albums

Facebook: 2.5 billion/month

My PC: 10GB/vacation

Image tagging

- Image retrieval systems use tags
- Human tagging
  - Accurate, widely used
  - Slow and boring
- Image based auto-tagging
  - Still many constraints

Smartphone, the wild card

- Today's smartphones have powerful built-in sensors
- People always carry their phones

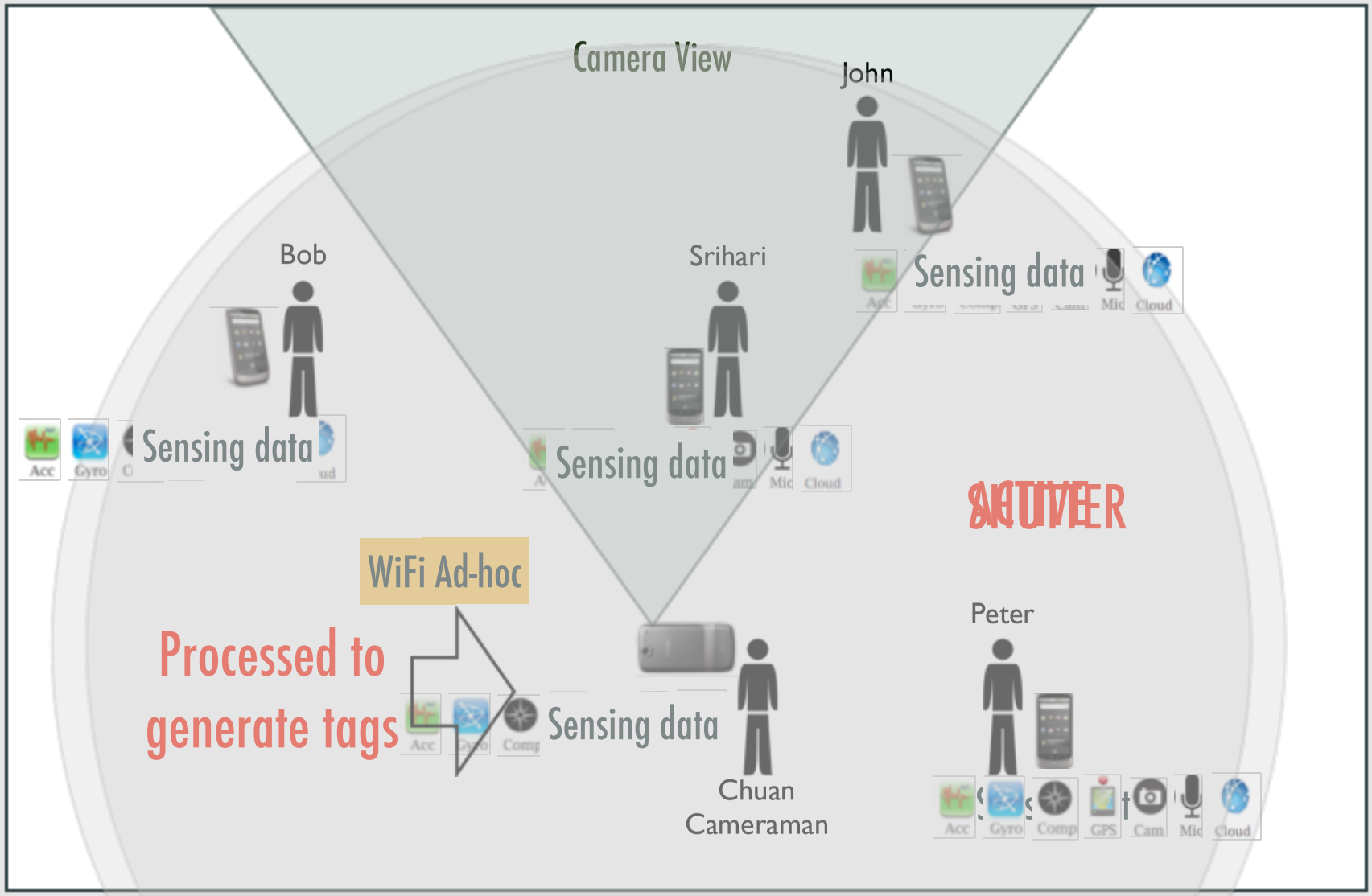
TagSense

- A system for auto-tagging, with smartphone sensors
- Leverages multiple sensing domains

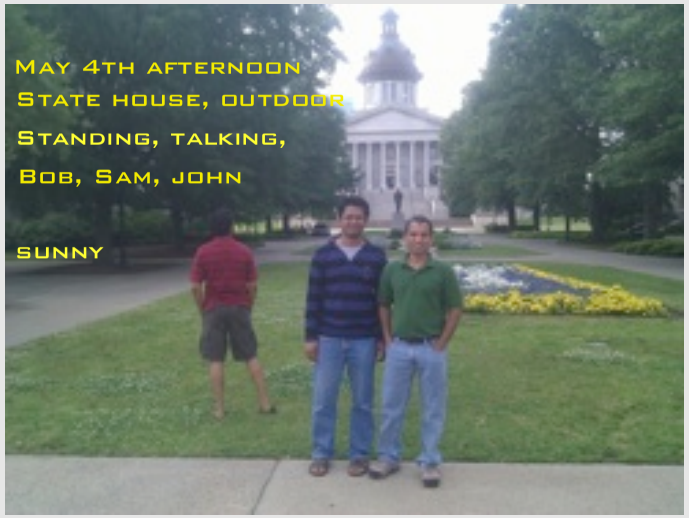
An example

What happens at picturing time

TagSense architecture



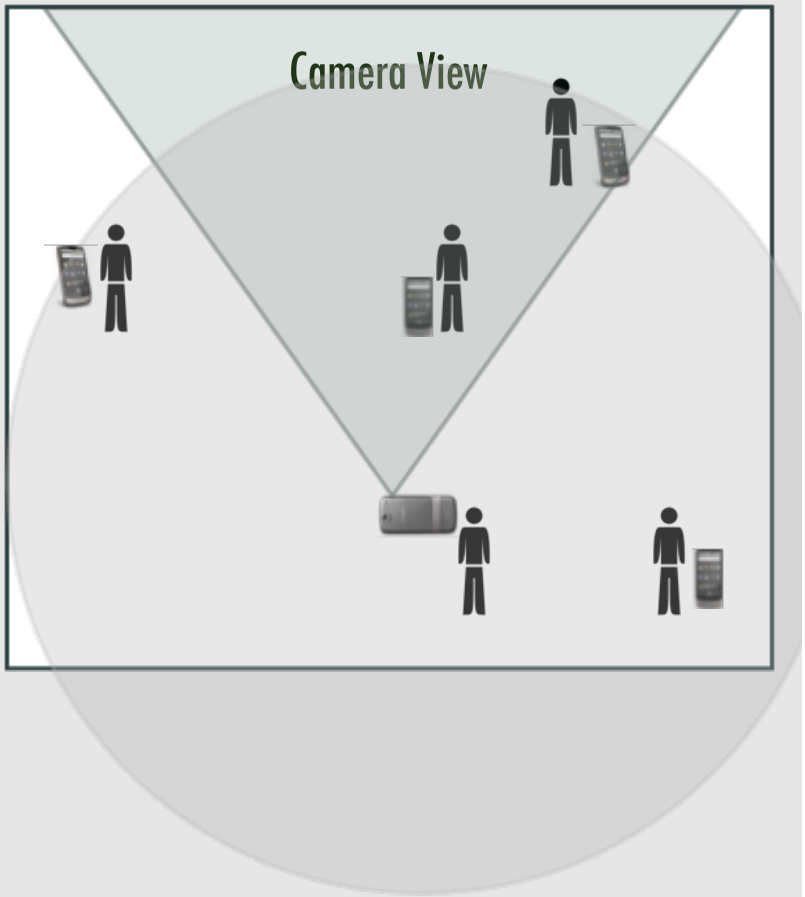
TagSense tag generation



- When? Clock + GPS + WiFi = May 4th afternoon
- Where? GPS + WiFi + Comp + Light s. = State house, outdoor
- What? Acc + Gyro + Mic + Cloud = Standing, talking, sunny
- Who? ??? = Bob, Sam, john

The challenge of 'Who'

- Include only those in camera view
- Why not use localization?
- New opportunities enabled by multi-dimensional sensing



Possible opportunities

1. Accelerometer based motion signatures
2. Complementary compass directions
3. Correlating visual and acceleration

1. Accelerometer based motion signatures

Move into posing

↓

Stable when posing

↓

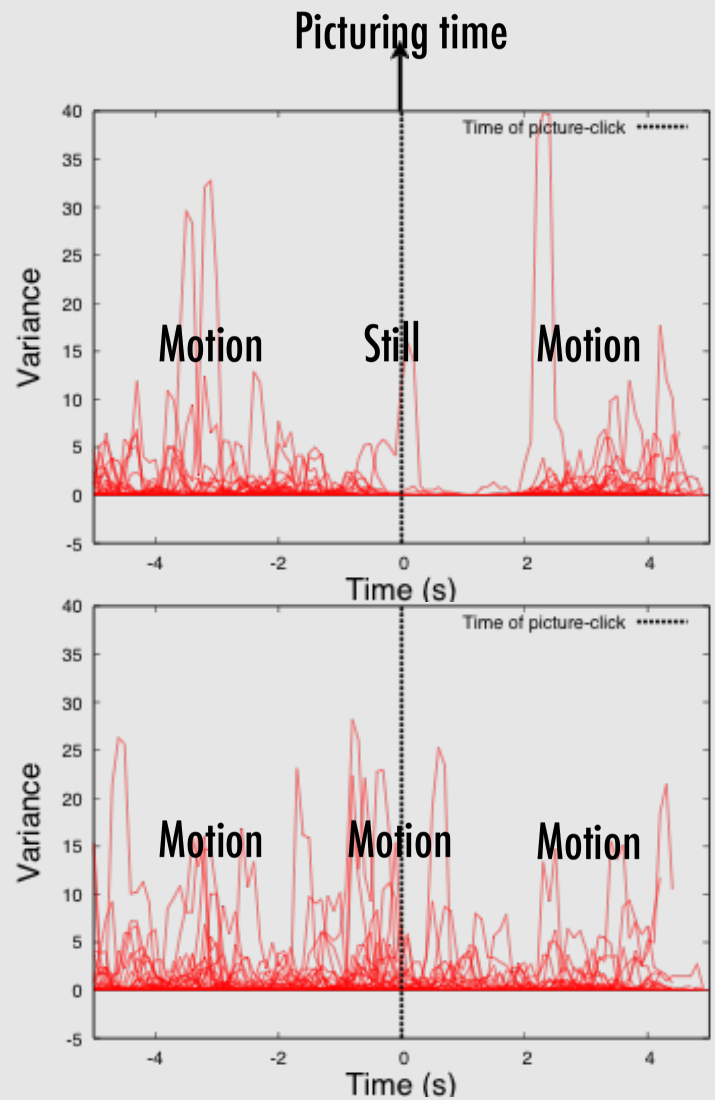
Move out of posing

Is Bob a special case?

Acceleration plotting for 50+ pictures

People inside the picture:

People outside the picture:



Possible opportunities

1. Accelerometer based motion signatures
2. Complementary compass directions
3. Correlating visual and acceleration

2. Complementary compass directions

People in the picture are likely to face the camera

Is someone facing the camera?

- Compass reading != User's orientation
- The diff: Personal Compass Offset (PCO)
- Need to calibrate PCO
  - Use posing picture for calibration

Possible opportunities

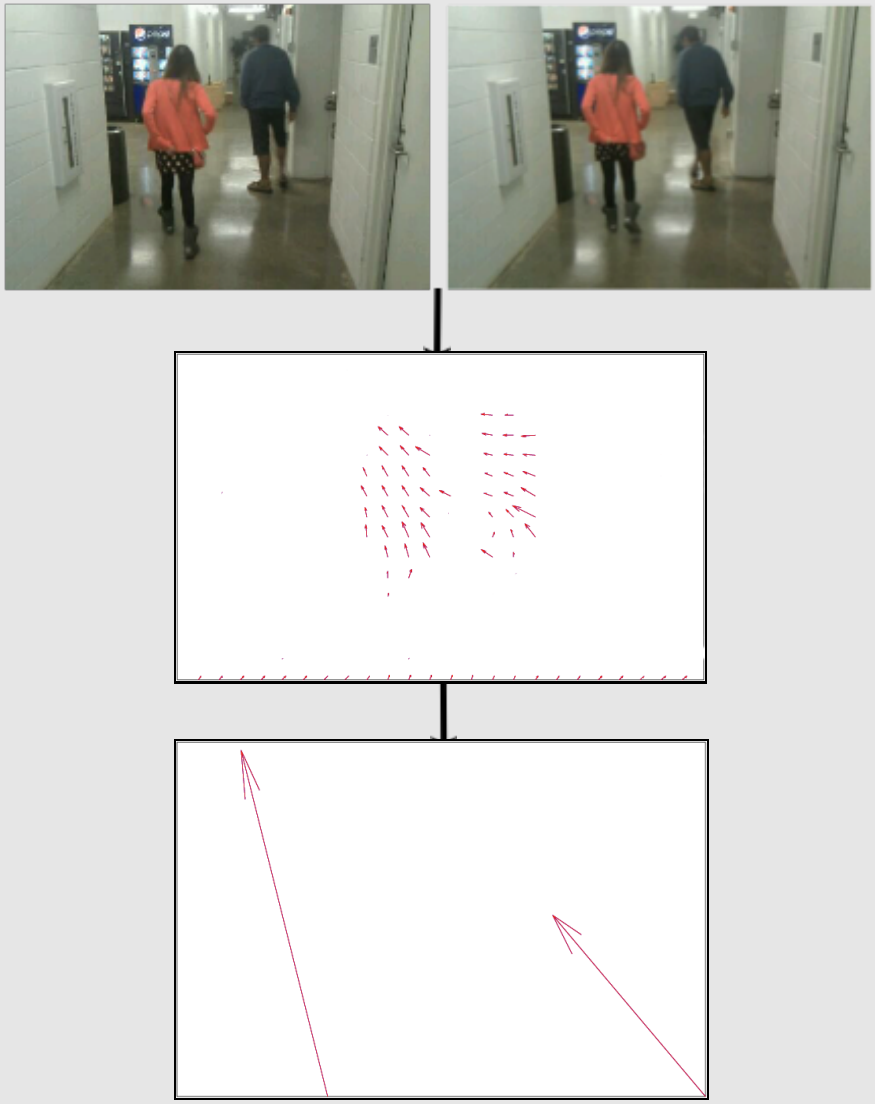
1. Accelerometer based motion signatures
2. Complementary compass directions
3. Correlating visual and acceleration

3. Correlating visual and acceleration

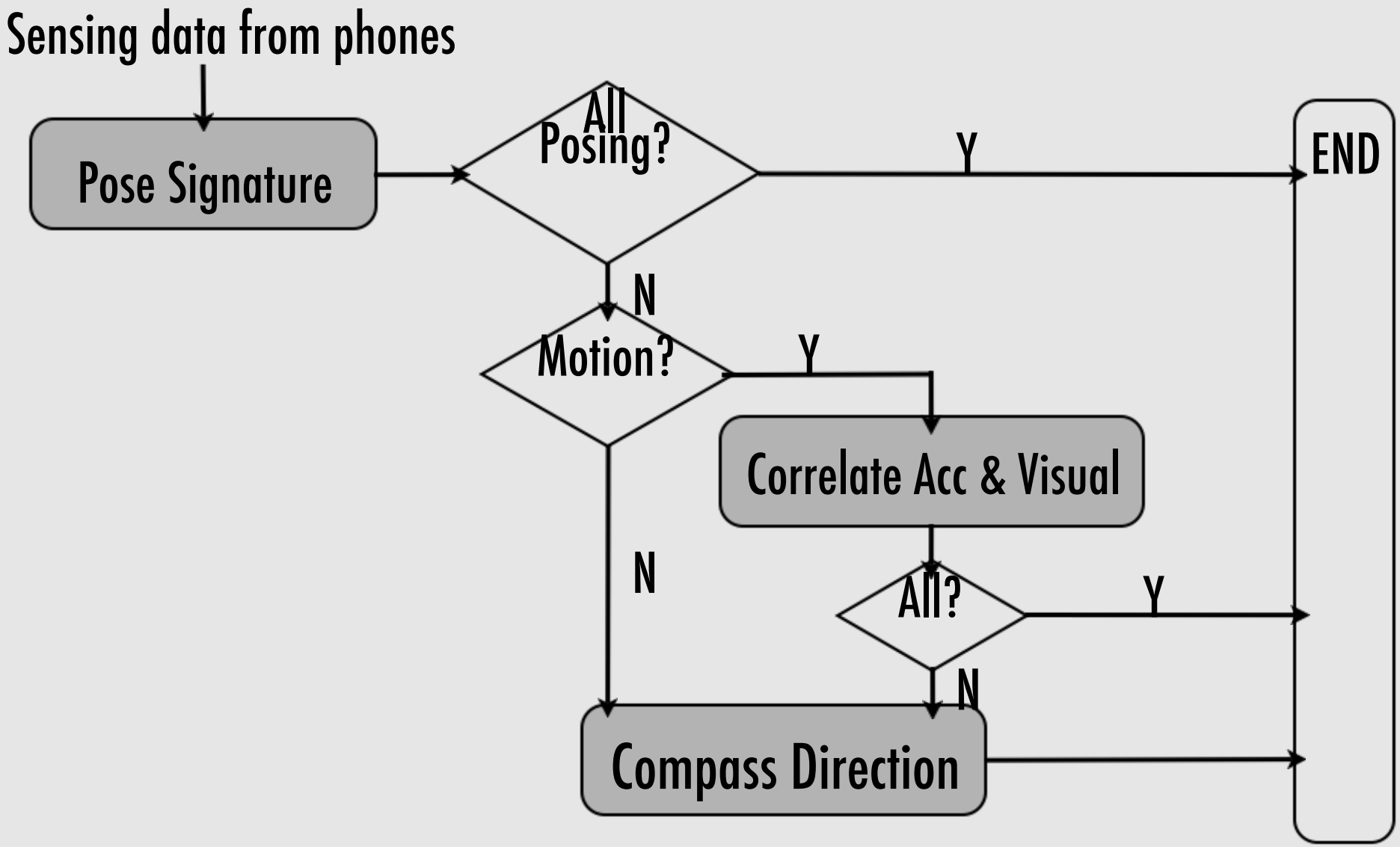
Motion vectors extracted from visual

Correlating visual and acceleration

- Taking several snapshots after shutter click
- Motion vector can be obtained by optical flow from adjacent snapshots
- Correlate with accelerometer readings to find who



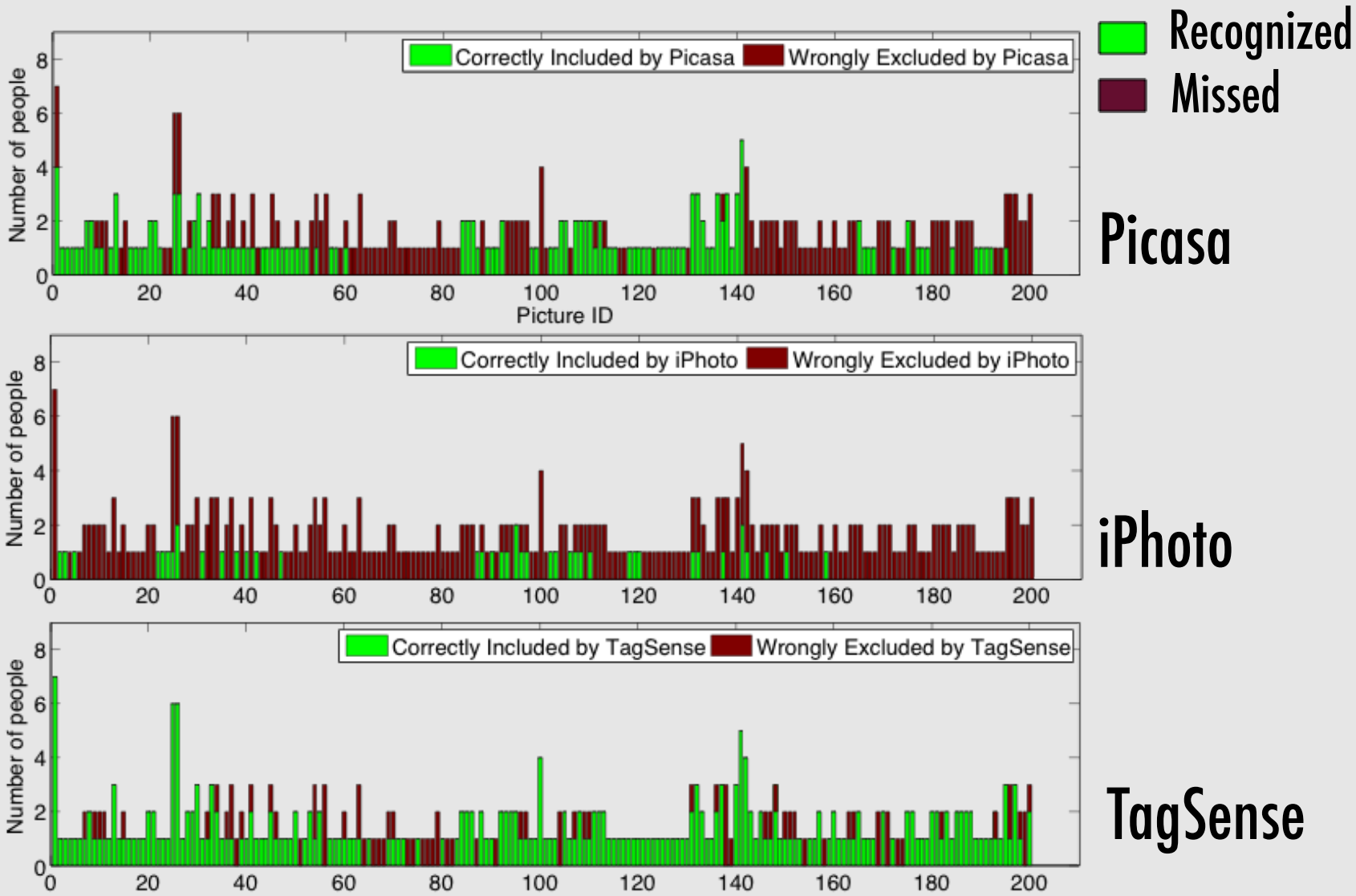
Putting all together



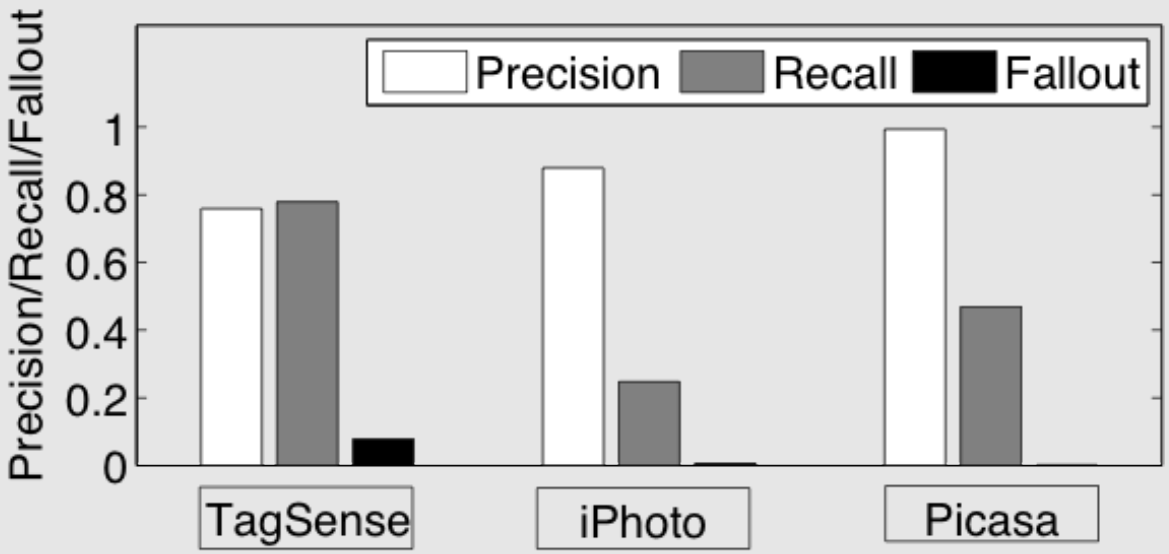
TagSense evaluation

- A prototype on Android Nexus One phones
- Evaluated TagSense with 200+ pictures
- Compare people tagging results with Picasa & iPhoto

Evaluation for tagging people



Evaluation for tagging people



$$\text{Precision} = \frac{|\text{People Inside} \cap \text{Tagged by TagSense}|}{|\text{Tagged by TagSense}|}$$

$$\text{Recall} = \frac{|\text{People Inside} \cap \text{Tagged by TagSense}|}{|\text{People Inside}|}$$

$$\text{Fallout} = \frac{|\text{People Outside} \cap \text{Tagged by TagSense}|}{|\text{People Outside}|}$$

<u>TagSense achieves better Recall with some sacrifice on Precision</u>

TagSense case 1

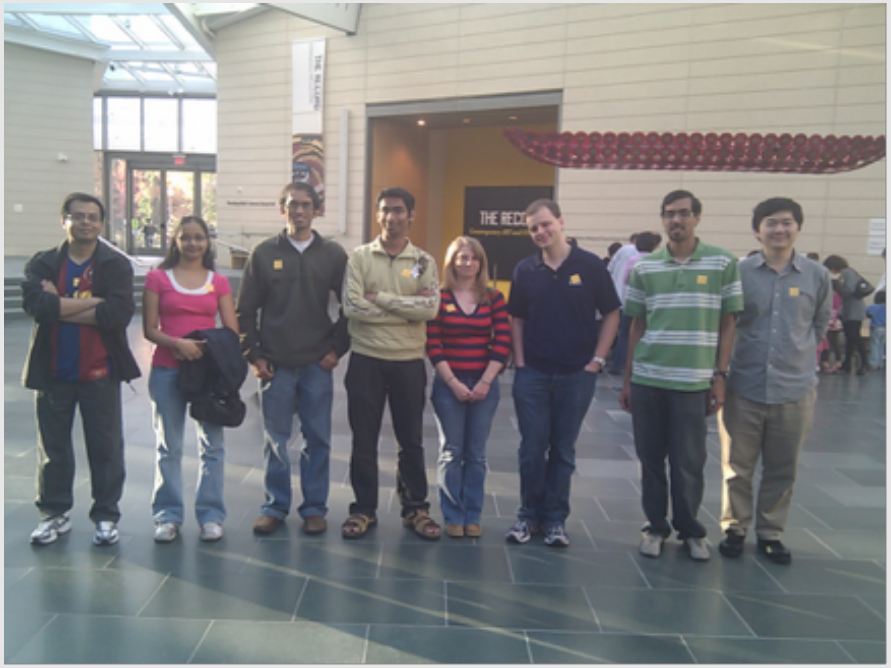
December 4th afternoon,

Hudson Hall, outdoor,

standing, snowing,

Xuan

TagSense case 2



November 21st afternoon,

Nasher Museum, indoor,

Romit, Sushma, Naveen,

Souvik, Justin, Vijay, Xuan,

standing, talking

TagSense case 3

November 21st noon,

Duke Wilson Gym, indoor,

playing, music,

Chuan, Romit

TagSense case 4 (failure case)

November 21st evening,

155 main street, indoor,

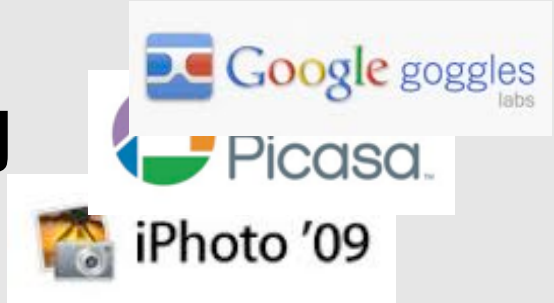
Talking,

(Not recognized)

<u>Image processing and smartphone sensing could be complementary</u>

Related works

- Image processing based auto-tagging



- Semi-auto tagging
  - Cloud based [Naaman JDCL2005] [Sarvas MobiSys2004]
  - Crowd-Sourcing based solution [CrowdSearch MobiSys2010]
- Special hardware
  - ContextCam[Shwetak UbiComp2004]

Limitations of TagSense

- TagSense cannot tag pictures taken in the past
- TagSense can not recognize people not carrying phones
- TagSense's current vocabulary is quite limited

TagSense revision & future work

Thanks & Questions?

- TagSense – emulating human recollection
  - Sensors are similar to human senses
  - Recorded sensing data is similar to human memory

- Future: tag generation is a work in progress
  - Short time scale, multiple sensing dimensions
  - Extensible: Face/Activity recognition techniques also fit in

    New sensors are being added to smartphones
  - A door to many possibilities: video-tagging, augmented reality, ...