

# TagSense: A Smartphone-based Approach to Automatic Image Tagging

Chuan Qin, Xuan Bao, Romit Roy Choudhury, Srihari Nelakuditi

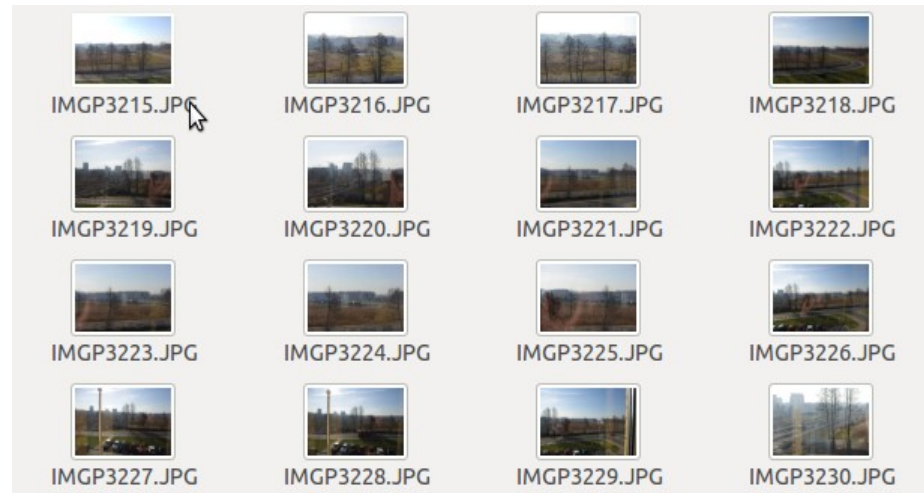
MobiSys 2011

Grzegorz Jabłoński

Distributed Systems course

# Image tagging

- Pictures and videos are undergoing huge changes
- Image retrieval
  - Image search
  - Personal albums
- Tagging videos



# Tagging

- Tags – people, place...
- Now
  - crowdsourcing
  - online gaming
- Computer based tagging
  - Faces
- Notion of tag?

GO

Timer: 00:00:00 of 60 min

Categorize Images

Requester: Tagasauris

Qualifications Required:

Total Earned: \$0.00  
Total HITS Submitted: 0

Duration: 60 minutes



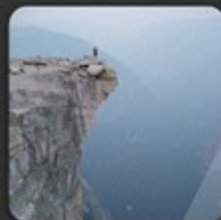
Day in the park



We do



Nature



Hiking Half Dome



Ella's birthday



Portugal



Summer in Australia



Surf photos



Tea time



Island Vacation



Weekend with Vivian



Hawaiian flowers



Snow day



Oregon dunes



Playdate with Grace

Search Zoom

Events Faces Places Albums Projects

Info Edit Create Add To Share

# Examples



- November 21st afternoon, Nasher Museum, indoor, Romit, Sushma, Naveen, Souvik, Justin, Vijay, Xuan, standing, talking
- Many people, smiling, standing

# Examples



- December 4th afternoon, Hudson Hall, outdoor, Xuan, standing, snowing
- One person, standing, snowing

# Examples



- November 21st noon, Duke Wilson Gym, indoor, Chuan, Romit, playing, music
- Two guys, playing, ping pong

# Use smartphones!

Two main advantages:

- Built-in sensors
- People carry their phones everywhere



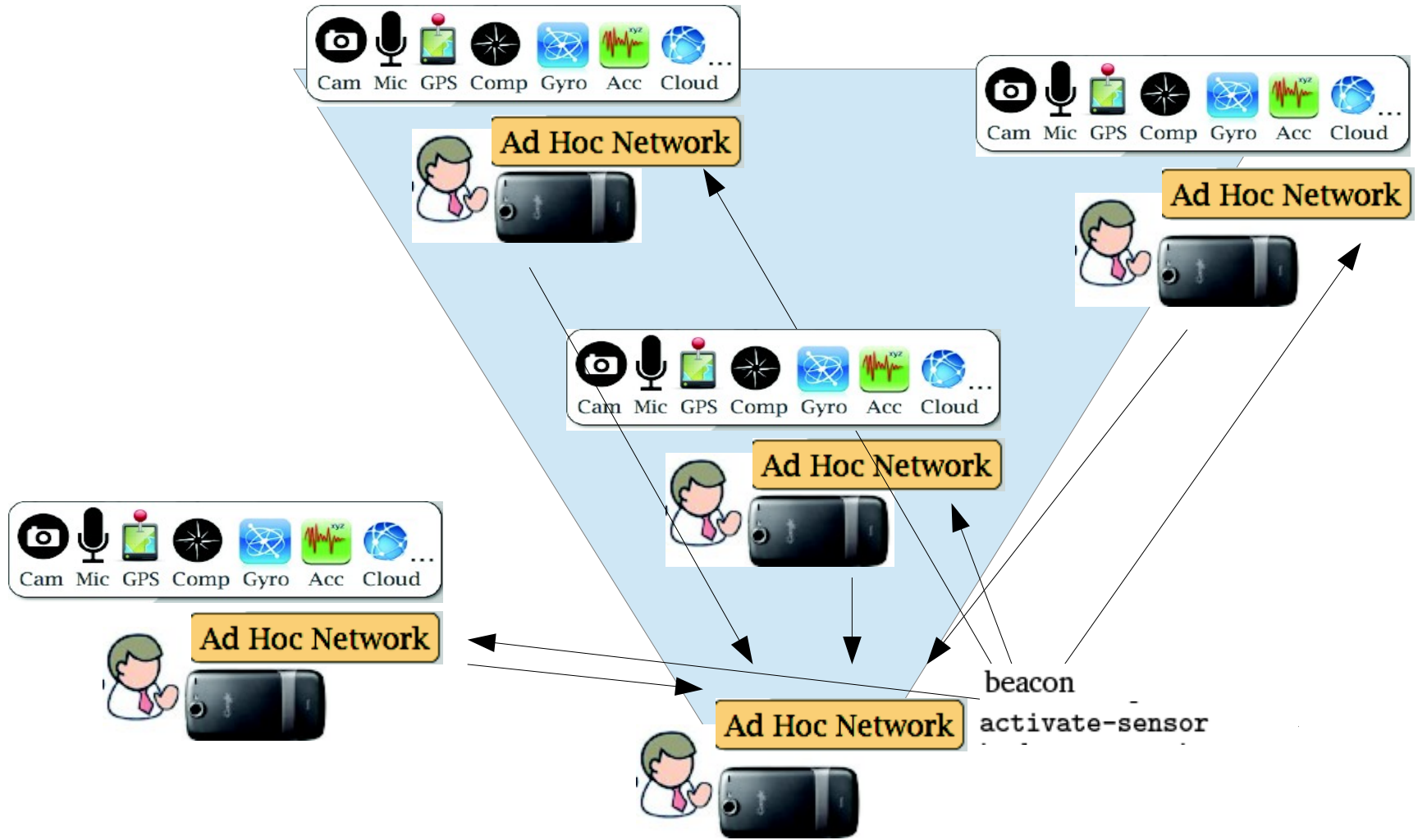
Why is it better?



# TagSense

- Computer based tagging
- Does not depend on faces
- Uses smartphones sensors and features
  - WiFi, accelerometer, compass, light sensor, camera, microphone, GPS, gyroscope
- Challenges
  - Who is in the picture?
  - Data mining
  - Power consumption

# System overview



# when-where-who-what

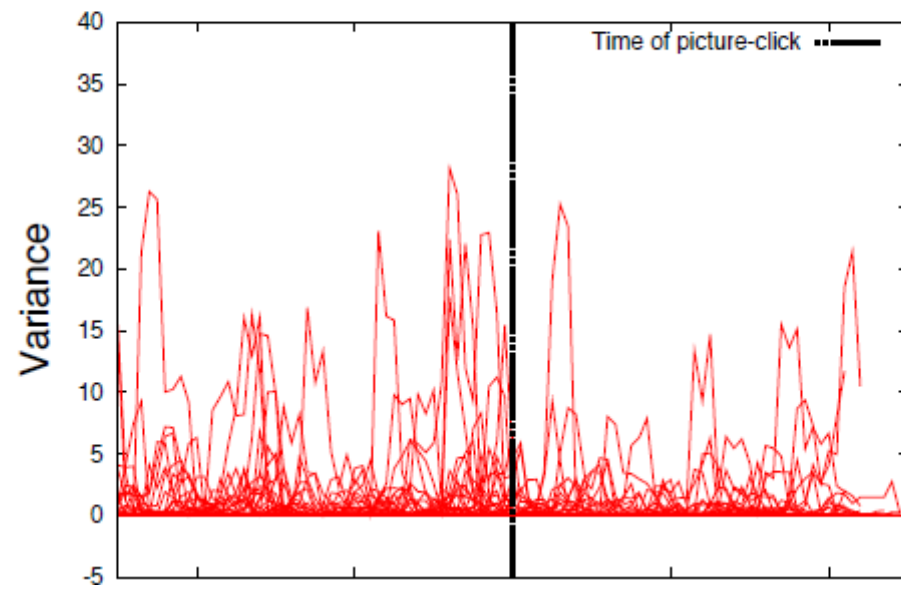
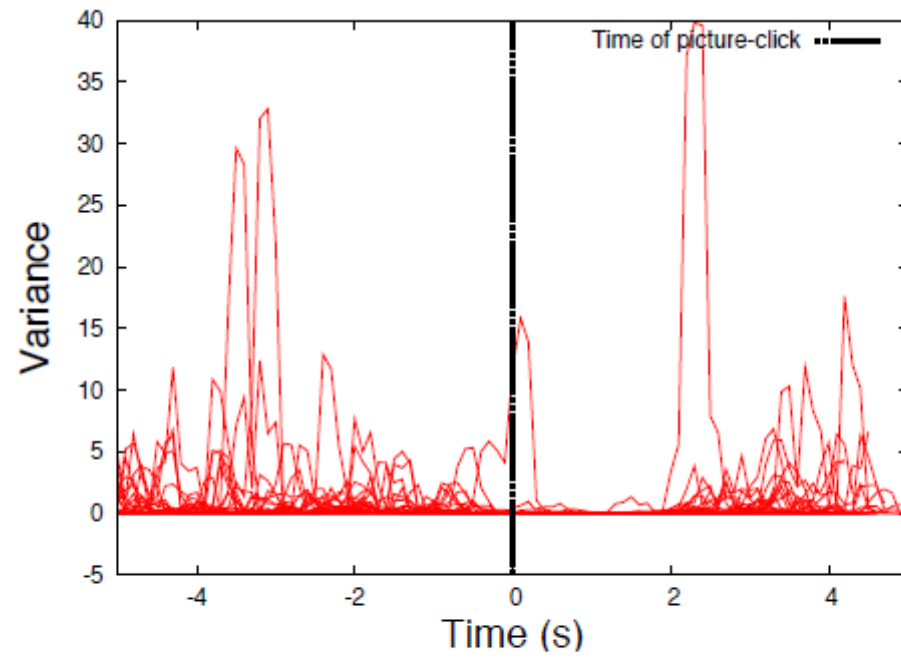
- Format:
  - <time, logical location,  
Name1 <activities for name1>,  
Name2 <activities for name2>, ... >

# Who?

- It is hard to tell who is in the picture
- Omnidirectional antenna is not enough
- Three solutions in TagSense:

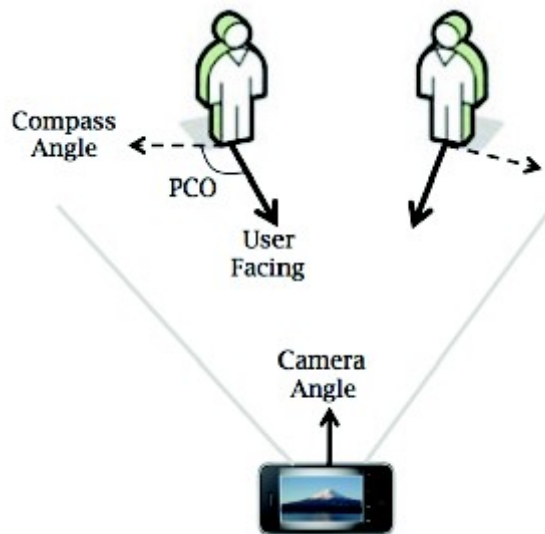
# Who? (1)

- Accelerometer
- How people behave?
- Motion signature



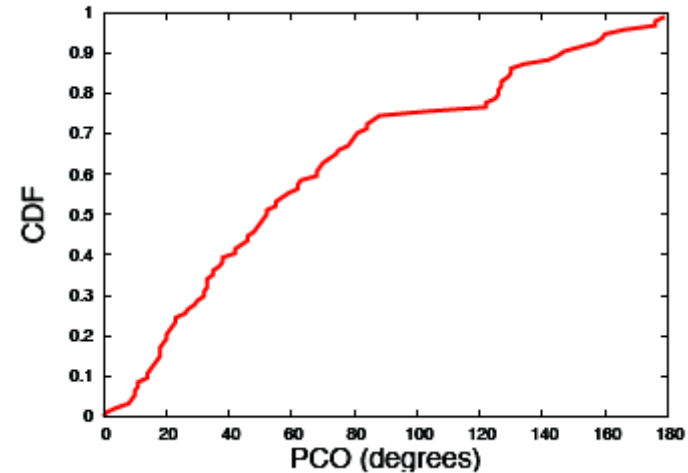
# Who? (2)

- Complementary Compass Directions
  - Signature is not enough
  - TagSense uses compass direction
- direction



# Who? (2)

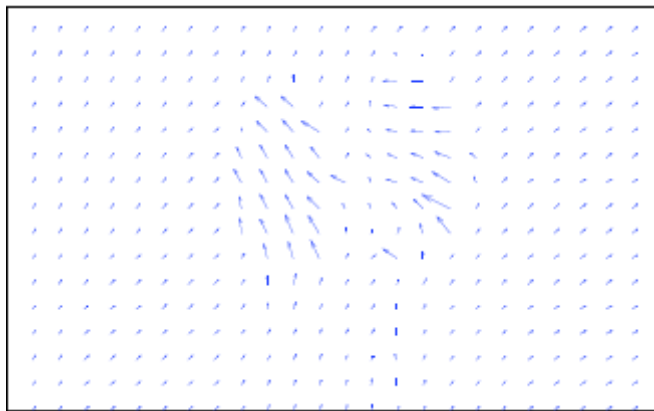
- Still not enough
- Recalibrate  
(whenever it is possible)



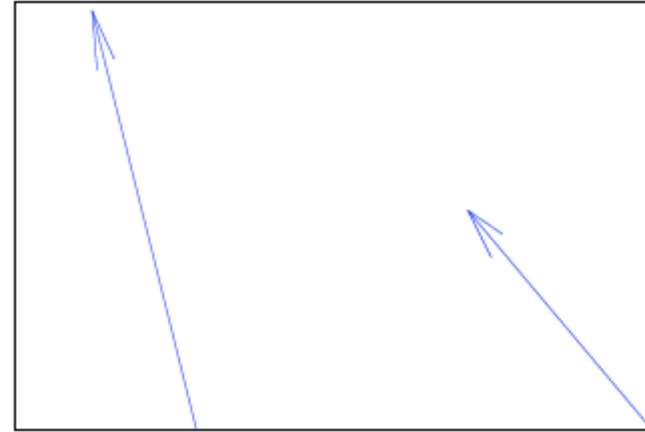


# Who? (3)

- Moving subjects



# Who? (3)



- TagSense matches optical velocity with accelerometer readings
- Use coarse grained properties
- Discussion:
  - No pinpointing
  - No kids
  - Assumes people face the camera

# What?

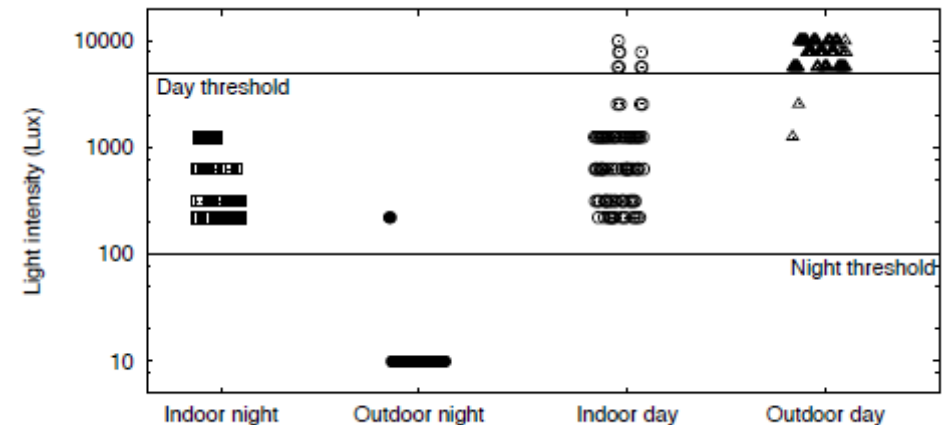
- Accelerometer:
  - Standing, Sitting, Walking, Jumping, Biking, Playing
- Acoustic:
  - Talking, Music, Silence

# Where?

- Reverse lookup on GPS position
- SurrondSense
- Indoor / Outdoor

- Location + phone  
compass is used to

tag picture backgrounds (Enkin, Google API)



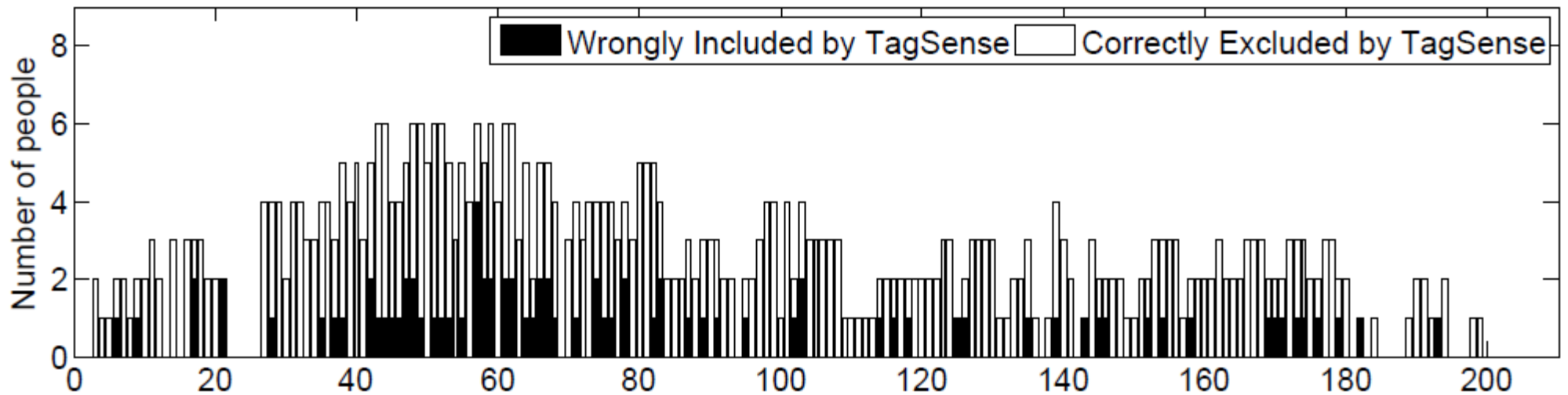
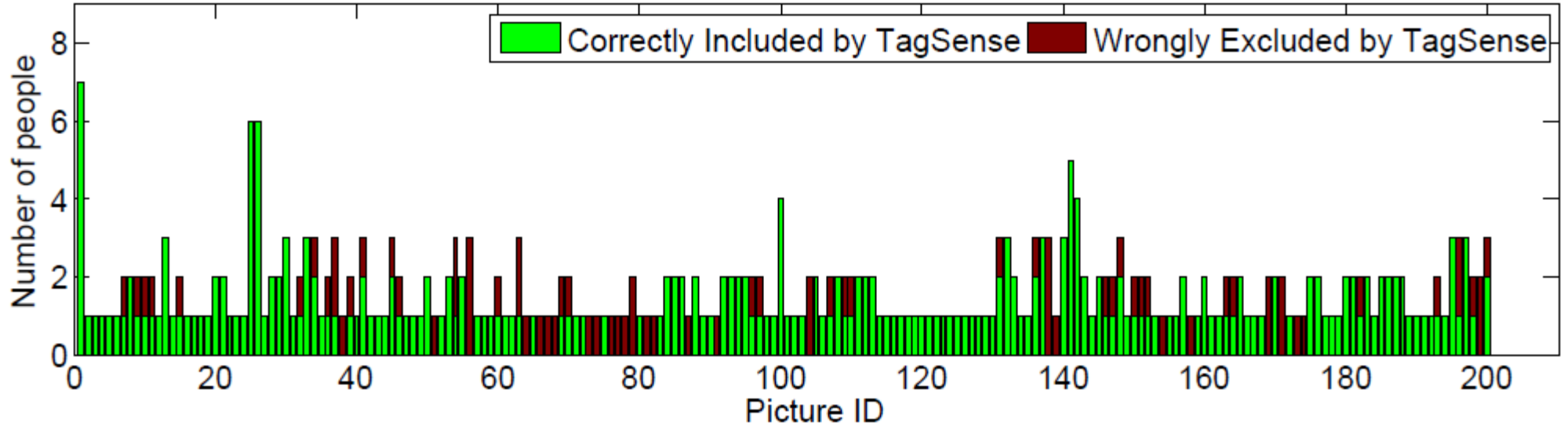
# When?

- Camera current time
- Fetch information from Internet weather service (outdoor only)
- Adds “at-night” tag after sunset

# Performance evaluation

- 8 phones
- Duke University's Wilson Gym
- Nasher Museum of Art
- Research lab in Hudson Hall
- Thanksgiving party

# Tagging people



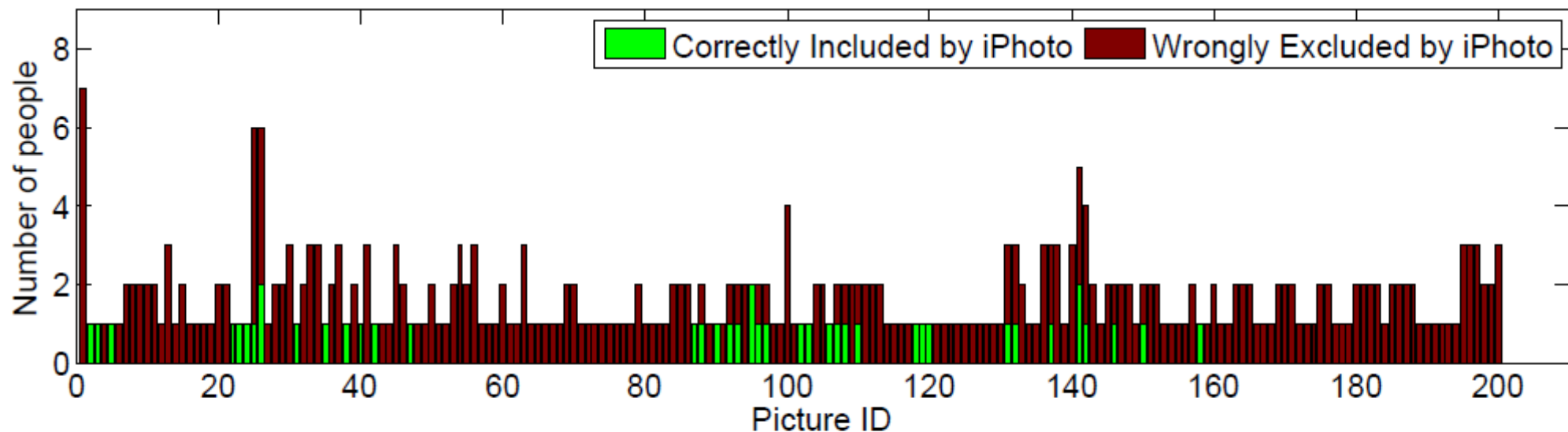
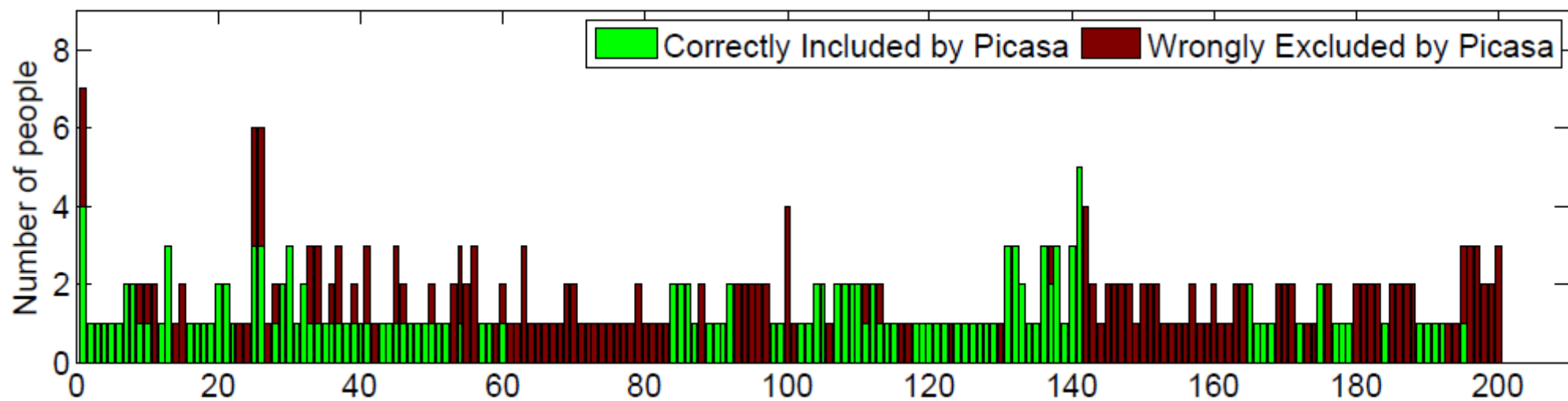


Figure 8: iPhoto wrongly excludes quite a few people. But only a few are wrongly included (graph not shown).





# Evaluation metrics

$$\textit{precision} = \frac{|People\ Inside \cap Tagged\ by\ TagSense|}{|Tagged\ by\ TagSense|}$$

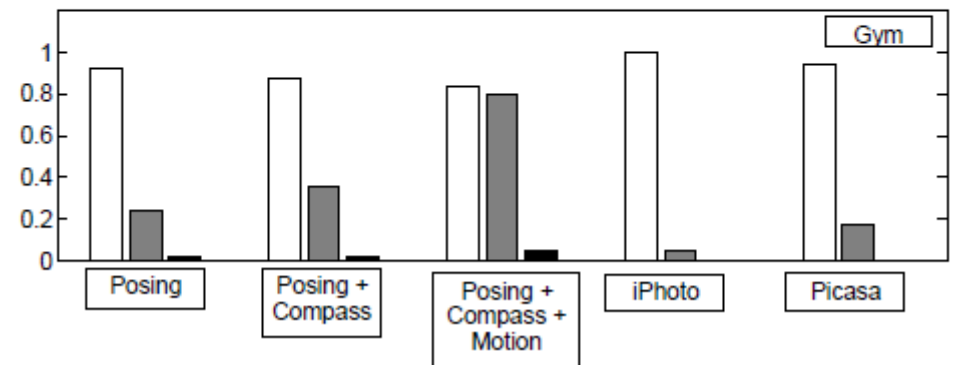
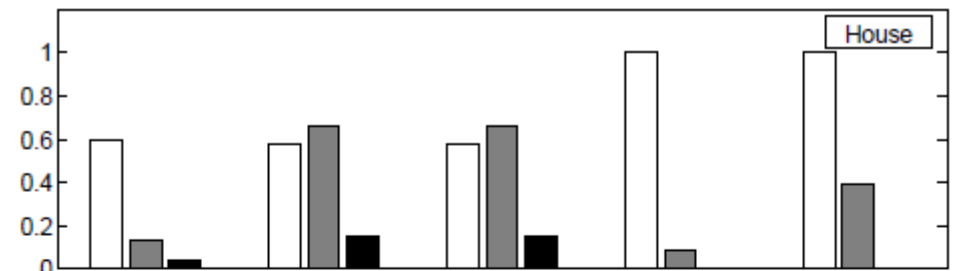
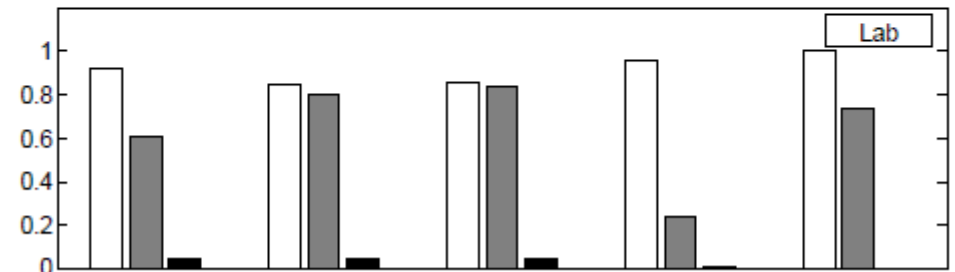
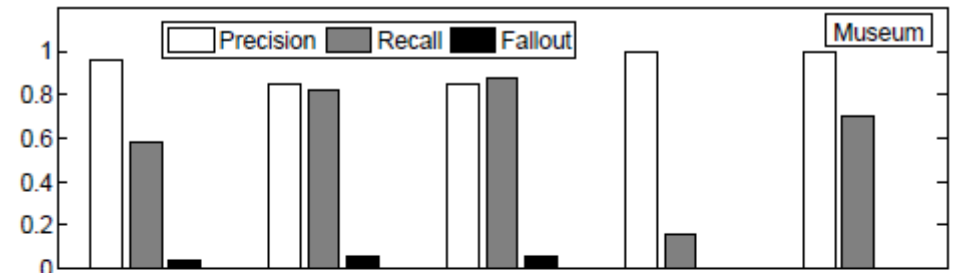
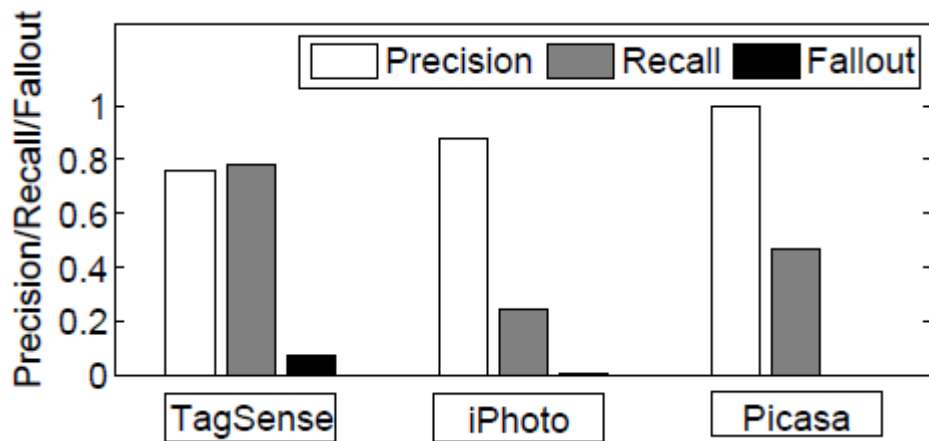
$$\textit{recall} = \frac{|People\ Inside \cap Tagged\ by\ TagSense|}{|People\ Inside|}$$

$$\textit{fall-out} = \frac{|People\ Outside \cap Tagged\ by\ TagSense|}{|People\ Outside|}$$

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

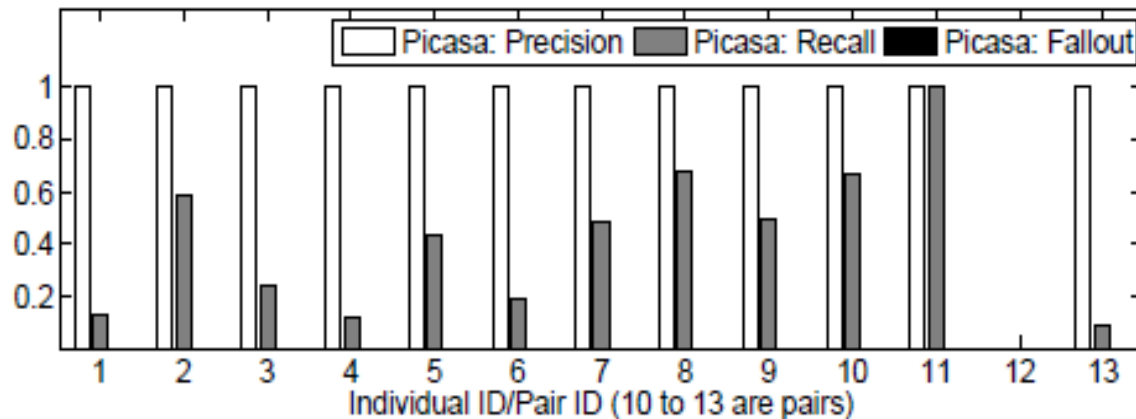
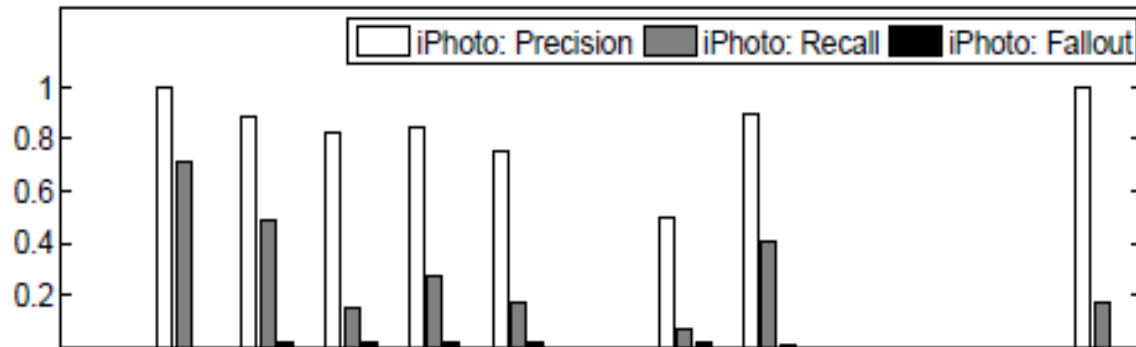
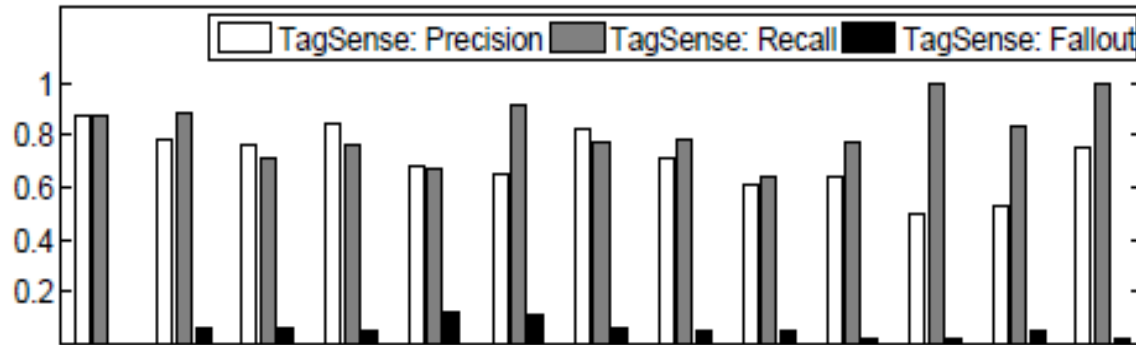
$$\text{recall} = \frac{|People\ Inside \cap Tagged\ by\ TagSense|}{|People\ Inside|}$$

$$\text{fall-out} = \frac{|People\ Outside \cap Tagged\ by\ TagSense|}{|People\ Outside|}$$



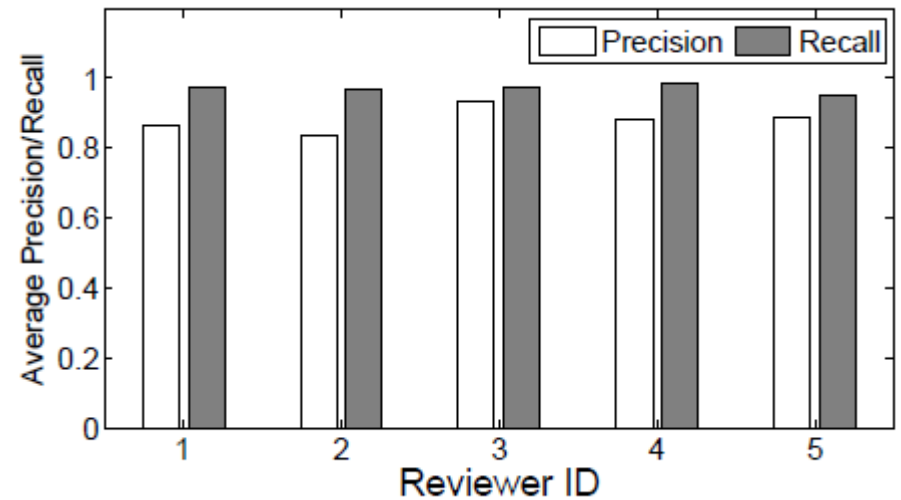
# Name based search

- Merge?



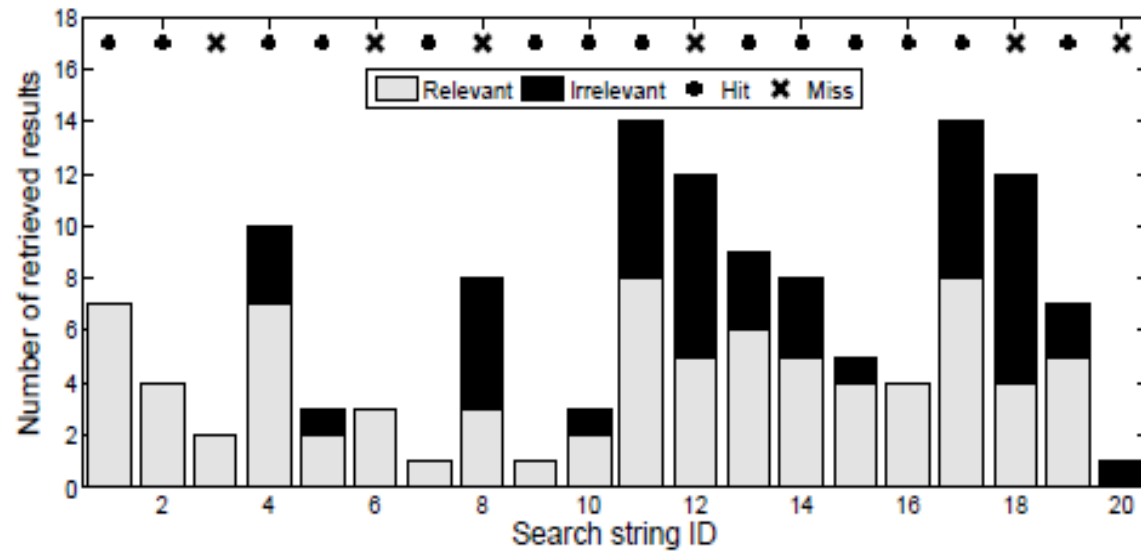
Individual ID/Pair ID (10 to 13 are pairs)

# Tagging Activities and Context



# Tag Based Image Search

- 200 tagged images, 5 volunteers
- 20 random pictures, volunteers asked to retrieve them



# Limitations

- Limited vocabulary
- Do not generate captions
- Cannot tag past pictures
- Requires group password
- Complex methods

# Related work

- Contextual metadata – similar images
- ContextCam (ultrasound receivers and emitters)
- SenseCam(change in light, body heat)
- SoundSense
- Activity recognition
- Image processing – Google Goggles

# Future

- Activity / context recognition
- Directional antennas
- Granularity of localization
- Smartphones replace cameras



Questions?