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

Smartphones
What happens at picturing time

Chuan starts camera application
TagSense architecture

[Diagram showing the TagSense architecture with various components labeled: Camera View, Sensing data, WiFi Ad-hoc, Processed to generate tags, Chuan, Cameraman, John, Srihari, Bob, Peter, and Chuan.]
**TagSense tag generation**

- **When?**
  - Clock + WiFi
  - = May 4th afternoon

- **Where?**
  - GPS + WiFi + 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
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

Sensing data from phones

Pose Signature

Posing? All?

Motion? All?

Correlate Acc & Visual

Compass Direction

END
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

Recognized
Missed

Picasa

iPhoto

TagSense

Correctly Included by Picasa
Wrongly Excluded by Picasa

Correctly Included by iPhoto
Wrongly Excluded by iPhoto

Correctly Included by TagSense
Wrongly Excluded by TagSense

Number of people
Picture ID
Evaluation for tagging people

TagSense achieves better Recall with some sacrifice on Precision

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

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

\[
\text{Fallout} = \frac{|\text{People Outside} \cap \text{Tagged by TagSense}|}{|\text{People Outside}|}
\]
TagSense case I

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)

Image processing and smartphone sensing could be complementary
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 cannot 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, ...