TagSense

A Smartphone-based Approach to Automatic Image Tagging

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

Presentation by Philippe von Bergen

Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

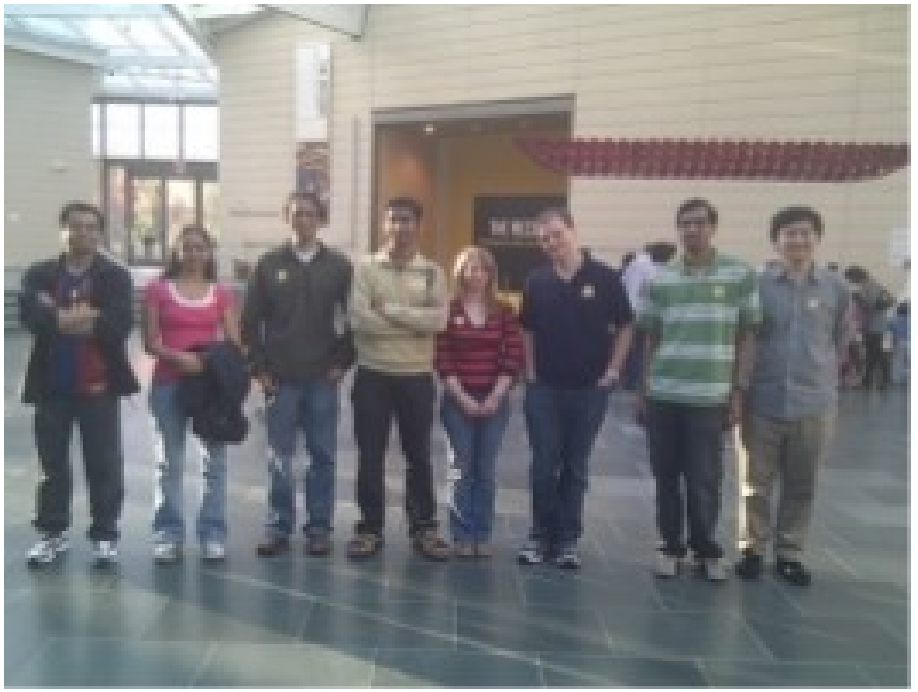
Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Introduction

- Automatic image tagging system
- Distributed
- Collaborative
- Multi-dimensional
- Who – What – Where – When

Problem Space (1/3)



- Date, time
- Location
- People present
- Action
- Ambience
- Tags:

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

Problem Space (2/3)

- Location: Location services, reverse look up
- Indoor/Outdoor: Light sensor
- Names: Present phones
- Action: Accelerometer
- Ambience: Microphone
- Date, time: Internal clock

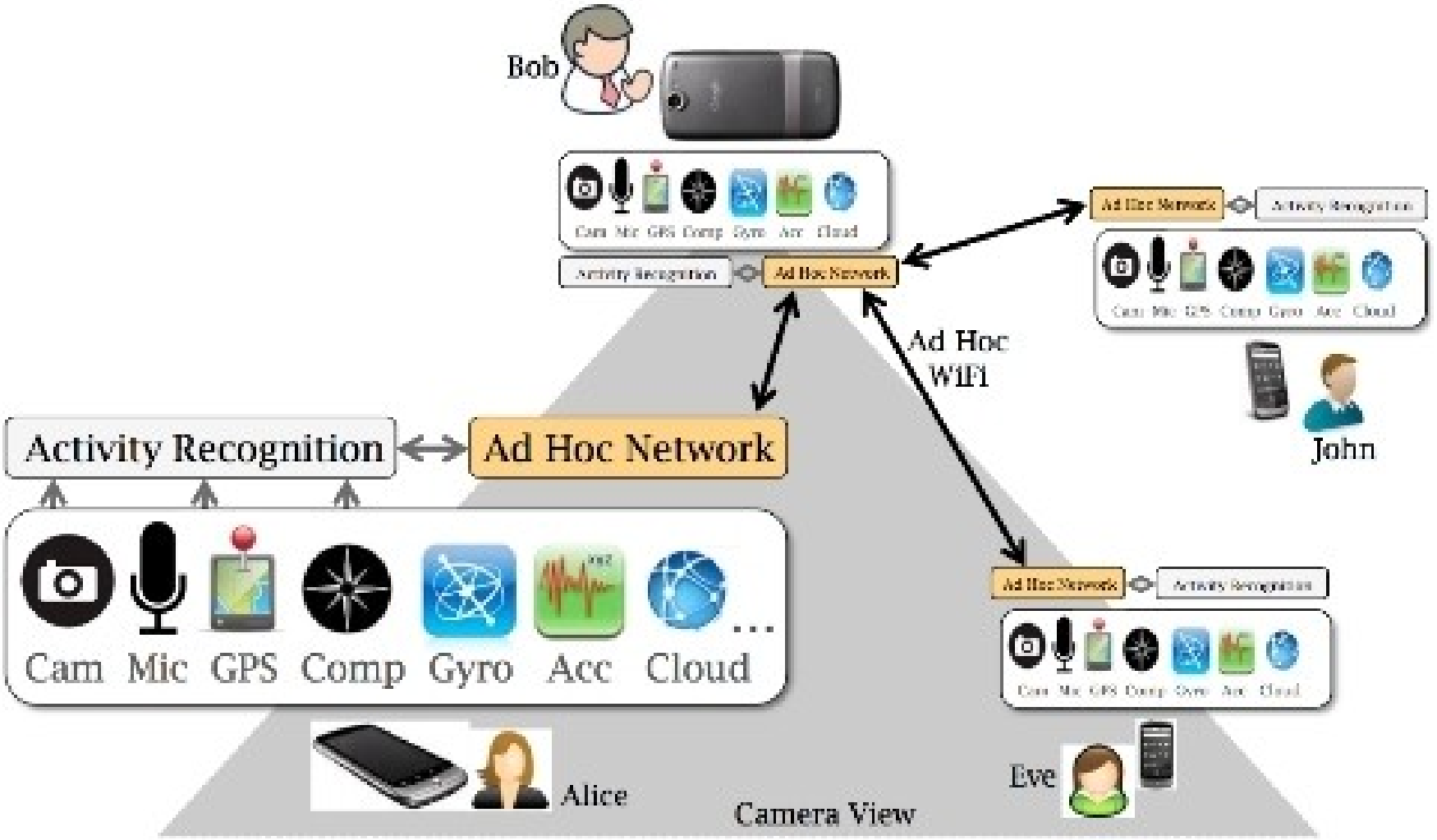
Problem Space (3/3)

- Humanly assigned tags complementary
- Complementary to existing solutions

  (Picasa, iPhoto, Google Goggles)
- TagSense needs electronic foot print

  → Not applicable to objects and subjects without devices

System (1/2) – Overview



System (2/2) – Cloud, Privacy

- Local recognition of tags, if possible
- Cloud service
  - CPU-intensive calculations
    (Laughter recognition)
  - External databases
    (Reverse GPS)
- Password and encryption to ensure privacy

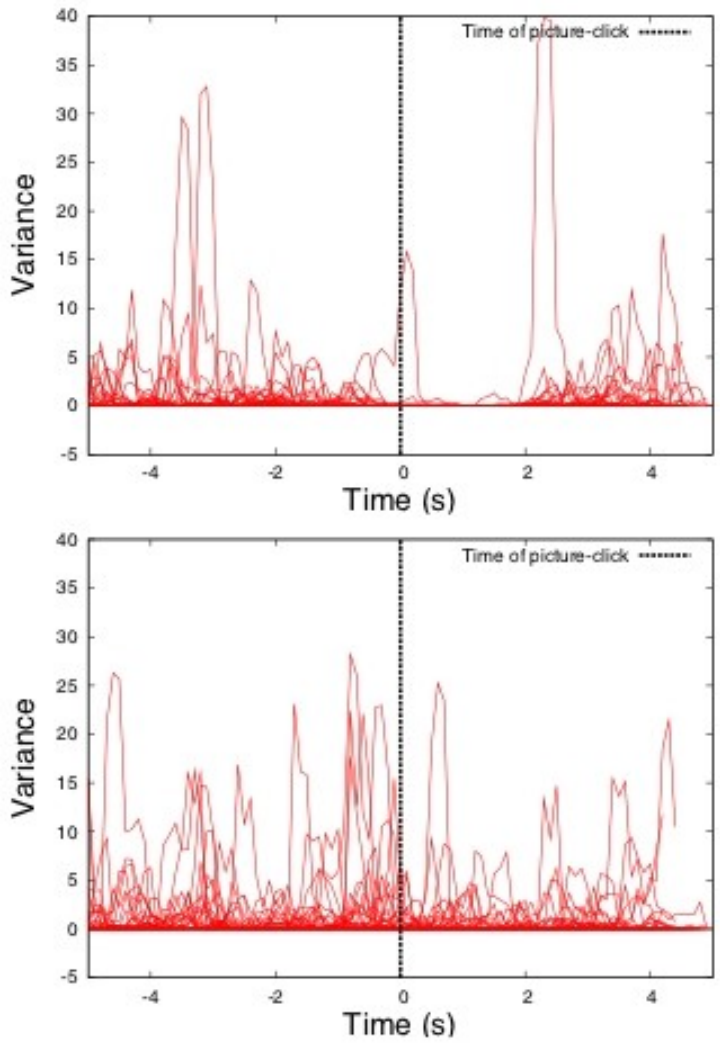
Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Who (1/6) – Overview

- Several systems
  - Accelerometer → Posing signature
  - Compass heading → Personal compass offset
  - Movement → Multiple snapshots and heuristics
- Serial application of all three techniques

Who (2/6) - Posing signatures



- Detection via accelerometer
- Several seconds of posing
- Used to detect people outside picture

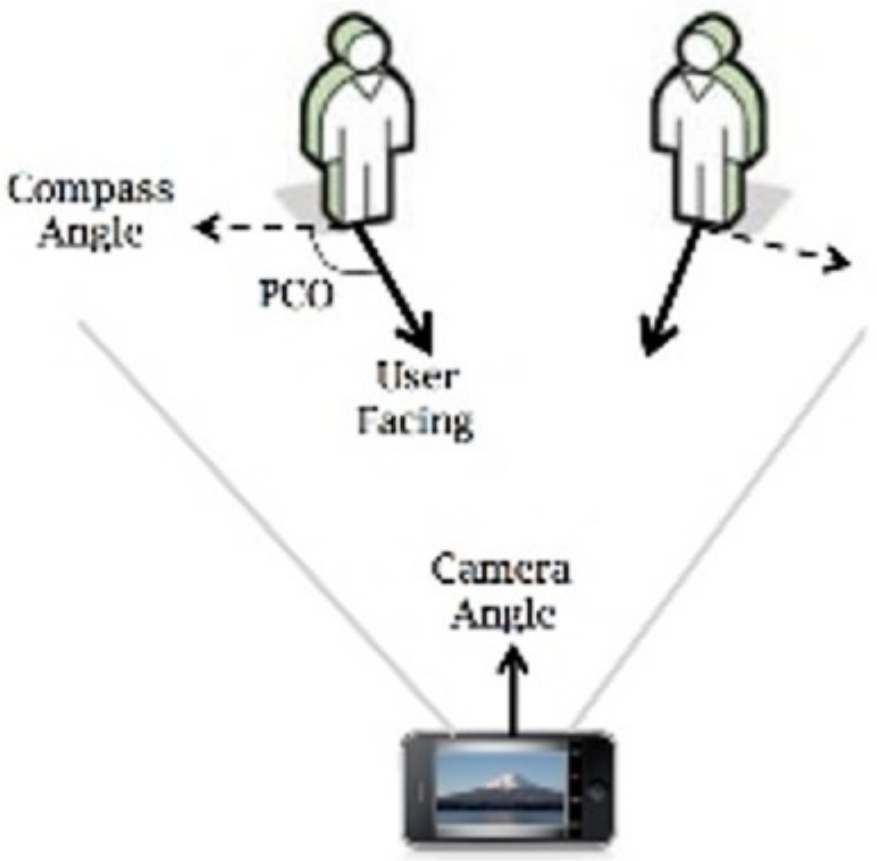
Who (3/6) – Compass direction

- Posing signature sufficient, but not necessary
- Using compass direction to detect people ”facing” the camera
- Personal compass offset compensating phone heading:

$$UserFacing = (CameraAngle + 180) \bmod 360$$

$$PCO = ((UserFacing + 360) - CompassAngle) \bmod 360$$

Who (4/6) – Compass direction

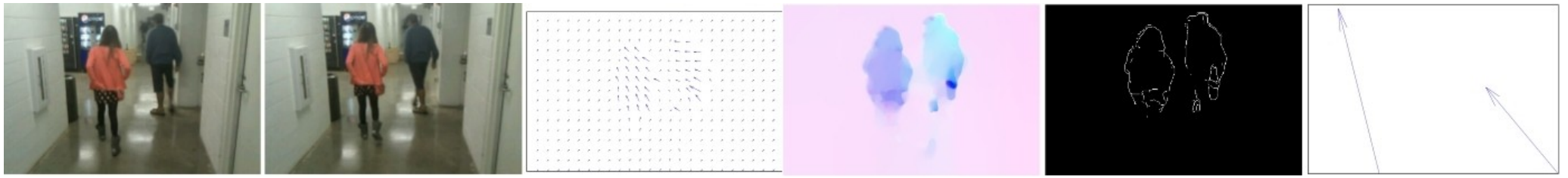


- Recalibrated using posing signature and anchor pictures
- Revisit pictures containing possible errors
- Partly offline

Who (5/6) – Moving subjects

- Multi-dimensional sensing heuristic
- Multiple snaphshots to detect pixel movement
- Compared to accelerometer data
- Coarse bucket matching

Who (6/6) – Moving subjects



1. Optical Flow
2. Camera movement substracted
3. Movement coloring
4. Edge finding
5. Motion vector based on center pixels
6. Coarse bucketing

What

- Distinct physical activities
  - Accelerometer data
  - Location information
- Ambience classified using microphone
  - Talking
  - Music
  - Silence

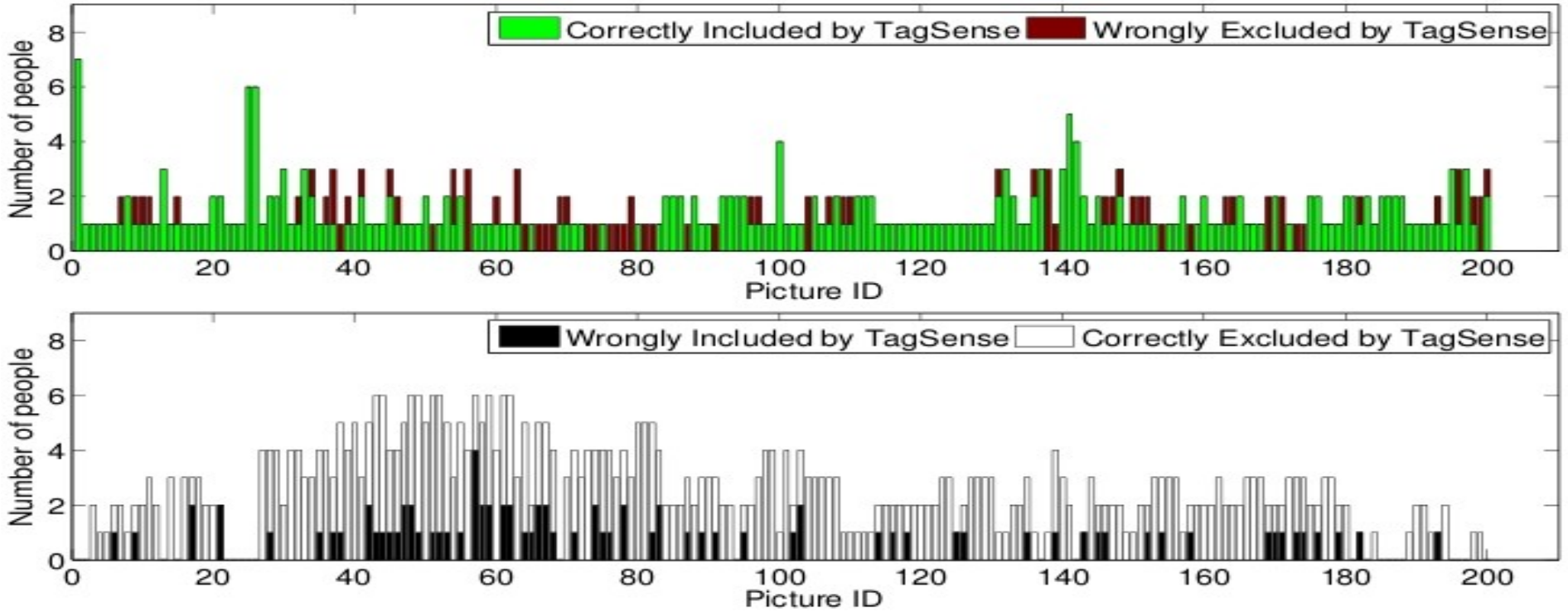
Where / When

- Location based on reverse GPS look up
- Indoor/Outdoor recognition using ambient light sensor
- Objects near or in picture detected using location/orientation database
- Date and time enriched with weather conditions

Overview

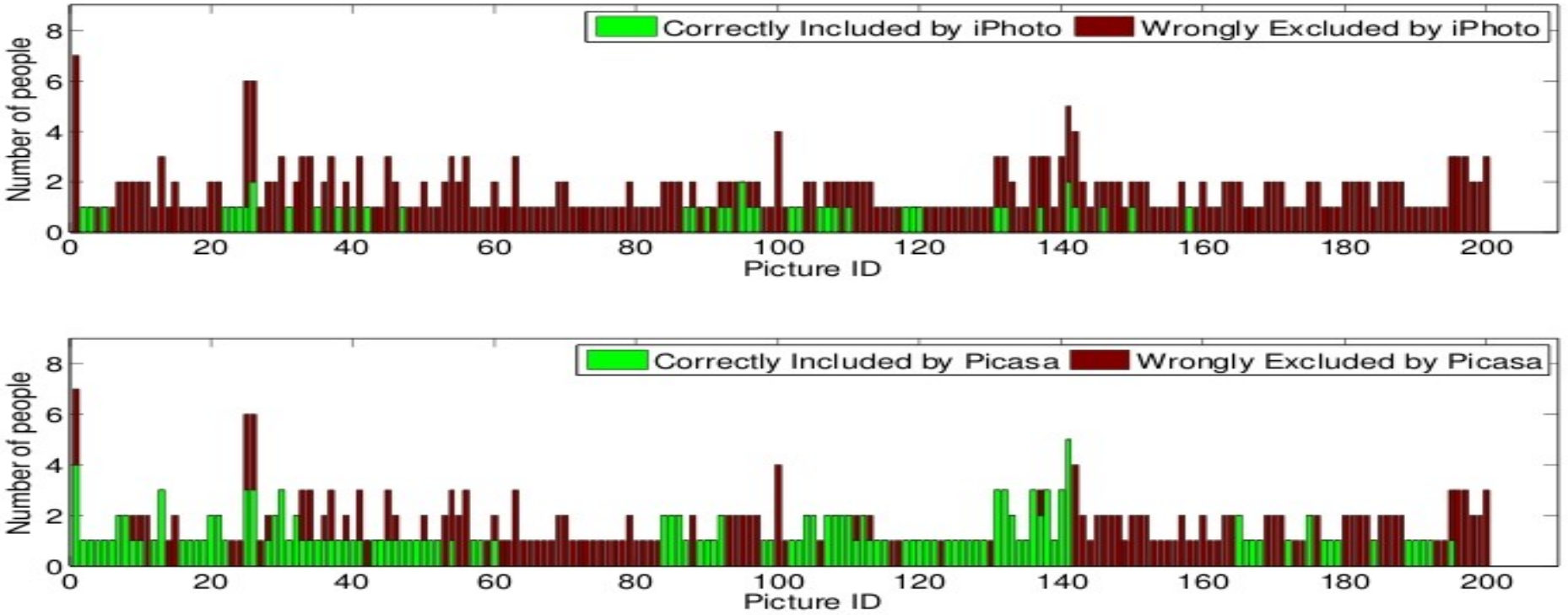
- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Evaluation (1/6) – Detection



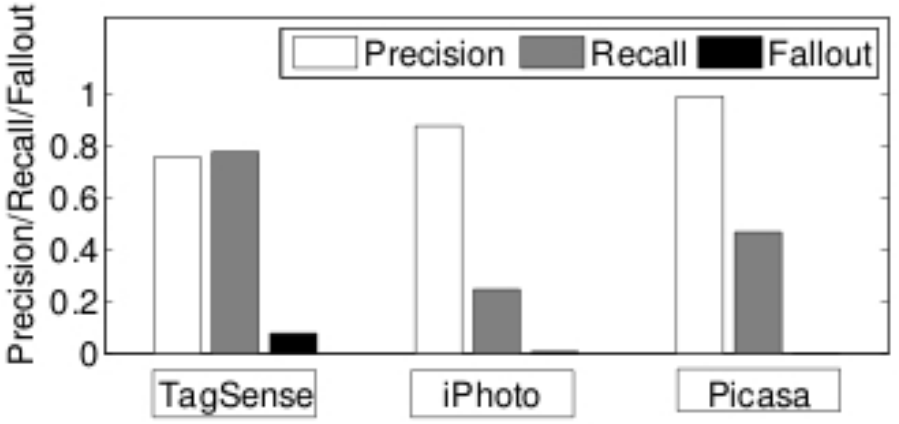
- Reasonably well performance on test set
- Precision needs to be increased

Evaluation (2/6) – Detection

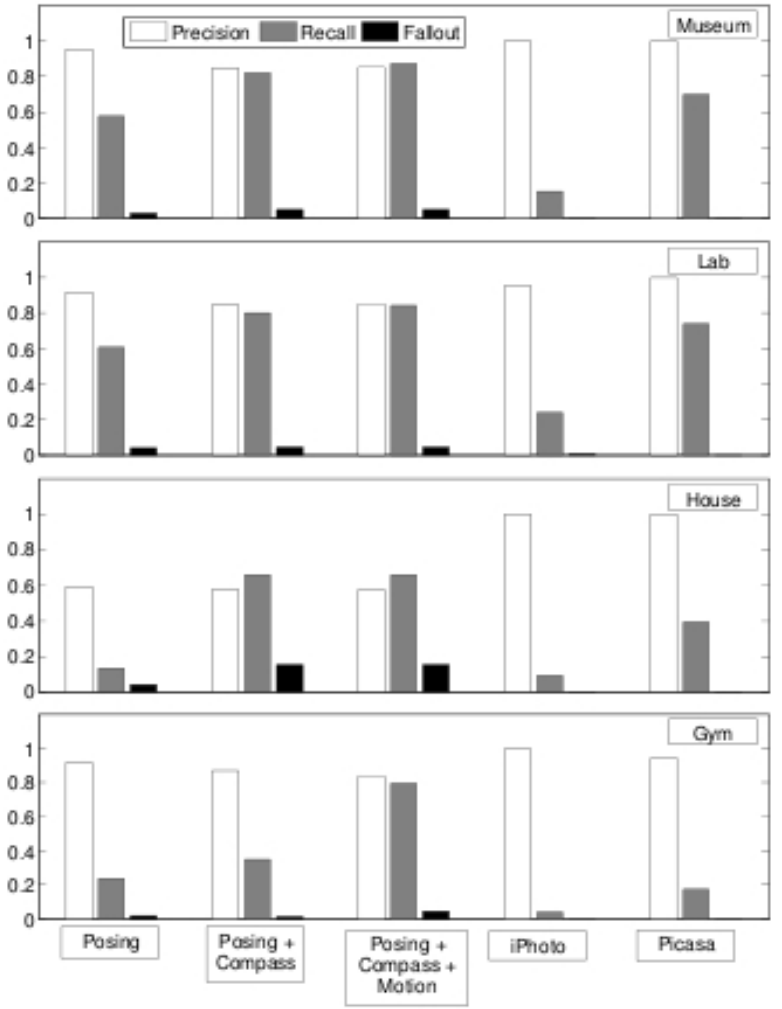


- More false negatives based on one training face
- More training might improve results

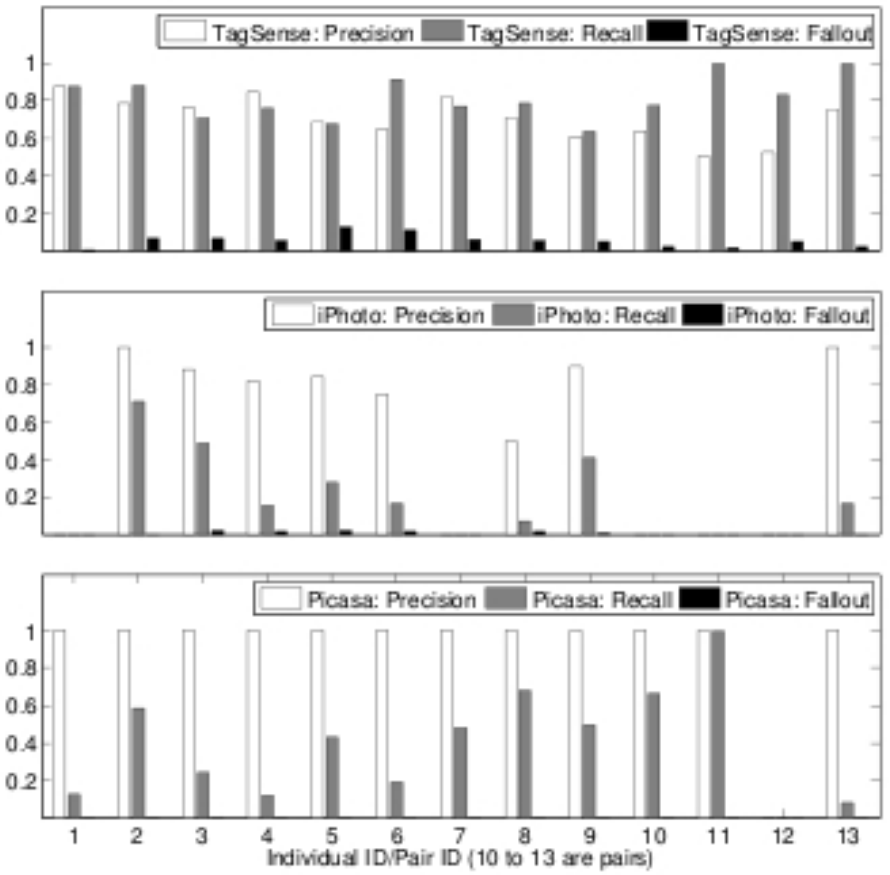
Evaluation (3/6) – Overall



- Better recall
- Increased fall-out
- Reduced precision

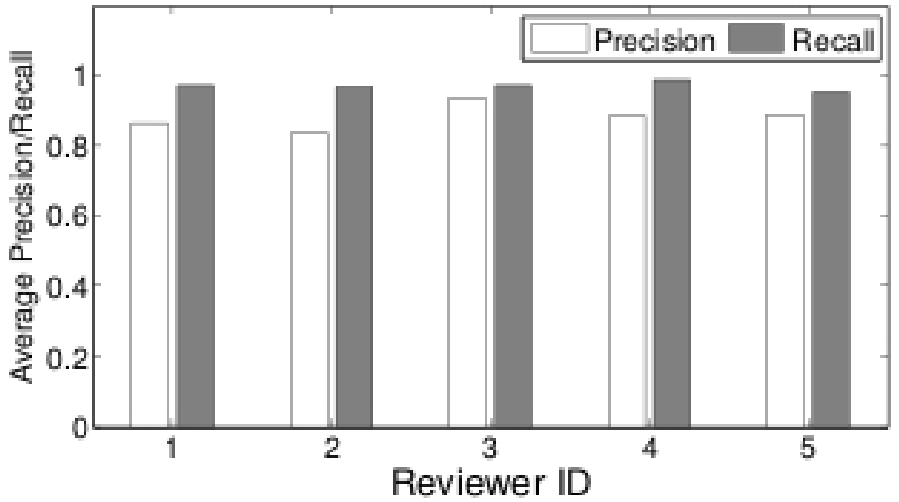


Evaluation (4/6) – Name search



- More consistent
- Better recall
- Improvements possible
- Combination might further increase precision

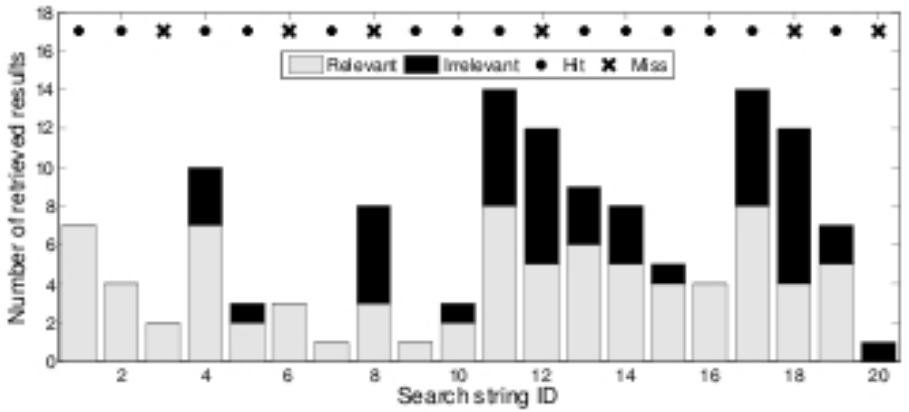
Evaluation (5/6) – Tagging



- Limited to TagSense vocabulary
- Precision and recall high
- Improvement needed for vocabulary

Evaluation (6/6) – Tag search

Name	Avg. Relevant	Avg. Irrelevant	Hit rate
User 1	2.75	4.85	0.85
User 2	5.6	1.8	0.65
User 3	4.05	2	0.5
User 4	4.05	2.35	0.7
User 5	2.55	1.6	0.55



- Volunteers search previously shown pictures
- Sufficient precision, depending on user and picture
- Average hit rate: 0.7

Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Limitations

- Limited vocabulary
- No captions can be generated
- Past pictures can not be tagged
- Cumbersome session management
- Complex system for people detection

Conclusion

- Leverages automatic tagging of picture
- Prototype implementation
- Evaluation shows lower precision, but higher recall and fall-out
- Complementary approach might produce best results

Contribution

- New, alternative, multi-dimensional approach to automatic image tagging
- TagSense architecture
- Evaluation of TagSense

Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Review (1/3) – Numbers

- Overall rating
  - Average: 1.9 (Accept)
  - Standard deviation: 0.7
- Confidence:
  - Average: 2.1 (Medium)
  - Standard deviation: 0.3
- Contribution:
  - Average: 3.9 (Strong)
  - Standard deviation: 0.54

Review (2/3) – Compliments

- PoC comparison to Picasa, iPhoto
- Invisible content captured
- Fair assumptions
- Off the shelf hardware used
- Privacy addressed and implemented
- Fallback methods for person recognition
- Aware of limitations
- Good presentation

Review (3/3) – Critiques

- Extended evaluation required
- Not applicable to old pictures
- Simplified assumptions
- All participants need the application
- Complicated session management
- CPU and power consumption not addressed
- Manual editing missing
- Redundancy

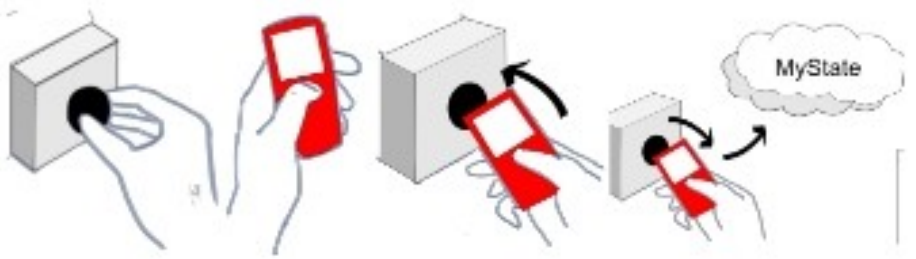
Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

MyState

Hardy, Rukzio, Holleis, Wagner
Lancaster University, University of Duisburg-Essen, DOCOMO Euro-Labs

- Physical tags containing textual information
- Placed at arbitrary location
- Posting content to Facebook Application



MyState & TagSense

- Reading RFID-Tags to determine
  - Objects
  - Locations
  - Additional tagging information, context
  - People not having a phone
- Limited to short ranged

What did you do today?

Farrahi, Gatica-Perez
EPFL

- Routines and behavior detection
- Based on GSM information
- Latent Dirichlet Allocation and Author Topic model

Wdydt? & TagSense

- Increase presence detection precision using a local cell tower representation
  - Use smart phones to triangulate positions
- Improved approach for moving subjects identification and activities

Overview

- Tagsense
  - Introduction, Problem Space, System
  - Who, What, Where, When
  - Evaluation
  - Limitations, Conclusion, Contributions
- Reviews
- MyState & What did you do today?
- Summary & Discussion

Summary

- TagSense uses distributed systems approach
- Sensors enrichen picture with tags
- Problem driven idea
- Use of behavior inference for tags
- Results could be used to determine behavior
- Performance improvement by combination with face recognition
- Limitations to overcome

Discussion

- Did they took realistic assumptions?
- How to combine TagSense with face recognition?
  - Or increase precision otherwise?
- Is machine learning able to increase precision?
- How to integrate TagSense with social media?
- How to simplify session management / privacy?
- What additional information could be tagged?