Courteous Glass

Jaeyeon Jung
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
jjung@microsoft.com

Matthai Philipose
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
matthaip@microsoft.com

Abstract

Small and always-on, wearable video cameras disrupt social norms that have been established for traditional hand-held video cameras, which explicitly signal when and which subjects are being recorded to people around the camera-holder. We first discuss privacy-related social cues that people employ when recording other people (as a camera-holder) or when being recorded by others (as a bystander or a subject). We then discuss how low-fidelity sensors such as far-infrared imagers can be used to capture these social cues and to control video cameras accordingly in order to respect the privacy of others. We present a few initial steps toward implementing a fully functioning wearable camera that recognizes social cues related to video privacy and generates signals that can be used by others to adjust their privacy expectations.

Author Keywords

Wearable video cameras; privacy

ACM Classification Keywords

H.5.1 [Information interfaces and presentation (e.g., HCI)]: Multimedia Information Systems.

Introduction

Wearable cameras enable many new applications with continuous recording and analysis of scenes that users are

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3047-3/14/09$15.00.
http://dx.doi.org/10.1145/2638728.2641711

facing in their daily lives. For example, Autographer [1] is a lightweight camera that can be worn around the neck. It takes pictures automatically, capturing spontaneous images that users can view later. Integrated with powerful computing and networking capability, Google Glass [2] provides much more than hands-free video recording such as augmented reality features and voice commands. There have been also various research prototypes that attempt to develop end-to-end applications using wearable cameras (e.g., a diet monitoring system [5]).

Privacy issues associated with traditional recording technologies (e.g., CCTV, phones with cameras) still remain, as demonstrated by the ease of finding spycam with a simple online search. As always-on and traveling wherever a wearer goes, these issues would get only worse with wearable cameras. However, preventing unauthorized recording is difficult even with legal enforcement and there are few technical solutions available (except a research prototype by Truong et al. [7]).

In this paper, we focus on well-intended users of wearable cameras and the privacy issues that they face when they encounter people in the field of view (FoV). We argue that these privacy issues arise because wearable cameras violate social norms that people developed around the use of hand-held cameras. First, designed to be small and almost covert, wearable cameras are hard to be noticed by people who are in FoV, thus depriving them from an opportunity to opt out from recording (by walking outside FoV or covering the face). Second, as these cameras are often left on, even the wearer may not be aware of recording and fail to ask the consent to the people being recorded. Indeed, Denning et al. report that 17 out of 31 participants they interviewed expressed the preference for someone to ask their permission before recording them with augmented reality devices [4]. Third, even if the wearer recognizes that some people in FoV expressed their preference not to be recorded, it may take some delay and maneuver to respond (e.g., one needs to hold a button for 6 seconds to switch off Autographer).

We propose that it may be viable to mitigate these privacy issues without reducing the benefits of continuous recording of the events of interest if we can use low-fidelity sensors, especially far-infrared (FIR) imagers to monitor for known social cues so as to determine when is okay to turn on/off RGB cameras which are used to run vision tasks. We argue that having a separate sensor dedicated for enforcing privacy rules would reduce the attack surface and as capturing only low-fidelity thermal images (rather than full fledged RGB images), FIR imagers can do so with the least privilege. We present an initial prototype and a direction for ongoing research to fully develop a wearable camera that is courteous.

Social Norms With Respect To Recording

In the US, different set of laws govern audio and video recordings respectively. Our limited review of these laws (e.g., http://www.rcfp.org/reporters-recording-guide) suggest that recording a private conversation requires consent from one or all parties but the requirement varies state by state. However, video surveillance is mostly allowed without consent in public places (e.g., shopping malls, city streets) although some states ban the use of video or still cameras where the subject has an expectation of privacy (e.g., bathrooms).

Although the legal landscape is a murky and changing, there are a few social norms that people have been using in order to be respectful for others' privacy preferences when shooting video:

- Opt-in: the camera holder obtains consent from the subject usually through a verbal engagement before taking a picture
- Notification: the camera explicitly signals recording (with red light or clicking sound)
- Off-the-record (OTR): the subject sends signals of the desire not to be recorded either trying to block the camera's FoV or getting out of the view.

As we discussed in the previous section, wearable cameras typically violate all these three norms as they are continuously capturing scenes to run various tasks whether people are in the scene or not. As a consequence, wearable cameras may inadvertently record and disclose peoples images against their wish, making them accountable for their actions and locations (e.g., if the image captured a person at a company for a confidential job interview). Note that our proposal is complementary to the solution proposed by Templeman et al. to block capturing blacklisted places [6].

To mitigate such privacy issues, we propose the following courtesy protocol for wearable cameras (Figure 1). We believe that the protocol would have a minimal impact on many tasks that are proposed to run on these cameras if the tasks do not involve people (e.g., reminding people to take a pill before eating, tracking the last location where a car key is left, keeping a record of meals taken). For the protocol to be effective, we assume a *privacy-preserving oracle* that tracks the presence of people and detects off-the-record gestures without recording video images.

```
 1:while (true) {
 2:  if (recording is on) {
 3:   if (a new person enters into FoV)
 4:    turn_off_recording ();
 5:   else if (OTR gesture is detected)
 6:    turn_off_recording ();
 7: } else {
 8:   if (no people are present in FoV)
 9:    turn on recording ();
10:   else if (a new person enters into FoV)
11:    launch the opt-in process ();
12: }
13:}
```

Figure 1: A pseudo code of the courtesy protocol: The goal is to turn on recording devices only when no people are present in FoV or people in FoV have consented for recording.

Now we turn to the question of how to implement this privacy-preserving oracle. First when the recording is on (lines 2 to 6), we may piggyback the tasks of detecting a new person (line 3) and off-the-record gestures (line 5) using standard computer vision algorithms. However, when the recording is off (lines 8 to 11), we need ways to detect people entering in FoV without relying on RGB cameras. Although in principle, one can re-purpose RGB cameras with software stacks that limit the recording functionality, we argue that software-based solutions are prone to attacks (e.g., [3]). To minimize the attack surface, we propose that a separate sensor should be used for enforcing the protocol with hardware-level isolation. Next, we outline our solution.

Courteous Glass

Figure 2 shows our initial hardware mockup. The wearable camera is integrated with a far-infrared (FIR) imager that acts as a privacy-preserving oracle discussed in the previous section. Although currently hardware-based camera isolation is tentatively represented with a webcam cover, we envision that a lightweight control can be easily added to move the cover electronically.

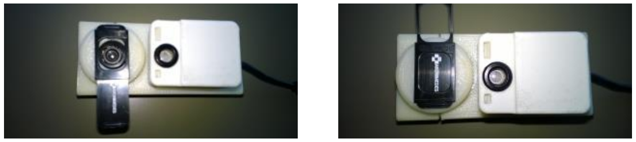

Figure 2: An initial hardware prototype of courteous glass that uses a FIR imager to implement the courtesy protocol shown in Figure 1. The RGB camera can be covered up (as shown in the right picture) when the FIR imager detects new people in FoV who have not consented for recording yet.

FIR Imagers
Our work is motivated in part by the drastic fall in cost of far-infrared (FIR) imagers. These imagers report a temperature reading at each pixel. The reading at a pixel is the average of the estimated temperatures of objects in its field of view. FIR imagers estimate temperature by assuming that objects are non-ideal lack-body radiators at equilibrium. At thermal equilibrium, ideal black bodies produce electromagnetic (EM) radiation in a spectrum specified by Plancks law. Given the spectrum, therefore, it is possible to estimate the temperature of the object. Non-ideal black bodies can be characterized roughly by an additional multiplicative parameter known at their emissivity. Most non-reflective non-transparent objects have an emissivity close to 1, making it possible to estimate the temperatures of most scenes by measuring the EM spectrum per pixel.

The technology for measuring the spectrum at each pixel has grown dramatically less expensive and less bulky over the past decade. As Figure 3 shows, technology has evolved from backpack-sized liquefied-gas cooled systems costing tens of thousands of dollars to through uncooled MEMS-based chips costing several hundred dollars to most recently, uncooled thermopile-based arrays that cost tens of dollars or less. We fully expect at low-resolution imagers to be priced for inclusion into mobile devices over the next few years.

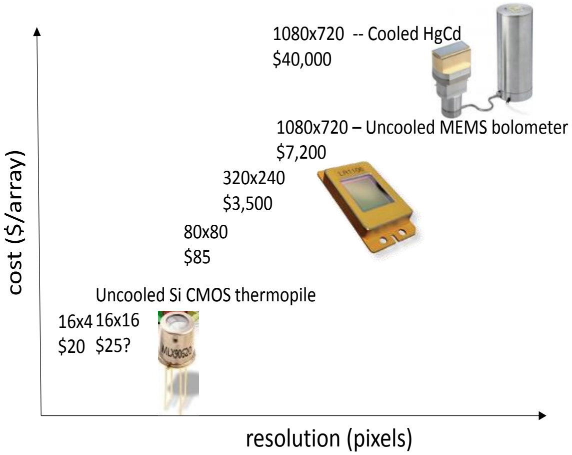

Figure 3: Technology trend over time for measuring the spectrum at each pixel

As Figure 4 shows, given output from an FIR sensor, it is often quite simple to detect whether a human is in the field of view while still not being able to tell who the human is. Over a wide ambient temperature, surface temperatures of faces range between 30 and 35C. Temperature-controlled buildings (e.g., offices, malls and homes) are usually set below 25C, so that the absolute temperature is an excellent means of detecting faces. It also shows that although details are missing, FIR images can capture simple gestures reasonably well, enabling an FIR sensor to run a task of detecting off-the-record signals in line 5 of Figure 1.

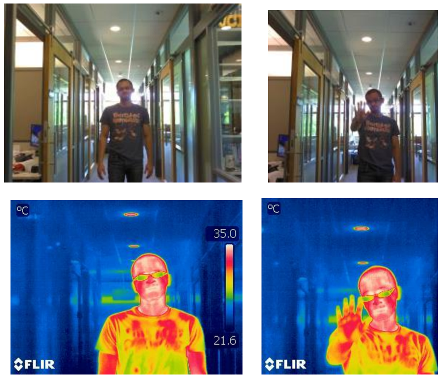

Figure 4: RGB images (top) vs. FIR images (botton): The image on the right shows that a off-the-record gesture (trying to block the camera) can be also detected using an FIR image.

We now examine in more detail the performance of FIR sensors as a "gating" mechanism for detecting people. Ideally, we want good recall *and* precision: we would like to maximize the fraction of true faces detected while minimizing the fraction of pixels falsely detected as faces. The performance of this scheme will of course depend on the precise face-detection algorithm applied to the raw sensor data. Here we consider a simple scheme where if the temperature of a pixel exceeds a fixed threshold, we infer that a face is present at that pixel else we infer no face.

Figure 5 shows the results from a simple experiment estimating the fraction of pixels from FIR footage collected from many thousand frames from warm outdoor (24C ambient temperature), cold outdoor (11C), cold indoor garage (14C), indoor office (21C) and indoor lobby (19C with people constantly coming in from cold outside) settings. We pick a single temperature threshold in these settings and compare the fraction of faces for which at least one pixel is above that threshold (x-axis) to the fraction of pixels that are falsely designated as faces (y-axis). At a threshold of 85F (29.4C), for instance, we detect 89% of faces while only allowing 3% of non-face pixels.

These numbers represent early results using a particularly simple detector. For instance, these pixels could go through another stage of face-detection based on temperature patterns for the face (as opposed to crude single-threshold rejection), yielding false positive faces once every hour or so. Further, the 11% of times when we failed to recognize faces, it was often because the faces were facing away from wearers. When faces directly face the wearers, we expect much higher detection rates, missing perhaps 0.1% of all face-to-face interactions (based on analyzing our current face-on data).

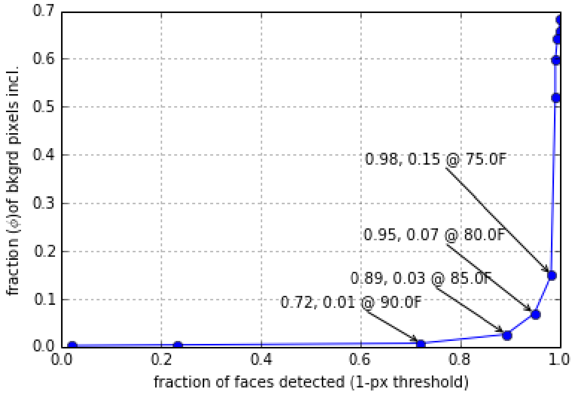

Figure 5: Face detection rates vs. settings

In summary, we believe that FIR sensors are suitable for implementing the courtesy protocol. Although adding an extra sensor would increase the building cost and energy use, we argue that the security and privacy benefits could outweigh them. For instance, with a separate sensor and the software stack for enforcing privacy rules, one can reduce the attack surface (especially if wearable cameras become popular and many apps become available for users to readily install on their devices). Moreover, since FIR sensors capture only low-fidelity thermal images, which might be insufficient for identification while sufficient for detecting faces and simple OTR gestures, by adopting these sensors, we can naturally achieve the principle of the least privilege.

Discussion

Wearable cameras are rapidly gaining popularity yet very little privacy solution exists. This work presents one idea of integrating low fidelity, less privacy invasive, FIR imagers to improve the social acceptability of wearable cameras. Although we believe that it is viable to implement the proposed privacy rules with FIR imagers, there remain several technical challenges: (1) how to improve the detection accuracy of FIR imagers in every environment (e.g., outside on hot summer days)? ; (2) how to detect other social norms related to video recording other than off-the-record gestures (since FIR can also measure breathing and heart rates, these could be used as useful features)?; and (3) what privacy rules to apply in a public setting (e.g., at a party, on a street) and how to automatically enforce such rules using the proposed system?

Along with solving technical challenges, we plan to explore ways to design user studies to test the usability of the prototype and to investigate the change of perception of people around the user of wearable cameras.

References

[1] Autographer. http://www.autographer.com.

[2] Google Glass. ttp://www.google.com/glass/.

[3] Brocker, M., and Chekoway, S. iseeyou: Disabling the macbook webcam indicator led. In *Proc. USENIX Security* (2014).

[4] Denning, T., Dehlawi, Z., and Kohno, T. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proc. CHI* (2014).

[5] Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., and Hansen, M. Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype. In *Proc. the Workshop on Embedded Networked Sensors* (2007).

[6] Templeman, R., Korayem, M., Crandall, D., and Kapadia, A. Placeavoider: Steering first-person cameras away from sensitive spaces. In *Proc. NDSS* (2014).

[7] Truong, K., Patel, S., Summet, J., and Abowd, G. Preventing camera recording by designing a capture-resistant environment. In *Proc. UbiComp* (2005).