Published in final edited form as:

Proc ACM Interact Mob Wearable Ubiquitous Technol. 2019 September ; 3(3): . doi:10.1145/3351230.

To Mask or Not to Mask? Balancing Privacy with Visual Confirmation Utility in Activity-Oriented Wearable Cameras

RAWAN ALHARBI,
Northwestern University

MARIAM TOLBA,
Northwestern University

LUCIA C. PETITO,
Northwestern University

JOSIAH HESTER,
Northwestern University

NABIL ALSHURAFA
Northwestern University

Abstract

Activity-oriented cameras are increasingly being used to provide visual confirmation of specific hand-related activities in real-world settings. However, recent studies have shown that bystander privacy concerns limit participant willingness to wear a camera. Researchers have investigated different image obfuscation methods as an approach to enhance bystander privacy; however, these methods may have varying effects on the visual confirmation utility of the image, which we define as the ability of a human viewer to interpret the activity of the wearer in the image. Visual confirmation utility is needed to annotate and validate hand-related activities for several behavioral-based applications, particularly in cases where a human in the loop method is needed to label (e.g., annotating gestures that cannot be automatically detected yet). We propose a new type of obfuscation, *activity-oriented partial obfuscation*, as a methodological contribution to researchers interested in obtaining visual confirmation of hand-related activities in the wild. We tested the effects of this approach by collecting ten diverse and realistic video scenarios that involved the wearer performing hand-related activities while bystanders performed activities that could be of concern if recorded. Then we conducted an online experiment with 367 participants to evaluate the effect of varying degrees of obfuscation on bystander privacy and visual confirmation utility. Our results show that activity-oriented partial obfuscation (1) maintains visual confirmation of the wearer's hand-related activity, especially when an object is present in the hand, and even when extreme filters are applied, while (2) significantly reducing bystander concerns and enhancing bystander privacy. Informed by our analysis, we further discuss the impact of the filter method used in activity-oriented partial obfuscation on bystander privacy and concerns.

Authors' addresses: Rawan Alharbi, Northwestern University, rawan.alharbi@northwestern.edu;.

1 INTRODUCTION

Wearable cameras are used as a tool to understand fine-grained human activities in the wild because of their ability to provide visual information that can be interpreted by humans [15, 45, 55] or machines [6, 43, 48]. Particularly in the ubiquitous computing (UbiComp) community, wearable cameras are increasingly being used to obtain visually confirmed annotations of wearers' activities in real-world settings, which is necessary to both understand human behavior at a fine-grained level, and build and validate non-visual wearable devices and their corresponding supervised machine learning algorithms to automate the detection of human activity [4, 8, 9, 61, 80]. However, the stream of images obtained from these wearable cameras embeds more details than needed; information that can expose bystanders—particularly in situations that can be embarrassing (e.g., fighting, binge drinking)—change wearer behavior, lead to wearer discomfort or stigma, trigger device abandonment [4, 18], and prevent the ability to understand naturally occurring behavior in real-world settings.

In real-world settings, bystander concerns emerge because they are not aware of the purpose or the scope of the recording (low situation awareness), which also leads to wearer worry about justifying the camera use to bystanders. Researchers have proposed many design solutions that can be implicitly communicated to the bystander in the wild for the purpose of increasing bystander situation awareness [36]. *Activity-oriented cameras* are one of the solutions that are currently used to communicate the scope and the intention of the recording by introducing a set of physical constraints (e.g., lens orientation, camera location) on the camera to restrict data collection to specific wearer activity [4]. Activity-oriented cameras, unlike egocentric cameras, are used to capture a specific type of activity. For example, researchers interested in obtaining information about human eating behavior have used an activity-oriented camera mounted on the chest with the lens pointing up toward the face [8] or a camera mounted on a hat with the lens pointing toward the mouth [9]. Although activity-oriented cameras have the potential to communicate camera scope and intention to bystanders, in some cases more information than intended can be captured (e.g., when using a fish-eye lens [4]), which might create a mismatch between the actual scope of camera recording and the scope implicitly communicated by the device. One way to close this gap is to reduce the extra information recorded by the camera by applying image obfuscation methods prior to data storage to increase trust in the device and prevent information misuse.

Image obfuscation using image transformation methods or filters (e.g., blur, edge, masking, in-painting, abstraction) can be applied to limit the information shared in the image to enhance perceived or actual privacy [11, 52]. These obfuscation filters are either applied to the entire image (*total obfuscation*) [21, 63] or to part of the image (*partial obfuscation*) [29, 42, 50]. Total obfuscation can enhance privacy, especially as the intensity of the obfuscation method increases [21]. However, in total obfuscation, a limit on the intensity of obfuscation is necessary in order to maintain the utility of the resulting video. Partial obfuscation, on the other hand, is applied to only parts of the image making it possible to increase the intensity of obfuscation to enhance privacy while still maintaining its utility for human viewers [29, 42, 50]. Previous work [29, 42, 50] applied partial obfuscation using a blocklist-based approach which allows one to define a list of items or objects, like a face, to be removed

from the image. However, rather than defining a list of specific objects to block, we use a type of partial obfuscation that we call *activity-oriented partial obfuscation* in which we obfuscate all pixels except the ones surrounding the wearer's activity of interest, in our case hand-related activities. In the context of activity-oriented wearable cameras, it is easier to define the activity of interest, rather than defining all possible objects to be blocked. Activity-oriented partial obfuscation lies on a spectrum between blocklist obfuscation and total obfuscation (see Figure 1), which can significantly affect the amount of contextual information lost, impacting both privacy and utility. Understanding the tradeoff between privacy concerns and utility (*privacy-utility tradeoff*) with wearable video cameras used in the wild can help us obtain a balance between the need for visual confirmation of fine-grained human behavior while enhancing bystander privacy.

Here, we focus on a specific type of image utility, *visual confirmation utility*, which is the ability of human viewers to recognize the activity of the wearer in the recorded video stream. We also define *bystander privacy* as a viewer's inability to determine the activity of the bystander. Motivated by existing research that seeks to collect information related to behaviors such as eating [8–10, 15, 25, 81], drinking [5], and smoking [53, 64], we assessed the effect of activity-oriented partial obfuscation on activity-oriented cameras designed to capture a group of hand-related activities that involves hand-to-head gestures. To the best of our knowledge, no one has investigated the effect of applying activity-oriented partial obfuscation with different filters on both visual confirmation utility and bystander privacy. In particular, **we aim to answer the following research questions (RQ) to improve the quality of data collection using wearable activity-oriented cameras, as it will shed light on how to enhance bystander privacy while maintaining the utility of the video collected in a real-world setting**.

RQ1: How do different activity-oriented partial obfuscation filters affect the visual confirmation utility of identifying hand-related activities that involve hand-to-head gestures by a human viewer?

In particular, we want to compare the accuracy of human labels obtained from viewing non-obfuscated videos with the accuracy of the labels derived from viewing the obfuscated videos with different filters. Hand-to-head gestures can be confounding to each other if fine-grained and some contextual information is lost. Therefore, this comparison can help us to determine if the visual confirmation utility is preserved, or not, after applying activity-oriented partial obfuscation to it with different filters. It will also help us to understand the limitations of activity-oriented partial obfuscation and the filters applied.

RQ2: How do different activity-oriented partial obfuscation methods affect (a) bystander privacy and (b) concerns?

Different obfuscation filters result in varying degrees of obfuscation intensity, varying the amount of information loss in the image, which impacts the ability of a viewer to determine the activity of a bystander, and as a result impacts bystander concerns. We want to compare the human viewer's ability to identify the bystander's activity in obfuscated videos of varying intensities compared with when no obfuscation is applied. Also, we want to understand the effectiveness and limitations of each obfuscation method on hiding

information about the bystander's activity and whether the method has an impact on bystander concerns.

RQ3: Which of the five obfuscation methods tested achieves greatest visual confirmation utility and bystander privacy while reducing concerns when collecting hand-to-head activities and why?

After answering RQ1 and RQ2, we investigate the privacy-utility tradeoff of each obfuscation method in order to shed light on which obfuscation methods are most optimal based on a tradeoff between visual confirmation and bystander privacy. This information will help guide future research in selecting the appropriate obfuscation method and intensity for capturing hand-related activities using an activity-oriented wearable camera.

To answer our research questions, we collected 10 video scenarios where an actor wore an activity-oriented camera aimed at capturing a specific group of hand-related activities that involves hand-to-head gestures that are of interest to the research community. The wearer performed hand-to-head gestures found to confound each other, along with other activities that do not contain hand-to-head gestures. Confounding activities can appear similar to each other if fine-grained information about the gesture or the context is lost due to obfuscation (i.e., activities that contain hand-to-head gestures such as eating, drinking, yawning, or calling), which allows us to test the effectiveness of partial obfuscation on preserving fine-grained information needed to distinguish confounding gestures. The wearer also performed other activities that do not contain hand-to-head gestures to provide insight into other hand-related gestures. To understand bystander privacy, wearer activities were captured in situations in which previous work suggests concern regarding bystander privacy [16, 19, 32]. We applied activity-oriented partial obfuscation with five commonly used filters (blur, blur-high [blurH], edge, edge-high [edgeH], and mask) on all pixels in the video except the ones related to the wearer's hand-to-head gesture. We used these scenarios in an online experiment (N=367) to study the effect of the five different filters used with the activity-oriented partial obfuscation methods on visual confirmation utility and bystander privacy and concerns. Data were analyzed quantitatively and qualitatively.

We show that activity-oriented obfuscation maintains visual confirmation utility for hand-related activities that involve an object in hand, even when extreme filters are applied to obfuscate the background. We also show that bystander concerns are significantly reduced as the intensity of the obfuscation increases. We noted that with activity-oriented partial obfuscation, participant-reported concerns stem from the perceived interpretation of bystander activity (regardless of whether it is the correct activity or not), especially when the filter intensity is low. Therefore, in the case of activity-oriented partial obfuscation, it is important to not only think about reducing information about the bystander's true activity but also to think about how the obfuscation method may impact the interpretation and potential misinterpretation of the activity being performed. It is also important to assess how misinterpretation of activity may be more concerning to the bystander than a correct interpretation of the actual activity recorded.

2 BACKGROUND AND RELATED WORK

2.1 Bystander Privacy Concerns in Wearable Cameras

Bystander concerns associated with wearable cameras are one of the main discomforts expressed by the wearer (even when no discomfort is apparent by bystanders) [4, 31, 32, 57] and by the bystanders themselves [19, 47]. Research in understanding bystander concerns of wearable cameras have inspired others to address bystander concerns using multiple methods. These methods include privacy-mediating systems that allow bystanders to opt-out of recordings [1, 38, 68], design approaches that increase bystander situation awareness (e.g., lens orientation) [4, 36], and system approaches that prevent recording when a specific context (e.g., bathroom) or activity (e.g., typing) is detected [24, 71, 78]. These approaches can complement approaches that aim to reduce some of the information shared using obfuscation [12, 30, 37, 51, 63, 72]. For example, if context-aware privacy systems detect a specific context in the video that is intended to remain hidden (e.g., being around friends) instead of using a binary (share or not to share) decision regarding the information collected, obfuscation methods can be utilized to remove the sensitive information that triggered the contextual privacy system (e.g., friends) and keep other valuable information related to the wearer activity.

Previous work in obfuscation has used a blocklist approach (i.e., where a list of items, such as faces, to be removed from the video are predefined) to determine the portion of the image to be obfuscated (*blocklist partial obfuscation*) in the context of sharing images in social media [29, 42, 51] and in crowd-sourcing platforms to obtain visually confirmed activity labels from stationary cameras [40]. However, in the case of wearable cameras, the construction of this blocklist is impractical as cameras are worn in more contexts than can be envisioned and as privacy depends on context [13, 62], making the block-by-default approach to obfuscation more practical [59]. Dimiccoli et al. [21] have investigated a block-by-default approach method where the entire area of the image is obfuscated (*total obfuscation*), and they showed that obfuscation reduced bystander concerns, especially when the intensity of the obfuscation method is high. However, such approaches also reduce the important and necessary fine-grained context in the video, limiting the utility of the recorded information. Another way to utilize the block-by-default method is to use the partial obfuscation-by-default approach where only a specific area of interest in the image is shown and the obfuscation method is applied to everything else by default [59]. In this paper, we use the wearer activity as the guide in obfuscation-by-default, and therefore we call it *activity-oriented partial obfuscation*. The area in the image where obfuscation is applied in the activity-oriented partial obfuscation case lies on a spectrum between total obfuscation and partial obfuscation by blocklist (see Figure 1). It has been shown that both the area obfuscated and the intensity of the obfuscation can have an effect on privacy [42, 49]. To the best of our knowledge, no one has investigated the effect of activity-oriented partial obfuscation using different obfuscation methods and intensities on bystander privacy. It is also not known, after viewing the obfuscation method, how viewers feel about being around someone wearing a camera capable of such a partial obfuscation method.

2.2 Privacy and Image Utility Tradeoff

Privacy-enhancing obfuscation methods can also affect image utility, which motivated researchers to study the tradeoff between privacy and image utility [13, 29, 42, 50]. In prior literature, image utility is either measured from the perspective of a human viewer or from the perspective of a machine trained to detect specific objects, such as faces. Image utility from a human perspective has been studied in domains where human input is crucial. For example, in the context of sharing images in social media, Hasan et al. [29] defined the utility of the image as sufficient information that the user (human) wants to convey to others while taking into consideration aesthetics. Orekondy et al. [50] defined image utility as the ability of the image to convey semantic information in the context of sharing images in social media independent of aesthetic. Another case where utility to humans is important is the case of human-in-the-loop systems. For example, in the context of sharing images in teleoperated robots, Butler et al. [13] defined image utility as the information sufficient for a human to teleoperate a robot effectively without causing harm.

Machine learning algorithms showed potential in automating some of the human viewers' tasks such as object and activity recognition [48], especially when a human-labeled dataset is available to train a model for the recognition task at hand. In cases where automatic activity detection is feasible, image utility is defined as the information sufficient enough for an algorithm to detect an object or to classify an activity. However, in order for a model to be developed, annotators must first accurately label video data. In this work, we are concerned with visual confirmation utility to ensure proper ability of annotators (i.e., human viewers) to accurately infer the activity of the wearer after viewing the obfuscated or non-obfuscated video. Understanding the subjective utility from a human viewer's perspective on the privacy-utility tradeoff when partial obfuscation is applied can also support human-in-the-loop systems and fully automated systems. This is done by understanding the obfuscation methods that can impact the annotator's ability to accurately provide fine-grained labels that require human visual confirmation, while at the same time understanding its impact on bystander privacy, which is reported to impact an individual's willingness to wear a camera.

To address privacy concerns in wearable cameras, methods have been created to reduce information about bystanders or objects in the environment from a human viewer's perspective. Due to the concern that images or videos will be misused by a human, privacy is often measured subjectively, and it is the reason many researchers try to understand privacy from a human perspective [2, 4, 16, 19, 32, 54, 57, 58]. This has allowed others to define automated measures of privacy, where an algorithm or model is trained to detect an object or situation that has been shown to raise privacy concerns (e.g., typing on a computer). While obfuscation methods have been reported to be a sufficient method for enhancing privacy by humans, in specific situations they have been shown to not be as effective in obfuscating the information from a machine [3, 44]. On the other hand, machine learning models may fail at preserving privacy as they have been shown to not detect the intended object in the scene when the test image distribution is not from the training set distribution (i.e., out of distribution) in both adversarial (i.e., fake images created to fool the model) [23, 27] and non-adversarial (i.e., real images that fool the model) [3] cases. More importantly, in previous work, participants report that their willingness to wear an activity-oriented camera

is mostly impacted by their concern for what bystanders would think if they view the video being recorded [4]. As such, while understanding bystander privacy concerns from a machine perspective are important, we focus this paper on the human subjective perspective of privacy, in hope that by understanding concerns related to specific activities, proper obfuscation methods can be selected depending on the activity-recognition goals of the study. Here, we describe how humans view the effect of partial obfuscation on protecting information related to the bystanders, especially the bystander activity. We believe that, by understanding the human perspective, we can help inform future research interested in designing automated measures for privacy in the case of partial obfuscation for wearable cameras in real-world settings.

2.3 Key Definitions

In this section, we provide key terms related to the privacy-utility tradeoff used in this paper.

Activity-Oriented Partial Obfuscation: Obfuscation that is applied on all pixels in the image that are not surrounding the wearer's hand.

Visual Confirmation Utility: The ability of a human viewer to recognize a hand-to-head activity of the person that is wearing the activity-oriented camera.

Bystander Privacy: The inability of a human viewer to recognize the bystander's activity.

Bystander Concerns: The participant's reported concerns toward being captured by a wearable camera in multiple private scenarios.

3 EXPERIMENT AND METHODS

We conducted an online experiment to study the effect of partial obfuscation using different image-transformation methods (filters) on preserving visual confirmation utility in wearable cameras while enhancing bystander privacy and reducing bystander concerns. We collected 10 video scenarios (Section 3.1) using an activity-oriented wearable camera. We applied activity-oriented partial obfuscation methods using five different filters (Section 3.2) on the collected videos: blurH, blur, edgeH, edge, and mask. Including the no obfuscation case (as-is), we had a total of six arms (cases). In the online experiment (Section 3.3), participants were randomly assigned to one of the six arms. The participants viewed each of the 10 videos/scenarios in one of the arms and subsequently answered questions about the wearer and the bystander in the video. Upon survey completion, each participant was compensated 3 USD. The median time for completing the experiment was 34 minutes. The study was approved by the Institutional Review Board of Northwestern University.

3.1 Collected Scenarios

To collect our scenarios, we used an activity-oriented camera. Unlike egocentric cameras, activity-oriented wearable cameras can reduce bystander discomfort around the wearer because the intention and the scope of the recording can be communicated naturally to bystanders by changing the direction of the lens. Activity-oriented cameras are designed with a focus on capturing a set of activities that will then define the location and lens

orientation of the camera. Therefore, in our case, we chose activities that involve hand-to-head gestures because of their relevance to the research community in understanding human behavior, such as eating and its confounding activities [5, 8–10, 81]. The activity-oriented wearable camera used was placed on the chest using a chest strap and had the lens pointed upward.

While wearing an activity-oriented camera, we asked the wearer to perform hand-to-head activities related to eating and its confounding activities (eating, drinking, biting nails, yawning, scratching, and wearing glasses) and a few other non-hand-to-head activities that can occur surrounding eating episodes (washing hands, talking to a bystander, answering a call, and typing on a computer) which can help us understand to what extent activity-oriented partial obfuscation can preserve coarse contextual information if needed. All scenarios/videos presented to the participants included both a wearer and at least one bystander. For the bystander activities, we asked the actors (bystanders) to perform sensitive activities that, if caught on camera, both the wearer and the bystander may be concerned. These sensitive activities were extracted from prior literature that studied privacy in the context of cameras [16, 19, 32]. The videos included six actors: the wearer and five actors playing the parts of bystanders. The average length of the videos presented to the participants was 9.3 seconds. Figure 2 shows a snapshot of the wearer and bystander activities in each of the 10 scenarios, while Table S1 (in supplementary materials) describes each collected video scenario.

3.2 Obfuscation Method

Figure 3 shows images from an activity-oriented camera [4] with multiple obfuscation methods applied. Although the lens is pointed upward, the camera records more information than needed due to the fish-eye lens, which creates a discrepancy between the communicated scope of the camera and the scope of the data collected. The fish-eye lens is essential for providing contextual information regarding the type of activity being performed by the wearer. Removing the fish-eye lens would result in images that limit viewing to only the head, not capturing hand-to-head gestures. To determine the wearer hand and head pixels, we used a low-cost, low-powered thermal infrared (IR) sensor array to aid in extracting the foreground from the camera image for hand-to-head gestures (see first column images in Figure 3). Other sensors, such as depth cameras, can also be used, but we decided to use a low-powered thermal sensor, which would be more practical for battery considerations, particularly if the wearable camera is intended to be worn all day to capture naturally occurring behaviors in the real world. We implemented a pipeline that can aid in semi-automating the segmentation of hand-to-head gestures in the image by using the foreground/background mask generated from the IR sensor array. We manually fine-tuned the pipeline at the initial frames of each scenario to overlay the mask with the image. In future studies, we intend to modify the pipeline to make it function in real-time in situ on the device; however, this is beyond the scope of this paper. Figure 3 shows an example of applying obfuscation methods for hand-to-head activities along with non-hand-to-head activities.

We selected common obfuscation methods with varying degrees of obfuscation intensity to apply to the background pixels detected in the image; these methods have varying filters that

can exhibit a varying effect on both utility and privacy [29, 42]. Because the background pixels might hold information that is related to the activity of the wearer, but not part of the hand-to-head gestures, it has the potential to affect the visual confirmation utility of the wearer's activities. For example, a stationary plate of food in front of the wearer is important context to visually confirm the wearer is eating, but it will be obfuscated by our pipeline since the stationary plate does not move toward the face. However, certain filters might still preserve some high-level coarse-grained information about the background that can aid in maintaining visual confirmation utility when some context information is needed. For example, blurring the background in an image retains some information about the colors, whereas applying canny edge detection maintains the shape of objects in the background by only displaying the outline of the object.

We applied five activity-oriented partial obfuscation methods: blurH, blur, edgeH, edge, and mask. We obtained the parameters of the obfuscation method used from prior literature [29] that mapped these parameters to subjective intensity levels (low, medium, or high). We blurred the background pixels using a normalized box filter with a 50 by 50 kernel for *blurH* and 25 by 25 kernel for *blur* (each pixel in the kernel gets equal weighting). We applied the canny edge–detection technique to detect edges [14] (low- and high-threshold pixel values are set to 400 and 600, respectively, for *edge* and 570 and 950 for *edgeH*). Masking the background with *mask* is the most extreme obfuscation method as it completely filters all background information by setting background pixels to a single color (e.g., black).

3.3 Recruitment, Assignment, and Survey Flow

We used Amazon Mechanical Turk (MTurk) to recruit participants. The study was advertised to "answer questions about wearable cameras and its data." We limited participation to individuals in countries where English is a dominant language to minimize language barriers since we had questions with open-text format. We also required that the worker have a positive reputation (95% approval rating and minimum accepted hits of 1000) to ensure the quality of the responses/feedback. A link to a Qualtric survey was posted in MTurk, and the link randomly assigned participants to one of the six arms (as-is, blur, blurH, edge, edgeH, mask). The participant first answered a demographics questionnaire and questions about their history with wearable cameras. The demographics questionnaire included questions about their ethnic/racial identity, education, and employment.

Participants were asked to imagine being around someone wearing a camera that might not be visible to them. They were told that the camera records videos without recording audio. A video sample was provided to them in order to picture the field of view of the camera. Then they were given the baseline questionnaire in order to assess concerns regarding being a bystander before seeing the video. Participants then viewed the 10 video scenarios with their assigned obfuscation method and answered questions about person A (the wearer) to measure visual confirmation utility (Section 3.6). After that, they viewed the same 10 videos, but the questions were focused on person B (the bystander) to measure enhancement in bystander privacy (Section 3.7) and bystander concerns (Section 3.8). Response options to the questions were a mixture of text responses (so as to not influence participant response) and Likert and numeric rating scales.

3.4 Pilot Studies and Sample Size Considerations

In preparation for our study, we performed three pilot studies. The first asked six student volunteers to complete the surveys for the as-is arm (no obfuscation) to assess how well participants can identify wearer and bystander activity. We modified two videos after this pilot to ensure bystander activities were clearly visible. The second pilot was released to 20 MTurkers to test the study survey questions and comprehension of the questions. In the second pilot, we learned not to ask questions about the bystander and wearer on the same web page, so as to not influence or confuse the viewer. Therefore, we separated the wearer and bystander videos and questions to prevent confusion. In pilot three, we released the experiment to 120 (30 as-is, 30 blur, 30 edge, and 30 mask) participants to estimate the sample size required to test our hypotheses. In pilot one, participants were not paid. However, in pilots two and three each participant was paid 3 USD.

The preliminary results we received from pilot three showed promise in our ability to reduce bystander concern levels and bystander identification. The participants ability to determine bystander activity remained high, and so we added two new obfuscation methods in our final study. We added an edgeH and blurH arm to see if increasing the intensity of the edges and the blur would further lower the recognition of the bystander activity while maintaining participants' ability to discern the wearer's activity.

We then used the results from pilot three to inform our study enrollment size. Here, instead of using a classic t-test, we chose to determine our needed sample size based on a variant called the *two one-sided test* [77] that allows us to determine whether a procedure (e.g., blur) is noninferior, or equivalent, to the gold standard (i.e., as-is). For more detail, see Section 3.10.1. After comparing the proportion of correct responses in the as-is arm (81%) with the blur arm (70%), we chose to power our study to a tolerability threshold (δ) of 0.3, which would indicate that the percentage of correct responses for any obfuscation method was not more than 30% worse than seen in the as-is arm. Assuming an α of 0.01, power of 80%, and a δ of 30%, we determined that 59 participants per arm would be needed [17]. We collected data from at least 60 participants per arm to account for potential problems in data collection.

3.5 Participants

A total of 367 participants completed the study. Data from 6 participants were omitted due to malfunction in the Qualtrics server that did not allow the participants to view some of the videos (n=4) or due to providing a meaningless response 45 minutes(same response to all questions; n=2). Therefore, 361 participants were included in the analysis (n: as-is, 63; blur, 62; blurH, 61; mask, 60; edge, 56; edgeH, 59). Mean (SD) participant age was 33.3 (9.8) years (range, 18–68 years). Approximately half of the participants reported having worn a wearable camera prior to participation in this study. The majority of the participants reported being employed (full-time, 54%; part-time, 18%; unemployed, 14%; student, 8%; multi-jobs, 5%; retired, 2%;). The majority of participants identified as white (77%); the remaining participants identified as black/African American (10%), Asian (6%), Hispanic/Latino (5%), and other (3%). At baseline, participants reported their level of concern toward being recorded using a wearable camera before viewing the recorded scenarios.

3.6 Visual Confirmation Utility (RQ1)

Participants were instructed to watch a series of videos that were captured from the activity-oriented camera and to answer questions about Person A's activity. They were reminded that Person A is the wearer at each phase, and in each video provided a green arrow pointing toward Person A. Each video was presented on a page of its own, and the order of the scenarios was randomized to minimize carry-over effects. Participants were told to answer questions to see how well they could identify Person A and their activity after viewing the video. They were also instructed to view the videos as many times as needed to answer the questions. Table 1 lists the questions we asked about Person A (the wearer) in order to assess visual confirmation utility. A.Q1 and A.Q2 were used to get participants to think more thoroughly about A.Q3, which determined the activity of the wearer. In cases where the wearer was not performing an activity that involved the hand (e.g., talking), we changed A.Q3 to be "What is Person A doing?" The participants response to A.Q3 allowed us to assess visual confirmation utility. The last question (A.Q6) was essential to assess whether the obfuscation method truly obfuscated the bystander or not. After coding participant responses, correct responses were labeled with a 1, and incorrect responses were labeled with a 0 (see Table S2 in supplementary materials for examples of accepted and rejected labels). All text responses were coded by two independent coders who later met to resolve conflicts. All disagreements were resolved by negotiated consensus.

3.7 Bystander Privacy (RQ2a)

Participants viewed the same 10 videos in this phase; however, they were asked questions about Person B's activity. They were reminded about who Person B was and that they would also see a red arrow pointing toward Person B. Each scenario was presented on a single page, and the order of the scenarios was randomized to minimize carry-over effects. Participants were instructed to answer questions about Person B after viewing the video. They were also instructed to view the videos as many times as needed, and they were asked questions only about Person B. Table 2 list the questions asked about Person B (the bystander).

Participants response to B.Q2 *"What is Person B doing?"* allowed us to assess the identification of bystander activity. After coding the participants' responses, we marked correct responses with a 1 and incorrect responses with a 0. See Table S4 in supplementary materials for examples of accepted and rejected labels. All text responses were coded by two independent coders and later met to resolve conflicts. All disagreements were resolved by negotiated consensus.

3.8 Bystander Concerns (RQ2b)

Table 3 lists the questions we used to understand potential bystander concerns. For each scenario, we asked participants in each arm to imagine that they were Person B in the video and to rate how concerned they would be if they were in Person B's place and the video was recorded (C.Q1). We further tried to qualitatively understand factors impacting those concerns using the open-ended responses to question C.Q2. One author read all responses to determine a codebook (see Table S5 in supplement materials), and then two authors coded the responses independently who later met to resolve conflicts. All disagreements were

resolved by negotiated consensus. In total, we coded 2,919 responses. We excluded responses from the as-is group as we were interested in understanding bystander concerns where obfuscation is applied.

3.8.1 Baseline Questions.—Baseline questions asked the participants to imagine that they were bystanders and then rate their concern level on a 5-point Likert scale (1=Not at all concerned, 5=Extremely concerned) if the camera caught them performing each of the following activities: eating, smoking, talking (no audio), changing garment of clothing, crying, shopping, coughing/sneezing, shouting/being angry at someone, drinking alcohol, being affectionate with someone else, performing physical activity, praying or performing spiritual acts, using the bathroom, and using a medical device or taking medication. Means and standard deviations of participant responses to baseline questions were similar across participant arms, indicating no major difference in concern of being recorded across the different arms (see Figure S1 in supplementary materials).

3.9 Privacy-Utility Tradeoff (RQ3)

To calculate the *tradeoff* (t) for a given obfuscation method, we first calculated *bystander lack of privacy* by averaging *bystander concerns* (b_c from C.Q1 in Table 3) with the accuracy of detecting *bystander activity* (b_a). Because the bystander concern score is on a scale from 1–5, we multiplied it by 0.2 to make sure both values were between 0 and 1. The resulting score (b_{ac}) combines bystander activity detection (i.e., breach of privacy) with bystander concerns ($b_{ac} = 0.5 * b_a + 0.5 * b_c$); therefore, bystander privacy score (p) is ($1 - b_{ac}$). *Visual conformation utility* is denoted as (u_{h2h}) when only the hand-to-head (H2H) activities are taken into consideration. We also considered the case of all wearer activities (u_{all}) as this will help in assessing how much of the non-H2H context is preserved. The resulting tradeoff, for either the H2H (t_{h2h}) or all activities (t_{all}) is the weighted sum of privacy (p) and visual confirmation (either u_{h2h} or u_{all}). We used three different weights combination to visualize this tradeoff: (1) equal weighting of both utility and privacy ($w_u = w_p = 0.5$), (2) greater weighting of utility ($w_u > w_p$, as an example we set $w_u = 0.75$ and $w_p = 0.25$), and (3) greater weighting of privacy ($w_u < w_p$, as an example we set $w_p = 0.75$ and $w_u = 0.25$). We calculated t_{all} and t_{h2h} for each obfuscation method across all individuals, and selected the obfuscation method with the maximum mean tradeoff across participants to be optimal.

$$t_{h2h} = w_u * u_{h2h} + w_p * p$$

$$t_{all} = w_u * u_{all} + w_p * p$$

3.10 Data Analysis

3.10.1 Statistical Analysis: RQ1 - Visual Confirmation Utility.—We first calculated the proportion of correct responses to A.Q3, *"What is Person A doing (with their right hand)?"*, from the as-is (no obfuscation), blur, blurH, edge, edgeH, and mask arms. Since we do not obfuscate the hand-related activities, ***we hypothesized that each of the five obfuscation arms would be equivalent (or noninferior) to the as-is arm in correctly***

identifying the activity of the wearer. To test the relative preservation of visual confirmation utility, we used a two one-sided equivalence test [39, 65, 77], a variant of classic null hypothesis testing. Here, the null hypothesis was that the as-is proportion of correct responses (p_a) would be superior to the proportion of correct responses in an obfuscation arm (p_o) by at least a margin of (δ) percent (i.e., $|p_a - p_o| \geq \delta$). The margin represents the *tolerability* of the noninferiority test. The alternative hypothesis was that the two proportions were equivalent up to the tolerance margin (i.e., $|p_a - p_o| < \delta$). In our main analysis, we specified our tolerance margin to be $\delta = 30\%$ and later performed exploratory sensitivity analyses where we decreased the tolerance to 15% and 20%. Equivalence tests were performed using Python statsmodels package (0.9.0)[66].

Additionally, we calculated the arm-specific: (1) averages of the confidence individuals had about their response to A.Q3 (A.Q4), (2) proportion of individuals who thought Person A was eating or drinking (A.Q5), and (3) proportion of individuals who could see any other person in the video (A.Q6). We also assessed the agreement between the response to A.Q5 and the true image contents by calculating the false positive rate.

3.10.2 Statistical Analysis: RQ2a - Privacy.—We calculated the arm-specific percentage of correctly identified bystander activities for each video (B.Q2). To get a picture of the overall accuracy within each obfuscation method, we calculated the average percentage of correctly identified bystander activities across all videos. Since the obfuscation methods attempt to distort background pixels that include bystanders, ***we hypothesized that all five obfuscation methods would result in significantly lower proportions of correct responses than the as-is method*** (i.e., preserving bystander privacy). We used a two-proportion z-test to separately test these hypotheses.

3.10.3 Statistical Analysis: RQ2b - Bystander Concerns.—For each obfuscation method, we compared bystander-reported concerns after viewing the video (C.Q1) with the as-is arm. Activity-oriented partial obfuscation was determined to have reduced bystander concerns if their reported concerns with obfuscation were significantly lower than the reported concerns with the as-is arm. ***We hypothesized that all five obfuscation methods would result in significantly lower mean bystander concerns compared with the as-is method***. We used a Mann–Whitney U test to test significance.

3.10.4 General Considerations.—Since we had three families of hypotheses to test (RQ1, RQ2a, and RQ2b), we grouped each research question as a family to control the probability of false rejection [67]. We set a global significance level of $\alpha = 0.05$. We used a Bonferroni correction [22] to compensate for multiple comparisons. The maximum number of hypotheses tested within each family was five (comparing each obfuscation method with as-is), so results from our hypothesis testing were considered significant if $p < .01$. All statistics were done using Python (3.6.8) with the following packages: statsmodels (0.9.0) [66] and scipy (1.1.0)[33].

4 FINDINGS

Here we answer the research questions outlined in the introduction to understand the effect of activity-oriented partial obfuscation on visual confirmation utility of a wearer's hand-to-head activities and on bystander privacy and concerns.

4.1 RQ1: Visual Confirmation Utility

In RQ1, we tested whether the visual confirmation utility could be preserved by comparing the no-obfuscation (as-is) case with the five obfuscation methods.

4.1.1 Successful Identification of Wearer Activity.—

We defined visual confirmation utility as the ability of a human viewer to successfully identify the hand-to-head activities of the wearer. Visual confirmation utility was maintained using medium obfuscation (blur, edge), high obfuscation (edgeH, blurH), and extremely high obfuscation (mask); on average, the proportion of correct wearer activity labels obtained from these partial obfuscation cases was not inferior to the correct proportion of activity labels for the as-is arm ($p<.001$, $\delta = 30$; Table 4; Figure 4). We performed post hoc exploratory sensitivity analysis to test smaller tolerance values ($\delta = 20$, $\delta = 15$). When $\delta = 20$, all obfuscations were significantly equivalent to as-is ($p<.001$ for blur, blurH, edge and mask while $p<.01$ for edgeH). When $\delta = 15$, all obfuscations, except for edgeH, were significantly equivalent to as-is ($p<.001$ for blur and edge while $p<.01$ for blurH and mask).

Upon investigating the accuracy of each activity in each arm, we noted that two factors mainly impacted viewer ability to identify hand-to-head gestures. The first was whether the activity contained an object in-hand or not. For example, bite and yawn were activities that were more difficult to identify compared with other hand-to-head activities, even in the as-is case. These gestures do not contain an object in the hand of the wearer, and they have similar characteristics to other confounding activities (e.g., scratching face, picking one's nose). However, hand-to-head activities that involve an object in-hand (e.g., calling with a phone or eating food) were not affected by obfuscation intensity. Interestingly, the mask arm performed best to recognize the wearer wearing glasses, which could be attributed to reduced stimuli from complete obfuscation of background activity, allowing participants to hone in on the object in-hand. The second factor negatively impacting the identification of hand-to-head gestures depended on whether the object in-hand was affected by the obfuscation. For example, drinking was at times confused with "eating a fruit." Cropping parts of objects also impacted the recognition of the wearer's activity, where it was confused with pulling the hair up.

4.1.2 Identification Confidence.—

Overall, as the intensity of the obfuscation methods increased, the self-reported confidence in the provided label decreased. Some participants indicated that audio could aid in identifying the wearer's activity that does not incorporate an object in hand.

We tested label confidence by asking the participants if the participant could confirm that the wearer's activity was eating or drinking, regardless of whether they correctly labeled the activity. For example, a participant might not have been able to label *washing hands*, but

they were able to confirm that it was not an *eating* activity. When participants were asked "Do you think the wearer is eating or drinking?" in the videos that did not contain an eating or drinking activity, the percentage of the *"Yes"* responses (incorrect activity or false positive) was similar across all arms (as-is = 2%, blur = 2%, edge = 1%, blurH= 3% edgeH = 3%, mask = 4%).

4.2 RQ2a: Bystander Privacy

In RQ2a, we tested the effect of obfuscation on bystander privacy by comparing accuracy of identifying the bystander activity in the obfuscation arm with accuracy in the non-obfuscation (as-is) arm.

All of the videos that the participants viewed had at least one bystander. When participants were asked if they saw any person other than the wearer, the percentage of the *"No"* responses (no bystander seen) increased as the obfuscation intensity increased (as-is = 2%, blur = 28%, edge = 44%, blurH= 50% edgeH = 76%, mask = 96%). Also, as the obfuscation intensity increased, we saw a drop in the accuracy of the bystander activity identification (see Figure 5). Table 5 shows a significant drop ($p<.001$) in average accuracy of identifying bystander activity in all arms when compared with the as-is. However, we noticed that some filters failed to obfuscate bystander activities when the bystander was performing high- or medium-intensity activities that cause significant motion in the scene (such as fighting or exercising).

4.3 RQ2b: Bystander Concerns

In RQ2b, we tested the effect of obfuscation on bystander concerns quantitatively by comparing the reported bystander concerns in each arm with the reported concerns in the as-is arm. We also qualitatively analyzed the factors that affect bystander privacy in each obfuscation arm to understand how activity-oriented partial obfuscation impacted bystander subjective concerns. Overall, bystander concerns were significantly reduced as the obfuscation level increased (Table 5). In the case of blur and edge, when the bystander activity involved a lot of movement (e.g., fighting), the concern level increased as the activity identification was higher in this case. When the obfuscation level increased in edgeH and mask, we saw a reduction in bystander concern level.

Participants were asked to indicate their concern level and provide a rationale for their choice to help in understanding factors that impact bystander concern level. The 2,919 responses indicated that the main factors that affected bystander concern levels were: (1) subjective obfuscation effectiveness, (2) identity leak concerns, (3) activity and context captured, (4) possibility of multiple interpretations of the activity, and (5) fundamental concerns with being recorded. Table 6 shows the distribution of these reasons in each arm in two groups of interest: low concern (low or no concern) and high concern (moderate concern or more). The majority of the reasons behind reporting a low concern or no concern levels was because of obfuscation that reduced information about the bystander identification or activity (n=1346, 63%) or because participants considered the activity they viewed in the video to be non-concerning (n=363, 17%). On the other hand, the majority of the reasons for

reporting a moderate or more concern was due to a fundamental concern with being recorded (n=419, 54%) and then the activity of the bystander (n=223, 29%).

4.3.1 Subjective Effectiveness of Obfuscation Lowers Bystander Concerns.

—In most cases (n=1244), participants mentioned obfuscation effectiveness as a reason when they report a no concern (i.e., they chose “Not at all concerned”). Subjective obfuscation effectiveness was expressed when participants reported obfuscation as a reason that explained their concern level. Participants reported that obfuscation was successful if visual information about a bystander was reduced to the point where “you can barely even make out that its a person let alone a specific person.” Other participants considered the obfuscation was effective if it hid information about specific parts of the body (e.g., “You cannot see the person’s face” and “Only their silhouette is shown”).

4.3.2 Obfuscation Can Still Leak Some Information About Bystander Identity.

—In blur, some participants expressed that, although the concern level was not high, they were still concerned because there were some features present. A participant reported, “you can make out the shape of the face and certain features” such as “they are wearing something red,” or “it shows body proportions and hairlines.” In the case of edge, participants indicated that the outlines presented were identifiable. One participant mentioned, “Even though it is still an outline, it would be an outline of me.” Revealing some information about the bystander could enable identification if the bystander was known. One participant mentioned, “if you knew who it was you could use your imagination to [piece] it together.” Others expressed concerns regarding the possibility to “reconstruct the identity of Person B later.”

4.3.3 Concern Depends on Wearer Activity, Bystander Activity, and the Context.—

Bystanders concern level in partially obfuscated images were influenced based on the interpretation of the activity and context regardless of whether identification of the activity was correct (n=383, 61%) or not (n=248, 39%). When judging the concern level toward being the bystander, some participant responses depended on the context provided by the environment and the wearer’s activity (n=170, 5%) or on their perceived perception of the bystander activity seen in the video (n=586, 20%). “No cameras in a bedroom or bathroom - no exceptions” was mentioned as an example of a context that could raise concern in the case of blurH partial obfuscation, while if the context “appears to be a public space” lowered concern. The wearer activity also influenced the concern level as P155 misidentified the bystander activity to be eating (instead of crying), and the concern level was low because “eating is normal.” Even in the case of successful identification, the expressed level of concern was influenced by participants’ subjective privacy view. Participant P107 mentioned, “Prayer is not something to be concerned about.”

4.3.4 Obfuscation Can Allow for Multiple Interpretations.—

When the obfuscation was not strong enough to obscure everything or when it was not clear enough to get definite answers, participants voiced some concerns because it can be interpreted or “taken [understood] a few ways”(P52). Also, this misinterpretation for the activity “could be taken out of context”(P115) and “used against someone”(P52). For example, when the bystander

was smoking, one participant noted that “without context, the fact that they are possibly smoking an unknown substance could be used against them”(P119). When the bystander was eating, P119 also indicated that “It could be construed that they were binge-drinking because it’s hard to tell what they are doing when they lift the object over their face.” Also, the absence of non-visual cues could allow for multiple explanations as P52 mentioned that in scenario 8 the bystanders in the obfuscated video “might actually be fighting with hostility, or just messing around between friends. With no audio, it’s also difficult to understand the context of why they are acting this way.”

4.3.5 Fundamental Concerns toward Video Recording in General.—

Some participants expressed fundamental concern towards self-recording as they did not feel comfortable being recorded regardless of the privacy measures. The concerns were about the consent of the bystander being recorded. Others expressed concerns about the justification for being recorded in the first place and about who sees the video or if it will be posted publicly. Of the 361 participants, 35 (9.7%) accounted for >50% of the fundamental concerns shown in Table 6, suggesting that some bystanders have issues being recorded regardless of obfuscation.

4.4 RQ3: Privacy-Utility Tradeoff

In RQ3, we analyzed the utility-privacy tradeoff for each activity-oriented partial obfuscation method. Figure 6 shows the privacy-utility tradeoff in each filter used by activity-oriented partial obfuscation while also taking bystander concerns into consideration. We analyzed the privacy-utility tradeoff in relation to the type of activities collected. If contextual information was not needed to distinguish or confirm the hand-to-head gesture, then partial obfuscation using *mask* provided a great balance in this privacy-utility tradeoff (see top row of Figure 6). However, if contextual information was essential for identifying the activity accurately and with confidence, then partial obfuscation using *blur* achieved the best balance (see bottom row of Figure 6)). In addition, this tradeoff should also be analyzed in the context of information control. That is, if the data are going to be public or if there is a high chance of a sensitive context, then mask is the best approach. However, if data sharing is controlled and if the chance of capturing sensitive data is low, then blur- or edge-based partial obfuscation methods should be considered.

5 DISCUSSION

5.1 Activity-Oriented Partial Obfuscation Can Preserve Visual Confirmation of Hand-Related Activities Even with Extreme Filter

Our results show that activity-oriented partial obfuscation, even under intense filters (e.g., mask), preserves the visual confirmation of a specific group of hand-related activities that may be confounded if textual information is lost. Activity-oriented partial obfuscation extreme filters performed better with activities that contained an object in the wearer’s hand (see Figure 7a) likely because the obfuscation removed distractions and allowed the viewer to focus on a narrow view of the image. This, in turn, allowed the viewer to hone in on the object in the wearer’s hand. However, for activities that did not involve an object in hand, extreme filters performed poorly due to loss of information, which led the participant to

interpret the image based on their own experience, a finding that is consistent with previous work [7, 42]. As we are interested in using activity-oriented cameras to capture hand-to-head movements, the drop in accuracy of identifying activities that do not involve hand-to-head gestures does not impact visual confirmation utility, but this could be indicative of a reduction in determining context of hand-to-head activities, such as typing on a computer.

The decrease in identification confidence reported by participants that were associated with increased obfuscation intensity was likely due to loss of contextual information, especially when there was no object in hand. In these cases without an object in hand, coarse-grained visual or non-visual contextual cues serve as an aid that can reduce uncertainty about the wearer's activity. For example, audio was mentioned by participants as a potential helpful cue that can help in distinguishing some activities that do not involve an object in hand (e.g., distinguishing a yawn from laughter or chewing gum from talking). Raw audio is considered to be even more invasive than images in the case of wearable cameras [4]. Therefore, collected audio should not affect bystander privacy. For example, the audio can be limited to non-speech [56]. Moreover, work in audio source separation can extract vocal sounds (sensitive speech information) and keep background sounds (contextual cues) by using microphone arrays [75] or neural networks [69]. These non-visual and non-speech sound cues provide context around the obfuscated images, which may increase confidence in identifying wearer activities while enhancing privacy. Previous work showed that context and high-level activity can be inferred using still images at low frequency, but it can miss capturing some fine-grained gestures [73, 74]. Our results show that intense partial obfuscation can preserve fine-grained gestures but with a loss to contextual information. Future work should investigate a mixed approach to obfuscation to provide some visual cues that aid wearer activity detection. For example, an intense filter (e.g., mask) could be used the majority of the time, and a less-intense filter (e.g., blur) could be triggered infrequently when a particular context is detected.

5.2 Spatio-Temporal Obfuscation Filters Enhance Bystander Privacy More than Spatial-Only Methods

Spatial obfuscation filters tested in our study (blur, blurH, edge, and edgeH) were shown to limit information about bystander activity when the bystander was stationary, similar to still images [29]. However, spatial obfuscation filters (both low- and high-intensity) are not effective in obscuring bystander activities involving movement (e.g., a person fighting or exercising in the background) as they fail to obfuscate information within consecutive frames (temporal), which causes a visible motion in the video (see Figure 7b). Protecting the identity of the bystander (biological biometric) without obfuscating their activities threatens the privacy of the bystander as identity might be revealed through a correlation or linkage attack by using multiple information sources (e.g., GPS) or by using behavioral biometrics that can identify the person (e.g., walking gait). Therefore, obscuring information about bystander activity adds another layer of protection for bystander privacy. Spatial obfuscation methods such as cartooning [30] and avatar (replacing the human with an avatar) [60] have also been shown to be useful in hiding information about bystanders [30, 42]. However, similar to other spatial obfuscation methods, those methods also fail to obfuscate the activity of the wearer since they do not protect against inference from temporal information. We

showed that partial obfuscation using mask provides spatio-temporal obfuscation on bystander information. In cases where blur or edge filters are preferred (as they provide more contextual information about the environment), additional methods, such as full-body in-painting (removing the person and blending into the scene background) using neural networks [79] can be used to obfuscate moving objects in the scene while preserving some high-level contextual information obtained from spatial obfuscation methods.

5.3 Obfuscation Methods Should Also Consider Alternative Interpretations of the Obfuscated Activity

Overall, we have shown that activity-oriented partial obfuscation enhances bystander privacy by reducing information about their activity. However, obfuscating the actual activity is not always enough to reduce bystander concerns, as we have shown that bystanders rely on their interpretation of the obfuscated activity regardless of whether it is the correct activity or not (see Figure 7c). This indicates that future automated privacy-enhancing methods should take into account the human ability to seek and see patterns even in random data (pareidolia [20]) in order to guard against both human and machine attacks on the bystander. Currently, automatic privacy-enhancing techniques rely on detecting certain objects and people [26, 37, 50] to obfuscate, but they do not take into consideration that the loss of contextual information due to obfuscation allows for other interpretations that can be more concerning to the bystander if framed in a negative way. It is unknown whether automatic methods can detect these alternative interpretations, but previous work has indicated that machines fail to detect patterns that are out of distribution [3, 23, 27, 46, 76]. Therefore, we recommend using spatio-temporal filters while using the activity-oriented partial obfuscation method as it is capable of minimizing alternative interpretations when compared with other spatial filters.

5.4 Activity-Oriented Partial Obfuscation Beyond Human Visual Confirmation

We show how activity-oriented partial obfuscation significantly lowers bystander concerns while preserving image utility in providing visual confirmation of the wearer's hand-related activities to a human viewer while obfuscating other private information in the video. Preserving visual confirmation of wearer activities while enhancing bystander privacy and concerns will not only aid the UbiComp community in obtaining activity labels for sensor data in a real-world setting but will also aid other domains that can benefit from visual observation such as digital visual ethnography [28, 45] or device validation studies [35]. Time series sensor data used to build machine learning models to predict human behavior are difficult to confirm visually or to verify by a human. Therefore, these wearable sensors often need visual confirmation of their trained data sets. In the case of sensor data, images from wearable cameras can be used as a method to establish groundtruth or visual confirmation of the human activity performed in real-world settings. We have shown that activity-oriented partial obfuscation can still maintain visual confirmation for hand-related activities while reducing information about bystanders and some contextual information. Activity-oriented partial obfuscation can also be used in the domain of automatic activity detection using image data. Previous work has shown that only 20% of image pixels are needed to automatically identify a wearer's hand-related activity in wearable cameras [41]. By utilizing cues in the images [43, 70] or by utilizing other sensing modalities [41], a

predictive model can learn to attend to a specific critical region that is more informative about the wearer's activity than the background information and can therefore achieve a higher performance rate in predicting the wearer's activity when compared with predictive models with no attention. Although only part of the image was needed for prediction, labels of the activities were obtained using the whole image. Through our experiment, we show that it is also possible to obtain groundtruth activity labels using reduced information, which means that researchers interested in hand-related activities, especially ones that involve an object in hand, can utilize activity-oriented partial obfuscation to collect data in the wild rather than using the full raw image.

5.5 Limitations

We collected 10 scenarios where the literature showed that a bystander would exhibit discomfort if these activities were recorded. Additional scenarios are present in the literature that were difficult to collect from actors (e.g., showing affection, changing clothing, participating in a sexual act). Viewers in the study were asked to imagine themselves in the situation of the bystanders. The responses of the participants might differ because they may perceive a lower risk than if they were the actual bystanders in the video. In future studies, we aim to deploy the system in the field to collect bystander responses in situ.

The experiment was conducted with MTurk, which is a platform that has been used by other researchers for similar purposes [21, 29, 42]. Participants recruited from MTurk are not necessarily representative of the general population as they tend to have higher privacy concerns [34]. Bystander concerns might differ among different populations. However, we believe that the visual confirmation of the wearer activities in our collected video scenarios is an objective measure that will be less variable among the general population. Our study was also validated by three pilot studies before carrying out the final study and was performed with a large sample size. Our work investigates bystander and privacy concerns, as well as visual confirmation utility, from a human viewer perspective only. Future work should investigate automatic privacy attacks that can be used against partial obfuscation as well as investigate potential utility of partially obfuscated images for automatic activity recognition using machine learning algorithms.

6 CONCLUSIONS

Research with wearable cameras uses total or partial obfuscation to limit extraneous and sensitive information, especially that of bystanders. Although obfuscation enhances privacy, it can also degrade visual confirmation utility. In this work, we investigated the privacy-utility tradeoff in the case of partial obfuscation in activity-oriented cameras through an online experiment. We applied activity-oriented partial obfuscation using five filters (mask, blur, blurH, edge, and edgeH) on 10 bystander-sensitive video scenarios. The video showed only the wearer's hand-related activity while obfuscating everything else. We analyzed the effect of activity-oriented partial obfuscation using different filters on the utility of the image to enable a human viewer to visually confirm wearer activity while reducing information about the background, including any bystanders. We compared the privacy and utility results of the different filters with no obfuscation (as-is). Our results showed that all filters used in

activity-oriented partial obfuscation significantly reduced bystander concerns. Mask performed the best in obscuring bystander activities because it applied a spatio-temporal obfuscation. Spatial obfuscation methods failed at obfuscating a bystander's activity when the bystander was moving even when the obfuscation is intense (blurH). In the case of mask, wearer activities that contained an object in hand were successfully identified. However, the loss of context made identifying activities that did not involve holding a recognizable object harder and provided lower confidence than other obfuscation methods. We found that if some information about context is necessary to identify the wearer activity, then the blur, edge, and blurH obfuscation methods strike a good balance in the privacy-utility tradeoff. However, spatial obfuscation methods should be coupled with an information-control mechanism to restrict data access, as bystander activities that involve movement are not concealed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under award number CNS-1915847. We would also like to acknowledge support by the National Institute of Diabetes and Digestive and Kidney Diseases under award number K25DK113242 (NIDDK), and NSF under award number CNS-1850496. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

REFERENCES

[1]. Aditya Paarijaat, Sen Rijurekha, Druschel Peter, Oh Seong Joon, Benenson Rodrigo, Fritz Mario, Schiele Bernt, Bhattacharjee Bobby, and Wu Tong Tong. 2016 I-pic: A platform for privacy-compliant image capture. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services ACM, 235–248.

[2]. Ahmed Tousif, Kapadia Apu, Potluri Venkatesh, and Swaminathan Manohar. 2018. Up to a Limit?: Privacy Concerns of Bystanders and Their Willingness to Share Additional Information with Visually Impaired Users of Assistive Technologies. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 3 (2018), 89.

[3]. Michael A Alcorn Qi Li, Gong Zhitao, Wang Chengfei, Mai Long, Ku Wei-Shinn, and Nguyen Anh. 2018. Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects. arXiv preprint arXiv:1811.11553 (2018).

[4]. Alharbi Rawan, Stump Tammy, Vafaie Nilofar, Pfammatter Angela, Spring Bonnie, and Alshurafa Nabil. 2018. I Can't Be Myself: Effects of Wearable Cameras on the Capture of Authentic Behavior in the Wild. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 3 (2018), 90.

[5]. Bae Sangwon, Ferreira Denzil, Suffoletto Brian, Puyana Juan C, Kurtz Ryan, Chung Tammy, and Dey Anind K. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 2 (2017), 5.

[6]. Bambach Sven, Lee Stefan, Crandall David J, and Yu Chen. 2015 Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In Proceedings of the IEEE International Conference on Computer Vision 1949–1957.

[7]. Baumer Eric PS, Xu Xiaotong, Chu Christine, Guha Shion, and Gay Geri K. 2017 When Subjects Interpret the Data: Social Media Non-use as a Case for Adapting the Delphi Method to CSCW. In

Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing ACM, 1527–1543.

[8]. Bedri Abdelkareem, Li Richard, Haynes Malcolm, Kosaraju Raj Prateek, Grover Ishaan, Prioleau Temiloluwa, Beh Min Yan, Goel Mayank, Starner Thad, and Abowd Gregory. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 3 (2017), 37.

[9]. Bi Shengjie, Wang Tao, Tobias Nicole, Nordrum Josephine, Wang Shang, Halvorsen George, Sen Sougata, Peterson Ronald, Odame Kofi, Caine Kelly, et al. 2018. Auracle: Detecting Eating Episodes with an Ear-mounted Sensor. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 3 (2018), 92.

[10]. Biel Joan-Isaac, Martin Nathalie, Labbe David, and Gatica-Perez Daniel. 2018. Bites 'n'Bits: Inferring Eating Behavior from Contextual Mobile Data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 4 (2018), 125.

[11]. Boyle Michael, Edwards Christopher, and Greenberg Saul. 2000 The effects of filtered video on awareness and privacy. In Proceedings of the 2000 ACM conference on Computer supported cooperative work ACM, 1–10.

[12]. Brkic Karla, Sikiric Ivan, Hrkac Tomislav, and Kalafatic Zoran. 2017 I know that person: Generative full body and face de-identification of people in images. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on IEEE 1319–1328.

[13]. Butler Daniel J, Huang Justin, Roesner Franziska, and Cakmak Maya. 2015 The privacy-utility tradeoff for remotely teleoperated robots. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction ACM, 27–34.

[14]. Canny J. 1986. A Computational Approach to Edge Detection. IEEE Trans. Pattern Anal. Mach. Intell 8, 6 (6 1986), 679–698. 10.1109/TPAMI.1986.4767851 [PubMed: 21869365]

[15]. Chen Jacqueline, Marshall Simon J, Wang Lu, Godbole Suneeta, Legge Amanda, Doherty Aiden, Kelly Paul, Oliver Melody, Patterson Ruth, Foster Charlie, et al. 2013 Using the SenseCam as an objective tool for evaluating eating patterns. In Proceedings of the 4th International SenseCam & Pervasive Imaging Conference ACM, 34–41.

[16]. Choe Eun Kyoung, Consolvo Sunny, Jung Jaeyeon, Harrison Beverly, and Kientz Julie A. 2011 Living in a glass house: a survey of private moments in the home. In Proceedings of the 13th international conference on Ubiquitous computing ACM, 41–44.

[17]. Chow Shein-Chung, Shao Jun, Wang Hansheng, and Lokhnygina Yuliya. 2017 Sample size calculations in clinical research. Chapman and Hall/CRC.

[18]. Cowan Rachel E, Fregly Benjamin J, Boninger Michael L, Chan Leighton, Rodgers Mary M, and Reinkensmeyer David J. 2012. Recent trends in assistive technology for mobility. Journal of neuroengineering and rehabilitation 9, 1 (2012), 20. [PubMed: 22520500]

[19]. Denning Tamara, Dehlawi Zakariya, and Kohno Tadayoshi. 2014 In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM, 2377–2386.

[20]. Lazzaro Paolo Di, Murra Daniele, and Schwortz Barrie. 2013. Pattern recognition after image processing of low-contrast images, the case of the Shroud of Turin. Pattern Recognition 46, 7 (2013), 1964–1970.

[21]. Dimiccoli Mariella, Marín Juan, and Thomaz Edison. 2018. Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 4 (2018), 132.

[22]. Dunn Olive Jean. 1961. Multiple comparisons among means. Journal of the American statistical association 56, 293 (1961), 52–64.

[23]. Eykholt Kevin, Evtimov Ivan, Fernandes Earlence, Li Bo, Rahmati Amir, Xiao Chaowei, Prakash Atul, Kohno Tadayoshi, and Song Dawn. 2018 Robust Physical-World Attacks on Deep Learning Visual Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1625–1634.

[24]. Fan Mingming, Adams Alexander Travis, and Truong Khai N. 2014 Public restroom detection on mobile phone via active probing In Proceedings of the 2014 ACM International Symposium on Wearable Computers. ACM, 27–34.

[25]. Fontana Juan M, Farooq Muhammad, and Sazonov Edward. 2014. Automatic ingestion monitor: a novel wearable device for monitoring of ingestive behavior. IEEE Transactions on Biomedical Engineering 61, 6 (2014), 1772–1779. [PubMed: 24845288]

[26]. Frome Andrea, Cheung German, Abdulkader Ahmad, Zennaro Marco, Wu Bo, Bissacco Alessandro, Adam Hartwig, Neven Hartmut, and Vincent Luc. 2009 Large-scale privacy protection in google street view. In 2009 IEEE 12th international conference on computer vision IEEE, 2373–2380.

[27]. Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron, and Bengio Yoshua. 2014 Generative adversarial nets. In Advances in neural information processing systems. 2672–2680.

[28]. Gouveia Rúben, Karapanos Evangelos, and Hassenzahl Marc. 2018 Activity Tracking in vivo. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems ACM, 362.

[29]. Hasan Rakibul, Hassan Eman, Li Yifang, Caine Kelly, Crandall David J, Hoyle Roberto, and Kapadia Apu. 2018 Viewer experience of obscuring scene elements in photos to enhance privacy. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems ACM, 47.

[30]. Hassan Eman T, Hasan Rakibul, Shaffer Patrick, Crandall David J, and Kapadia Apu. 2017 Cartooning for Enhanced Privacy in Lifelogging and Streaming Videos.. In CVPR Workshops 1333–1342.

[31]. Hoyle Roberto, Templeman Robert, Anthony Denise, Crandall David, and Kapadia Apu. 2015 Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In Proceedings of the 33rd Annual ACM conference on human factors in computing systems ACM, 1645–1648.

[32]. Hoyle Roberto, Templeman Robert, Armes Steven, Anthony Denise, Crandall David, and Kapadia Apu. 2014 Privacy behaviors of lifeloggers using wearable cameras. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing ACM, 571–582.

[33]. Jones Eric, Oliphant Travis, Peterson Pearu, et al. 2001–. SciPy: Open source scientific tools for Python. http://www.scipy.org/ [Online; accessed <today>].

[34]. Kang Ruogu, Brown Stephanie, Dabbish Laura, and Kiesler Sara. 2014 Privacy Attitudes of Mechanical Turk Workers and the U.S. Public In 10th Symposium On Usable Privacy and Security (SOUPS 2014). USENIX Association, Menlo Park, CA, 37–49. https://www.usenix.org/conference/soups2014/proceedings/presentation/kang

[35]. Kim Youngdeok, Barry Vaughn W, and Kang Minsoo. 2015. Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. Measurement in Physical Education and Exercise Science 19, 3 (2015), 125–137.

[36]. Koelle Marion, Wolf Katrin, and Boll Susanne. 2018 Beyond LED Status Lights-Design Requirements of Privacy Notices for Body-worn Cameras. In Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction ACM, 177–187.

[37]. Korayem Mohammed, Templeman Robert, and Chen Dennis. 2016. Enhancing Lifelogging Privacy by Detecting Screens. (2016), 10–15.

[38]. Krombholz Katharina, Dabrowski Adrian, Smith Matthew, and Weippl Edgar. 2017 Exploring Design Directions for Wearable Privacy. In Network and Distributed System Security Symposium (NDSS). DOI: 10.14722/usec.

[39]. Lakens Daniël. 2017. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Social Psychological and Personality Science 8, 4 (2017), 355–362. [PubMed: 28736600]

[40]. Lasecki Walter S, Song Young Chol, Kautz Henry, and Bigham Jeffrey P. 2013 Real-time crowd labeling for deployable activity recognition. In Proceedings of the 2013 conference on Computer supported cooperative work ACM, 1203–1212.

[41]. Li Yin, Liu Miao, and Rehg James M. 2018 In the eye of beholder: Joint learning of gaze and actions in first person video. In Proceedings of the European Conference on Computer Vision (ECCV) 619–635.

[42]. Li Yifang, Vishwamitra Nishant, Knijnenburg Bart P., Hu Hongxin, and Caine Kelly. 2017. Effectiveness and Users' Experience of Obfuscation As a Privacy-Enhancing Technology for Sharing Photos. Proc. ACM Hum.-Comput. Interact 1, CSCW, Article 67 (12 2017), 24 pages. 10.1145/3134702

[43]. Ma Minghuang, Fan Haoqi, and Kitani Kris M. 2016 Going deeper into first-person activity recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1894–1903.

[44]. McPherson Richard, Shokri Reza, and Shmatikov Vitaly. 2016. Defeating image obfuscation with deep learning. arXiv preprint arXiv:1609.00408 (2016).

[45]. Ng Kher Hui, Shipp Victoria, Mortier Richard, Benford Steve, Flintham Martin, and Rodden Tom. 2015. Understanding food consumption lifecycles using wearable cameras. Personal and Ubiquitous Computing 19, 7 (2015), 1183–1195.

[46]. Nguyen Anh, Yosinski Jason, and Clune Jeff. 2015 Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 427–436.

[47]. Nguyen David H., Kobsa Alfred, Hayes Gillian R., Marcu Gabriela, Hayes Gillian R., Truong Khai N., Scott James, Langheinrich Marc, and Roduner Christof. 2008. Encountering SenseCam: Personal Recording Technologies in Everyday Life. UbiComp August 2015 (2008), 182 10.1145/1620545.1620571

[48]. Nguyen Thi-Hoa-Cuc, Nebel Jean-Christophe, Florez-Revuelta Francisco, et al. 2016. Recognition of activities of daily living with egocentric vision: A review. Sensors 16, 1 (2016), 72.

[49]. Oh Seong Joon, Benenson Rodrigo, Fritz Mario, and Schiele Bernt. 2016 Faceless person recognition: Privacy implications in social media. In European Conference on Computer Vision Springer, 19–35.

[50]. Orekondy Tribhuvanesh, Fritz Mario, and Schiele Bernt. 2018 Connecting Pixels to Privacy and Utility: Automatic Redaction of Private Information in Images. In Conference on Computer Vision and Pattern Recognition (CVPR).

[51]. Orekondy Tribhuvanesh, Fritz Mario, and Schiele Bernt. 2018 Connecting pixels to privacy and utility: Automatic redaction of private information in images. In Conference on Computer Vision and Pattern Recognition (CVPR).

[52]. Padilla-López José Ramón, Chaaraoui Alexandros Andre, and Flórez-Revuelta Francisco. 2015. Visual privacy protection methods: A survey. Expert Systems with Applications 42, 9 (2015), 4177–4195.

[53]. Parate Abhinav, Chiu Meng-Chieh, Chadowitz Chaniel, Ganesan Deepak, and Kalogerakis Evangelos. 2014 Risq: Recognizing smoking gestures with inertial sensors on a wristband. In Proceedings of the 12th annual international conference on Mobile systems, applications, and services ACM, 149–161.

[54]. Park Sangkeun, Kim Joohyun, Mizouni Rabeb, and Lee Uichin. 2016 Motives and concerns of dashcam video sharing. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems ACM, 4758–4769.

[55]. Pizza Stefania, Brown Barry, McMillan Donald, and Lampinen Airi. 2016 Smartwatch in vivo. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems ACM, 5456–5469.

[56]. Portelo Jose, Bugalho Miguel, Trancoso Isabel, Neto Joao, Abad Alberto, and Serralheiro Antonio. 2009 Non-speech audio event detection. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on IEEE, 1973–1976.

[57]. Price Blaine A, Stuart Avelie, Calikli Gul, Mccormick Ciaran, Mehta Vikram, Hutton Luke, Arosha K Bandara Mark Levine, and Nuseibeh Bashar. 2017. Logging you, logging me: A replicable study of privacy and sharing behaviour in groups of visual lifeloggers. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 2 (2017), 22.

[58]. Rashidi Yasmeen, Ahmed Tousif, Patel Felicia, Fath Emily, Kapadia Apu, Nippert-Eng Christena, and Su Norman Makoto. 2018 "You don't want to be the next meme": College Students'

Workarounds to Manage Privacy in the Era of Pervasive Photography In Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018). USENIX Association, Baltimore, MD 143–157.

[59]. Raval Nisarg, Srivastava Animesh, Razeen Ali, Lebeck Kiron, Machanavajjhala Ashwin, and Cox Lanodn P. 2016 What you mark is what apps see. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services ACM, 249–261.

[60]. Rhee Chi-Hyoung and LEE C. 2013. Cartoon-like avatar generation using facial component matching. Int. J. of Multimedia and Ubiquitous Engineering 8, 4 (2013), 69–78.

[61]. Roggen Daniel, Calatroni Alberto, Rossi Mirco, Holleczek Thomas, Förster Kilian, Tröster Gerhard, Lukowicz Paul, Bannach David, Pirkl Gerald, Ferscha Alois, et al. 2010 Collecting complex activity datasets in highly rich networked sensor environments. In Networked Sensing Systems (INSS), 2010 Seventh International Conference on IEEE, 233–240.

[62]. Rueben Matthew, Bernieri Frank J, Grimm Cindy M, and Smart William D. 2017 Framing Effects on Privacy Concerns about a Home Telepresence Robot. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction ACM, 435–444.

[63]. Ryoo Michael S, Rothrock Brandon, Fleming Charles, and Yang Hyun Jong. 2017 Privacy-Preserving Human Activity Recognition from Extreme Low Resolution.. In AAAI 4255–4262.

[64]. Saleheen Nazir, Ali Amin Ahsan, Hossain Syed Monowar, Sarker Hillol, Chatterjee Soujanya, Marlin Benjamin, Ertin Emre, Al'Absi Mustafa, and Kumar Santosh. 2015 puffMarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing ACM, 999–1010.

[65]. Schuirmann Donald J. 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of pharmacokinetics and biopharmaceutics 15, 6 (1987), 657–680. [PubMed: 3450848]

[66]. Seabold Skipper and Perktold Josef. 2010 Statsmodels: Econometric and statistical modeling with python. In 9th Python in Science Conference.

[67]. Shaffer Juliet Popper. 1995. Multiple hypothesis testing. Annual review of psychology 46, 1 (1995), 561–584.

[68]. Shu Jiayu, Zheng Rui, and Hui Pan. 2016. Cardea: Context-aware visual privacy protection from pervasive cameras. arXiv preprint arXiv:1610.00889 (2016).

[69]. Simpson Andrew JR, Roma Gerard, and Plumbley Mark D. 2015 Deep karaoke: Extracting vocals from musical mixtures using a convolutional deep neural network. In International Conference on Latent Variable Analysis and Signal Separation Springer, 429–436.

[70]. Singh Suriya, Arora Chetan, and Jawahar CV. 2016 First person action recognition using deep learned descriptors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2620–2628.

[71]. Steil Julian, Koelle Marion, Heuten Wilko, Boll Susanne, and Bulling Andreas. [n. d.]. PrivacEye: Privacy-Preserving Head-Mounted Eye Tracking Using Egocentric Scene Image and Eye Movement Features. ([n. d.]).

[72]. Sun Qianru, Ma Liqian, Seong Joon Oh Luc Van Gool, Schiele Bernt, and Fritz Mario. 2018 Natural and effective obfuscation by head inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 5050–5059.

[73]. Thomaz Edison, Essa Irfan, and Abowd Gregory D. 2015 A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing ACM, 1029–1040.

[74]. Thomaz Edison, Parnami Aman, Essa Irfan, and Abowd Gregory D. 2013 Feasibility of identifying eating moments from first-person images leveraging human computation. In Proceedings of the 4th International SenseCam & Pervasive Imaging Conference ACM, 26–33.

[75]. Valin J-M, Rouat Jean, and Michaud François. 2004 Enhanced robot audition based on microphone array source separation with post-filter. In Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, Vol. 3. IEEE, 2123–2128.

[76]. Ahn Luis Von, Blum Manuel, and Langford John. 2004. Telling humans and computers apart automatically. Commun. ACM 47, 2 (2004), 56–60.

[77]. Walker Esteban and Nowacki Amy S. 2011. Understanding equivalence and noninferiority testing. Journal of general internal medicine 26, 2 (2011), 192–196. [PubMed: 20857339]

[78]. Wu Muchen, Pathak Parth H, and Mohapatra Prasant. 2015 Enabling privacy-preserving first-person cameras using low-power sensors. In Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on IEEE, 444–452.

[79]. Yu Jiahui, Lin Zhe, Yang Jimei, Shen Xiaohui, Lu Xin, and Huang Thomas S. 2018. Free-Form Image Inpainting with Gated Convolution. arXiv preprint arXiv:1806.03589 (2018).

[80]. Zhang Shibo, Alharbi Rawan, Nicholson Matthew, and Alshurafa Nabil. 2017 When generalized eating detection machine learning models fail in the field. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers ACM, 613–622.

[81]. Zhang Shibo, Alharbi Rawan, Stogin William, Pourhomayun Mohamad, Spring Bonnie, and Alshurafa Nabil. 2016 Food watch: detecting and characterizing eating episodes through feeding gestures. In Proceedings of the 11th EAI International Conference on Body Area Networks ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 91–96.

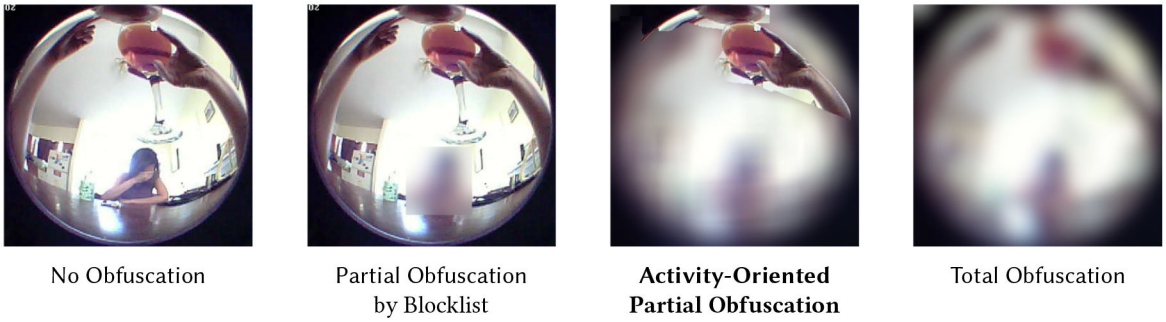


Fig. 1.

This illustration shows the differences between activity-oriented partial obfuscation by default (method used in our study) and other existing obfuscation methods, using the blur filter as an example. Activity-oriented partial obfuscation preserves information of a specific group of hand-related activities while obfuscating everything else by default.

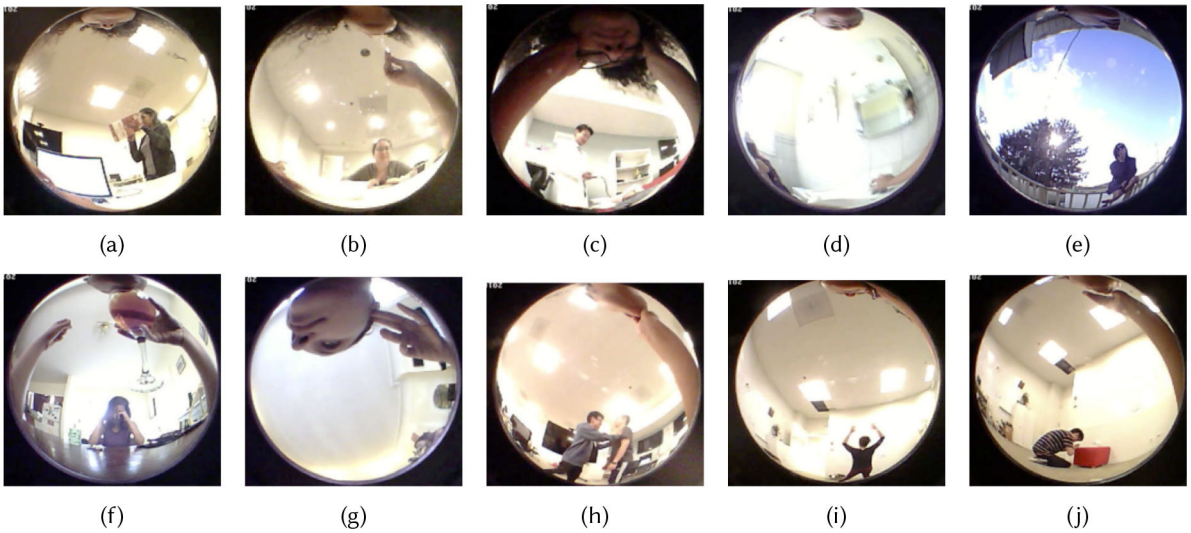


Fig. 2.

Snapshots of the collected video scenarios showing the variety of wearer activities captured around bystanders performing activities that can raise concerns if captured on video. Wearer/bystander: (a) typing/eating, (b) eating/talking, (c) wearing glasses/lying down (sick), (d) washing hands/sitting (on a toilet), (e) talking/smoking, (f) drinking/crying, (g) scratching/drinking, (h) biting nails/fighting, (i) calling/exercising, and (j) yawning/praying.

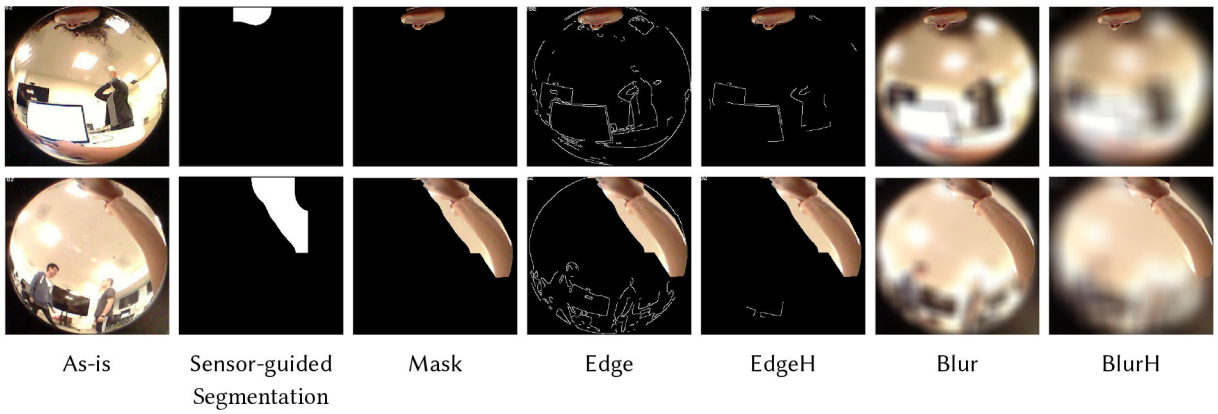


Fig. 3.

Activity-oriented partial obfuscation captures a set of hand-related activities, defined as hand-to-head activities that are of interest to the research community. Top row: example of obfuscation applied to activities that do not contain hand-to-head gestures (typing). Bottom row: example of partial obfuscation applied on hand-to-head gestures (wearer biting nails).

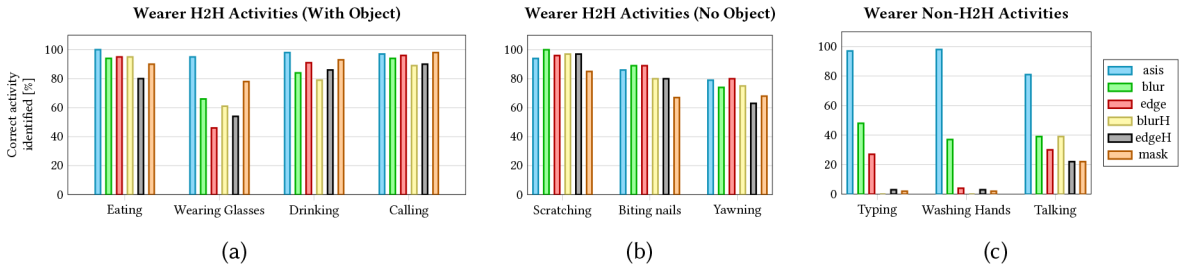
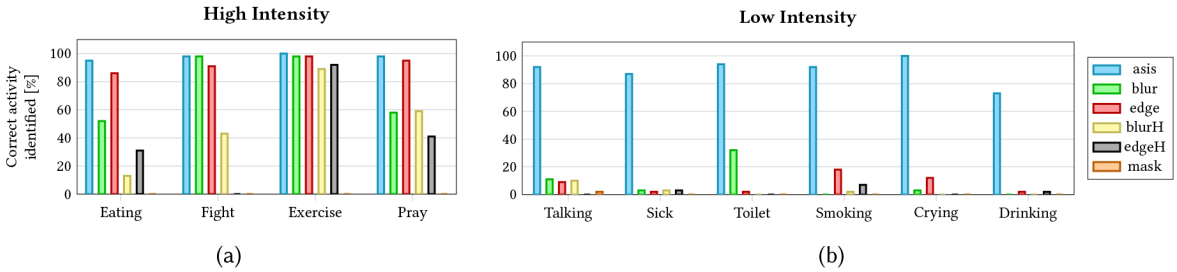


Fig. 4. Visual confirmation utility:

Percentage of correctly identified wearer activities using each obfuscation method across each scenario/video. Activities are grouped as: (a) hand-to-head activities (H2H) with an object in hand, (b) H2H activities without an object in hand, and (c) non-H2H activities. Detailed accuracy reports and analyses for H2H and non-H2H activities are provided in Table 4.



(a) (b)

Fig. 5. Enhancing bystander privacy:

Percentage of correctly identified bystander activities (lower percentage means greater improvement in bystander privacy) using each obfuscation method. Some obfuscation methods (such as blur and edge) fail to obfuscate bystander activities when the intensity or the amount of motion is high (Exercise and Fight in a) but succeed in obfuscating the activity when it is low in intensity (b).

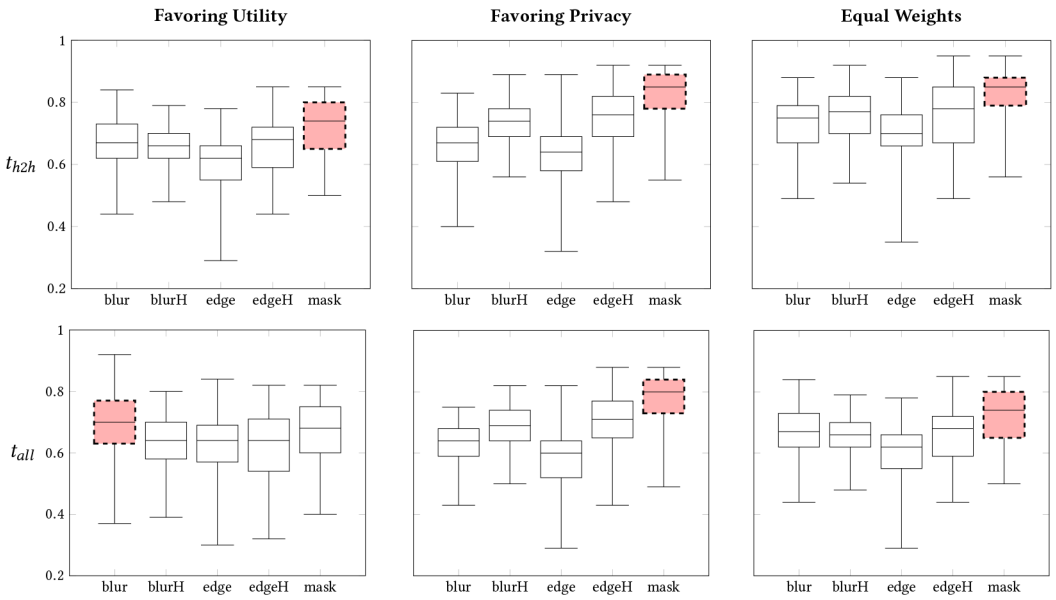
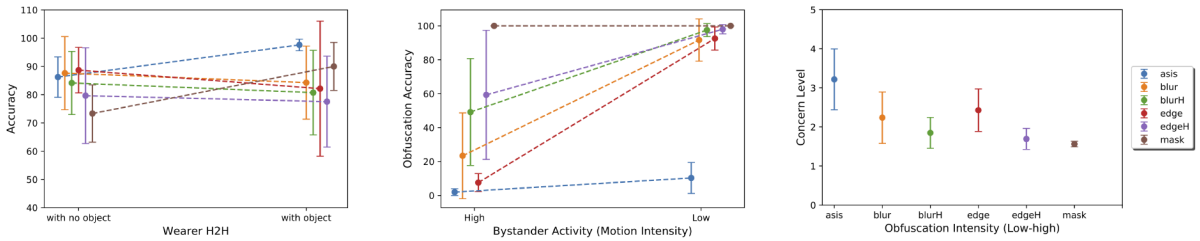


Fig. 6.

Privacy-utility tradeoff scores are depicted. The tradeoff is the weighted sum of bystander privacy (inverse of the ability to detect bystander activity and bystander concerns) and the accuracy of visual confirmation of the wearer's activity. The top row shows the tradeoff (t_{h2h}) when the utility is related to the wearer hand-to-head activities (i.e., contextual information needed to identify the activity is not obfuscated). The bottom row shows the tradeoff (t_{all}) when utility for both hand-to-head (H2H) and non-H2H activities (i.e., the case when important contextual information related to the activity) might be obfuscated. We also show the tradeoff score using different weights: (1) favoring utility, (2) favoring privacy, and (3) equal weights. Partial obfuscation using *mask* provides the best balance in privacy-utility tradeoff when the context is not needed to identify the activity, while *blur* provides the best balance in this tradeoff when context helps in identifying the wearer activity.



(a) (b) (c)

Fig. 7.

The main takeaways of this paper: (a) Activity-oriented obfuscation maintains visual confirmation utility for hand-related activities that involve an object in hand, even when extreme filters (e.g., blurH, edgeH, and mask) are applied to obfuscate the background. (b) Spatio-temporal obfuscation filters (e.g., mask) provide greater bystander privacy than spatial-only methods, in both high- and low-intensity activities (c) Bystander concerns can be significantly reduced using activity-oriented partial obfuscation. Low obfuscation intensity leads to higher variability in bystander concerns as concerns stem from the perceived interpretation of bystander activity (regardless of whether it is the correct activity or not). The error bars represent the standard deviation.

Table 1.

Survey questions about person A (the wearer)

	Question	Response type
A.Q1	Describe Person A's environment, context, or location.	open text
A.Q2	Where is Person A's right hand?	open text
A.Q3	What is Person A doing (with their right hand)?	open text
A.Q4	How confident are you in your answer about what Person A is doing (with their right hand)?	0–10 scale [0=random guess, 10=high confidence]
A.Q5	Do you think Person A is eating or drinking?	Yes, No, Maybe, or I don't know
A.Q6	Do you see any other person in the video other than Person A?	Yes, No, Maybe, or I don't know

Table 2.

Survey questions about person B (the bystander)

	Question	Response type
B.Q1	Describe Person B's environment, context, or location.	Open text
B.Q2	What is Person B doing?	Open text
B.Q3	How confident are you in your answer about Person B's activity?	0–10 scale [0=random guess, 10=high confidence]

Table 3.

Survey questions to understand potential bystander concerns

	Question	Response type
Baseline	Please rate how concerned you would be if such a camera captured you doing the following activities.	5-point Likert scale [1=not at all,5=extremely]
C.Q1	IMAGINE that YOU are Person B in the video you just viewed above, how concerned would you be if you were in Person B's place recorded in the video?	5-point Likert scale [1=not at all,5=extremely]
C.Q2	Why?	Open text

Table 4.

Accuracy of labeling wearer activity.

		Accuracy					
		as-is (ref)	**blur**	**blurH**	**edge**	**edgeH**	**mask**
H2H	Eating	100.0	93.55	95.08	94.64	79.66	90.00
	Wearing Glasses	95.24	66.13	60.66	46.43	54.24	78.33
	Drinking	98.41	83.87	78.69	91.07	86.44	93.33
	Scratching	93.65	100.0	96.72	96.43	96.61	85.00
	Nail biting	85.71	88.71	80.33	89.29	79.66	66.67
	Calling	96.83	93.55	88.52	96.43	89.83	98.33
	Yawning	79.37	74.19	75.41	80.36	62.71	68.33
	Average (H2H)	**92.74**	**85.71** (p<.001)	**82.2** (p<.001)	**84.95** (p<.001)	**78.45** (p<.001)	**82.86** (p<.001)
Non-H2H	Typing	96.83	48.39	0.00	26.79	03.39	01.67
	Washing	98.41	37.10	0.00	03.57	03.39	01.67
	Talking	80.95	38.71	39.34	30.36	22.03	21.67
	Average (All)	92.54	72.42	61.48	65.54	57.80	60.50

Accuracy is reported as percentages.

All p-values are reported after Bonferroni correction.

$\delta = 30$ for the equivalence test.

Table 5.

Accuracy of labeling the bystander activity in the video and participant-reported bystander concerns if they were in place of the bystander in the recorded video.

		Accuracy						Concern Level					
		as-is (ref)	**blur**	**blurH**	**edge**	**edgeH**	**mask**	**as-is (ref)**	**blur**	**blurH**	**edge**	**edgeH**	**mask**
High Intensity	Eating	95.20	51.60	13.10	85.70	30.50	0.00	2.87	2.21	2.03	2.61	2.00	1.52
	Fight	98.4	98.4	42.6	91.1	0.00	0.00	3.67	3.85	2.69	3.66	1.63	1.52
	Exercise	100.0	98.4	88.5	98.2	91.5	0.00	2.87	2.23	2.10	2.59	2.10	1.58
	Pray	98.4	58.1	59.0	94.6	40.7	0.00	3.40	2.52	2.13	2.89	2.02	1.63
	Talking	92.1	11.3	9.80	8.90	0.00	0.00	2.03	1.87	1.67	2.14	1.49	1.60
	Sick	87.3	3.20	3.30	1.80	3.40	0.00	3.57	1.68	1.44	1.96	1.53	1.48
Low Intensity	Bathroom	93.7	32.3	0.00	1.80	0.00	0.00	4.51	2.63	1.59	1.82	1.49	1.53
	Smoking	92.1	0.00	1.60	17.9	6.80	0.00	2.86	1.84	1.59	2.21	1.83	1.50
	Crying	100.0	3.20	0.00	12.5	0.00	0.00	4.10	1.90	1.84	2.36	1.37	1.72
	Drinking	73.0	0.00	0.00	1.80	1.70	0.00	2.29	1.65	1.41	2.04	1.47	1.55
	Average	**93.02**	**35.65** (p<.001)	**21.80** (p<.001)	**41.43** (p<.001)	**17.46** (p<.001)	**0.00** (p<.001)	**3.20**	**2.20** (p<.001)	**1.80** (p<.001)	**2.40** (p<.001)	**1.70** (p<.001)	**1.60** (p<.001)

Accuracy is reported as percentages.

Concern level has a scale of 1–5.

Table 6.

Distribution of the coded reasons in two groups of interest: Low Concern and High Concern.

	Low Concern						High Concern					
	blur	**blurH**	**edge**	**edgeH**	**mask**	**Total**	**blur**	**blurH**	**edge**	**edgeH**	**mask**	**Total**
Activity	140	82	82	59	0	**363**	81	46	81	15	0	**223**
Context	33	26	9	6	50	124	17	4	5	4	16	46
Fundamental	59	42	64	47	8	220	95	69	122	75	58	**419**
Identity	11	9	16	20	1	57	4	7	3	3	0	17
Interpretation	1	2	7	0	1	11	15	3	9	7	0	34
Obfuscation	155	292	151	338	413	**1349**	5	8	2	4	10	29
Other	3	7	7	2	2	21	0	3	2	0	1	6
All	402	460	336	472	475	**2145**	217	140	224	108	85	**774**

Low Concern: concern level < 3 (i.e., “Not at all concerned” or “A little concerned”).

High Concern: concern level ≥ 3 (i.e., “Moderately concerned”, “Concerned”, or “Extremely concerned”).