Sharing Multimedia on the Internet and the Impact for Online Privacy

Dr. Gerald Friedland
Senior Research Scientist
International Computer Science Institute
Berkeley, CA
friedland@icsi.berkeley.edu

Question
On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist)?

- a) Never
- b) About once a month or less
- c) About once a week
- d) About once a day
- e) More than once a day

A Popular Introduction to the Problem

Question
How would you judge the issue raised by Colbert?

- a) It's a comedy. I don't worry about any of this.
- b) There is some truth to it but its mostly exaggerated.
- c) It’s a comedy depiction of the reality but most of the stuff is becoming an issue.
- d) He only touched a small part of the problem. The actual issues are even more serious.

Our Observations

- Many Internet sites and mobile apps encourage sharing of data too easily and users follow.
- Users and engineers often unaware of (hidden) search and retrieval possibilities of shared data.
- Local privacy protection ineffective against inference across web-sites.

Social Cause

- People want to post on the Internet and like a highly-personalized web experience.
- Industry is improving search and retrieval techniques so that people can find the posts.
- Governments improve search and retrieval to do forensics and intelligence gathering.
Let's focus

- The previous described issues are a problem with any type of public or semi-public posts and are not specific to a certain type of information, e.g. text, image, or video.

- However, let's focus on multimedia data: images, audio, video.

Multimedia in the Internet is Growing

- YouTube claims 65k video uploads per day
- Flickr claims 1M images uploads per day
- Twitter: up to 120M messages per day => Twitpic, yfrog, plixi & co: 1M

Computer Science Problem

- More multimedia data = Higher demand for retrieval and organization tools
- Image, video retrieval hard =>
- Solution: Workarounds...

Workaround: Manual Tagging

Workaround: Geotagging

Geo-Tagging

Allows easier clustering of photo and video series as well as additional services.
Support for Geo-Tags

Social media portals provide programmatic interfaces to connect geo-tags with metadata, accounts, and web content.

<table>
<thead>
<tr>
<th>Portal</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>YouTube</td>
<td>3.0</td>
<td>3M</td>
</tr>
<tr>
<td>Flickr</td>
<td>4.5</td>
<td>180M</td>
</tr>
</tbody>
</table>

Allows easy search, retrieval, and ad placement.

Issues of Tracking using Geo-Tagging

“Be careful when using social location sharing services, such as FourSquare.”

Question

Did you know about geo-tagging and its potential?
- a) I had never heard about geo-tagging before.
- b) I knew about geo-tagging but never thought about what it could be used for.
- c) I knew about geo-tagging and knew the potential for photo organization and retrieval.
- d) I know about geo-tagging, its use and the privacy risks.
- e) I only heard about privacy risks of geo-tagging but never really thought about what it is good for.

Scientific Approach: Can you do real harm?

- Cybercasing: Using online (location-based) data and services to mount real-world attacks.

- Three Case Studies:

  Case Study 1: Twitter

  - Pictures located
  - From an undisclosed celebrity we found:
    - Home location (several pics)
    - Where the kids go to school
    - The place where he/she walks the dog
    - “Secret” office

  Celebs unaware of Geo-Tagging

  Working with the very talented Adam Hamilton on creating a new album.}
  [120 Petz, 501]

  Source: ABC News

  Click here to log in or

  Working with the very talented Adam Hamilton on creating a new album. [120 Petz, 501]
Celebs unaware of Geotagging

EXIF IFD
- Compression (0x111) = JPEG compression (6)
- X-Resolution (0x11A) = 721x1 (2/3)
- Y-Resolution (0x11B) = 721x1 (2/3)
- YX-Resolution Unit (0x11C) = Inches (2)
- YC-Resolution Unit (0x11D) = centimeters / center of pixel array (3)
- Embedded thumbnail image

EXIF GPS IFD
- GPS Version ID (0x01) = 0x02.0a02.0t02.0t02.0t
- GPS Latitude Reference (0x03) = N
- GPS Latitude (0x04) = 34.13.12.13.4 (degrees, minutes, seconds) = 34° 12' 3" = 34.200833°
- GPS Longitude Reference (0x06) = W
- GPS Longitude (0x07) = 118.12.12.13.4 (degrees, minutes, seconds) = 118° 12' 3" = 118.200833°

Case Study 2: Craigslist

“For Sale” section of Bay Area Craigslist.com:
4 days: 68729 pictures total, 1.3% geo-tagged
- Many ads with geo-location otherwise anonymized
- Sometimes selling high-valued goods, e.g. cars, diamonds
- Sometimes “call Sunday after 6pm”
- Multiple photos allow interpolation of coordinates for higher accuracy

Case Study 3: YouTube

- Once data is published, the Internet keeps it (in potentially many copies).
- Programmatic YouTube interface is easy to use and allow quick retrieval of large amounts of data

Can we find people on vacation in YouTube?

Cybercasing on YouTube

Experiment: Cybercasing using YouTube (240 lines in Python)
Cybercasing on YouTube

Input parameters

Location: 37.869885, -122.270539
Radius: 100 km
Keywords: kids
Distance: 1000 km
Time-frame: this_week

Output

Initial videos: 1000 (max_res)
User hull: ~50k videos
Potential hits: 106
Cybercasing targets: >12

Corollary

People are unaware of
1. geo-tagging
2. high resolution of sensors
3. large amount of geo-tagged data
4. easy-to-use APIs allow fast retrieval
5. resulting inference possibilities


The Threat is Real!

Bits
Business + Innovation + Technology + Society
September 12, 2010 10:21 AM
Burglars Picked Houses Based on Facebook Updates
Glenn K. Hillton

Question

Do you think geo-tagging should be illegal?

a) No, people just have to be more careful. The possibilities still outweigh the risks.
b) Maybe it should be regulated somehow to make sure no harm can be done.
c) Yes, absolutely this information is too dangerous.
Technical Question: Is this really about geo-tags?

Multimodal Location Estimation

We infer location of a Video based on content and context:
- Allows faster search, inference, and intelligence gathering even without GPS.
- Use geo-tagged data as training data


Internet Elapsed Time

Ongoing Work:
The Berkeley Multimodal Location Estimation Project

http://mmle.icsi.berkeley.edu

ICSI’s Evaluation Results


YouTube Cybercasing Revisited

<table>
<thead>
<tr>
<th>Old Experiment</th>
<th>No Geotags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Videos</td>
<td>1000 (max)</td>
</tr>
<tr>
<td>User Hull</td>
<td>~50k</td>
</tr>
<tr>
<td>Potential Hits</td>
<td>106</td>
</tr>
<tr>
<td>Actual Targets</td>
<td>&gt;12</td>
</tr>
</tbody>
</table>

YouTube Cybercasing with Multimodal Location Estimation vs using Geotags

But...

Is this really only about geo-location?

No, it's about the privacy implications of Internet search and (multimedia) retrieval in general.

Another Multimedia Example

Idea: Can one link videos across accounts? (e.g. YouTube linked to Facebook vs anonymized dating site)

Let's try an off-the-shelf speaker verification system: ALIZE (GNU GPL)

User ID on Flickr videos

Det curves for userid 312 videos 11,330 trials

Condition 1

EER = 31.4%

Persona Linking using Internet Videos

Result:

On average having 20 videos in the test set leads to a 99.2% chance for a true positive match!


Solutions that don’t work

• I blur my faces (audio and image artifacts can still find you)
• I only share with my friends (but who and with what app do they share with?)
• I don’t do social networking (others may do it for you)

Question

And now? What do you think has to be done?

a) Nothing can be done. Privacy is dead.
b) We need to educate people about this and try to save privacy. (fight)
c) I will really think before I post, and I agree with b).
d) I will really think before I post, and I agree with a).
e) I won’t post anything anymore! (flee)
My Personal Advice

Think before you post:
- Make sure you know who can read your post and you choose material appropriate for the audience.
- Make sure you know what you are posting: Is there hidden data included in your post? Are you allowed to reveal the information? Are you offending anybody?
- The Internet keeps data forever and in potentially many copies. Your need for privacy will change, however.
- Perform regular searches to find out what was posted about you by others.

Thank You! Questions?