Content-based Privacy for Consumer-Produced Multimedia

Adam Janin, Gerald Friedland, and Robin Sommer

TR-12-012

October 2012

Abstract

The following text summarizes our proposed article for the Brave New Topics session. Based on preliminary experiments, we argue that privacy-measures against global inference on multimedia data are in dire need and therefore require a research field to come up with brave new solutions.

This work was partially supported by funding provided to ICSI through National Science Foundation grant CNS:1065240 ("Understanding and Managing the Impact of Global Inference on Online Privacy"). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors or originators and do not necessarily reflect the views of the National Science Foundation.

1 Introduction

The growth of multimedia as demonstrated by social networking sites such as Facebook and YouTube combined with advances in multimedia content analysis (face recognition, speaker verification, location estimation, etc.) provides novel opportunities for the unethical use of multimedia. In small scale or in isolation multimedia analytics have always been a powerful but reasonably contained privacy threat. However, when linked together and used on an Internet scale, the threat can be enormous and pervasive. The multimedia community therefore has an *obligation* to understand and attempt to mitigate these risks.

Imagine a future where multimedia query engines *just work*. You can search by topic, location, person, camera identity, or time even when the uploader did not explicitly include such information. An unscrupulous attacker could query for videos recently recorded at resorts and then find videos taken with the same camera in nearby wealthy residential neighborhoods. This would produce an ideal "hit list" of targets who are likely away from home, which the thief could then refine. As reported in previous work, cybercasing already occurs, but with a multimedia query engine simple methods of anonymizing posts and suppressing metadata will no longer be enough. Rather, the multimedia community needs to come up with methods to identify when information (such as the "identity" of the camera) is being unintentionally leaked and develop mitigation techniques to reduce the potential harm.

In this paper we outline existing and future multimedia content analysis and linking techniques that could support unethical use, describe possible attack vectors, and outline mitigation techniques.

2 Privacy Risks

In this section, we describe some existing and future multimedia analytic techniques that pose a privacy risk. This is by no means an exhaustive list.

Location Estimation Multimedia location estimation formed the genesis of our interest in privacy in multimedia, and was reported in previous work. Using multimodal methods, state-of-the-art algorithms can estimate about 15 % of Flickr videos with an accuracy better than 10 m and over 50 % with an accuracy better than 1 km. This extends the amount of exactly trackable multimedia by a significant factor without requiring actual GPS sensors.

Time Estimation The date and time that a multimedia document was recorded can be estimated using cues such as sun location or measuring shadow lengths. More powerfully, if you can determine that Video A was recorded at the same time and place as Video B, and you know or can infer Video A's time, you now know Video B's time. Just excluding time/date metadata from *your* vacation video does not protect you if somebody else includes it in theirs.

Person Detection In the image realm, this is usually called face detection; in audio, speaker recognition. While the uploader can take active methods to anonymize the foreground participants if privacy is an issue (e.g. replacing their

face with a black box, replacing their audio with a bleep sound), the privacy of background participants is problematic because the uploader may not care about incidental privacy breaches of the background participants.

Object Detection Detecting a person with an iPhone might make them a more desirable robbery target. Marketers could target people based on the furniture quality in the background of a video. Note that mitigation techniques are particularly problematic with object detection, since one cannot simply remove *all* objects from a multimedia document without severely impacting the document's content.

Environmental Acoustic Noise Uploaders often recognize the need to obscure faces. However, when recording video data they often forget that the audio track includes a unique signature that might break their anonymity. This has been shown in several studies, including our previous work. Also, the combination of such linking methods with other methods such as location estimation leads to even more powerful privacy invading possibilities.

Sensor Detection It may be possible to uniquely identify what camera was used to record a video or what microphone was used to record audio based on the artifacts of the sensor. For example, lens aberration may be unique to a particular camera; the exact frequency response of a microphone might be used to narrow down the possible microphones. This provides a whole new avenue of linking, completely bypassing other means of anonymization.

3D Recordings Time-of-flight cameras, light field camera, stereo cameras, and microphone arrays are all becoming more pervasive. It is clear that similar devices will continue to be developed. Each comes with its own sets of issues, and have the potential to capture even more unwanted data. Since this trend will only accelerate, it is necessary for the multimedia community to address these issues.

Exotic Sensors Everything from air pressure sensors to heart rate monitors are becoming more common, and it is likely data from these sensors will be incorporated into multimedia documents much as GPS is now. Since users often have no real notion of what is being collected or how accurate it is, they have little or no intuition on the privacy implications. A prominent historic example is GPS – it was only recently that the profound privacy implications of geotagging became commonly known.

3 Application Interfaces

Today, one can readily access much of the *structured* information available online via programmatic interfaces: major services like Google, Facebook, Twitter, Flickr, YouTube, and LinkedIn, all offer extensive APIs that make automatic retrieval trivial. These APIs often offer more comprehensive access than the corresponding web interface, and their availability is the primary driver behind the wide range of 3rd party "apps" that constitute a key part of today's social networking space.

We contend that as multimedia retrieval technology matures, it will eventu-

ally become part of such APIs, making the capabilities available to everybody able to write a few lines of Python code. For example, Google already provides simple forms of image and video search, and rumors say they have face recognition ready for mass deployment as part of their Goggles service. Facebook has already integrated face recognition into their platform, and though it is not yet exposed via the Facebook API, third party companies such as face.com are already providing programmable access to face recognition of Facebook content.

Having large-scale multimedia retrieval at one's fingertips provides an opportunity for amazing next-generation online services. However, we believe that it will also open up a new dimension of privacy threats that our community has not yet understood.

4 Attacks

Availability of Internet-scale multimedia retrieval capabilities allows a wide range of attacks that threaten users' privacy. Whereas today's search queries remain limited to mostly textual information, attackers will eventually query for audio and video *content*. Criminals could leverage that to reliably locate promising targets. For example, they may first identify individuals owning high-value goods within a target area and then pinpoint times when their victims' homes are unattended.

Another threat are background checks becoming much more invasive than today: many companies have strong incentives to examine their customers' private life for specifics impacting business decisions. An insurer, for example, might want to double-check that customers receiving disability payments are not skiing on Facebook photos. Likewise, an employer seeking new hires might want to check a candidate's Twitter followers for potentially embarrassing information.

A whole new realm of marketing techniques are enabled by multimedia retrieval and linking. A company could extract all videos of people wearing branded merchandise, cluster them by location and time, and target that location for direct marketing. The privacy implications of such broad and automatic analysis have been insufficiently studied.

As a final example, the new capabilities make *stalking* easier by providing the means to not only quickly locate the victims, but also profile their typical behavior patterns.

5 Mitigation Techniques

Countering such attacks is not straight-forward since filtering out sensitive information from audio and video content is fundamentally harder than with structured text data.

A major challenge for conserving privacy in consumer produced videos is the development of methods to identify the content-wise important information

from the semantically background information. It is this background data that has the highest risk of incidentally leaking private information.

We believe machine learning will play a key role in detecting such leaks. For example, one can label who is an "extra" in a movie by the number of times they appear and the number of lines they speak. The extras form the semantic background to the movie – they are noticeable, but not directly relevant. A machine learning algorithm could use "star" vs. "extra" as ground truth, and learn models to distinguish the two. Applied to consumer-produced videos, the system could then identify foreground vs. background participants using the trained model.

Once the information that is breaking privacy is identified, it must also be removed or distorted sufficiently to reduce the threat. This is difficult with most existing multimedia analysis algorithms, since they are statistical in nature. If we understood the specific cues the statistical methods learn, we could obscure those cues, hopefully without distorting the rest of the content. For example, if the background semantic "bird call of a Nene" is detected, you are leaking location information (Hawaii). Just damping that sound may be enough to obscure the location. This sort of cue detection is in the nascent stages for some methods (e.g. concept detection as in TrecVID MED), and nearly non-existent for others. It is incumbent on the multimedia community to develop an understanding of the cues so that mitigation techniques can be developed.

For other methods, more direct mitigation may be possible. For example, an upload tool could blur semantically background faces in a video. A query tool could refuse to perform speech recognition and indexing on background voices. A key component of such a system would be to ensure, possibly with the interaction of the uploader, that foreground content is not compromised.

Independent of any technological protection, we believe a key ingredient to comprehensive mitigation must be user *education*. Users can take steps to protect themselves once they realize the power that modern content analysis tools yield in the hands of adversaries. They might then even choose not to post certain content in the first place.

6 Conclusion

In summary, we believe the diversity of attacks and the complexity of solving the privacy issues with multimedia content will require creative thinking of a community of researchers and therefore spawn of a new field in multimedia content analysis. We believe it is not only a Brave New Topic, but a necessary new field.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1065240. Any opinions, findings, and conclusions or

recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] G. Friedland, J. Choi, H. Lei, and A. Janin. Multimodal location estimation on Flickr videos. In *Proceedings of the ACM International Workshop on Social Media (WSM)*, Scottsdale, Arizona, USA, November 2011.

[2] G. Friedland, G. Maier, R. Sommer, and N. Weaver. Sherlock holmes' evil twin: on the impact of global inference for online privacy. In *Proceedings of the 2011 workshop on New security paradigms workshop*, NSPW '11, pages 105–114, New York, NY, USA, 2011. ACM.

[3] G. Friedland and R. Sommer. Cybercasing the joint: on the privacy implications of geo-tagging. In *Proceedings of the 5th USENIX conference on Hot topics in security (HotSec)*, Washington, D.C., USA, August 2010.

[4] H. Lei, J. Choi, A. Janin, and G. Friedland. User verification: Matching the uploaders of videos across accounts. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 2404 –2407, may 2011.