

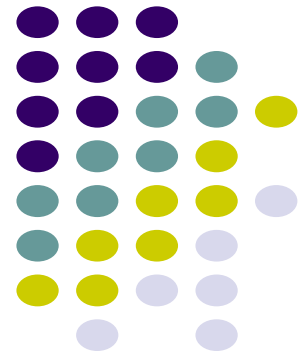
Ubiquitous and Mobile Computing

CS 403x: *TagSense*

Ying Lu

Computer Science Dept.

Worcester Polytechnic Institute (WPI)



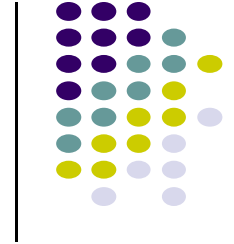
Background



Under the explosion of videos and pictures, users are expecting to search and browse pictures as they do for textual content.

Create Tags!!!!

Problem



How to generate tags for pictures automatically?

Significance



- Search
- Tag video -- user able to move video slider to specific location

Related Work



Photo tags:

Apple iPhoto, Google Picasa

Location recognition:

Mobile Media Metadata (MMM), ContextCam

Mobile sensing:

SenseCam, SoundSense and SurroundSense

TagSense



- Sensor-assisted
- Automate
- “When-where-who-what”
- Describe an on-going occasion in multi-dimensions



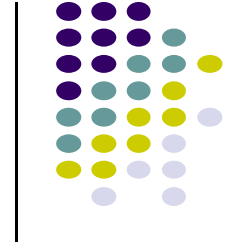
Figure 1: Three example pictures. TagSense tags each picture with the time, location, individual-name, and basic activity. Face recognition by iPhoto and Picasa can tag people in the left picture, less so in the middle, and not so in the right picture. An arbitrary human (emulating crowd-sourcing) is able to label with semantically rich tags, however cannot name individuals. Google Goggles, relying on image processing, offers poor tags.

Picture 1: November 21st afternoon, Nasher Museum, indoor, Romit, Sushma, Naveen, Souvik, Justin, Vijay, Xuan, standing, talking.

Picture 2: December 4th afternoon, Hudson Hall, outdoor, Xuan, standing, snowing.

Picture 3: November 21st noon, Duke Wilson Gym, indoor, Chuan, Romit, playing, music.

Methodology



WHO

- Accelerometer based motion signatures
- Complementary compass directions
- Moving Subjects: optical motion vector

WHAT

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

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

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



Methodology

WHERE

- GPS
- light sensor (indoor/outdoor)

WHEN

- time
- weather



Figure 1: Three example pictures. TagSense tags each picture with the time, location, individual-name, and basic activity. Face recognition by iPhoto and Picasa can tag people in the left picture, less so in the middle, and not so in the right picture. An arbitrary human (emulating crowd-sourcing) is able to label with semantically rich tags, however cannot name individuals. Google Goggles, relying on image processing, offers poor tags.

Picture 1: November 21st afternoon, Nasher Museum, indoor, Romit, Sushma, Naveen, Souvik, Justin, Vijay, Xuan, standing, talking.

Picture 2: December 4th afternoon, Hudson Hall, outdoor, Xuan, standing, snowing.

Picture 3: November 21st noon, Duke Wilson Gym, indoor, Chuan, Romit, playing, music.



Limitations and tradeoffs

- Power for sensors
- Tag vocabulary limit
- Does not generate captions
- Only works for taking pictures
- Requires extra steps



Results

Tagging

Correctly include: TagSense, Picasa > iPhoto

Correctly exclude: TagSense > iPhoto > Picasa

Overall:

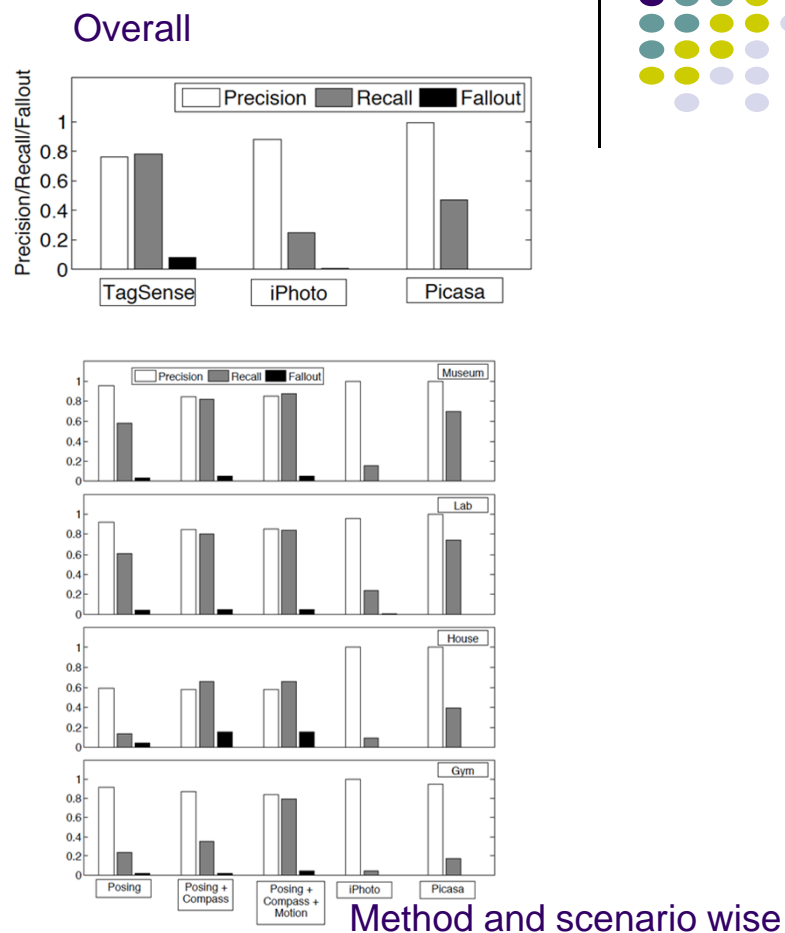
Precision: Picasa > iPhoto > TagSense

Recall: TagSense > Picasa > iPhoto

Method and scenario wise:

Recall: TagSense > Picasa > iPhoto

Precision: Picasa ≈ iPhoto > TagSense





Conclusions

Automatic tag generating by fetching sensor data at capture

Future work:

Utilize more sensors (antennas, granularity of localization, etc)

References

Qin, C et al. (2011). TagSense: A smartphone-based approach to automatic image tagging. MobiSys.

