Chapter 7
Content-based Privacy for Consumer-Produced Multimedia

Gerald Friedland, Adam Janin, Howard Lei, Jaeyoung Choi, Robin Sommer

Abstract We contend that current and future advances in Internet scale multimedia analytics, global inference, and linking can circumvent traditional security and privacy barriers. We therefore are in dire need of a new research field to address this issue and come up with new solutions. We present the privacy risks, attack vectors, details for a preliminary experiment on account linking, and describe mitigation and educational techniques that will help address the issues.

7.1 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 these risks, mitigate the effects, and educate the public on the issues.

Imagine a future where multimedia query engines *just work*. You can search by topic, location, person, camera identity, and 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 (see Section 7.2), 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 must work to educate the public about the risks of inferencing at the Internet scale, invent 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.

After defining the topic and presenting prior work (Section 7.2), we outline existing and future multimedia content analysis and linking techniques that could support unethical use and describe possible attack vectors (Section 7.3). Next, we describe some preliminary experiments providing evidence that multimedia analytics can circumvent one aspect of privacy by linking accounts (Section 7.4). Finally, we outline mitigation and educational techniques (Section 7.5) and conclude that this is a new topic to be explored (Section 7.6).

7.2 Definition and Prior Work

Privacy is a concept that is hard to define. As a consequence, many definitions exist, including "privacy is the right to be left alone" [35] and more modern definitions, such as U.S. President Barak Obama's "Framework for Protecting Privacy" [24]. Merriam Webster defines privacy as "a) the quality or state of being apart from company or observation and b) freedom from unauthorized intrusion" [41]. While all of the definitions aim at the same goal, they are too broad for our engineering purposes. Therefore, in this paper we restrict ourselves to a more technical definition. We define privacy as "practically securing the implications of communication", which sets it apart from the field of *secure communication*, which is "securing the properties of the communication itself" [42] through methods of cryptography, steganography, identity hiding, and other well-known computer science topics. In other words, our privacy research is not about securing a communication line between several parties; it is to make sure that publicly available information conveys only the data the author intended. We acknowledge that this goal, like the aims of secure communication, will most likely never be achieved perfectly. However, improvements in methods can make communication "more private". Given that even our narrower definition is still a very broad goal, we will limit ourselves to attack vectors that pose an actual criminal threat and/or directly influence life-changing decisions.

While the scientific community has investigated correlation between different data sets in terms of privacy implications, most of these efforts have focused on de-anonymizing or compromising a single data set with the help of auxiliary information. Except for the few exceptions described below, efforts have mostly concentrated on structured data, ignoring multimedia content analysis.

7.2.1 Work on Structured Data

In 1997, Sweeney [37] showed that anonymously published medical records can be de-anonymized when correlated with external data, triggering a large body of

follow-up work on designing anonymous statistical databases as well as understanding their limitations [13, 14, 38, 10, 1].

More relevant to the multimedia community, Narayanan et al. present an algorithm and proof for de-anonymizing sparse datasets [31]. They apply their algorithm to anonymized Netflix movie ratings: given knowledge of a subset a person has rated (e.g. learned from a lunch conversation or public ratings), the system is able to identify *all* movies in the database that the user has rated. In [32], the same idea is used to de-anonymize a social network graph by leveraging a graph from a second network with real identities as auxiliary data. Researchers from Parc investigated inference using web search engines in order to analyze whether anonymized (or obfuscated) private documents that are going to be released publicly can be de-anonymized [36, 8]. They do not consider multimedia content nor inference between information that is already publicly available.

Griffith et al. [20] correlate public birth, death and marriage records from the state of Texas to derive the mother's maiden name of more than 4 million Texans. Balduzzi et al. [3] automatically query 8 social networks with a list of 10 million e-mail addresses to retrieve the associated user profiles. They then correlate that profile information across the networks and are able to identify mismatches between them. (i.e. they find users who chose different names, age, etc. in different networks). More generally, Bishop et al. [4] discuss the need to go beyond "closed worlds" when sanitizing a data set and consider external knowledge explicitly.

With geo-location information being a popular key to image and video retrieval, another area of related research is locational privacy. The Electronic Frontier Foundation published an overview of locational privacy aspects [5]. Locational privacy in vehicular systems, e.g. toll collection, is addressed in [34, 23]. Zhong et al. [43] present protocols for secure privacy preserving location sharing. The upcoming HTML 5 standard will include APIs to query a client's location. The *Cree.py* [21] application uses geolocation data from social networks and media hosting services to track a person's movements.

Several web sites highlight the potential of information leakage users might not be aware of:

`Sleeptime.org` estimates sleep patterns of Twitter users.

`Stolencamerafinder.co.uk` crawls for digital camera serial numbers in online photos in order to find pictures taken with stolen cameras.

`Icanstalku.com` published geotags found in tweets.

`pleaserobme.com` used status updates form social networks to locate users who were currently not at home but had published their home address.

7.2.2 Work on Multimedia Data

The above section was presented to outline current work on structured data. History has shown that work on multimedia data follows in the footsteps of structured data with a delay (for example, work on compression, messaging capabilities, or even

World Wide Web content itself). As a result, we see an initial growth in multimedia articles that present work on privacy. We see this early work as evidence for our hypothesis of a new field of research.

In a recent effort [19], we analyzed the privacy implications of *geotagging*, i.e. high-accuracy location information attached as meta-data to audio, image, and video files. Specifically, we examined the risk that such geotags pose for what we termed "cybercasing": using online data and services to mount real-world attacks. Moreover, we showed that geo-tags are not needed as they can be replaced by multimedia analytical location estimation techniques [17].

In [28] Lukas et al. propose a method for the problem of digital camera identification from images based on the sensor's pattern noise. For each camera under investigation, they first determine its reference pattern noise, which serves as a unique identification fingerprint. This is achieved by averaging the noise obtained from multiple images using a denoising filter. To identify the camera from a given image, they consider the reference pattern noise as a spread-spectrum watermark, whose presence in the image is established by using a correlation detector. Experiments on approximately 320 images taken with nine consumer digital cameras are used to estimate false alarm rates and false rejection rates.

Many researchers have worked on automatic video blurring (for example [11, 27, 16]); however, [33] showed that many of the proposed techniques are not effective. In response to this problem, [12] has presented an initial framework to validate video privacy.

7.3 Privacy Risks and Possible Attacks

In this section, we describe some existing and future multimedia analytic techniques that pose a privacy risk including how these risks could be exploited. 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 (see Section 7.2). Using multimodal methods, state-of-the-art algorithms can estimate the location of about 40 % of Flickr videos with an accuracy better than 100 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 known as 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 an iphone in a person's hand 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 (Section 7.2). 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 is already possible to narrow down or even 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, pixel noise is 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.

We outline a small number of specific attacks that can now or could shortly be used to invade privacy in detrimental ways using Internet scale multimedia analytics and linking.

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 eventually 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 rumor has it that face recognition is 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.

The 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 is 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 refuse payment to a customer receiving disability where the insurer finds Facebook photos of the customer skiing. Likewise, an employer seeking new hires might check a candidate's Twitter followers for potentially embarrassing information that could be used against the company in the future and refuse to hire such candidates.

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.

The new capabilities make *stalking* easier by providing the means to not only quickly locate the victims, but also profile their typical behavior patterns, friends, relatives, and acquaintances.

7.3.1 Example

In this section, we will exemplify the power of multimedia retrieval in combination with structured-data retrieval in a mockup scenario adopted from [18].

Consider the following business: Fred works for Schooner Holdings and wishes to gain (possibly illicit) inside information on future profits at the chipmaker Letin. Fred hires Eve, who runs an "expert network". Eve puts Fred in touch with Bob, a Letin employee. In the process of consulting for Fred, Bob is encouraged to reveal information about Letin's upcoming products.[1]

[1] In many countries, this practice is possibly illegal but exists in a gray area and is seemingly routine practice. The Galleon insider trading trial [9] was based largely on the use of expert network consultants.

Currently, the greatest limit on this process is Eve *finding* experts like Bob who (perhaps unknowingly) possess potential insider information and are willing to act as consultants. Eve would greatly improve her business if she could find "corruptibles": individuals in the business of interest who might be favorable to legitimate or illegitimate offers.

Thus Eve starts searching social networks for individuals who are compatible with her desired level of (il)legality. She instructs her crawler to begin with LinkedIn and web searches, crawling the names and contact information for personnel at companies of interest.

Then her crawler shifts to Facebook, Twitter, other social networks, and blogs, beginning with all candidates found in the first pass. This crawler does not just look at the candidates but also at friends of candidates.

She also searches any media, including images and videos, for links to other people that the social network might not provide directly. Face recognition for example can provide probable connections to other profiles. She also examines media for any compromising material, such as illegal acts, drug paraphernalia, or party photos. Eve knows that her automated content analysis does not need to be perfect: she leverages crowdsourcing services like Mechanical Turk [2] to validate potential candidate matches using human labor at a very low cost.

Eve's crawler also queries further public and semi-public records. There are commercial services that map an email address to a mailing address. Her crawler uses these to discover where candidates live and how much their property is worth (e.g. by using Zillow.com's access to property tax data and sales history).

With all this data, Eve's crawler can now create "inference chains" which estimate the probability that any given candidate in her set has a potential weakness, enabling Eve to search for possible points of corruptibility. An individual who is dating someone with a reputation as a gold digger, or who purchased their house at the height of the real estate bubble, might have financial problems. Such candidates could be honestly corrupted by offering consulting positions, allowing Eve to expand her expert network.

Eve might also contract with those operating outside the law. Then blackmail becomes an attractive option, especially if considering guilt by association. Someone with a security clearance may be vulnerable if his associates are drug abusers, or if he is having an affair that can be inferred through social patterns.

Nothing in the preceding scenario is unrealistic: every step Eve takes can be constructed using today's technology. It is simply a matter of putting all the pieces together to collect and analyze the reams of data which exist on today's social networks and other databases.

Unfortunately, there is also hardly any protection in place against somebody like Eve. Furthermore, while structured data still plays a dominant role in this scenario, it is easy to see how multimedia data will blur the boundaries even more. For example, if we assume that face recognition technology reaches close to perfection, user names will no longer provide a boundary as long as a face photo is part of the website. Moreover, speaker recognition, location estimation, and other techniques described in Section **??** will add even more possibilities. Finally, note that the meth-

ods need not be perfect — Eve needs only a small number of likely hits to follow up on to allow nefarious actions to proceed.

7.4 Preliminary Experiments

This section presents technical details on a preliminary experiment on matching user accounts based on consumer-produced videos, demonstrating that multimedia retrieval can circumvent traditional security and privacy barriers, such as the assumption that different account names will separate the same persona.

Consider the following scenario: A professor at a University is proud to present lectures to a very large audience on a public video distribution site. These lectures contain her voice, her face, and her name in the credits, making their authorship anything but anonymous. At the same time, she is dating online and follows the dating site's suggestion to provide introduction videos of herself to make her profile more personable. The suggestion comes with the assurance that, unless the author of the introduction video identifies herself, the video will remain anonymous.

In the following, we will provide evidence that this promise of anonymity is hard to keep in the face of increasingly accurate multimedia retrieval technologies.

7.4.1 Dataset

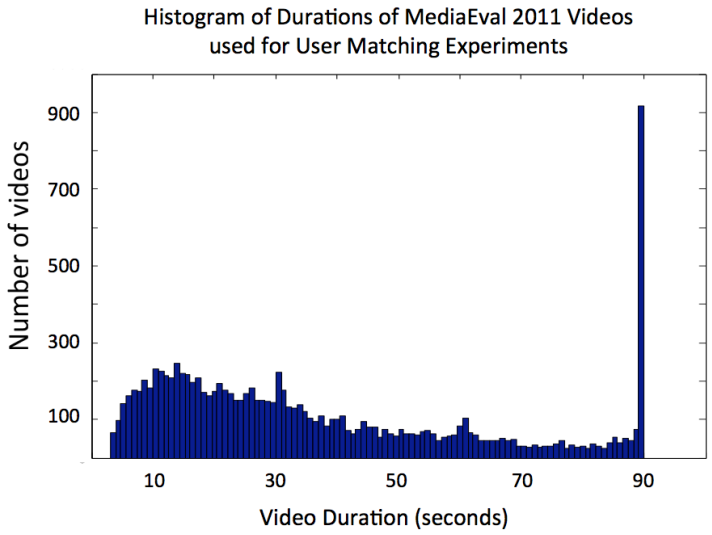

Fig. 7.1 A histogram visualizing the duration of the videos of the data set used in our experiments.

We begin by describing the data sets used in this experiment. The audio tracks are extracted from the videos distributed as training and test sets for the Placing Task of MediaEval 2011 [30], a multimedia benchmark evaluation. The Placing Task involves automatically estimating the location of each test video using one or more of: metadata (e.g. textual description, tags), visual/audio contents, and social information. The videos are not pre-filtered or pre-selected in any way to make the data set more relevant to the user-verification task, and are therefore likely representative of videos selected at random.

A total of 10,857 Creative Commons licensed Flickr videos, uploaded by 2,943 Flickr users, were used in our experiments. Flickr requires that an uploaded video must be created by its uploader (if a user violates this policy, Flickr sends a warning and removes the video). This policy generally ensures that each uploader's set of videos is "personal" in the sense that they were created by the same person and therefore likely have certain characteristics in common, such as editing style, recording device, or frequently recorded scenes/environments, etc.

From a by-hand examination of 123 short-duration videos from the data set, we found that most of videos' audio tracks are quite "wild". 59.3 % of the videos are home-video style with ambient noises. 47.2 % of the videos had heavy ambient noises such as crowds chatting in the background, traffic noise, and wind blowing into microphone. 25.2 % of the videos contained music, either played in the background of the recorded scene, or inserted at the editing phase. 59.3 % of the videos did not contain any form of human speech at all, and even for the ones that contained human speech, 64 % were from multiple subjects and crowds in the background speaking to one another, often at the same time. Although we found that 10.5 % of videos contained audio of the person behind the camera, there is no guarantee that the owner of the voice is the actual uploader; it is possible that all videos from the same uploader were recorded by different people (such as family members).

Figure 7.1 displays a histogram of the lengths of the 10,857 videos used in our dataset. All videos are limited to 90 seconds, accounting for the peak at 90 seconds. 71.8 % of videos have less than 50 seconds of playtime, while 50 % have less than 30 seconds of playtime.

7.4.2 Technical Approaches

This section describes the multimodal user verification experiments based on audio and a set of five visual features. Note that the task of user verification is to determine if two videos are uploaded by the same user or different users. The i-vector-based approach [7], which is currently the state-of-the-art in the field of speaker recognition, is used to perform classification and audio-visual feature combinations. The approach involves extracting a set of low-dimensional vectors to represent the user identity of each video. The vectors can be derived from either the audio or visual features.

To extract the audio-based low-dimensional vectors, which are known as the i-vectors in [7], a total variability matrix T is first trained to model the variability (both user-, acoustic environment-, and acoustic channel-related) of the high-dimensional Baum-Welch statistics obtained from the MFCC C0-C19+Δ+$\Delta\Delta$ (60 dimensions total) audio feature vectors of each video. The matrix acts as a projection matrix used to obtain the low-dimensional vectors, which characterize the user of each video based on its audio. Specifically, for each video, the audio track is first extracted, and a vector of first-order Baum-Welch statistics M of the audio feature vectors, centered around the means of a GMM world model, is obtained. The statistics can be decomposed as follows:

$$M = m + T\omega \tag{7.1}$$

where m is the GMM world model mean vector, and ω is the low-dimensional vector. The GMM contains 1,024 mixtures, and each mixture contains 60 mean dimensions corresponding to the dimensionality of the MFCC features. Hence, the total dimensionality of M is 61,440, which the T-matrix projects onto a set of 400 dimensions to form the low-dimensional audio-based vectors.

The visual-based low-dimensional vectors are obtained from the result of a Principal Components Analysis (PCA) projection of a set of pre-extracted visual features onto a small set of its eigen-dimensions. The visual features are extracted using the open source library LIRE [29]. The features used include Tamura (TAM), Gabor (GAB), Auto Color Correlogram (ACC), Color and Edge Directivity Descriptor (CEDD), and Fuzzy Color and Texture Histogram (FCTH). The TAM feature is a texture-based feature. For our experiments, 24 dimensions are used to represent the low-dimensional vectors for the GAB, ACC, CEDD, and FCTH features, and 12 dimensions are used for the TAM feature.

The audio and visual features are combined by concatenating the corresponding low-dimensional vectors (in this way, the combined-feature experiments use more parameters than the standalone-feature experiments). The system performs a Within-Class Covariance Normalization (WCCN) [22] on the resulting vectors, which whitens their covariance via a linear projection matrix. A generative Probabilistic Linear Discriminant Analysis (pLDA) [26] log-likelihood ratio is then used to obtain a similarity score between the low-dimensional vectors of each training and test video. The generative pLDA log-likelihood ratio for similarity score computation is shown below:

$$\begin{aligned} score(\omega_1, \omega_2) = logN\left(\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}; \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{tot} & \Sigma_{bc}\\ \Sigma_{bc} & \Sigma_{tot}\end{bmatrix}\right) \\ - logN\left(\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}; \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{tot} & 0\\ 0 & \Sigma_{tot}\end{bmatrix}\right) \end{aligned}$$

where ω_1 and ω_2 are the vectors for a pair of training and test videos, $N(\cdot)$ is the normal Gaussian probability density function, and Σ_{tot} and Σ_{bc} are the total and between-class scatter matrices computed from the training vectors. Hence, one user-

similarity score is obtained for each training versus test video using the above approach.

The Brno University of Technology's (BUT's) Joint Factor Analysis Matlab demo [15] is used to assist in the system development, and the open-source ALIZE toolkit [6] is used to train the UBM. The HTK Library [25] is used for MFCC feature extraction.

7.4.3 Experiments and Results

A set of 1,268 Flickr users in the corpus were designated as training users, and 2,851 were designated as test users, with roughly 1,200 users in common with the training users. Each training user is associated with one video in the training set, and 4,869 videos are associated with the 2,784 test users. Overall, a set of 6,251 videos were used for training and testing. A separate set of 146 users with 4,605 videos were used to train the T-matrix, PCA projection matrices, and the total and between-class scatter matrices used in the system. 2,302 videos from the 146 users were used to train the GMM world model. A total of 6 million similarity scores were computed between video pairs from the training and test users, with 3,385 of the scores from pairs with the same user. Table 7.1 shows the Equal Error Rate (EER), and the Miss Rates at 1 % and 0.1 % False Positive (FP) rates for the 6 million scores of the system. A Miss occurs when a pair of same-user videos are classified as having different users, and a FP (false positive) occurs when different-user videos are classified as having same users, given a particular scoring threshold. For the Miss rate at 1 % FP, the threshold is set such that 1 % of the different-user pairs are classified as having same users. User verification results for both audio and visual features, standalone and in combination, are shown. Also shown are the number of dimensions used in the low-dimensional vectors used to compute the user-similarity scores for each feature, or combination of features.

Results in Table 7.1 indicate that the audio-based MFCC feature has the best standalone performance - 26.1 % EER, 65.6 % Miss at 1 % FP, and 86.6 % Miss at 0.1 % FP. If the MFCC features are combined with the top-four standalone visual features in terms of EER (ACC,CEDD,GAB,FCTH, and TAM), then the performance improves to 24.0 % EER, 59.2 % Miss at 1 % FP, and 78.4 % Miss at 0.1 % FP. This represents an 8.0 % relative EER improvement, a 9.8 % relative improvement of Miss at 1 % FP, and a 9.5 % relative improvement of Miss at 0.1 % FP. The results demonstrate the effectiveness of combining the audio and visual modalities for this task. The standalone visual features perform significantly worse than the MFCC feature.

Table 7.1 User matching results for audio and visual features standalone and in combination. Similarity scores were computed on 6 million pairs of videos, with a total of 1,268 training users and 2,784 test users as described in Section 7.4.3.

Feature	EER	Miss at 1% FP	Miss at 0.1% FP	Vector Dims
ACC	35.1%	84.9%	96.0%	24
CEDD	35.0%	82.1%	91.4%	24
FCTH	34.9%	82.2%	91.5%	24
GAB	44.4%	97.3%	99.6%	24
TAM	33.9%	87.6%	98.8%	12
GAB+CEDD+ ACC+FCTH+ TAM	33.0%	76.6%	91.5%	108
ACC+CEDD+ FCTH+TAM	32.4%	74.9%	89.7%	84
MFCC	26.1%	65.6%	86.6%	400
MFCC+ACC+ CEDD+GAB+ FCTH+TAM	24.1%	60.0%	79.3%	508
MFCC+CEDD+ ACC+FCTH+ TAM	24.0%	59.2%	78.4%	484

7.4.4 Summary of Experimental Results

The outcome of the above experiment for user matching is certainly not yet a reason for panic as user matching based on content is still very preliminary. However, given that our best approach was able to match random, short consumer-produced videos with an equal error rate of 24 % (compared to 50 % for chance), it means that a future can be foreseen where attacks like this become feasible. Moreover, many attacks are not targeted at matching one particular user. When finding victims from a large pool, the miss and false alarm rates are more important. The above experiments show that at 1 % false alarm, we would only miss about 60 % of the true positives. Given a scenario where the 1 % false alarm does not represent many videos, one can search through the 40 % of the non-missed true positives for a pair of videos containing the same user uploader.

7.5 New Topics For Research

Countering the attacks described above is not straight-forward since filtering out sensitive information from audio and video content is fundamentally harder than with structured text data. We therefore propose a new topic in multimedia devoted to considering both privacy research as well as education.

7.5.1 Mitigation Research

A major challenge for conserving privacy in consumer produced videos is the development of methods to identify the foreground information information that the user considers important from the background information. It is this background data that has the highest risk of incidentally leaking private information.

We believe that machine learning will play a key role in detecting such unnoticed information 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 (however, this might not be enough, see also discussion in Section 7.2). A query tool could refuse to perform speech recognition and indexing on background voices. This would be very similar to today's common practice for copy machines to refuse the copying of bank notes. A key component of such a system would be to ensure, possibly with the interaction of the uploader, that foreground content is not compromised.

7.5.2 Education on Privacy

Independent of any technological protection, we believe a key ingredient to comprehensive mitigation must be *education*. University electrical engineering and computer science curricula usually include an abundance of material on how to improve retrieval based on the underlying multimedia content analysis but only rarely talk about the negative impacts of these technologies. Privacy content is mostly limited to traditional topics in secure communications such as steganography, encryption, and other well-known techniques and/or even removed from consideration, as ethical concerns are considered not to be part of engineering. Therefore, even when acknowledged as a problem, many new technologists lack the knowledge of how

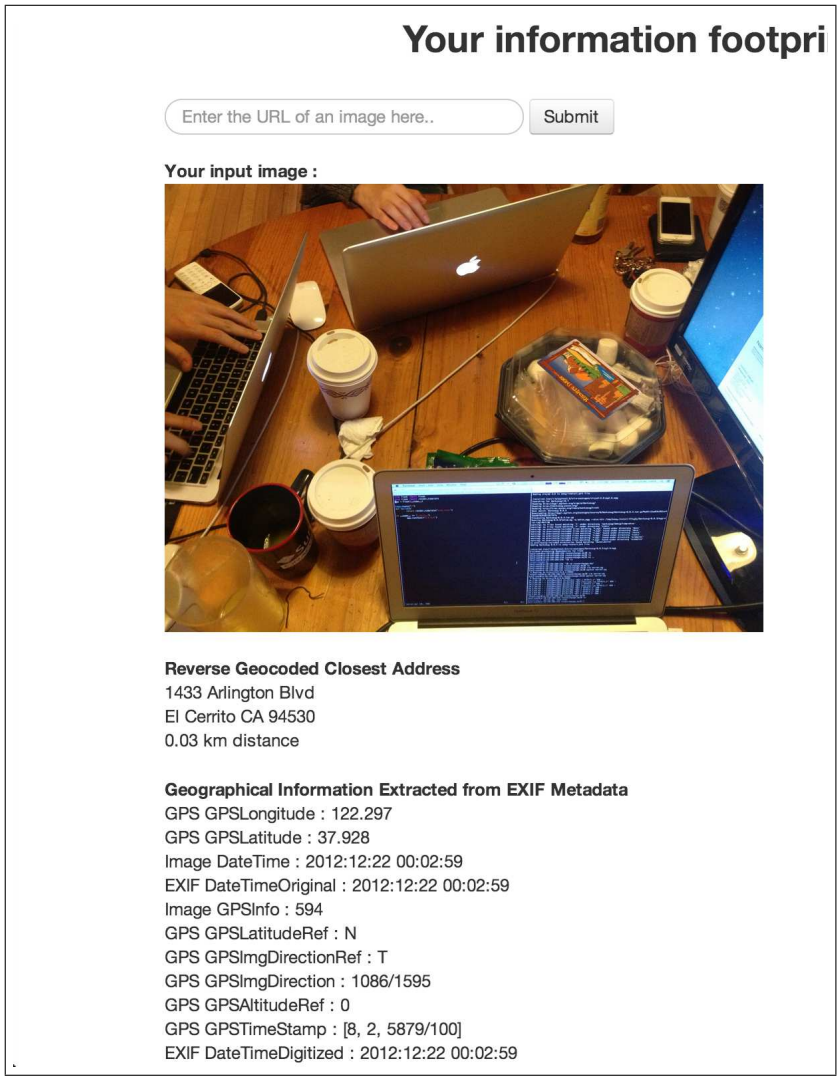

Fig. 7.2 Education is part of the new topic. A mockup of an educational browser tool showing that online image often includes meta-data that allows inference beyond the content of the image.

to react to society's concerns and even mitigate easy-to-address risks. An argument often heard from students is: "We'll deal with privacy and social issues later – right now we need to focus on development." The truth, however, is that, for example, if privacy and security had been a concern in the early stages of developing the Internet, many of today's issues, such as spam and phishing email, would most likely be much less of a problem. Undergraduate and graduate engineering education cur-

riculums should therefore include a strong component on privacy that makes future technologists aware of the societal implications of their research and development.

The second line of education should concern users, especially young people. Among the groups most affected by privacy concerns are high-school students [40]. They are the most frequent users of social-networking sites and apps, but often do not have a full understanding of the potential consequences their current online activities might have later in their lives. For example, a Facebook posting that a high-schooler's friends think is "cool" might be seen by a much larger audience than she or he expected– including perhaps future employers who wouldn't agree with the high-schooler's judgement. In addition, not understanding–or not thinking about– the consequences of posting often leads to oversharing information about other people, including friends and relatives. Consequently, 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.

Figure 7.2 shows a preliminary mockup for a teaching tool that we created as part of a project for social media privacy education for teenagers [39]. The input for the web-based tool is an arbitrary image that has been published on the web. The image is than analyzed for EXIF data. If found, the data is displayed textually. Furthermore, if the EXIF data contains geo-tags, the location for the image is shown on a map and all Twitter feeds that belong to that location are also shown. We saw that people are often shocked, how much information an indoor image like the one shown conveys and at the same time, how much can be inferred from that location, e.g. when a photo that does not contain any faces actually maps back to their own twitter feeds.

Building effective educational components that transfer knowledge on privacy protection and the consequences of multimedia retrieval to younger adults who are not yet capable of understanding deep research results constitutes a new domain for research. Here, educational research needs to team up with HCI and other multimedia-related fields to attack this part of the new topic. The question is how to enable educators to master an up-to-date, scientifically-informed understanding of privacy, without having to rely on (often exaggerated) newspaper articles.

7.6 Conclusion

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. The article surveyed the field and showed that awareness of the issue is focused on structured data but does not extend to multimedia retrieval. Using a scenario, a taxonomy of attacks, and a preliminary experiment, we outlined how multimedia retrieval adds a new quality to privacy and security research. We believe that mitigation is both a question of research as well

as education. 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 a new field in multimedia content analysis. We believe web-scale multimedia privacy is not only a new topic, but also 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. We thank Michael Ellsworth for his rewording suggestions.

References

1. C. Aggarwal. On k-anonymity and the curse of dimensionality. In *Proceedings of the International Conference on Very Large Data Bases*, 2005.
2. Amazon.com Mechanical Turk, https://www.mturk.com/mturk/welcome.
3. M. Balduzzi, C. Platzer, T. Holz, E. Kirda, D. Balzarotti, and C. Kruegel. Abusing social networks for automated user profiling. In *RAID'2010, 13th International Symposium on Recent Advances in Intrusion Detection*, 09 2010.
4. M. Bishop, J. Cummins, S. Peisert, A. Singh, B. Bhumiratana, D. Agarwal, D. Frincke, and M. Hogarth. Relationships and Data Sanitization: A Study in Scarlet. In *Proc. Workshop on New Security Paradigms*, 2010.
5. A. Blumberg and P. Eckersley. On locational privacy, and how to avoid losing it forever. *Electronic Frontier Foundation*.
6. J. Bonastre, F. Wils, and S. Meignier. Alize, a free toolkit for speaker recognition. In *Proceedings of ICASSP*, volume 1, pages 737–740, 2005.
7. L. Burget, P. Oldřich, C. Sandro, O. G., P. M., and N. Brümmer. Discriminantly trained probabilistic linear discriminant analysis for speaker verification. In *Proceedings of ICASSP*, Brno, Czech Republic, 2011.
8. R. Chow, P. Golle, and J. Staddon. Detecting privacy leaks using corpus-based association rules. In *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008.
9. P. Cohan. Why executives risk their job to tip a hedge fund. `http://meet-the-street.blogspot.com/2009/10/expert-networks-what-every-iro-needs-to.html`, 2009.
10. I. Dinur and K. Nissim. Revealing information while preserving privacy. In *ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, 2003.
11. F. Dufaux and T. Ebrahimi. Scrambling for video surveillance with privacy. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 160–160, 2006.
12. F. Dufaux and T. Ebrahimi. A framework for the validation of privacy protection solutions in video surveillance. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 66–71, 2010.
13. C. Dwork. Differential privacy. In *33rd International Colloquium on Automata, Languages, and Programming (ICALP)*, 2006.

14. C. Dwork. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*, volume 4978 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin / Heidelberg, 2008.
15. Joint factor analysis matlab demo. http://speech.fit.vutbr.cz/software/joint-factor-analysis-matlab-demo/.
16. J. Fan, H. Luo, M.-S. Hacid, and E. Bertino. A novel approach for privacy-preserving video sharing. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, pages 609–616, New York, NY, USA, 2005. ACM.
17. G. Friedland and J. Choi. Semantic computing and privacy: a case study using inferred geolocation. *International Journal of Semantic Computing*, 5(01):79–93, 2011.
18. 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 New Security Paradigms Workshop (NSPW)*, September 2011.
19. G. Friedland and R. Sommer. Cybercasing the Joint: On the Privacy Implications of Geo-Tagging. In *Proc. USENIX Workshop on Hot Topics in Security*, August 2010.
20. V. Griffith and M. Jakobsson. Messin' with texas deriving mother's maiden names using public records. In *Proceedings of the International Conference on Applied Cryptography and Network Security (ACNS)*, 2005.
21. H-Security. Cree.py application knows where you've been. `http://www.h-online.com/security/news/item/Cree-py-application-knows-where-you-ve-been-1217981.html`.
22. A. O. Hatch. Generalized linear kernels for one-versus-all classification: Application to speaker recognition. In *Proceedings of ICASSP*, Toulouse, France, 2006.
23. B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *MobiSys '08: Proceeding of the 6th International Conference on Mobile Systems, Applications, and Services*, 2008.
24. T. W. House. Consumer Data Privacy in a Networked World. `http://www.whitehouse.gov/sites/default/files/privacy-final.pdf`, 2012.
25. Hmm toolkit (htk). http://htk.eng.cam.ac.uk/.
26. S. Ioffe. Probabilistic linear discriminant analysis. In *Proceedings of ECCV*, pages 531–542, 2006.
27. T. Koshimizu, T. Toriyama, and N. Babaguchi. Factors on the sense of privacy in video surveillance. In *Proceedings of the 3rd ACM workshop on Continuous archival and retrival of personal experences*, CARPE '06, pages 35–44, New York, NY, USA, 2006. ACM.
28. J. Lukas, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *Information Forensics and Security, IEEE Transactions on*, 1(2):205–214, 2006.
29. L. Mathias and S. A. Chatzichristofis. Lire: Lucene Image Retrieval – An Extensible Java CBIR Library. In *In proceedings of the 16th ACM International Conference on Multimedia*, pages 1085–1088, October 2008.
30. Mediaeval web site. http://www.multimediaeval.org.
31. A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2008.
32. A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2009.
33. C. Neustaedter, S. Greenberg, and M. Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Trans. Comput.-Hum. Interact.*, 13(1):1–36, Mar. 2006.
34. R. A. Popa, H. Balakrishnan, and A. Blumberg. VPriv: Protecting Privacy in Location-Based Vehicular Services. In *Proceedings of the USENIX Security Symposium*, 2009.
35. D. Prosser. Privacy. In *California Law Review*, 48, page 383, 1960.
36. J. Staddon, P. Golle, and B. Zimny. Web-based inference detection. In *Proceedings of 16th USENIX Security Symposium*, 2007.
37. L. Sweeney. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine, and Ethics*, 25(2–3), 1997.

38. L. Sweeney. k-anonymity: A model for protecting privacy. *Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 2002.
39. Tools for teaching privacy to k12 and undergraduate students. http://teachingprivacy.icsi.berkeley.edu.
40. Facebook users by age. http://en.wikipedia.org/wiki/Facebook#Reception.
41. M. Webster. Privacy. `http://www.merriam-webster.com/dictionary/privacy`, 2013.
42. Wikipedia. Secure Communication. `http://en.wikipedia.org/wiki/Secure_communication`, 2013.
43. G. Zhong, I. Goldberg, and U. Hengartner. Louis, lester and pierre: Three protocols for location privacy. In *Proceedings of the Privacy Enhancing Technologies Symposium*, 2007.