



# Sharing Multimedia on the Internet and the Impact for Online Privacy



**Dr. Gerald Friedland**  
**Director Audio and Multimedia Lab**  
**International Computer Science Institute**  
**Berkeley, CA**  
**[friedland@icsi.berkeley.edu](mailto:friedland@icsi.berkeley.edu)**

10100101  
101CS101  
10100101

# 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.



## 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 100k video uploads per day, or 48h 72h per minute.
- Flickr claims 1M images uploads per day
- Twitter: up to 120M messages per day => Twitpic, yfrog, plixi & co: 1M images per day



# Resulting Problem

- More multimedia data = Higher demand for retrieval and organization tools.
- But multimedia retrieval is hard
  - Researchers work on making retrieval better.
  - Industry develops workarounds to make retrieval easier right away.



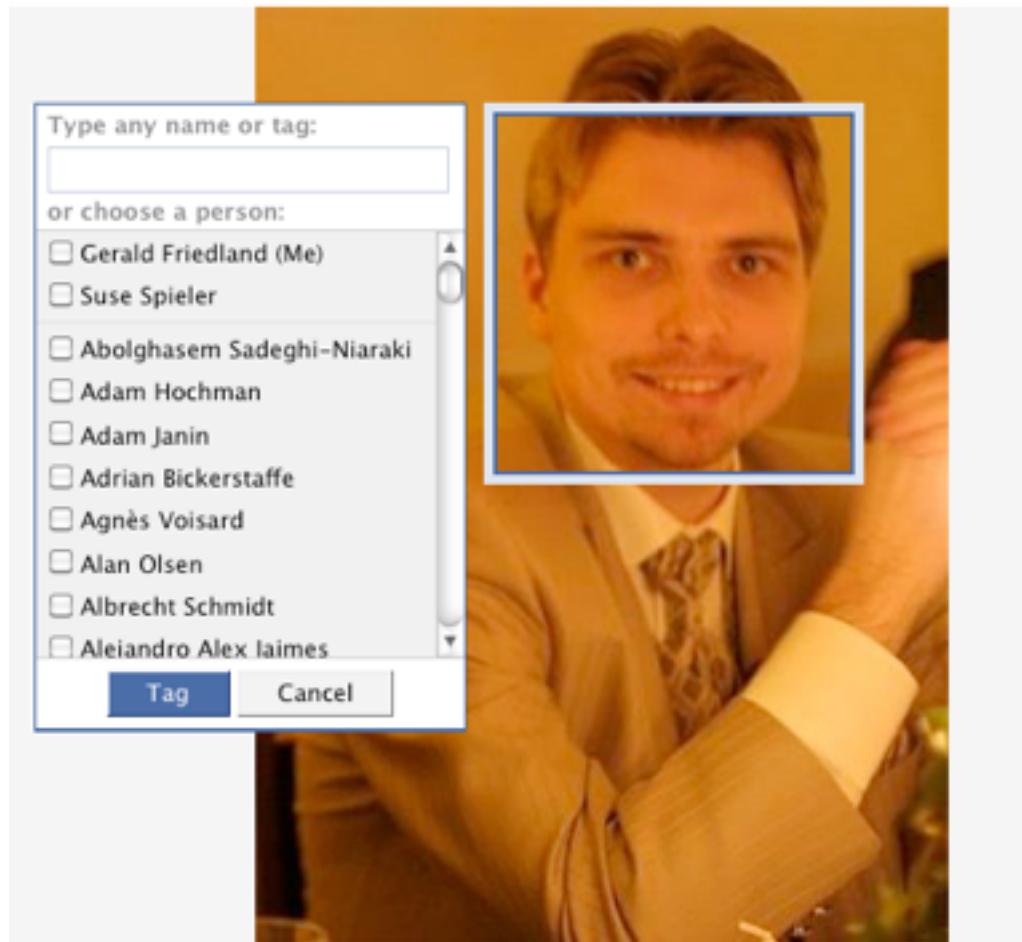
# Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.
- Let's take a look at some retrieval approaches:
  - Geotagging
  - Multimodal Location Estimation
  - Audio-based user matching



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Workaround: Manual Tagging



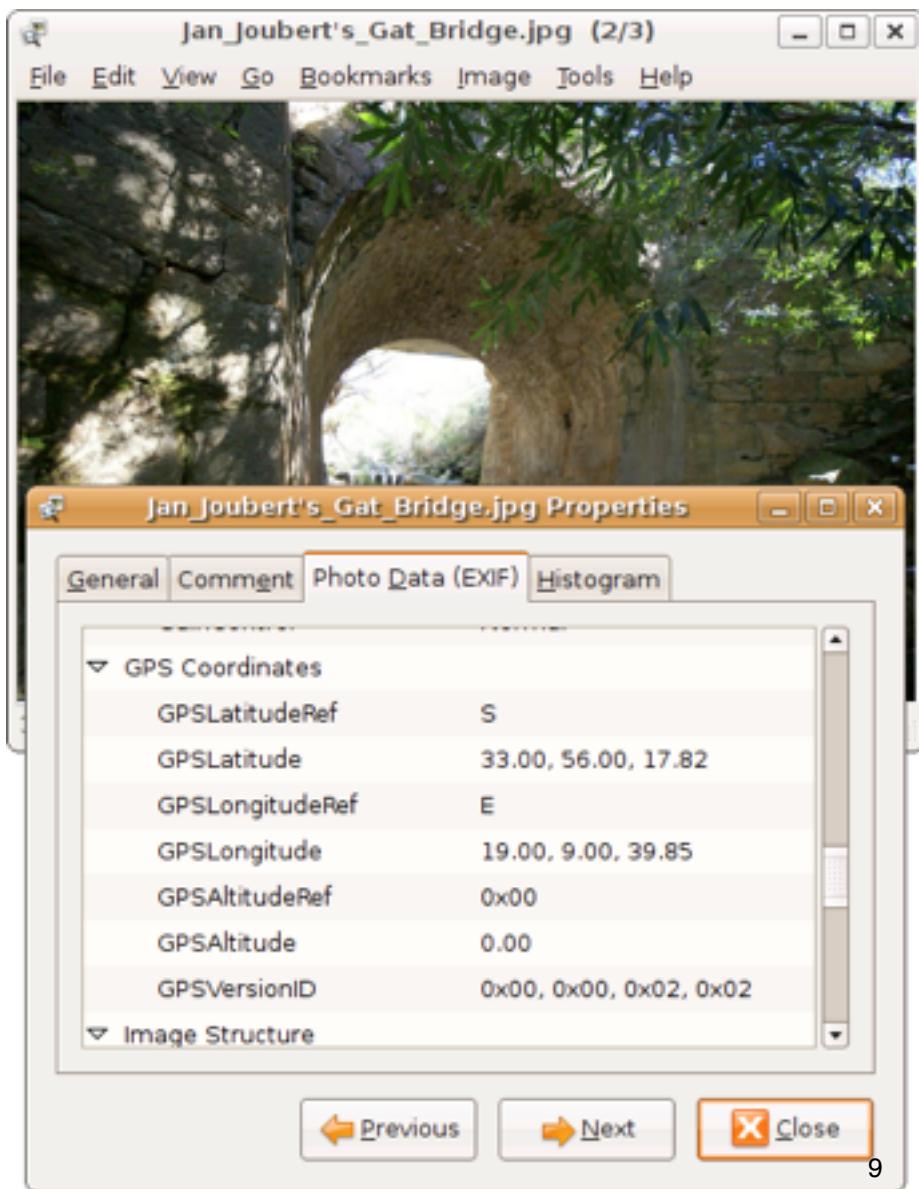


INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE



Source: Wikipedia

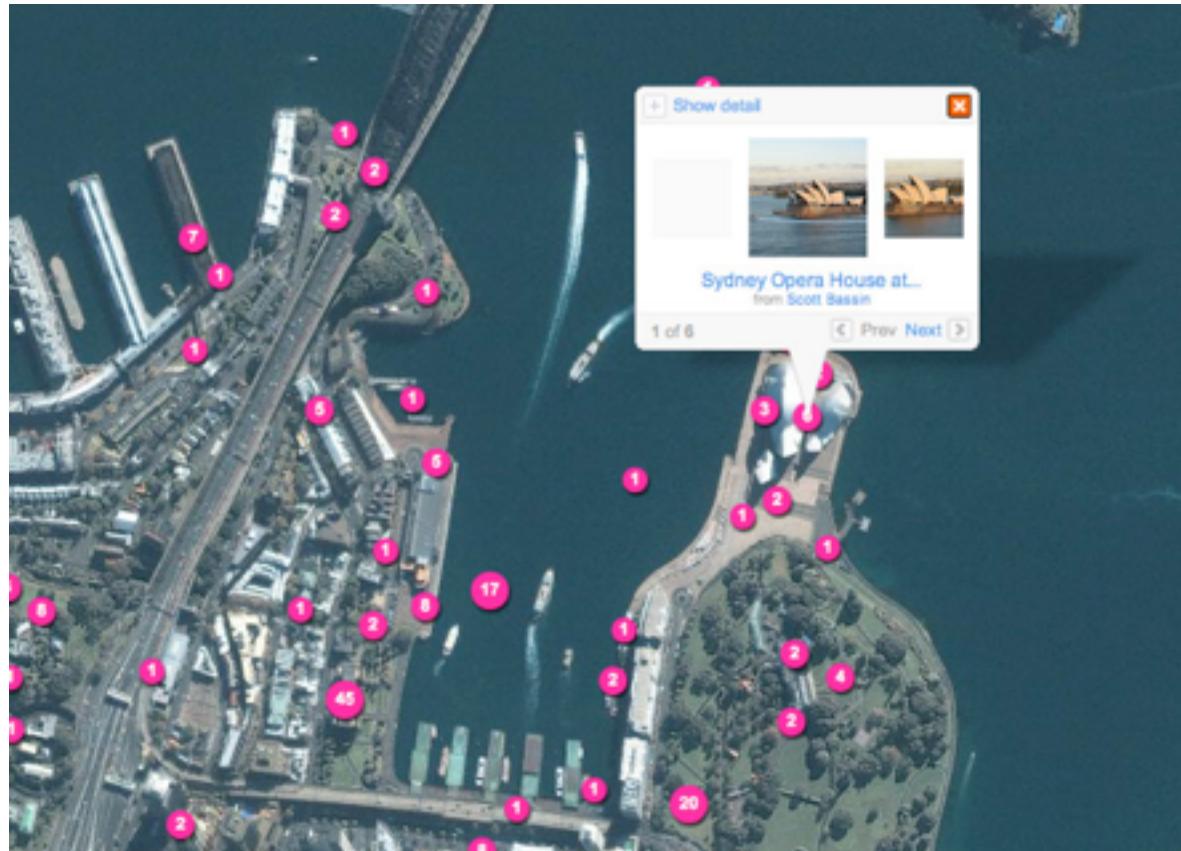
# Workaround: Geotagging





INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Geo-Tagging



**Allows easier clustering of photo and video series as well as additional services.**



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Support for Geo-Tags

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

| Portal             | %   | Total |
|--------------------|-----|-------|
| YouTube (estimate) | 3.0 | 3M    |
| Flickr             | 4.5 | 180M  |

Allows easy search, retrieval, and ad placement.



# Related Work

**PLEASE ROB ME**

Raising awareness about over-sharing

Check out our [guest blog post](#) on the CDT website.

A map with two red location pins, each marked with a white 'X'.

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



# Related Work

## I Can Stalk U

Raising awareness about inadvertent information sharing

Home

How

Why

About Us

Contact Us

If you came from the Social Media White Noise Podcast, see  
[our response](#)

What are people *really* saying in their tweets?



albeeta: I am currently nearby 10456 Riverside Dr Los Angeles CA

less then a minute ago · [Map Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to albeeta](#)



chrisbeanland: I am currently nearby http://maps.google.com/?q=51.5385,-0.133

less then a minute ago · [Map Location](#) · [View Tweet](#) · [View Picture](#) · [Reply to chrisbeanland](#)

### Links

- Mayhemic Labs
- PaulDotCom
- SANS ISC
- Electronic Frontier Foundation
- Center for Democracy & Technology

How did you find  
me?

Mayhemic Labs, June 2010: “Are you aware that Tweets are geo-tagged?”



# 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.



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Hypothesis

Since geo-tagging is a workaround for multimedia retrieval, it allows us to peek into a future where multimedia retrieval works.

What if multimedia retrieval actually worked?



# Can you do real harm?

- **Cybercasing:** Using online (location-based) data and services to enable real-world attacks.
- Three Case Studies:



G. Friedland and R. Sommer: "Cybercasing the Joint: On the Privacy Implications of Geotagging", Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C, August 2010. <sup>16</sup>



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Case Study 1: Twitter

- Pictures in Tweets can be geo-located
- From a tweet we found:
  - Home location
  - Where they work
  - The place they go to
  - “Secret” office





INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Celebs unaware of Geo-Tagging



[Click here to login or](#)



Working with the very talented Adam Hamilton on creating a new album. My best, Bill

Source: ABC News <sup>18</sup>



# Celebs unaware of Geotagging

## EXIF IFD1

- Compression {0x0103} = JPEG compression (6)
- X-Resolution {0x011A} = 4718592/65536 ==> 72
- Y-Resolution {0x011B} = 4718592/65536 ==> 72
- X/Y-Resolution Unit {0x0128} = inch (2)
- Y/Cb/Cr Positioning (Subsampling) {0x0213} = centered / center of pixel array (1)
- Embedded thumbnail image:



## EXIF GPS IFD

- GPS Version ID {0x00} = 0x02,0x02,0x00,0x00
- GPS Latitude Reference {0x01} = N
- GPS Latitude {0x02} = 34/1,12/1,3/1 [degrees, minutes, seconds] ==> 34° 12' 3" == 34.200833°
- GPS Longitude Reference {0x03} = W
- GPS Longitude {0x04} = [degrees, minutes, seconds] ==> [redacted] == [redacted]



# Google Maps shows Address...

A Google Maps satellite view showing a residential and commercial area in Sherman Oaks, Los Angeles. The map displays streets like W. Sherman Way, De Soto Ave, and Gault St. Key landmarks include Canyon Park Branch Library, El Paseo Mexican Restaurant, and Moksha Indian Cuisine. A red location marker labeled 'A' is placed near the intersection of De Soto Ave and W. Sherman Way, with a green arrow pointing towards it from the bottom right. The interface includes standard Google Map controls (zoom, orientation, search) and navigation links (Get Directions, My Maps, Print, Send).



# Case Study 2: Craigslist

“For Sale” section of Bay Area Craigslist.com:  
4 days: 68729 pictures total, 1.3% geo-tagged

| #   | Model       | # | Model                 |
|-----|-------------|---|-----------------------|
| 414 | iPhone 3G   | 6 | Canon PowerShot SD780 |
| 287 | iPhone 3GS  | 3 | MB200                 |
| 98  | iPhone      | 2 | LG LOTUS              |
| 32  | Droid       | 2 | HERO200               |
| 26  | SGH-T929    | 2 | BlackBerry 9530       |
| 20  | Nexus One   | 1 | RAPH800               |
| 9   | SPH-M900    | 1 | N96                   |
| 9   | RDC-i700    | 1 | DMC-ZS7               |
| 6   | T-Mobile G1 | 1 | BlackBerry 9630       |



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

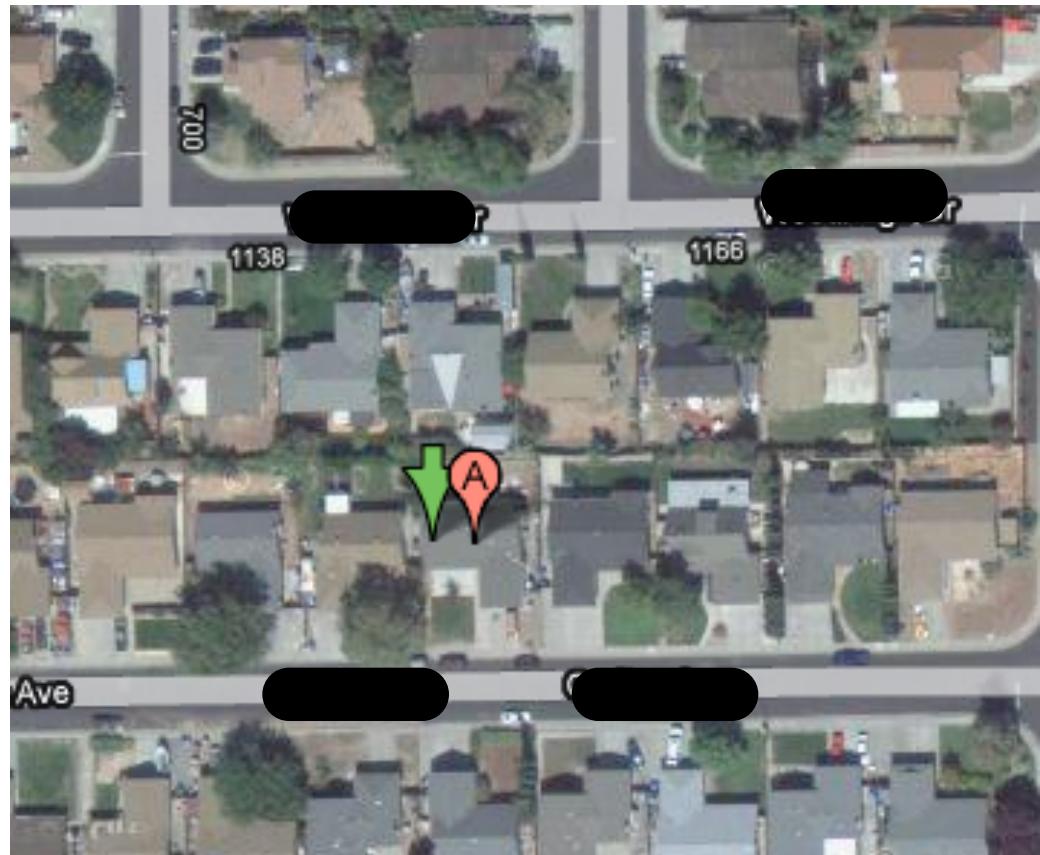
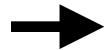
# People are Unaware of Geo-Tagging

- 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



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Craigslist: Real Example





# Case Study 3: YouTube

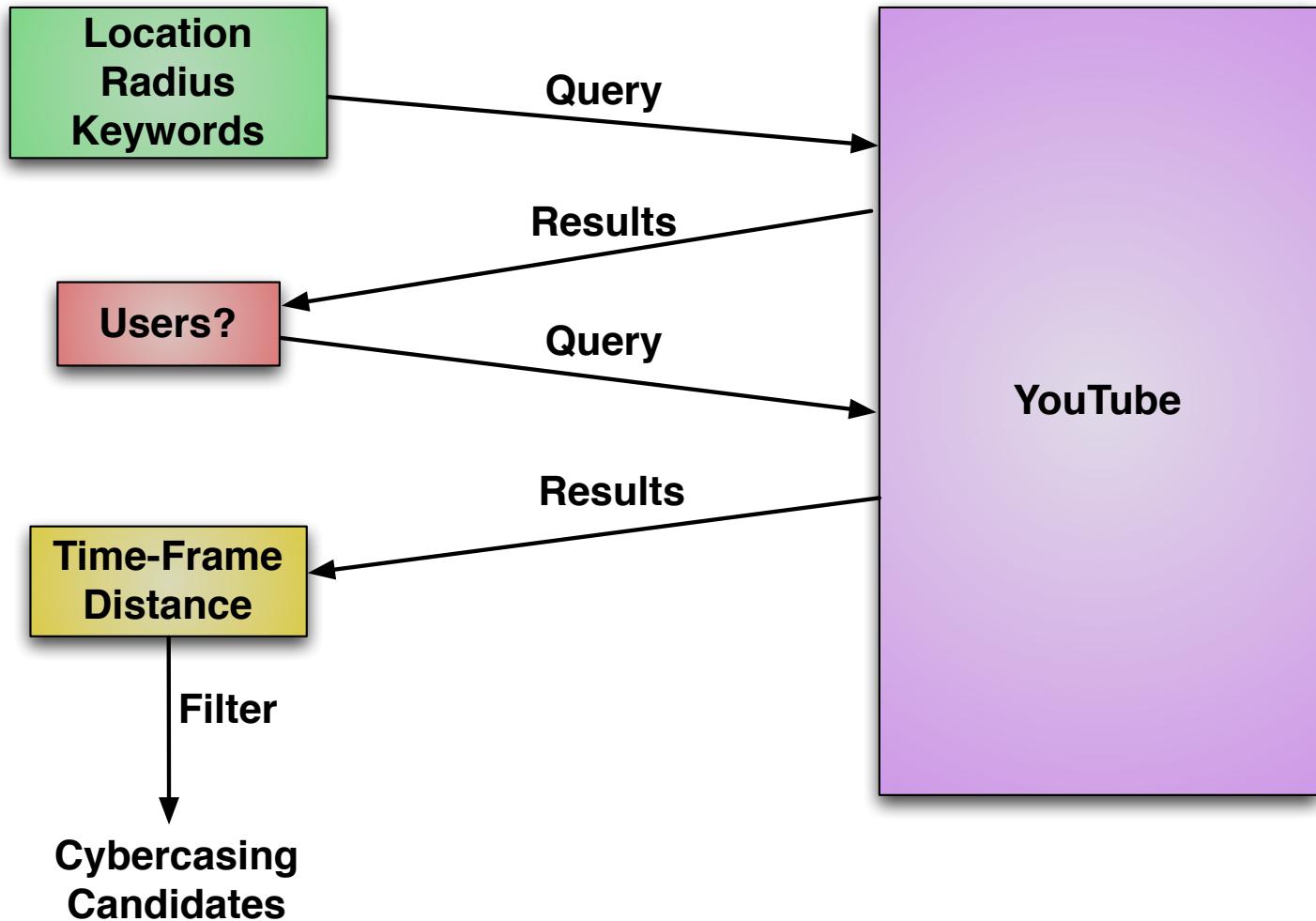
Recall:

- Once data is published, the Internet keeps it (in potentially many copies).
- APIs are 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 the YouTube API (240 lines in Python)





INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Cybercasing on YouTube

## Input parameters

**Location:** 37.869885, -122.270539

**Radius:** 100km

**Keywords:** kids

**Distance:** 1000km

**Time-frame:** this\_week

10100101  
101CS101  
10100101

# Cybercasing on YouTube

First Day of [REDACTED] Vacation

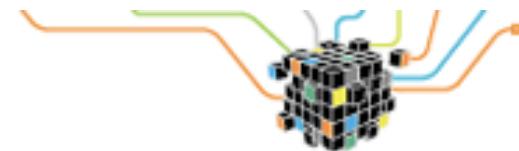
[REDACTED] videos  Subscribe





INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# The Threat is Real!



## Bits

Business • Innovation • Technology • Society

September 12, 2010, 10:24 AM

### Burglars Picked Houses Based on Facebook Updates

By NICK BILTON

**News Feed**      Top News • Most Recent 63

---

Share: Status Question Photo Link Video

Going to the beach for the weekend! (Someone else will be home though so think again Facebook Bandits!)

Lock Share

Illustration by Nick Bilton/The New York Times



# 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.



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

**But...**

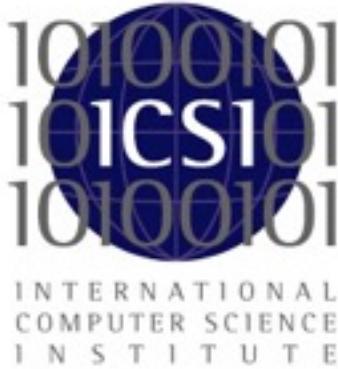
Is this really about geo-tags?  
(remember: hypothesis)



# Ongoing Work: The Berkeley Multimodal Location Estimation Project



<http://mmle.icsi.berkeley.edu>



# Multimodal Location Estimation

We infer location of a Video based on visual, audio, and tags:

- Use geo-tagged data as training data
- Allows faster search, inference, and intelligence gathering even without GPS.

**G. Friedland, O. Vinyals, and T. Darrell: "Multimodal Location Estimation," pp. 1245-1251, ACM Multimedia, Florence, Italy, October 2010.**

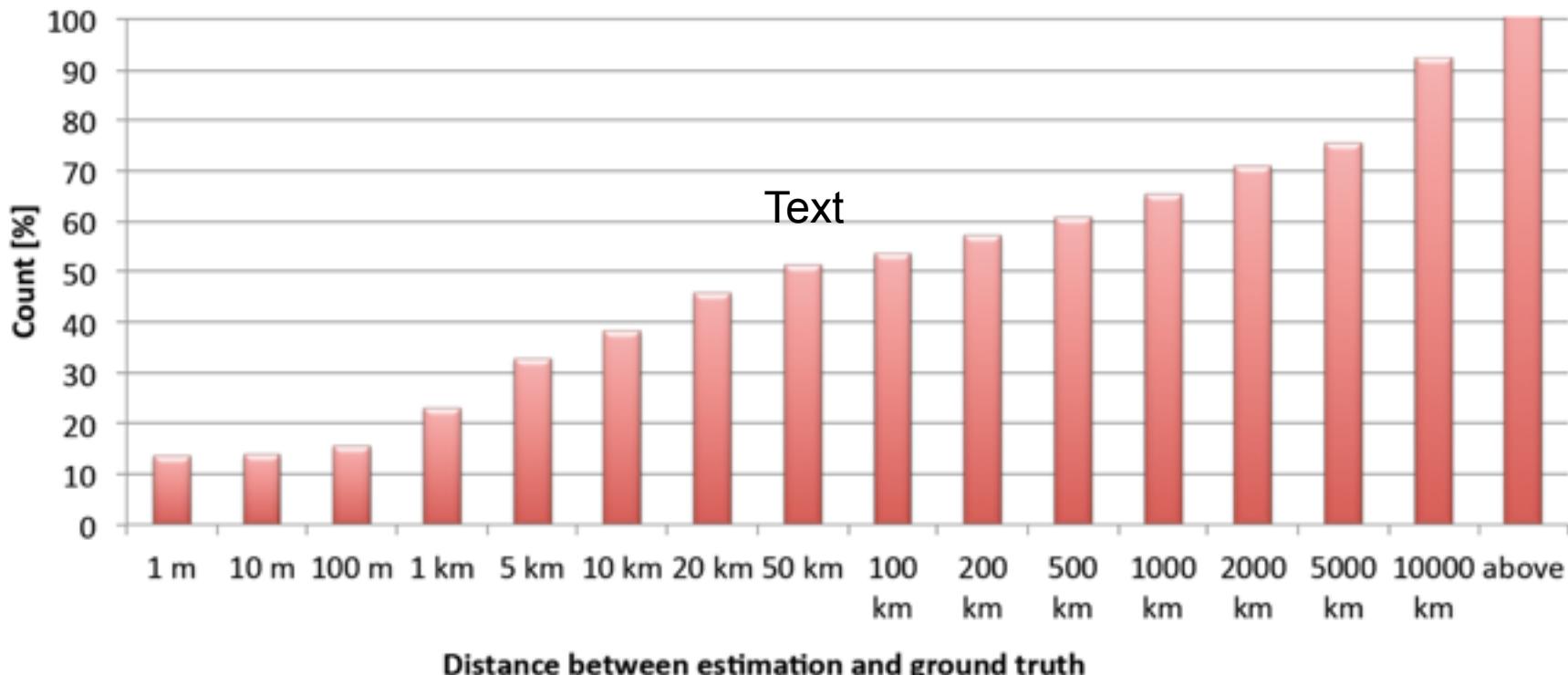


# MediaEval Benchmark

MediaEval Benchmarking Initiative for Multimedia Evaluation

The "multi" in multimedia: speech, audio, visual content, tags, users, context

**ICSI/UCB Estimation System at Placing Task 2012 (Cumulative)**



J. Choi, G. Friedland, V. Ekambaram, K. Ramchandran: "Multimodal Location Estimation of Consumer Media: Dealing with Sparse Training Data," in Proceedings of IEEE ICME 2012, Melbourne, Australia, July 2012.

# YouTube Cybercasing Revisited

|                | Old Experiment | No Geotags |
|----------------|----------------|------------|
| Initial Videos | 1000 (max)     | 107        |
| User Hull      | ~50k           | ~2000      |
| Potential Hits | 106            | 112        |
| Actual Targets | >12            | >12        |

YouTube Cybercasing with Geo-Tags vs  
Multimodal Location Estimation



# Question

Do you think research about geo-location should be abandoned?

- a) No, of course not.
- b) No, but regulated.
- c) Yes, absolutely.



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

**But...**

Is this really only about geo-location?

No, it's about the privacy implications of multimedia retrieval in general.



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# Example

**Idea: Can one link videos across accounts?**

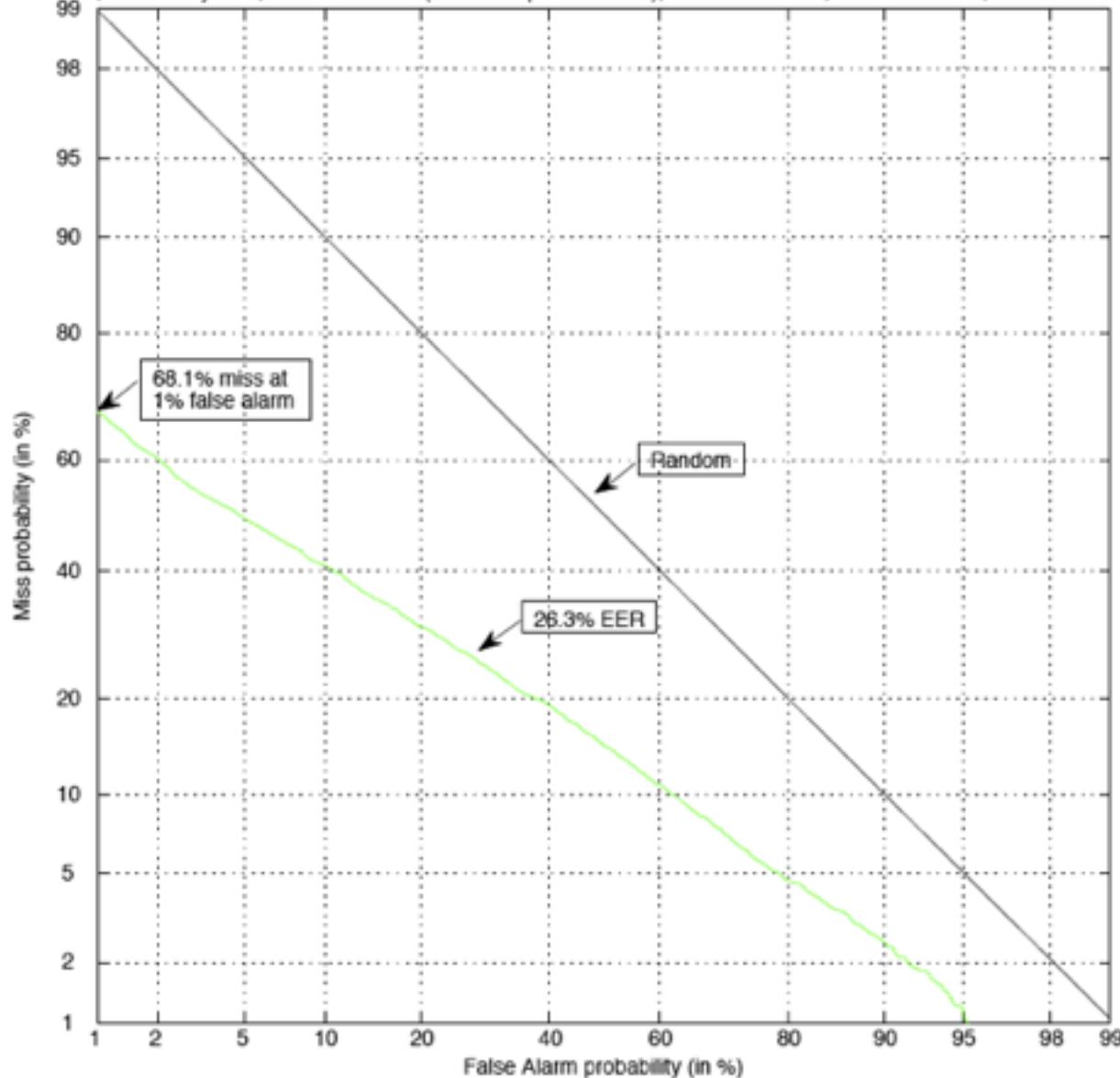
(e.g. YouTube linked to Facebook vs anonymized dating site)



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# User ID on Flickr videos

User-ID, i-vector system, 6.4 million trials (3287 true positive trials), 1239 train users, 2784 test users, 60–16000 Hz MFCC





# Persona Linking using Internet Videos

**Result:**

On average having 20 videos in the test set leads to a 99% chance for a true positive match!

**H. Lei, J. Choi, A. Janin, and G. Friedland: “Persona Linking: Matching Uploaders of Videos Across Accounts”, at IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Prague, May 2011**



# 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)



INTERNATIONAL  
COMPUTER SCIENCE  
INSTITUTE

# What to do about it?

<http://teachingprivacy.icsi.berkeley.edu>

Need volunteers for CalDay presentation!



# Thank You!

# Questions?

Work together with:  
Robin Sommer, Jaeyoung Choi, Luke  
Gottlieb, Howard Lei, Adam Janin, Oriol  
Vinyals, Trevor Darrel, Dan Garcia, K.  
Ramchandran, E. Venketsan, and  
others.