Technological Approaches for Addressing Privacy Concerns When Recognizing Eating Behaviors with Wearable Cameras

Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, Gregory D. Abowd

School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

First-person point-of-view (FPPOV) images taken by wearable cameras can be used to better understand people's eating habits. Human computation is a way to provide effective analysis of FPPOV images in cases where algorithmic approaches currently fail. However, privacy is a serious concern. We provide a framework, the **privacy-saliency matrix**, for understanding the balance between the eating information in an image and its potential privacy concerns. Using data gathered by 5 participants wearing a lanyard-mounted smartphone, we show how the framework can be used to quantitatively assess the effectiveness of four automated techniques (face detection, image cropping, location filtering and motion filtering) at reducing the privacy-infringing content of images while still maintaining evidence of eating behaviors throughout the day.

Author Keywords

Privacy; health; dietary assessment; eating behavior; food; diet; wearable; point of view; privacy-saliency matrix.

ACM Classification Keywords

K.4.1 Computers and Society: Public Policy Issues

INTRODUCTION

Over the last few years, wearable cameras have emerged as a new way to capture and record a wide variety of experiences from a first-person point-of-view (FPPOV) perspective. Due in large part to improvements in camera, battery and storage technologies, wearable cameras can be now packaged in a form factor that allows them to be lightweight, unobtrusive, and easy to mount or carry without restricting the wearer's activity. The first truly usable wearable camera was the Microsoft SenseCam (now available as the Vicon Revue), originally designed to explore the domain of passive media capture and personal data management [9]. Recent consumer wearable cameras include the GoPro, Replay and Contour, which are designed specifically for sports activities, and the Looxcie and Memoto, aimed at recording everyday moments for archival and future review. Google's Glass Project has also fueled the excitement for general purpose consumer wearable cameras.

Motivated by the wearable camera's ability to directly and continuously observe and record real-world settings, researchers have begun to explore the potential of FPPOV images in a number of domains, such as autism support [15], travel behavior [12], and activity recognition [19]. Our work is situated in the context of eating behavior recognition, since one of the most promising applications of first-person perspective photos has been automated dietary analysis [18]. A variety of specialized devices have been designed for this purpose, such as the e-Button [4], and various approaches centered on FPPOV have been studied, from assisting people with the recall of their meals over a 24hr period [2] to automatically recognizing an individual's eating activities throughout the day.

Accompanying the exciting possibilities of wearable cameras are a host of technical and nontechnical challenges. One of the fundamental issues is how to process the volumes of data (continuous video or sequence of images taken over the course of a day). Analyzing these images for activities of interest involves reviewing photos manually, a tedious, time-consuming and error-prone task. Although much progress has been achieved in the area of computer vision, state-of-the-art algorithms do not yet yield the automated performance on real-world images that is required for many practical applications. In light of these limitations, an approach that has been embraced by researchers is to use human computation techniques to analyze FPPOV images, using services like Amazon Mechanical Turk or other commercial solutions.

Although human computation has proved to be a viable image analysis alternative to manual or algorithmic techniques, it introduces some challenges of its own, in particular privacy. Privacy is always front and center when it comes to collecting FPPOV images from wearable cameras [8]. These images might capture sensitive information of the person wearing the camera or reveal the identity of others who are

UbiComp'13, September 8–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2493432.2493509

Figure 1. Participants in the study were provided with a mobile phone and a lanyard for wearing around the neck. A custom application on the phone took first-person point-of-view photos every 30 seconds.

captured in the photos as well. If these images were to remain in the exclusive possession of the individual wearing the camera, privacy risks would be kept to a minimum. However, if they are uploaded to be inspected by non-trustworthy third-parties, such as Amazon Mechanical Turk workers, additional precautionary steps must be taken to reduce or, ideally, eliminate the possibility of privacy violations.

Although privacy in FPPOV images has been recognized by the community as an area that deserves furthers exploration, additional studies are needed to examine techniques for mitigating the privacy risk that is inherent in this type of image capture. In this paper, we present a framework, the **privacy-saliency matrix**, to guide our understanding of removing imagery that poses a threat to privacy while retaining (the subset of) imagery that is salient to the analysis of the image, in our case eating behavior. To demonstrate the use of the framework, we quantify how four simple yet practical automated techniques — face detection, image cropping, location filtering, and motion filtering — address the privacy challenge. While we do not expect any of these techniques to be perfect, it is the goal of this work to examine just how good they are in practice, and at what cost.

We present an empirical study in which FPPOV imagery from 5 participants over an average of 3 days each is coded for the saliency of each image with respect to eating behaviors as well as the potential for privacy concerns. We then apply the four automated techniques to see how much of the privacy-sensitive content can be removed while still retaining the salient images for identifying eating behaviors. This is achieved through the privacy-saliency matrix.

RELATED WORK

There is no question that individuals carrying wearable cameras are vulnerable from a privacy perspective, but arguably the group of individuals who are often at most risk when it comes to privacy are secondary participants. These individuals are captured in FPPOV photos as the wearer of the camera frequents physical spaces such as restaurants, parks and public transportation. To make matters worse, these individuals are typically unaware that they are being photographed. Generally, in public spaces in the United States, photos may be taken of others, but there are instances where an individual might have a "reasonable expectation of privacy." In this case, photography or any kind of recording, is a social and potentially legal concern. Langheinrich suggests a set of principles for guiding system design, based on a set of fair information practices common in most privacy legislation in use today: notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, and access and recourse [14]. Several of these principles are challenged with FPPOV imagery used on a continual basis.

A useful way to characterize the dimensions of the problem is through Boyle and Greenberg's theory of privacy [5], which was originally conceived in terms of video-augmented media spaces. It suggests that privacy can be compromised in terms of three modalities of control: solitude, confidentiality and autonomy. Solitude refers to control over interactions between the self and the environment. When someone would like to be left alone, but is forced into some type of interaction, loss of control is established and privacy is compromised. Confidentiality has to do with control over what others know about oneself. An invasion of privacy might occur when personal photos or documents expose an individual in public, for example. Finally, autonomy is the control of one's own behavior and actions, including behaviors around the definition of self and identity. In the context of FPPOV images of everyday experiences, privacy revolves primarily around confidentiality and autonomy issues. When an individual is photographed in public by someone else, the image is kept in the possession of the photographer, leaving no control for the photographed.

Nguyen et al. recognized the need to further understand the impact of recording technologies in and of everyday life [17]. They conducted an extensive cross-cultural study examining how individuals who might be captured in such images perceive and react to this type of photography. Using a paratyping technique [1], they were able to get highly contextualized feedback from hundreds of people regarding the acceptance, characteristics, processes, and policies surrounding FPPOV images. One of the key findings of this research was that people would prefer to be notified about the recordings, if at all possible, but would be unlikely to confront the individual wearing the camera. Also, they would like to be asked for permission in case their images are to be shared with others.

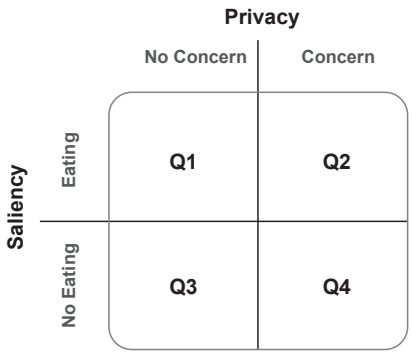

Figure 2. Privacy-saliency matrix provides a framework for studying the balance between privacy concerns and evidence of eating in images.

An effort to address the ethical considerations of wearable cameras in the context of health research was put forth by Kelly et. al [13]. They proposed a set of guidelines in the form of a framework whose aim is to protect all involved when FPPOV photos are being recorded. Extending well beyond privacy concerns, the framework touches on issues ranging from how to educate study participants during the process of gathering informed consent, to best practices that should be followed to keep the autonomy of third-parties.

THE PRIVACY-SALIENCY MATRIX

One of the most constructive ways to address privacy and technology is to make explicit the balance between the positive value proposition of a technology and the negative impact on privacy concerns. Iachello and Abowd portrayed this kind of analysis in ubicomp as a proportionality argument [10]. For FPPOV imagery, the balance is between whether an image contains information considered to be a privacy concern and if that image contains information salient to a particular task at hand, such as eating. For a set of images, we can visualize this balance in a 2-by-2 matrix, the privacy-saliency matrix (see Figure 2).

The two dimensions of the matrix, as the name suggests, reflect the presence of privacy concerns and content salience. In this work, content salience corresponds to evidence of eating behavior or not. Any FPPOV image taken throughout the daily life of an individual can be uniquely placed into a single quadrant of the matrix. Images in **Quadrant 1 (Q1)** contain evidence of eating and exhibit no privacy concerns. For example, these images show people eating by themselves or the camera only captures evidence of the food in front of a person and not any evidence of others who might be around. Images in **Quadrant 2 (Q2)** contain evidence of eating behavior but also exhibit some information that would be considered a privacy concern. Usually, these photos capture people eating with others who can be identified (e.g. friends or family also eating across the table, or strangers who are nearby). Images in **Quadrant 3 (Q3)** do not reveal any eating behavior, nor do they pose any privacy threat. Sending these images to a human computation service is not a problem for privacy reasons, but having too many of them makes the human computation task more expensive and, depending on the information task being presented to the workers, more susceptible to misclassifications. Images in **Quadrant 4 (Q4)** similarly do not reveal any eating behavior, but they do pose a privacy threat.

The privacy-saliency matrix makes it clear how we can understand the opportunities for technology to address the privacy concern for using human computation to identify eating for FPPOV imagery. It also provides a way to quantitatively assess the impact of any given technique or set of techniques. In the context of eating activities, these techniques can be assessed by the following guidelines:

- **Keep images in Q1**: We would like to keep as many images in Q1 as possible, since they show an eating activity without privacy concerns.
- **Eliminate images in Q3 and Q4**: Images in Q3 and Q4 can be eliminated completely since they do not depict an eating activity. As we described above, it is important to remove Q4 images because of privacy concerns. Removing images from Q3 has other non-privacy advantages.
- **Move images from Q2 to Q1**: It would be advantageous to to keep the images in Q2, since they also capture an eating activity. The issue with Q2 images is that they contain one or more elements that pose a privacy risk. The ideal scenario would be to purge the visual component that constitutes that privacy risk while keeping the rest of the image, and thus the evidence of eating behavior, intact. In effect, this corresponds to moving the image from Q2 to Q1.
- **Eliminate images in Q2**: Depending on the approach, it might not be possible to fully suppress the privacy risks of images in Q2 and move them to Q1. A less desirable alternative is to simply delete these images, since they cannot be reviewed by human computation workers. In this case, we want some assurance that the episode of eating evidence by that image removed from Q2 is reflected by an image in Q1 already. For example, if taking pictures every 30 seconds during a meal, it is likely that images within some temporal window of another image might reveal the same eating behavior. This may not hold for shorter duration eating activities, like a snack.

It is important to note that since the ultimate goal is to optimize the multi-variate balance between privacy and content salience for a given application, single-objective measures such as precision and recall are not adequate. The field of multi-objective optimization, also known as Pareto optimization, is concerned with reaching optimality of more than one objective function, and thus comes closest to addressing the privacy-saliency compromise we present in this paper [16]. One way to solve a Pareto optimization problem is by supporting an expert decision maker decide on a solution path to pursue and this is, in effect, the role that we see the privacy-saliency matrix play.

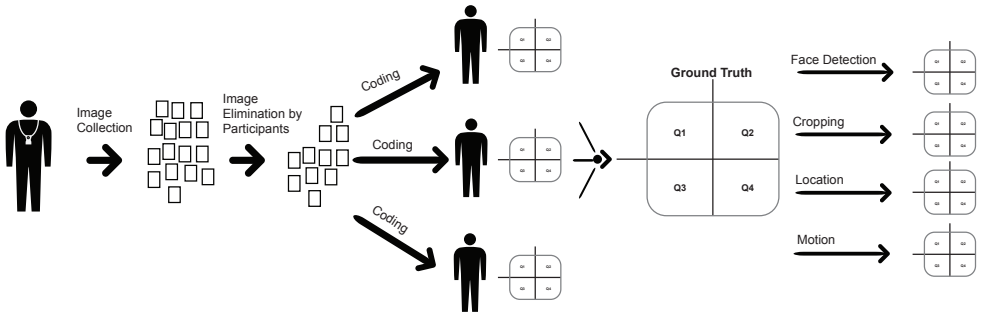

Figure 3. A high-level view of the user study, image coding, and evaluation process. Once participants reviewed and released their images for analysis, the images were coded for evidence of eating behaviors and privacy concerns. Four privacy mitigation techniques were applied on the images separately, and each of the resulting matrices were compared to the privacy-saliency matrix reflecting the images' ground truth.

Participant	Age	Gender	# of Images
P1	31	Male	1230
P2	24	Male	5360
P3	21	Male	2528
P4	23	Male	1958
P5	25	Male	3346

Table 1. We recruited 5 participants to be part of the study. A total of 14,422 first-person point-of-view images were captured and analyzed.

USER STUDY

An IRB-approved user study was conducted with graduate student participants (n = 5, all male) from our university (Table 1). The only criteria that we set for participating in the study was that participants had to be familiar with the operation of a smartphone device and be able and willing to recharge the phone every night. Participants were asked to wear the phone for 3 days.

After going over the study protocol, participants were provided with an iPhone 3GS smartphone preloaded with a custom application that took geo-tagged photos automatically every 30 seconds using the phone's camera. Additionally, the application also received events from the phone's accelerometer sensor continuously and saved it on the device. The sensor data was collected together with the images at the conclusion of the study.

Participants were asked to wear the device as much as possible for the duration of the study; we told them that they could turn off the phone, or take it off, if they did not feel comfortable wearing the device in certain places or situations. All images captured by the mobile application were saved in the phone's default photo library, so participants could review and delete photos whenever they wished. Finally, at the end of the study, participants had the opportunity to review, delete, and get a copy of all captured photos before releasing the images to us. In total the number of FPPOV images collected across all participants was 14422.

METHOD

The methodology for evaluating our privacy mitigating techniques for FPPOV images in the context of eating activity recognition was comprised of two phases. Figure 3 shows the overall workflow. In the first phase, the images were individually coded for evidence of eating behavior and also for privacy threats using the privacy-saliency matrix. The goal was to establish a ground truth baseline for the image set so that we could confidently measure the impact of each automated technique on an image-by-image basis. In the second phase, all images were processed with one of the 4 techniques proposed (i.e. face detection, image cropping, location filtering and motion filtering), and results were compared to the baseline.

Privacy-saliency coding of initial images

The images were reviewed by 3 coders, 2 of whom are authors. To reduce the learning effect caused by reviewing FPPOV images in sequential order, we developed a custom image annotation application that arranged images randomly. Coders viewed images on a grid, and tagged them according to privacy and saliency (as defined on a codebook) using keyboard shortcuts for efficiency. The criteria for a privacy concern was the presence of a human head in the image or any body part thereof (e.g. hair, eye, nose). The head could belong to the participant himself or someone else who happened to be photographed. Evidence of eating behavior was determined to be one or more visual cues that indicated that the participant was engaged in an eating activity, such as the presence of silverware, food on a plate, food in hand, others eating nearby, the identification of a restaurant, etc.

The inter-rater agreement amongst coders on the total of 14,422 images was calculated to be 0.73 (Fleiss' kappa), indicating general agreement. In the case of disagreement, we treated privacy and saliency differently. If any one of the three coders thought that there was a privacy concern in the image, the image was considered to have a privacy concern. The overall categorization on the eating dimension was based on a majority vote by the coders.

PRIVACY MITIGATION TECHNIQUES

In this section, we describe in more detail the four techniques that we implemented with the goal of automating balancing privacy against saliency: face detection, cropping, location filtering and motion filtering.

Face Detection

It is relatively common for faces to be captured in FPPOV images. When this occurs, the identity of the individuals whose faces were recorded is completely revealed, a worst-case scenario in terms of privacy. Ideally, we would like to be able to flag all FPPOV images that contain faces, the images found in Q2 and Q4 in the privacy-saliency matrix, so that they can be either deleted or filtered further. For the analysis in this paper, we simply assume all flagged images are deleted.

We evaluated the performance of two face detection algorithms with respect to its impact in the distribution of images in the privacy-saliency matrix, (1) the one available in the Core Image framework of Mac OS X (10.7 and above), and (2) the set of Haar's cascade classifiers available through the OpenCV library [6]. For the Core Image detector, we implemented an application that leveraged the framework's API. The Haar classifiers consisted of groups of Haar-like features that were learned using Viola and Jones' boosted cascade approach (AdaBoost) for encoding the contrast and spatial relationship of facial features within a window. The Haar Cascade Classifiers were trained on hundreds of face images at similar orientations. Following training, the classifiers were applied to images at multiple scales using a sliding window.

Image Cropping

Recognizing eating behavior in a passive, objective and automated fashion is a hard problem amplified by the fact that eating is often a social activity. Taking photos from a first-person perspective will generally result in images that include other people, such as those sitting across the table or sharing the same environment (e.g. restaurant), a clear privacy risk. This is a typical case where it would be desirable to crop FPPOV images to exclude undesirable elements in the scene (e.g. faces) while retaining the salient content (e.g. evident of eating activity). In the matrix representation previously discussed, this corresponds to the 'Move images from Q2 to Q1' scenario.

The first cropping technique we considered is perhaps the simplest, and hinges on the observation that when people eat, they usually have a plate or food container right in front of them. Thus, when taking photos from a first-person perspective, the bottom-half region of the images is likely relevant to the evidence of eating (Figure 8). The top-half region of the image is usually where faces are located, and can be discarded. We later consider how the position of the POV camera impacts cropping.

We implemented an application in Objective-C for Mac OS X that cropped the bottom-half of participants' images, shrinking the image height in half. Image cropping not only has a desirable effect of eliminating privacy risks, it also has an undesirable potential side effect of deleting the evidence of eating behavior. Therefore, to calculate exactly how this technique performed, all cropped images were coded again for evidence of eating behavior and privacy. Like before, 3 coders reviewed and tagged the images, two of whom are authors. The inter-rater agreement amongst coders in this session was calculated to be 0.8 (Fleiss' kappa).

Location Filtering

The top of Figure 7 shows the geo-location distribution of images for one participant. Red areas of the graph indicate where eating behavior was found in the ground truth coding, and gray areas of the graph are images with no eating behavior. What this plot suggests is that eating activity is localized in space, and this is evident from all of the participants in our study. This empirical evidence reinforces our intuition that routines, such as eating can often be inferred from location data [3, 11]. Most of the eating behavior can be mapped to a small number of locations, such as home and work. Naturally, presence in locations such as restaurants and to a lesser degree bars, are highly correlated with the activity of eating as well. The central idea of this technique is to reduce privacy exposure by considering only the photos that maximize the chance of an eating behavior being recorded. In the privacy-saliency matrix, this technique is aligned with the goal of eliminating photos in Q3 and Q4, whose images do not show evidence of eating activity.

This approach leverages the latitude and longitude metadata embedded in each one of the images captured by participants over the duration of the study. To demonstrate the value and performance of this technique, we show how we can eliminate a significant number of images simply on the basis of their geo-spatial physical distance from the closest image that depicts an eating activity. This distance is calculated from the latitude and longitude of two points using the Haversine formula:

$$d = 2r \arcsin(\sqrt{\sin^2(\tfrac{\Delta\phi}{2}) + \cos(\phi_1)cos(\phi_2)\sin^2(\tfrac{\Delta\lambda}{2})})$$

In a practical application of location filtering, we would infer the likely locations of eating in two ways. First of all, when collecting location and FPPOV images for a longer period of time, previous work shows that it is possible to infer where home and work are for an individual based on location traces alone [3, 11]. Secondly, discovering that an individual is or was at a restaurant can be easily done by looking up the individual's coordinates on a location database, such as the ones provided by Foursquare, Google Places and Yelp. By combining these two methods, we argue that further locations could be feasibly inferred through a semi-supervised learning approach.

Motion Filtering

It is more likely that people are eating when they are not moving. Based on this insight, we implemented a filter that disregards images when the level of motion of the individual wearing the camera around the time the images were taken exceeds a predefined threshold. The objective was to elimi-

nate images from Q3 and Q4 in the privacy-saliency matrix, which do not convey any information as far as eating activities are concerned.

To collect movement data at the time FPPOV photos were shot, we instrumented our image capture application to continuously log the stream of accelerometer events for as long as the application was running. This enabled us to compile sensor data at the moment images were captured and also several seconds before and after. The level of motion, set for each image, was calculated to be the standard deviation of the composite 3-axis accelerometer data (i.e. x, y, and z) over the minute the photo was taken:

$$M_s = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}|(|x_n| + |y_n| + |z_n|) - \mu|^2} * 100$$

where N is sampling rate times number of seconds in a minute. The normalized score value M_s ranged from 0 to 65 and the threshold for eating activities was set to 8. This was determined empirically, based on the distribution of FPPOV images of our study participants. As shown in Figure 6, the distribution of motion intensity for eating images has a range of 1-21 only, which is distinct from the distribution of motion intensity seen in non-eating images. Additionally we verified that these distributions are significantly different with a Kolmogorov-Smirnov test ($p < 0.001$).

RESULTS

A total of 14,422 images were captured in our 5-person study. Figure 4 shows the ground truth coding in terms of the privacy-saliency matrix of the raw FPPOV images. We show the resulting privacy-saliency matrix after each of the four automated techniques are applied to those images.

We ran two face detection algorithms on the participants' images, the one available through the Mac OS X's Core Image framework and the set of of Haar's cascade classifiers available through the OpenCV library. The Haar classifiers outperformed the Core Image detector by an order of magnitude, therefore we are just reporting results with respect to this classifier. As shown in Figure 4, Q2 and Q4 in the privacy-saliency matrix saw the largest decrease in the number of images, in the range of 35% to 42%. Around 13% to 14% of the images in Q1 and Q3 were flagged for containing faces, which is indicative that the face detection algorithm generated false positives, since the images in these quadrants were previously screened for faces by human coders.

Note that we did not measure the performance of the algorithm with respect to its ability to recognize faces. Instead, by assuming the removal of images from the quadrants when the algorithm detected faces in them, we measured how the application of the algorithm modified the distribution of images in the privacy-saliency matrix. One of the reasons why the face detection method did not perform better is because first-person perspective images are often blurry and do not capture faces looking directly at the camera frequently. Nevertheless, as FPPOV images become more popular, it is likely that we will see the development of face detection and other computer vision techniques that are optimized for this type of photography. Also, the privacy criteria that we employed while coding the images was the presence of a human head or any visible part thereof, such as hair, nose, eyes, etc, and not a face. In light of this, many of the images assigned to Q2 and Q4 in the matrix could have never been flagged by face detectors.

With regards to cropping the bottom-half region of the images, it had a positive effect in that it reduced the number of photos with privacy concerns. The number of images in Q2 and Q4 fell around 67% and 30% respectively, as shown in Figure 4. More importantly, the intended effect of having images transition from Q2 to Q1 materialized. Out of 174 images in Q2, 75 moved to Q1. This represents a best case scenario since many images depicting eating activities but compromised by privacy threats had those threats removed with cropping. A smaller but still significant number of images (48) moved from Q2 to Q3. This can be interpreted from two perspectives. On one hand, 48 images that presented privacy issues before no longer did after cropping. This meant that they could be examined by human computation workers without the risk of a privacy violation, for example. On the other hand, the evidence of eating activities in the images is no longer present, so from the point of view of eating behavior recognition, these images do not hold any useful information anymore.

Location filtering proved to be an effective approach for removing images that do not include evidence of eating activity. When considering photos within a radius of 0.2 km of a known eating location, images in Q3 and Q4 fell by 46% and 40.89% respectively. However, as previously discussed, the condition under which these results were obtained is when all eating locations are known. If that is the case, all instances of eating activity are accounted for, and thus there is no loss of images in Q1 and Q2 (no percentage change in the number of images). Unfortunately, all collected location data for one of the participants was corrupted and had to be discarded. This required us to generate ground truth quadrant numbers for the privacy-saliency matrix with 4 participants instead of 5. This is the reason why the numbers in Q1 and Q2 differ from those in the ground-truth privacy-saliency matrix in Figure 4.

Motion filtering performed similarly to location filtering in terms of the reduction of images in Q3 and Q4. Q3 saw a decrease of 35.57% in its images and the number of images in Q4 fell by 41.49%. Because of the need to establish a range in the motion score under which an eating behavior is most likely to occur, it is always the case that some images representing eating activities end up outside of that range and are disregarded. This is why the privacy-saliency matrix for motion filtering shows a decrease in the number of images in Q1 (24.47%) and Q2 (20.69%). Without a doubt, this decrease is undesirable, but it is less pronounced than the loss of images in Q3 and Q4. Overall, the collective loss of images in all quadrants, affecting Q3 and Q4 to a higher degree, underscores the trade-off between capturing activities of interest and mitigating privacy concerns that lies at the core of this paper.

Figure 4. The privacy-saliency matrices showing the coded distribution of images before the application of the privacy mitigation techniques (ground truth) and after. Note that due to corrupted data, the location filter could be applied to images from 4 participants only. The matrix in the bottom-right corner shows how images transitioned from one quadrant to another after cropping. The arrows in green show transitions that we consider 'good' (e.g. reduction of images with privacy concerns), while red arrows highlight transitions that we consider 'bad' (e.g. removal of evidence of eating behavior).

ADDITIONAL PRIVACY RISKS

Though we followed a strict criteria of marking all the images that had any part of the head as a privacy threat, we discovered several other categories of threats while coding the images. In some instances, information captured in an image could be linked back to an individual. For example, personal id, credit card number, cell phone usage, email screen. In other cases, the display of jewelry, tattoos, clothes could help an acquaintance identify an individual. Furthermore, a silhouette could provide enough information for a friend or family member to infer identity. A non-obvious threat emerged as a result of analysis of one participant's images of a meeting where under-table shots had potential of providing compromising information about secondary participants.

Our IRB mandated us to mark all images that contained any personally identifiable information like face, accessories, and tattoos. Although we found the IRB requirements to be restrictive, our findings suggested a more complex definition of privacy, one that begs understanding of the relationship between the secondary participants and third party that looks at the images. For example, an email of a person becomes more important than the jewelry or tattoo when an image is shown to a third person. However, it is not easy to establish that relationship when an image becomes publicly available hence most stringent rules should be imposed in those cases. But in the cases where access is limited to a set of third party members such as coders or Mechanical Turkers, some criteria could be overlooked without compromising privacy.

An important and somewhat paradoxical condition that our work does not take into account is when the recording of an eating activity represents a privacy violation. In a survey focusing on the activities and habits that people do at home that they would not want recorded, Choe et al. found that the "cooking and eating" category ranked third, behind the self-appearance and intimacy categories [7]. This finding underscores the complexity of the privacy-saliency balance, in particular when there is an overlap between the two.

THE IMPACT OF CAMERA POSITION

In this paper, we studied techniques that address concerns in FPPOV images when these images are taken at chest level with a mobile phone camera. There are several other locations where a wearable camera could be mounted in the body. To gain a better understanding of the impact of camera position on the privacy-saliency balance, one of the authors of the paper wore two wearable cameras, in the head and neck, for a period of 5 days. The head camera employed in this informal study was the Contour ROAM Model 1600, and the neck camera was custom designed through the combination of a Logitech Webcam 210, a Raspberry Pi Model B and a PowerGen 8400mAh External Battery (Figure 5). The number of images collected for the head and neck cameras

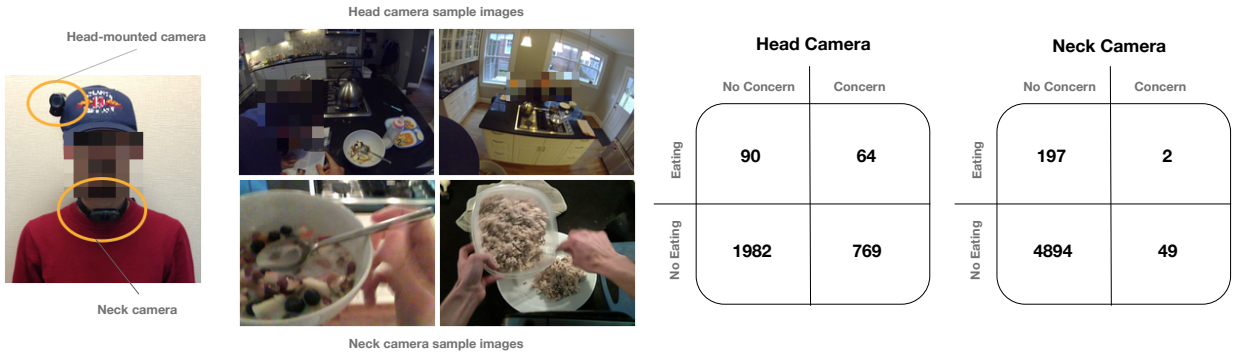

Figure 5. We also collected first-person point-of-view images from two additional camera positions, head and neck, and coded them for privacy concerns and evidence of eating behavior. The neck camera images were particularly interesting, since they often captured the individual eating but not the features that typically characterize privacy concerns.

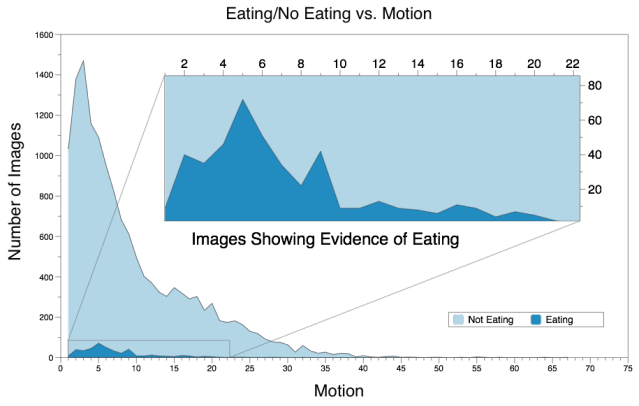

Figure 6. We computed a measure of human motion intensity by leveraging accelerometer data from the mobile phone camera. By adding up the number of images in each quadrant of the privacy-saliency matrix by level of motion, it is possible to see that the most eating activities are contained within a region of motion that range from 1 to 21.

were 2,905 and 5,144 respectively. Both cameras were set to take photos every 30 seconds, but battery life prevented them from operating throughout an entire day. In particular, since the Contour camera is designed to capture specific moments in sports activities, it could only record photos for 3 hours at a time before having to be recharged. This is why there was a discrepancy in the number of images collected by both cameras. The author who wore the cameras for the informal study coded all captured images using the same methodology and criteria employed with the images collected in the 5-participant study.

Thanks to its wide angle lens and high-mounting position, the head camera captured images that reflected not only what was immediately in front of the person, but also most of the surrounding context. On the positive side, all eating activities were recorded, even when the individual was having a small snack and not sitting on a dining table, which usually makes it easy to identify an eating activity. On the other hand, any person in the vicinity of the individual wearing the head camera was also captured in the photos, which resulted in many images being flagged for privacy concerns. This was quantified in the ratio of total images to images with privacy concerns for the head camera, 3.48. For comparison, this same ratio for the chest-level FPPOV images collected in the 5-participant study was 5.45.

The FPPOV images from the neck camera were very promising. We found the set of neck images to contain fewer blurry photos compared to the chest and head images, most likely because the neck camera did not move as much as the other cameras. In terms of orientation, the inclination caused by the neck collar made the camera point down when worn, capturing what was always in front of the individual, but at a downward angle. This proved to be an excellent way to remove many of the privacy issues that affect FPPOV images, since the photos did not record people's faces, which were always above the viewport of the camera. The ratio of total images to images with privacy concerns for the neck camera amounted to 100.8 (approximately 1 image with privacy concerns for every 100 images). Additionally, the angled camera pointed straight at what people were eating in front of them. These results, though preliminary, provide evidence that we should investigate camera position in more detail in future work.

FUTURE WORK

Based on this initial exploration of the space, we see opportunities for future work both in terms of extending the research we have done so far and in improving our methodology.

First of all, it is imperative that we validate our methodology and results with a larger cohort of participants, ideally from different backgrounds, demographics and social-economic status. In this study, all participants were graduate students; follow up studies will include subjects from all walks of life. More specifically, it would be highly relevant to involve participants who are personally motivated in keeping a food journal and also those who might have a different or broader perspective on privacy. Second, we estimated ground truth for eating behaviors and privacy with a group of coders that included two authors. In an effort to improve the validity of

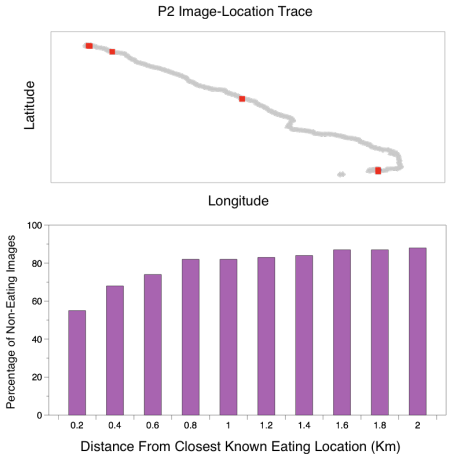

Figure 7. The top chart shows a location trace of one of the participants in the study. Each point in the trace corresponds to a FPPOV image automatically taken with the wearable camera. From the distribution of photos, it is possible to see that photos with evidence of eating activity (red squares) are clustered around a few locations only. The bottom chart illustrates the positive correlation between the number of images depicting non-eating activities and the distance between the location the image was taken and the closest known eating location.

Figure 8. Several images that contain evidence of eating behavior might pose a privacy concern. By cropping a portion of the image, it is often possible to eliminate privacy issues.

our work, our aim is to compose a larger and more diverse coding panel in the future. Third, we would like to study in more detail how to leverage established multi-objective optimization techniques in our privacy-saliency framework.

In terms of the privacy techniques applied to the FPPOV images, we chose simple yet practical approaches that contributed to the challenge of privacy mitigation while illustrating a real application of privacy-saliency matrix. In our view, the techniques that we examined can be seen as building blocks, and it is likely that bringing them together will lead to more powerful and effective mechanisms for reducing privacy risks. For instance, the location and motion filters used in combination can provide more effective removal of images not related to eating. We did not implement a true geocode-based location filter in our study, but had we done so, the location information alone might highlight images when the user was walking by a restaurant. In that case, the motion filter would be useful to remove that image. Similarly motion at work or in the home might help to remove images that are not likely to contain eating in those locations.

Face detection is exactly that — it detects a face. However, the appearance of a head at any orientation relative to the camera could be considered a privacy risk. So we should consider using a more general head detection solution. Similarly, we might want to detect any imagery on human skin to detect distinguishing characteristics like tattoos or jewelry. Non-human imagery, like computer screen images, credit card numbers or house numbers could also be detected and be criteria for deletion or cropping of imagery. We intend to explore these techniques more rigorously and develop a framework for combining and assessing their effectiveness.

Finally, we discussed promising results from an informal study where cameras were placed in the head and neck of one of the authors, capturing FPPOV images from those perspectives for a period of 5 days. The neck camera proved to be particularly good at not capturing images with privacy issues, because of its position and orientation on the body. We plan to run a more rigorous study in the future and report on the balance between privacy and observing eating behaviors from those camera angles.

CONCLUSIONS

The overarching impetus for this work was the need and desire to tackle the privacy problem in FPPOV imagery. Although privacy in FPPOV images has been recognized by the community as an area that deserves furthers exploration, few studies have investigated this domain in a practical and quantifiable manner.

The contributions of this paper are threefold. Firstly, we introduce a formulation around FPPOV imagery for studying and quantifying the balance between a set of images that might pose a privacy concern versus images that contain information salient to a particular task of interest. By making this balance evident, the representation informs what steps one might take to achieve specific goals, such as retaining images with certain characteristics while discarding others. The formulation centers on a 2-by-2 matrix that we call the privacy-saliency matrix. In our specific case, the privacy-saliency matrix makes it clear how we can understand the opportunities for technology to address the privacy concern for using human computation to identify eating.

Secondly, we show how sensor data such as geolocation and data streams collected from mobile phone accelerometers can be combined with FPPOV images in the specific service of addressing privacy concerns when the goal is to to recognize behaviors of interest (e.g. eating) with wearable cameras. The location and motion filtering techniques demonstrated in this paper successfully leveraged sensor data to

determine the likelihood that an eating activity was occurring, thus reducing the need to consider photos that might introduce privacy issues.

Thirdly, using the privacy-saliency matrix representation, we evaluated the performance of four simple techniques (i.e. face detection, cropping, location filtering, and motion filtering) at mitigating the privacy risks incurred when capturing FPPOV images. This evaluation was based on an empirical study in which FPPOV imagery from 5 participants was collected over an average of 3 days, totaling 14,422 images. As expected, none of these techniques are particularly good at optimizing both the privacy and saliency of images to desired levels, but they expose the need for mechanisms that support reasoning about this optimization, which we believe our proposed framework does.

ACKNOWLEDGEMENTS

We would like to thank the Intel Science and Technology Center for Pervasive Computing (ISTC-PC) for supporting this work.

REFERENCES

1. G. D. Abowd, G. R. Hayes, G. Iachello, J. A. Kientz, S. N. Patel, M. M. Stevens, and K. N. Truong. Prototypes and Paratypes: Designing Mobile and Ubiquitous Computing Applications. *IEEE pervasive computing*, 4(4), Oct. 2005.
2. L. Arab, D. Estrin, D. H. Kim, J. Burke, and J. Goldman. Feasibility testing of an automated image-capture method to aid dietary recall. *European journal of clinical nutrition*, 2011.
3. D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Computing*, 7(5):275–286, Oct. 2003.
4. Y. Bai, C. Li, Y. Yue, W. Jia, J. Li, Z.-H. Mao, and M. Sun. Designing a wearable computer for lifestyle evaluation. In *Bioengineering Conference (NEBEC), 2012 38th Annual Northeast*, pages 93–94, 2012.
5. M. Boyle and S. Greenberg. The language of privacy: Learning from video media space analysis and design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):328–370, 2005.
6. G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
7. E. K. Choe, S. Consolvo, J. Jung, B. Harrison, and J. A. Kientz. Living in a glass house: a survey of private moments in the home. *Proceedings of the 13th international conference on Ubiquitous computing*, pages 41–44, 2011.
8. A. R. A. Doherty, S. E. S. Hodges, A. C. A. King, A. F. A. Smeaton, E. E. Berry, C. J. A. C. Moulin, S. S. Lindley, P. P. Kelly, and C. C. Foster. Wearable cameras in health: the state of the art and future possibilities. *American journal of preventive medicine*, 44(3):320–323, Mar. 2013.
9. J. Gemmell, L. Williams, K. Wood, R. Lueder, and G. Bell. Passive capture and ensuing issues for a personal lifetime store. *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 48–55, 2004.
10. G. Iachello and G. D. Abowd. Privacy and proportionality: adapting legal evaluation techniques to inform design in ubiquitous computing. In *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 91–100. ACM, Apr. 2005.
11. J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. In *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, WMASH '04, pages 110–118, New York, NY, USA, 2004. ACM.
12. P. Kelly, A. Doherty, E. Berry, S. Hodges, A. M. Batterham, and C. Foster. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):44, May 2011.
13. P. Kelly, S. J. Marshall, H. Badland, J. Kerr, M. Oliver, A. R. Doherty, and C. Foster. An ethical framework for automated, wearable cameras in health behavior research. *American journal of preventive medicine*, 44(3):314–319, Mar. 2013.
14. M. Langheinrich. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. pages 273–291. Springer Berlin Heidelberg, Berlin, Heidelberg, Oct. 2001.
15. G. Marcu, A. K. Dey, and S. Kiesler. Parent-driven use of wearable cameras for autism support: A field study with families. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 401–410, 2012.
16. K. Miettinen. *Nonlinear Multiobjective Optimization*. Springer, New York, 1999.
17. D. H. Nguyen, G. Marcu, G. R. Hayes, K. N. Truong, J. Scott, M. Langheinrich, and C. Roduner. Encountering SenseCam: personal recording technologies in everyday life. *Proceedings of the 11th international conference on Ubiquitous computing*, pages 165–174, 2009.
18. M. Sun, J. D. Fernstrom, W. Jia, S. A. Hackworth, N. Yao, Y. Li, C. Li, M. H. Fernstrom, and R. J. Sclabassi. A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association*, 110(1):45, 2010.
19. H. Zhang, L. Li, W. Jia, J. D. Fernstrom, R. J. Sclabassi, and M. Sun. Recognizing physical activity from ego-motion of a camera. *Proceedings of IEEE EMBS*, 2010:5569–5572, 2010.