Sensitive Lifelogs: A Privacy Analysis of Photos from Wearable Cameras

Roberto Hoyle[†] **Robert Templeman**[†] **Denise Anthony**[‡] **David Crandall**[†] **Apu Kapadia**[†]
[†]School of Informatics and Computing
Indiana University Bloomington
Bloomington, IN, USA
[‡]Department of Sociology
Dartmouth College
Hanover, NH, USA
{rjhoyle, retemple, djcran, kapadia}@indiana.edu, Denise.L.Anthony@dartmouth.edu

ABSTRACT
While media reports about wearable cameras have focused on the privacy concerns of bystanders, the perspectives of the 'lifeloggers' themselves have not been adequately studied. We report on additional analysis of our previous in-situ lifelogging study in which 36 participants wore a camera for a week and then reviewed the images to specify privacy and sharing preferences. In this Note, we analyze the photos themselves, seeking to understand what makes a photo private, what participants said about their images, and what we can learn about privacy in this new and very different context where photos are captured *automatically* by one's wearable camera. We find that these devices record many moments that may not be captured by traditional (deliberate) photography, with camera owners concerned about impression management and protecting private information of both themselves and bystanders.

Author Keywords
Lifelogging; wearable cameras; privacy

ACM Classification Keywords
K.4.2. Social Issues; K.4.1. Public Policy Issues: Privacy

INTRODUCTION
With several wearable cameras recently hitting the market, including the Narrative Clip,[1] Autographer,[2] and Google Glass[3] (Figure 1), the concept of 'lifelogging' has started to gain mainstream traction. Lifelogging cameras can automatically capture images throughout the day from a first-person view, easily collecting hundreds of images per hour. Lifelogging presents a new form of photography with different challenges: whereas the role of traditional photographers is to choose which moments to photograph and then to deliberately compose and take each picture, lifeloggers generate a deluge of automatically-taken photos, many of which are poorly composed, blurry, or have no useful content, and must curate them to decide which are worth keeping or sharing.

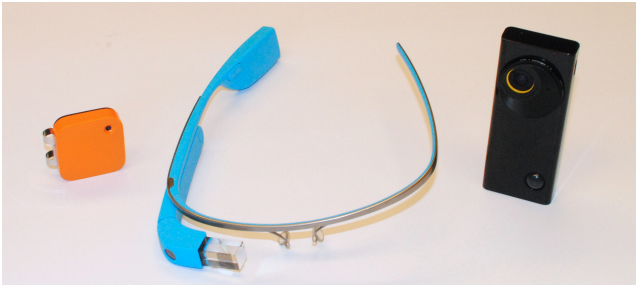

Figure 1. Lifelogging and wearable cameras, from left: Narrative Clip, Google Glass, and Autographer.

Lifelogging devices also raise important privacy concerns for both their wearers and the bystanders that are captured in the images. They may take photos in private, intimate settings that most people would not choose to photograph (e.g., in a bathroom or bedroom, or of a computer monitor). Bystanders are also typically unaware that they are being photographed by these discreet cameras. After images are captured, lifeloggers may upload them to a photo-management service on the cloud and/or share them with others. The very large number of images captured makes it difficult to ensure that photos are curated properly and this can lead to privacy leaks (e.g., by accidentally sharing a bathroom photo among a batch of vacation images, or a photo that captured part of a private document hidden in a corner of the image).

To better understand sharing behaviors and preferences of people with lifelogging cameras, we previously conducted a user study (N=36) where participants wore lifelogging cameras for a week and answered survey questions on a subset of images about how and why they would or would not share their photos [3]. We collected 14,477 images and obtained detailed sharing reasons for 1,015 images. We found that people exerted a high degree of physical discipline in private situations by stowing away the camera, pausing the collection, or immediately deleting private photos on the device rather than trying to delete them afterwards. We analyzed their survey responses, finding that many factors (like location, certain ob-

[1] `http://www.getnarrative.com`
[2] `http://www.autographer.com`
[3] `http://www.google.com/glass/start`

CHI 2015, April 18 - 23, 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04. . . $15.00
http://dx.doi.org/10.1145/2702123.2702183

jects, and a sense of propriety for the privacy of bystanders) affect sharing decisions.

While the previous analysis provided a quantitative assessment of sharing behaviors, it did not analyze the contents of the images themselves, and thus could not connect image content with privacy concerns. In this Note, we fill this gap by annotating and analyzing the same rich dataset. Beyond the stated sharing preferences of the participants, we seek to understand 1) what makes a lifelogging photo sensitive; 2) what participants say about what makes an image private; and 3) what we can learn about privacy by looking at the images and the participants' reasons. To answer these questions, we had five researchers independently code images for a) the reasons participants gave for sharing or not sharing photos (combined with information gleaned from the actual image) and b) the subject matter within the images. The coders then met to resolve all differences and discuss the findings.

Our results show that a large number of lifelogging photos capture parts of people's lives that are *not normally documented by regular photography*. These include a variety of private information, including credit and ATM cards, emails on computer screens, answer sheets for exams, academic transcripts, and so on. We also observed that participants were concerned with their 'impression management' [2], avoiding capturing photos of "bad habits" like picking their nails or using nicotine. Our findings thus highlight the need for mechanisms to manage and curate images captured by lifelogging devices, and shed light on the important privacy challenges created by this new form of photography.

METHODOLOGY

We analyzed the image data from our previous study [3], which considered 36 participants' use of lifelogging devices over the course of a week but did not analyze image content as we do here. Each participant wore a smartphone on a lanyard around the neck that took a photograph every five minutes between 8am and 10pm. Participants were given *in-situ* controls to pause the study for up to an hour, or to retroactively undo recording by erasing up to the last hour of data. At the end of each day, they reviewed and deleted any images that they wanted to remove from the study (e.g., those with nudity), then marked images that had no recognizable content and explained any use of the *in-situ* controls. Participants were then asked to describe sharing preferences for each photo, choosing among: Close Friends and Family; Other Friends and Family; Coworkers, Classmates, and Acquaintances; and/or Everybody Else. They were asked to give reasons why they chose to share or not share for a subset of images.

Dataset

The lifelogging dataset consists of 14,477 images from 36 participants, out of which 8,590 were labeled as "unusable" (having no recognizable content) by participants and were not analyzed further. Of the remaining images, 1,015 include freeform text explaining the reason behind the sharing decision. We identified images having some sensitive content by taking those that were not shared with at least one group. We further split these into two sets: a "private" set of images that were not shared with anyone (N=75), and a "semi-private" set which were shared with at least one group (N=186).

	Semi-private Photos				Private Photos	
Reasons given for	Sharing		Not sharing		Not sharing	
No good reason	140	(76%)	44	(24%)	49	(65%)
Image quality	37	(20%)	3	(2%)	6	(8%)
Person in the photo	30	(16%)	66	(36%)	13	(17%)
Person in photo's appearance	22	(12%)	0	(0%)	2	(3%)
Interest/Concern about self	14	(8%)	53	(29%)	17	(23%)
Interest/Concern about others	21	(11%)	9	(5%)	2	(3%)
Presentation management	13	(7%)	16	(9%)	9	(12%)
Screens or monitors visible	0	(0%)	18	(10%)	3	(4%)
Objects in image	1	(0.5%)	49	(27%)	20	(27%)
(written) Information visible	0	(0%)	15	(8%)	4	(5%)
Activity shown	3	(2%)	10	(5%)	7	(9%)
Location of image	1	(0.5%)	49	(27%)	12	(16%)
TOTAL # of Images	185 (100%)				75 (100%)	

Table 1. Reasons given by participants for sharing or not sharing photos, by whether photos were semi-private (shared with some but not all groups) or completely private (shared with no groups). More than one answer is possible, so columns add to more than 100%.

Coding of the Images

These 261 images were coded along two dimensions: (1) their visual content, and (2) the reasons provided by the user for sharing or not sharing them. Using an initial set of approximately 35 codes, five coders independently coded a random sample of 10 images from among the set of 261 images. We then discussed discrepancies until we developed a clearly defined set of 16 content and 15 reason codes. We then proceeded to code all 261 images, with at least three coders independently coding each one. We achieved initial 94.5% agreement on the image codes. Discrepancies were then discussed until consensus was reached on all codes for each image.

Ethical considerations

Our analysis was approved by our IRB. Privacy concerns prohibit publishing images, so we only describe them here.

FINDINGS

The results of our coding are summarized in Tables 1 and 2. A large number of images were simply shared by default ("for no good reason"), highlighting the need for effective mechanisms to manage images, since otherwise images with subtle privacy violations may be shared accidentally. Our analysis found private information, impression management, and respect for others' privacy were key reasons that participants kept images private, and we now discuss each of these here.

Private Information

Our analysis shows that private information in images was a key reason that users chose not to share, and a major specific concern was computer and smartphone screens. For 10% of semi-private and 4% of private photos, participants explicitly mentioned screens, and private photos were also significantly more likely to have screens than semi-private ones (private=0.55, semi-private=0.38, ANOVA F-test = 5.9 (df=1), $p < 0.02$), indicating that their presence is sensitive. Beyond screens, users indicated that 27% of photos were not shared because of "objects within the photo." We suspect that a large number of these concerns are related to text content; according to our coding, 62.3% of these photos had monitors, 37.7% had documents, and 59.0% had text clearly visible.

Content of photos	Semi-private		Private	
Image quality (in focus)	122	(66%)	42	(56%)
# People in the photo = none	97	(53%)	56	(77%)
= 1	55	(30%)	15	(22%)
= 2	20	(11%)	1	(1.4%)
= 3-5	1	1 (6%)	0	(0%)
Faces visible in photo	51	(28%)	9	(12%)
Screens/Monitors visible	71	(38%)	41	(55%)
Any written text visible	86	(47%)	33	(44%)
Bottles/cups/alcohol visible	35	(19%)	3	(4%)
Location				
Indoor	154	(83%)	63	(84%)
Bedroom	9	(6%)	19	(30%)
Bathroom	8	(5%)	2	(3%)
Outdoor	15	(8%)	3	(4%)
Sidewalk/street	8	(53%)	1	(33%)
Other/not clear	7	(47%)	2	(66%)
In-Transit	14	(8%)	8	(11%)
Car	3	(21%)	3	(38%)
Public bus	10	(71%)	4	(50%)
TOTAL # of Images	185	(100%)	75	(100%)

Table 2. Content of images by whether photos were semi-private (shared with some but not all groups) or private (shared with no groups). More than one answer is possible, so columns add to more than 100%.

We highlight several examples to show the diversity of private content in images. In one, P36[4] is seen writing a course evaluation for a class, with answers clearly visible. In another, P50 is seen writing a note, and chose not to share because *"the page in the picture is actually part of a surprise for a close friend, so I wouldn't want her seeing it. I also wouldn't necessarily want certain people seeing it because it contains something that could be deemed 'unsanctioned' "* (which, we believe, refers to a reference to alcohol). One participant's camera captured an image of their ATM card, and they chose not to share it for fear of the card number being stolen. P37 decided not to share an image because it depicted a friend's apartment number.

In contrast, others seemed unconcerned about photos that we would have considered private. One user captured many images at work that included numerous clearly legible academic transcripts, and marked them to be publicly shared.[5] This participant did restrict sharing of other images, so he or she was aware of other privacy issues.

Impression Management

Many users chose to share or not share based on 'impression management,' in order to portray traits that they want to present to the world. As described by Goffman, *"...control is achieved largely by influencing the definition of the situation which the others come to formulate, and he can influence this definition by expressing himself in such a way as to give them the kind of impression that will lead them to act voluntarily in accordance with his own plan"* [2]. Impression management could involve sharing an image to show a positive trait, or suppressing an image that shows a negative one.

[4]Participant IDs ranged from 32 – 68 in the dataset.

[5]These images were shared publicly, so were not included in the pictures that we coded earlier. We highlight them here as it shows an interesting user behavior. As all sharing in the study was hypothetical, no transcripts were leaked, and we removed these images from our dataset after discovering them.

In our dataset, we identified impression management as the reason why 7% of semi-private images were shared, and why 9% of semi-private and 12% of private photos were not shared. P53, for example, chose not to share an image of watching a movie, explaining that they were *"not studying when I should be."* P50 did not share a photo of their bedroom with coworkers, because *"I would just feel weird if employers or coworkers saw a picture of my bedroom."* P40 avoided sharing a picture of an electronic cigarette because *"I don't want my family to know that I have a nicotine addiction,"* but shared it with coworkers, classmates, and acquaintances. P55 did not share one of nail picking , because *"I am performing a bad habit and would not want people I know to see."*

On the other hand, several participants specifically chose to share photos in order to promote a positive impression. For instance, P45 shared several images depicting computer work because it showed "studying." Another participant captured a photo of a coffee mug with a crude slogan on it; they chose not to share the photo with the public, but did share it with friends because *"they would understand my sense of humor."*

Other reasons for sharing that commonly co-occurred with impression management were "Someone in Photo" (22.2%), "Physical Appearance" (18.5%), and "Image Quality" (14.8%). Images shared because of impression management often had text visible (59.3% of the images), computer screens (66.7%), and documents (40.7%). Reasons not to share that were related to impression management included "Object in the image" (45.7%), "Activity in the image" (45.7%), and "Someone in Photo" (34.3%), and often had computer screens (75%), text (25%), or documents (12.5%). It thus appears that participants tried to prevent association with certain objects and activities, and to associate themselves with high-quality photos and good physical appearance.

Privacy of Bystanders

We found that lifeloggers based sharing decisions not only out of concern for their own privacy, but also out of concern for others. In particular, 36% of semi-private photos and 17% of private photos were not shared because of "someone in the photo." In some of these cases, the participant was referring to themselves, because of a photo in the bathroom mirror. However, our content coding showed that 68% and 31% of these photos contained someone other than the participant, and specifically mentioned the privacy of another person for 12% and 8% respectively. Thus participants avoided sharing a large number of photos because of bystanders in the image.

Privacy seems to be somewhat related to the presence and number of people within a photo. Private photos were less likely to have any people in them, as only 23.3% of private photos had any people compared to 47% of semi-private photos, (χ^2-stat $= 12.2, df = 1, p < 0.001$). Similarly, they had significantly fewer people than photos shared semi-privately (based on analysis of number of people in a photo regressed on sharing status of private vs. semi-private, accounting for clustering of observations within users: β coeff. $= -0.429$, robust std. err. $= 0.100, p < 0.001$). These results hold if analysis is restricted to only photos with any people in them.

A wide range of images attracted concern about other people's privacy. P60 captured one of a woman whom they were chatting with, and chose not to share it because *"my friend should have a right to her privacy."* P42 chose not to share a photo of a woman sitting in a chair at home, saying *"my mother would be embarrassed by the photo taken without her knowing."* P55 captured a woman in the background of a photo but said *"I do not know this person and am unwilling to share their face publicly."* Finally, for a photo of a dorm room showing a bunk bed and some wall decorations, P38 indicated *"this picture is of someone else's bedroom. It is private, and should not be shared without their permission."*

DISCUSSION AND LIMITATIONS

Our long-term goal is to develop automated mechanisms that will help curate a user's lifelogging pictures, scanning for private content based on user-supplied privacy policies. Jana *et al.* [4] and Thomaz *et al.* [9] preserve the privacy of the lifelogger by extracting non-visual semantic information from images, but do not help users find which photos to share, which is a major goal of lifelogging. Rawassizadeh [7] postulated that lifelogging data would be shared and studied how to model sharing patterns. O'Hara and Tuffield [6] propose that the primary use of lifelogging data is for sharing and public consumption, and that privacy concerns that had been the focus of lifelogging research were not targeted appropriately. In their view lifeloggers collect images for sharing, making privacy-preserving transformations counterproductive.

Our study finds that lifeloggers are interested in suppressing private information, typically corresponding to specific objects (especially computer displays and physical documents) and activities. They are also interested in sharing because of presentation management, as well as protecting the privacy of bystanders appearing in a photo. The large number of images that were shared for no good reason, though, suggests that users may be overwhelmed by the number of collected images. Fortunately, most images are banal, without sensitive content, partially since users tended to use *in-situ* physical control over devices in sensitive situations [3]. However, the constant presence of such devices in intimate settings leaves the possibility that sensitive images may leak, underscoring the need for automated techniques to manage privacy. These sharing decisions are likely to be context-dependent and highly subjective based on an individual, suggesting that automatic privacy-preserving mechanisms need to be able to analyze images to detect people and certain objects, places, and activities. Initial work has examined detecting sensitive places and objects in lifelogging streams [8, 5].

The images analyzed in this study were collected over only a week and from participants who are not regular lifeloggers. The consequences of widespread long-term lifelogging are unclear, and are an interesting area for research. Allen [1] postulated that the effects of unforgettable memories could be detrimental. The chore of curating lifelogging images may prove so overwhelming that people give up and stop lifelogging, or perhaps privacy attitudes will change so that people compromise their expectations of privacy.

CONCLUSION

Our analysis of a large-scale lifelogging dataset showed traits that had only been hypothesized before. Not only were lifeloggers concerned about protecting their own private information and managing their own image, but they also considered the privacy of bystanders. We reported on the content of sensitive images with the hope that this may help to develop image analysis programs to gauge the sensitivity of images automatically, in order to help users curate the large amount of data that lifelogging devices collect.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grants CNS-1408730, CNS-1407788, CNS-1016603, CNS-1252697, and IIS-1253549. This work was also partially funded by a Google Research Award and by the Office of the Vice Provost for Research at Indiana University Bloomington through the Faculty Research Support Program.

REFERENCES

1. Allen, A. Dredging up the past: Lifelogging, memory, and surveillance. *University of Chicago Law Review* (2008).
2. Goffman, E. *The presentation of self in everyday life*. Doubleday, Garden City, NY, 1959.
3. Hoyle, R., Templeman, R., Armes, S., Anthony, D., Crandall, D., and Kapadia, A. Privacy behaviors of lifeloggers using wearable cameras. In *ACM Int'l Joint Conference on Pervasive and Ubiquitous Computing* (2014), 571–582.
4. Jana, S., Narayanan, A., and Shmatikov, V. A Scanner Darkly: Protecting User Privacy from Perceptual Applications. *IEEE Symposium on Security and Privacy* (2013), 349–363.
5. Korayem, M., Templeman, R., Chen, D., Crandall, D., and Kapadia, A. Screenavoider: Protecting computer screens from ubiquitous cameras. In *CoRR arXiv Technical Report arXiv:1412.0008* (2014).
6. O'Hara, K., Tuffield, M. M., and Shadbolt, N. Lifelogging: Privacy and empowerment with memories for life. *Identity in the Information Society 1*, 1 (Dec. 2008), 155–172.
7. Rawassizadeh, R. Towards sharing life-log information with society. *Behaviour & Information Technology 31*, 11 (Nov. 2012), 1057–1067.
8. Templeman, R., Korayem, M., Crandall, D., and Kapadia, A. PlaceAvoider: Steering First-Person Cameras away from Sensitive Spaces. In *21st Annual Network and Distributed System Security Symposium (NDSS)* (2014).
9. Thomaz, E., Parnami, A., Bidwell, J., Essa, I., and Abowd, G. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In *ACM Int'l Joint Conference on Pervasive and Ubiquitous Computing* (2013).