Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation

MARIELLA DIMICCOLI*, University of Barcelona and Computer Vision Center, Spain
JUAN MARÍN, University of Barcelona, Spain
EDISON THOMAZ, University of Texas at Austin, USA

Recent advances in wearable camera technology and computer vision algorithms have greatly enhanced the automatic capture and recognition of human activities in real-world settings. While the appeal and utility of wearable camera devices for human-behavior understanding is indisputable, privacy concerns have limited the broader adoption of this method. To mitigate this problem, we propose a deep learning-based approach that recognizes everyday activities in egocentric photos that have been intentionally degraded in quality to preserve the privacy of bystanders. An evaluation on 2 annotated datasets collected in the field with a combined total of 84,078 egocentric photos showed activity recognition performance with accuracy between 79% and 88% across 17 and 21 activity classes when the images were subjected to blurring (mean filter k=20). To confirm that image degradation does indeed raise the perception of bystander privacy, we conducted a crowd sourced validation study with 640 participants; it showed a statistically significant positive relationship between the amount of image degradation and participants' willingness to be captured by wearable cameras. This work contributes to the field of privacy-sensitive activity recognition with egocentric photos by highlighting the trade-off between perceived bystander privacy protection and activity recognition performance.

CCS Concepts: • **Security and privacy → Privacy protections**; • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Computing methodologies → Activity recognition and understanding**; *Supervised learning by classification*;

Additional Key Words and Phrases: Wearable Cameras, Egocentric Vision, Privacy, Activity Recognition, Image Degradation

ACM Reference Format:
Mariella Dimiccoli, Juan Marín, and Edison Thomaz. 2017. Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 132 (December 2017), 18 pages. https://doi.org/10.1145/3161190

1 INTRODUCTION

Understanding human behavior is a central theme in ubiquitous computing and human-computer interaction [48]. For many years, the vision of systems that capture the complexities of human life, anticipate our intentions and adapt to accommodate our needs has been all but a promise. However, in recent years, new sensing and

*This is the corresponding author

Authors' addresses: Mariella Dimiccoli, University of Barcelona and Computer Vision Center, Gran via de les corts Catalanes 585, Barcelona, 08007, Spain, mariella.dimiccoli@cvc.uab.es; Juan Marín, University of Barcelona, Barcelona, 08007, Spain; Edison Thomaz, University of Texas at Austin, Department of Electrical and Computer Engineering, Austin, TX, 78703, USA.

2474-9567/2017/12-ART132 $$15.00
https://doi.org/10.1145/3161190

computational approaches based on wearable cameras and computer vision techniques have emerged, unlocking new opportunities in the *capture* and *understanding* of human activities in naturalistic settings. Photographs captured from a first-person perspective, also referred to as *egocentric*, record everyday moments automatically, in high-fidelity, and with an unprecedented richness in detail. As a result, wearable cameras have become increasingly popular in a variety of activity recognition applications over the last decade including cognitive and autism support [20, 30], mobility behavior profiling [25], dietary monitoring [41, 46], law enforcement accountability [9], and many others. Recently, computational methods based on Convolutional Neural Networks (CNN) [23, 28] have been largely successful at translating these photos into activity recognition models [6, 8, 39].

Despite the benefits and unique attributes of first-person photography, privacy concerns and a general discomfort with always-on automatic photo-capture devices have limited the mainstream adoption of applications and services that rely on wearable cameras for activity recognition. These concerns are justified; the continuous recording of every experiential moment, whether at home, at work, around family, or in public spaces, implicates not only those doing the recording but anyone who happens to be photographed or video-recorded. In particular, this technique has greatly enhanced the vulnerability of bystanders [14]. Egocentric photos might catch bystanders in embarrassing situations, undesirable poses, or reveal information they would rather not have on record.

In this work, we demonstrate a CNN-based computational method for recognizing everyday human activities with egocentric photos while mitigating bystander privacy concerns. Our approach is based on the hypothesis that by *intentionally degrading the quality of egocentric photos, such as by obfuscating them with a blurring filter, we can increase the willingness of bystanders to be captured by wearable cameras for activity recognition applications*. We make the assumption that willingness to be captured in an egocentric photograph is a proxy for perceived privacy protection. To validate this hypothesis, we first conducted a crowd-sourced study with 640 survey participants and proved that there is a statistically significant positive correlation between amount of image degradation applied to egocentric images and perception of privacy. Our analysis showed that it is possible to increase the percentage of bystanders willing to be captured by wearable camera photographs by 17.5% when a blurring filter is applied to these photos.

Next, we analyzed how well a deep learning-based classifier can recognize activities from egocentric images that have been subjected to different levels and types of filtering (i.e., mean, Gaussian, and bilateral blur). More concretely, we applied a slight variant of the late fusion ensemble methods demonstrated by Castro et al.[8] and Cartas et al. [7], and examined activity recognition accuracies when our models were trained both with the original and altered photographs. For this analysis, we compiled an annotated dataset of 39,176 first-person photos and also leveraged a publicly available dataset of 44,902 first-person images for a combined total of 84,078 photos.

To our knowledge, our work represents the first quantitative study investigating the trade-off between perceived bystander privacy protection and activity recognition performance with egocentric photos. Unlike prior work that flags specific sensitive objects or places such as faces, screens or bathrooms, we degrade the quality of entire photo set using image filters. Consequently, sensitive content is always degraded as if it were detected with a recall equal to one; this is a considerable advantage with respect to selective blurring strategies that are prone to errors. Additionally, blurred images can still be annotated and used to fine-tune the system for a given user. Since sensitive content has been blurred, the images could be annotated even by a third person, which is desirable when the user is an older or a person with cognitive or visual disability. Finally, in addition to the proposed technique, we also contribute the image dataset [1] that we collected in our analysis to encourage the community to build upon our work.

[1] Available at http://users.ece.utexas.edu/~ethomaz/publications.html

2 RELATED WORK

2.1 Activity Recognition and Wearable Cameras

Many efforts have applied crowdsourcing and computer vision techniques to scale the recognition of activities from first-person perspective photos [6, 7, 17, 34, 38, 46, 51]. For instance, Castro et al. showed accuracies above 80% when predicting a person's activities from egocentric images and contextual information using a standard CNN [8]. The dataset used for this analysis, which was compiled over a 6-month period in completely free-living conditions, represented 19 types of activity for a total of 40,103 photos. However, beyond a few approaches that can identify specific privacy-sensitive objects (i.e., computer screens) [26], lack of privacy protection remains an impediment to the adoption of wearable cameras in activity recognition research.

2.2 Privacy in Activity Recognition

Once images have been manipulated, questions arise as to whether they can be used by activity recognition algorithms and how effective they are at mitigating privacy risks. Kelly et al. proposed a set of ethical principles and formalizations for dealing with this issue in health research when wearable cameras are in use[24]. Ryoo et al. [37] explored the problem with extreme low-resolution images (e.g., 16x12 pixels) by introducing the concept of Inverse Super Resolution (ISR). ISR is based on the assumption that multiple low-resolution (LR) images may contain a comparable amount of information to a single high-resolution (HR) image. Since the process of generating LR images is achieved by learning the optimal set of motion transforms by relying on motion and convolutional features, applicability is restricted to videos.

To help quantify the effectiveness of privacy-protecting computational approaches while also examining the negative impact of these approaches on the recognition task, Thomaz et al. [45] developed a general framework called privacy-saliency matrix. It was evaluated with four techniques (i.e., face detection, image cropping, location filtering and motion filtering) on a dataset gathered by 5 participants in free living conditions. Similarly, Korshunov et al. examined the balance between the intelligibility of activities and privacy preservation in a video surveillance setting. In this work, several privacy protection techniques were tested, from pixelation to masking [27].

2.3 Privacy in Photo Sharing

Due to the rise in photos taken with mobile devices and the popularity of social media, several efforts have investigated privacy concerns in the context of photo sharing. Tanaka et al. applied a price sensitivity measure to data collected in a online survey to study this question [43]. Besmer and Lipford conducted focus groups to understand privacy perceptions of photo sharing in Facebook [4]. They found that privacy concerns in this domain center on identity and impression management. Finally, Ra et al. developed a privacy-preserving photo sharing system that can be used by service providers [36]. These efforts are relevant to our work because they illustrate attempts to mitigate privacy risks with image degradation (e.g., face blurring upon face detection). However, these research problems are not set in the context of wearable imaging.

2.4 Selective Filtering

Selectively eliminating or identifying elements that pose privacy threats in photographs is an approach that has been explored by researchers. For example, Korayem et al. demonstrated the feasibility of automatically detecting computer screens in first-person photographs [26]. This work was motivated by prior studies highlighting exposure to private information that is displayed in screens, and thus get captured by wearable cameras [21, 22]. While the authors point out that Americans now spend more time on digital devices than watching TV, their work focuses on the detection of desktop and laptop computer displays, but not to phones, tablets, or other electronic devices.

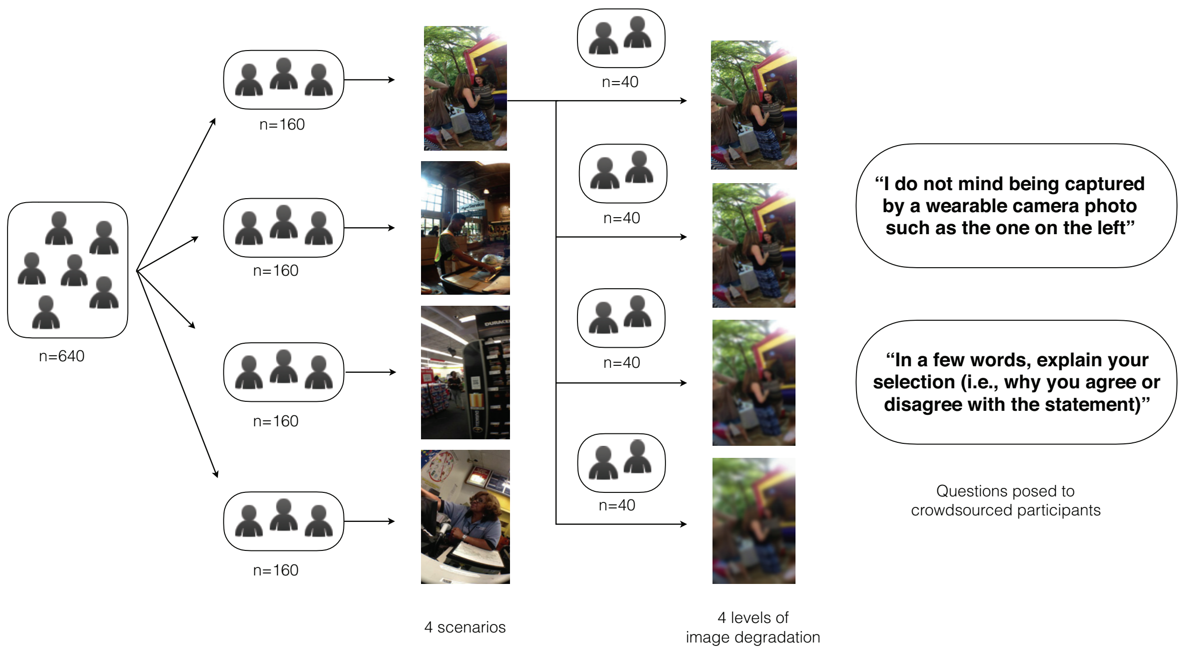

Fig. 1. Survey participants (n=640) recruited with Amazon's Mechanical Turk were evenly divided into 4 groups and 4 sub-groups based on different scenarios of egocentric image capture and amount of degradation (i.e., Gaussian blur) applied to the images. Participants indicated whether they would be comfortable if captured by a wearable camera across 16 conditions on a 5-point Likert scale, and were asked to justify their answer in free text form.

Along the same lines methodologically, Templeman et al. proposed a technique that allowed those wearing on-body cameras to explicitly flag sensitive physical spaces such as bathrooms [44]. The approach requires individuals to take photographs of these sensitive places, which are used in the construction of visual models of rooms that should not be captured. The weakness of this approach is that sensitive places in less visited environments such as bars and restaurants have a greater probability of not being detected by the system because of the lack of similar images in the training set.

One common technique to protect the privacy of subjects visible in video sequences is to blur or pixelize sensitive areas of the image, such for example, a person's face [5, 16]. This approach has been traditionally used in news broadcasts. While it might be possible to detect a face even while it is moving, individual identity is preserved as the individual cannot be recognized.

A fundamental limitation of automated selective filtering as a way to address privacy concerns is that it is virtually impossible to guarantee that all privacy threats will be correctly identified and subsequently eliminated with current algorithms, such as face detection.

2.5 Understanding Bystander Perceptions

There is an extensive body of prior work examining bystander's reactions to CCTV recordings and cameras in public spaces [2, 3, 15, 31, 32]. These studies are relevant to our work but they are neither focused on wearable camera photos nor on the effect of photo degradation on bystanders' perceptions of privacy.

Gaussian Filter					
	No Filter (NF)	$k = 11, \sigma = 2$	$k = 25, \sigma = 4.1$	$k = 45, \sigma = 7.1$	NF vs. k=45
Strongly Agree	12.5%	12.02%	13.12%	17.5%	**+5%**
Agree	23.75%	30.38%	36.87%	36.25%	**+12.5%**
Neutral	9.37%	9.49%	9.37%	8.12%	-1.25%
Disagree	30.62%	32.27%	26.25%	23.75%	**-6.87%**
Strongly Disagree	23.75%	15.82%	14.37%	14.37%	**-9.38%**

Table 1. As degradation was increasingly added to egocentric images (i.e., by increasing the kernel size of the Gaussian blur filter), the more comfortable survey participants felt about being captured by wearable cameras.

Hoyle et al. investigated privacy perceptions of on-body cameras from the perspective of those wearing the device [22]. The study led to several revealing findings, from factors that determine the privacy sensitivity of a photograph to how camera wearers feel about the privacy of bystanders. In related work, the authors also studied a dataset of 14,477 images from 36 participants to further understand privacy and sharing settings [21].

In a series of in-situ studies at cafés, Denning et al. interviewed 31 bystanders regarding their reactions to a co-located augmented reality device [11]. Findings revealed that bystanders were either neutral or negative to the presence of the device, but would be more accepting if they could be asked for permission whenever they were being recorded.

The study by Nguyen et al. [33] is particularly relevant to our effort. Using paratyping [1], the researchers assessed bystander reactions to a wearable camera also in realistic settings. Our work is different in that we are interested in bystander reactions to the images recorded by the wearable devices under different levels of degradation, and not just the presence of the device itself.

3 IMAGE DEGRADATION AND THE PERCEPTION OF PRIVACY

As shown in the previous section, researchers have explored a myriad of approaches to understand and mitigate privacy concerns when cameras are used in public settings. However, none of the prior work has specifically studied the effect of photo quality degradation on bystander's level of comfort when captured by wearable cameras. To address this question, we conducted a survey by recruiting crowd sourcing workers. Based on the survey results, we computed a correlation between the kernel size of a Gaussian filter, which determines how much blurring is applied to a photograph, and how comfortable bystanders feel if captured in a photograph with and without blurring. Additionally, we also compiled unstructured textual feedback from participants reflecting their views on the study scenarios presented.

We chose blurring for image degradation in our work because Gross et al. [19] demonstrated that privacy protection afforded by severe blurring is more effective than pixelation for sensitive applications [49]. Additionally, blurring has been successfully used as a privacy-protecting measure in a variety of high-profile services such as the Google Street View, where faces and car license plates are intentionally and selectively obfuscated.

3.1 Survey

We recruited 640 participants using Amazon's Mechanical Turk platform. No specific qualifications were required for the recruited participants; they came from the global pool of Mechanical Turk workers (i.e., native English speaking ability was not enforced) and were paid 20 cents for completing the survey task. The survey was split in two phases of 320 turkers each; the second phase was administered 2 months after the first phase. The survey began with a brief introduction to wearable cameras and how this class of devices typically operates,

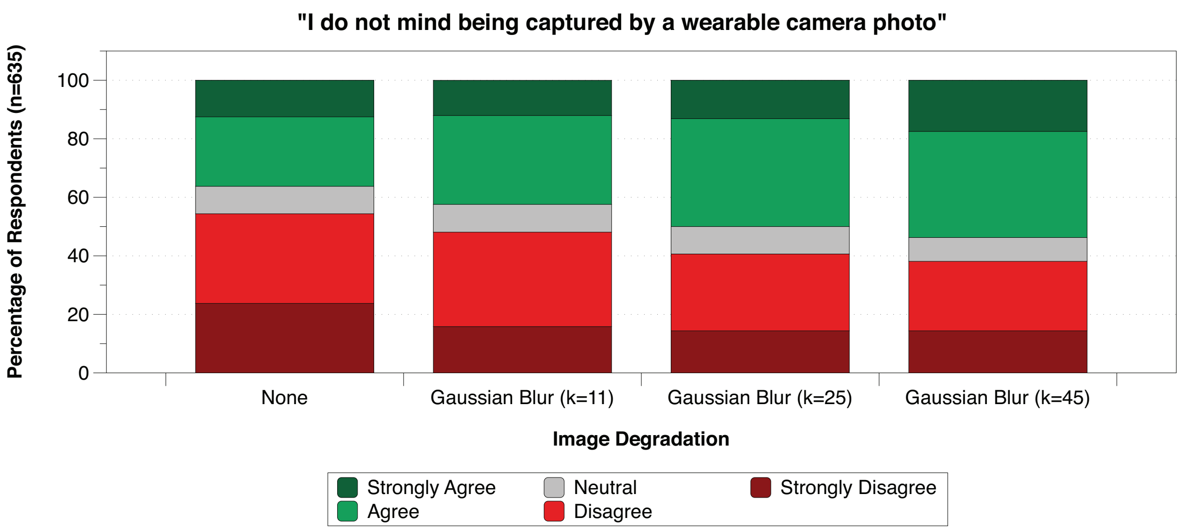

Fig. 2. Percentage of respondents to the statement *"I do not mind being captured by a wearable camera"* for different levels of image degradation. We surveyed 640 crowd workers using Amazon's Mechanical Turk platform and divided them into 4 groups. For 3 of the groups, a different amount of Gaussian blur (i.e., *k=11*, *k=25*, *k=45*) was applied to the sample wearable camera photo shown to the participants

.

capturing photos automatically at regular intervals. Next, we asked survey participants to consider a hypothetical scenario where they find themselves in a public or semi-public physical space with other people, such as at a party or grocery store. Within this setting, an individual that is unknown to the survey participant is wearing a body-mounted camera, which is taking first-person photographs automatically. The survey participant was then told that he or she is captured in one of the photographs. Here is the description for one of the scenarios: *"Imagine you work at the post-office and a customer approaches you wearing a body-mounted camera; the wearable camera is taking photos automatically every 30 seconds. The camera also has a blurring filter for privacy protection. In at least one of the photos, shown on the left, you are captured while doing your job."*

To illustrate what the photograph might look like, a real photo taken with a wearable camera in a realistic setting was shown to the participant. The picture depicted an individual performing some task (e.g. bagging groceries at the supermarket, talking to a friend at a party), unaware that a photograph was being shot. For realism, the task and location of the photograph matched the setting of the hypothetical scenario. Since our goal was to assess the impact of image degradation on the perception of privacy, a filter was applied to the photograph depending on the group the survey participant was assigned to. Using a 5-point Likert-scale, we asked participants to indicate how much they minded being captured by a wearable camera photograph, such as the one presented to them. More specifically, they were asked to indicate how they felt about the statement: *"I do not mind being captured by a wearable camera photo such as the one on the left"*. Possible response choices were *"Strongly Agree"*, *"Agree"*, *"Neutral"*, *"Disagree"*, and *"Strongly Disagree"*.

In the second phase of the survey, we included two additional questions aimed at giving us more insight into the profile of survey participants. We asked them to enter their age in a text field and to react to a statement

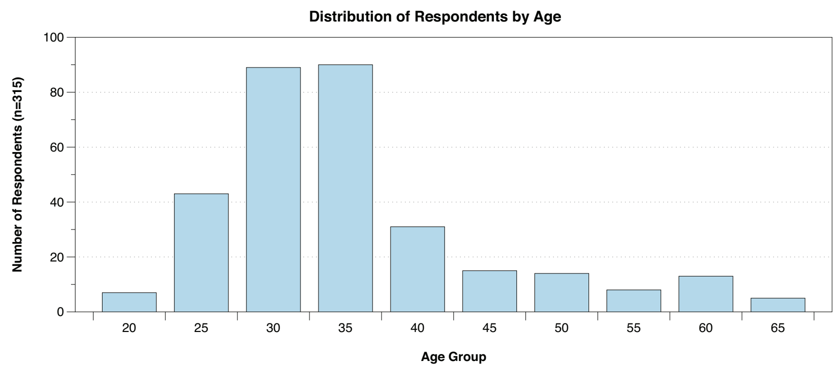

Fig. 3. The distribution of crowd-sourced participants by age. Most of them were between 25 and 40 years of age.

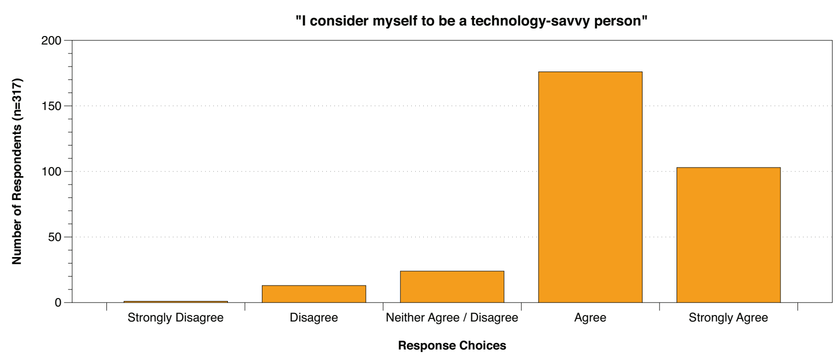

Fig. 4. How the crowd-sourced participants responded to the statement *"I consider myself to be a technology-savvy person"*. Most participants either agreed or strongly agreed.

about how comfortable they were with technology (*"I consider myself to be a technology-savvyy person"*). Like for the previous statement, responses were recorded on a 5-point Likert-scale.

As shown in Figure 1, we divided the 640 crowd workers into 4 groups of 80 participants each. For 3 of the groups, a different amount of Gaussian blur (i.e., *k=11*, *k=25*, *k=45*) was applied to the wearable camera photo shown to the participants in the task scenario. The remaining group, the control group, was shown a photo without any blurring. Since the setting in which a wearable camera photo is captured might influence how people perceive the privacy risk of a photograph, we further divided each of the 4 groups into 4 sub-groups of 20 participants each. The sub-groups presented participants with different wearable photo capture scenarios (i.e., party, grocery store, post office, and shopping); the photos presented to participants always matched the respective sub-group scenario. These scenarios were picked to loosely match the types of interpersonal ties described by Granovetter: strong (i.e., family and friends), weak (i.e., acquaintances), absent (i.e., "nodding" relationships), or none [18]. Finally, the photos used in this study were taken from the FPPWild-4 dataset described in 4.1.

3.2 Results

Overall, the results of the survey matched our expectations. As degradation was increasingly added to egocentric images (i.e., by increasing the kernel size of the Gaussian blur filter), the more comfortable survey participants felt about being captured by wearable cameras. More specifically, we found that there is a statistically significant positive relationship between the amount of image degradation, measured by the kernel size of the Gaussian filter, and participants' willingness to be captured by wearable cameras. This was particularly true when combining the "Strongly Agree" and "Agree" responses ($r(2)=.99$, $p<.05$). As shown in Figure 2, it is possible to visualize both the rise of "Strongly Agree" and "Agree" choices as the photos becomes blurrier. This is also reflected in Table 1; when combining the "Strongly Agree" and "Agree" entries, there was a 17.5% percentage increase in positive responses to being captured in egocentric photos with blurring versus without (i.e., NF vs k=45). When examining the negative responses the same way, a decrease of 16.25% was observed. Due to data entry errors, a small number of survey responses had to be discarded from the analysis.

We also studied how the survey responses varied depending on the scenarios that were presented. In other words, considering that all images undergo the same degradation process, we were interested in the sensitivity of the responses to the scenario itself. This analysis was performed with survey data from 320 participants. The measurement unit we used was the gain in responses towards willingness-to-be-photographed when considering the original images and the most degraded images (k=45). The most significant change was observed in the shopping scenario, with a gain of 10 positive responses, followed by the post-office scenario with a gain of 6 responses. The party and grocery store scenario saw virtually no change in positive responses. A reason for this might be the distance to the bystander when the picture was taken, which can have an amplifying effect on the level of image degradation applied; in the shopping scenario, the bystander is about 10ft away from the camera. Another reason for little or no change in positive responses could be the setting; in the party scenario, even when maximum degradation is applied, bystanders might not feel that it is appropriate to have photographs of them taken due to the intimacy of the occasion. We reserve these research questions for further studies.

Participants were also encouraged to explain their selection in a free-form text field associated with the task. As expected, many participants expressed worry when no filters were applied to photographs. One participant wrote: *"You can clearly see me, and someone could find out where I work"*. Other responses were *"I don't want photos of me being taken without my permission with no reasonable cause"* and *"I find this to be a complete invasion of one's personal space and privacy. As such, I would be very upset to have my photo taken as such"*. However, and to our surprise, a share of participants were very accepting of the photos even though they were not degraded in any way: *"It does not bother me to have my picture taken by a wearable camera. I have nothing to hide"*, and *"It's in a public place and I'd just be doing my job. I don't mind this type of photo taking"*. Interestingly, one participant suggested that degrading the photo would be an acceptable compromise: *"I would prefer if the picture was blurred"*.

At the other end of the spectrum, when Gaussian blur was applied to the photos with *k=45*, many participants did not mind being captured in wearable camera photographs: *"The blurring of the photo provides anonymity"*, *"This one is blurry enough that I'm indifferent to it"*, and *"This is so blurry I can't even tell if I'm a man or woman, so I'm fine with it"*. One response emphasized the importance of the filter: *"As long as it has a blurring filter, I don't particularly mind it capturing me or other bystanders. Without this filter it would be completely unacceptable for someone else to have that much control over my privacy"*. But for some participants, even a highly blurred photograph was not enough as a privacy-protecting measure:*"I still think it's imposing on my privacy. The blurring of the image makes it seem a little bit better, but the overall consensus is that my privacy is bring intruded on and I do not like it"*

Examining the age of survey participants and how tech savvy they consider themselves (Figures 3 and 4), it is possible to see that the distributions are skewed towards a younger population that is comfortable with computer

Fig. 5. Mobile phone worn on a lanyard and programmed to work as a wearable camera.

technology. This is not surprising considering that participants were recruited through Amazon's Mechanical Turk and must be taken into account when considering the generalizability of our results.

4 COMPUTATIONAL APPROACH

The crowd-sourced survey validated our thesis that image degradation does improve bystander's perception of privacy. However, prior work in activity recognition has not investigated whether degraded images can be processed as effectively as non-degraded photos in discriminating activities in naturalistic settings. We address this research question by developing and evaluating an activity recognition approach that operates on egocentric images using deep learning methods. In this section, we present this approach and provide more details about the datasets we used for training and evaluation.

4.1 Data Collection and Annotation

To study the impact of image degradation on activity recognition performance, we needed a large set of images collected under free living condition that offered good coverage of an individual's daily activities. Public datasets widely used for activity recognition in the computer vision community such as GTEA, GTEA+, GTEA Gaze [12, 13, 29], ADL [35] have been acquired by wearing a video camera for a short period of time. As a result, they include a limited set of higher-level activities (e.g., cooking, eating) and a much larger number of short term actions (e.g., pick up bread, open drawer). Moreover, they are usually captured either in home or office environments with few if any bystanders around. Indeed, the only person visible in most photos of these datasets is almost always the individual wearing the camera (e.g., looking at himself in the mirror). These characteristics render these datasets unsuitable for our study. One of the exceptions is the public NTCIR-12 dataset. It consists of 89,593 images belonging to three users with different lifestyles. The images were captured by a OMG Life Autographer wearable camera with a frame rate of 2 pictures per minute over a period of about a month per each user. Face blurring was applied to all photos. We used a subset of 44,902 images (approximately 15,000 per user) that were manually annotated with 21 activity labels by Cartas et al. [7] to test our computational approach.

To increase the external validity of our study, we complemented the NTCIR-12 public dataset by compiling another dataset: 39,176 first-person photos collected over a period of 4 months. We called this dataset FPPWild-4.The FPPWild-4 photos were captured by a 39-year old married male with kids living in the center of a metropolitan

Fig. 6. Examples of annotated images in our dataset with their corresponding activity label.

city in the United States. The individual wore a mobile phone on a lanyard that was programmed to operate like a wearable camera and take a front-facing photograph automatically every 60 seconds (Figure 5). Wearing the camera everyday, and in naturalistic settings, ensured the ecological validity of the study. At his discretion, the participant was instructed to temporarily stop logging during the day when capturing photographs was perceived to be inappropriate or not allowed.

To obtain ground truth annotation for the FPPWild-4 dataset, we had the individual wearing the camera manually annotate the images. To facilitate this task, we designed a web-based annotation tool that allowed captured photographs to be reviewed, labeled and deleted. The tool greatly reduced the time and effort required to filter photos with privacy threats, and streamlined the annotation step by permitting groups of photos to be assigned to a label at once. For each photo or photo group, the individual could pick one label out of 17 choices such as “working", “exercising", “eating", and “shopping". The complete set of activity labels included in the web-based annotation app, shown in Table 1, was chosen prior to data collection and in consultation with the subject. The 17 labels reflected the individual’s lifestyle, and represent routines and activities that are commonplace in everyday life. If a particular activity could not be visually inferred from a photo, the participant had the choice to leave the photo without a label. Unlabeled and deleted images were not part of the 39,176 photo dataset. Finally, no photographs were collected while the participant was asleep; photos were typically collected between the hours of 8AM and 9PM, every day of the week.

4.2 Image Degradation by Image Processing

To degrade images, we employed 3 commonly used *blurring* filters: a Mean Filter, a Gaussian Filter and a Bilateral Filter. In general, these filters operate by attenuating high frequencies more than low frequencies, which results in the removal of image details and noise.

Dataset Summary			
Activity	# of Photos	Activity	# of Photos
Working	13682	Cooking	764
Driving	1225	Doing Chores	723
Eating	4753	Presentation	844
Socializing	9185	Biking	691
Meeting	1339	Cleaning	637
Watching TV	1261	Chatting	113
Exercising	500	Shopping	606
Playing with dogs	1149	Resting	116
Reading	1408	**Total**	**39176**

Table 2. Our dataset summary: activity labels and corresponding number of annotated images.

- **Mean Filter**: Its principle of operation hinges on replacing each pixel value in an image with the mean value of pixels, including itself, in a neighborhood whose shape and size is specified by the user. Commonly, a 3×3 square kernel is applied to remove the pixel values which are unrepresentative of their surroundings, whereas a 5×5 square kernel is used to obtain a more severe smoothing.
- **Gaussian Filter**: It is similar to the mean filter, but it uses a different kernel that represents the shape of a Gaussian. Applying a Gaussian filter to an image is the same as convolving the image with a Gaussian function. The degree of smoothing is determined by the standard deviation of the Gaussian. Larger standard deviation Gaussians require larger convolution kernels in order to be accurately represented. In this work we computed the standard deviation σ as a function of the kernel size k with this formula $\sigma = 0.3 * ((k-1) * 0.5 - 1) + 0.8$.
- **Bilateral Filter**: It operates by replacing the intensity value at each pixel in an image by a weighted average of intensity values from nearby pixels. The weights depend not only on Euclidean distance of pixels, as in the case of a Gaussian filter, but also on the visual differences (e.g. color intensity). Therefore, it is particularly indicated to blur the image while preserving edges. In fact, pixels having color intensity similar to the pixel value to be replaced have larger weights than equally distant pixels having very different color intensity values. A bilateral filter is fully specified by two parameters, say k and δ, that determine the spatial and the range domain of the filter.

As it can be seen in Figure 7, we considered three different levels of degradation for each filter. The choice of filter parameters was driven by two criteria, (1) the desire to preserve bystander privacy, and (2) the goal of maximizing classifier performance. All filter used square kernels.

4.3 Model Training and Testing

For automatic activity recognition, we used a CNN approach similar to Castro et al. [8]. However, instead of using the AlexNet architecture, we used the GoogLeNet architecture [42] since its "depth" makes it more suitable for capturing complex object relations that characterize activities.

The network was initialized with the weights obtained after training on ImageNet [10] for the task of object detection. The dataset was split into 70% for training, 15% for validation and 15% for testing. All layers of the network were fine-tuned by using 0.001 base learning rate with a step policy, a momentum of 0.9, a 0.5 drop rate each 10 epochs. In Table 3, we report the global accuracy achieved after 20 epochs for 3 types of train-test

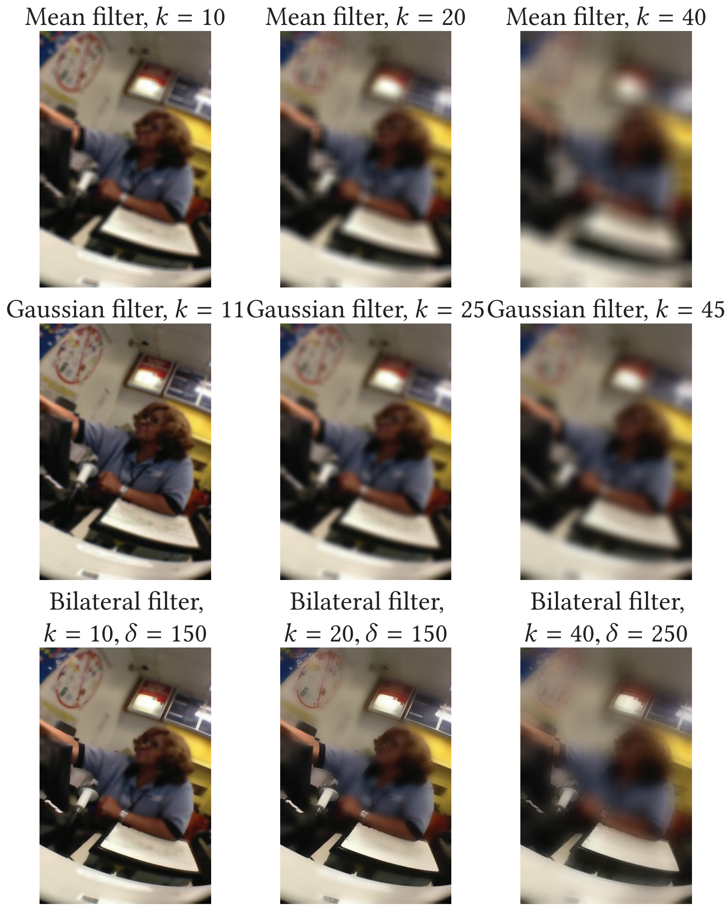

Fig. 7. Examples of images degraded using the mean filter (first row), the Gaussian filter (second row) and the bilateral filter (third row).

paradigms: (1) using the original images for training and test, (2) using different types and levels of degradation for training and testing, and finally (3) using the original images for training and the degraded images for testing.

Since the dataset is made of images belonging to a single person, and typically, people tend to follow routines in their daily lives, metadata such as the the day of the week and the time of the day convey useful a priori information to improve classification performance. As in [8], we used a Late Ensemble Method to combine the activity probabilities obtained using the CNN with temporal metadata, and used the resulting feature vector as input to a random forest classifier.

Since the NTCR-12 dataset has been acquired by three users with different lifestyle, it is not possible to take advantage of their time information and color histogram. In addition the task is more challenging on this dataset because we must deal with an increased intra-class variability with a much smaller number of images per user. To face these problems, instead of using time and color as contextual information, Cartas et al. [7] used the output of a fully connected layer of the CNN. In this work, we used a similar approach. All layers of the network were fine-tuned by using 0.000067 base learning rate with a step policy, a momentum of 0.9, a 0.5 drop rate each 10

Training and testing on the original images			
CNN	88%		
CNN+RF	88%		
Training and testing on blurred image			
Mean	k = 10	k = 20	k = 40
CNN	87%	86%	45%
CNN+RF	89%	88%	71%
Gaussian	$k = 11, \sigma = 2$	$k = 25, \sigma = 4.1$	$k = 45, \sigma = 7.1$
CNN	88%	87%	62%
CNN+RF	89%	89%	78%
Bilateral	$k = 10, \delta = 150$	$k = 20, \delta = 150$	$k = 40, \delta = 250$
CNN	88%	82%	66%
CNN+RF	89%	88%	66%
Training on the original images, testing on blurred image			
Mean	$k = 10$	$k = 20$	$k = 40$
CNN	83%	70%	45%
CNN+RF	83%	71%	46%
Gaussian	$k = 11, \sigma = 2$	$k = 25, \sigma = 4.1$	$k = 45, \sigma = 7.1$
CNN	86%	79%	66%
CNN+RF	86%	79%	67%
Bilateral	$k = 10, \delta = 150$	$k = 20, \delta = 150$	$k = 40, \delta = 250$
CNN	85%	81%	65%
CNN+RF	86%	81%	66%

Table 3. **Global accuracy values obtained with the dataset we compiled for this experiment, i.e., FPPWild-4.**

epochs. In Table 3, we report the global accuracy achieved after 20 epochs for 3 types of train-test paradigms specified above.

5 RESULTS

Overall, we were encouraged by the results of our evaluation, shown in Table 3 and Table 4 . When training and testing on the blurred images, accuracy loss for all filters with kernel sizes k=10 and k=20 was minimal. There was a much more significant drop in accuracy when k=40, when image degradation was so high that it became difficult to identify object boundaries and edges.

When training on the original images and testing on the blurred images, the pattern was similar, except that a more accentuated drop in accuracy was observed with kernel sizes k=10 and k=20. This was especially true for the Mean filter; with k=20, accuracy fell to 70% (CNN) and 71% (CNN+RF), a drop of 18% if compared to training and testing with the original images. Similarly to the train-test model on blurred images, when k=40, accuracies fell dramatically.

Finally, as it can be observed in Table 3, since the Bilateral filter is good at preserving edges, the drop in performance when training with the original images is very small compared to when training with the blurred images. This drop increases when using a Gaussian filter, and becomes very large when using the Mean filter.

Training and testing on the original images			
CNN	76%		
CNN+RF	85%		
Training and testing on blurred image			
Mean	k = 10	k = 20	k = 40
CNN	78%	70%	57%
CNN+RF	81%	79%	74%
Gaussian	$k = 11, \sigma = 2$	$k = 25, \sigma = 4.1$	$k = 45, \sigma = 7.1$
CNN	80%	73%	65%
CNN+RF	83%	79%	76%
Bilateral	$k = 10, \delta = 150$	$k = 20, \delta = 150$	$k = 40, \delta = 250$
CNN	81%	80%	56%
CNN+RF	83%	82%	74%
Training on the original images, testing on blurred image			
Mean	$k = 10$	$k = 20$	$k = 40$
CNN	68%	67%	21%
CNN+RF	54%	25%	11%
Gaussian	$k = 11, \sigma = 2$	$k = 25, \sigma = 4.1$	$k = 45, \sigma = 7.1$
CNN	61%	31%	16%
CNN+RF	66%	42%	16%
Bilateral	$k = 10, \delta = 150$	$k = 20, \delta = 150$	$k = 40, \delta = 250$
CNN	61%	58%	51%
CNN+RF	69%	63%	46%

Table 4. **Global accuracy values obtained on a subset of the NTCIR-12 dataset.**

6 DISCUSSION

To reiterate, the goal of this study is to understand how to mitigate bystander privacy concerns by intentionally degrading the quality of egocentric photographs that are used for activity recognition. A key finding of this work is that it is technically possible to maintain recognition accuracies by operating on images that have been blurred with a kernel size k=20. This was particularly true when training and testing on the blurred images belonging to a single person. And perhaps more importantly, from a privacy perspective, it can be seen in Table 1 that images degraded with a blurring filter with a kernel size k=20, does indeed make individuals more comfortable about being captured in wearable camera photos.

In this section, we present relevant topics associated with our approach, starting with a discussion of application categories our approach is suitable for, and ending with the limitations of our study.

6.1 Method Applicability

In light of the results obtained in the studies, it is fitting to highlight practical scenarios for which our approach is and is *not* appropriate. Wearable camera usage typically falls under two categories. The most common utilization of wearable cameras is as a lifelogging device. Wearable cameras such as the Narrative[2] have been designed to

[2]http://getnarrative.com

passively capture memories and experiences that would be difficult or impossible to record (e.g., take photos while surfing or snowboarding). For this type of experience capture, individuals want to record and save high-quality photographs and videos, either for personal use or to share with others. Therefore, saving a degraded version of a photograph or video in this context is *highly undesirable*, even though it might pose privacy concerns to those wearing the camera or bystanders. Therefore, our approach is *not* suitable for this class of applications.

The other application category of wearable cameras characterizes these devices as sensors. In this case, photographs are considered to be a signal from which features and descriptors are passively extracted and analyzed to achieve a specific goal, such as to recognize a set of human behaviors automatically. In practice, a system like this could be used in a myriad of ways, such as to inform a home monitoring system that an older adult is no longer able to perform activities of daily living successfully. Here, the quality of the images as judged by a human is not critical; the images might be blurred or pixelated. What is important is that (1) useful image features can be extracted from the photos and that (2) the photos are intelligible enough to an annotator such that ground truth labels can be assigned to them during a classifier training phase.

While the characterization of photographs as signals might, in principle, not seem to pose any privacy concerns, that is not the case. Training, improving and personalizing a classifier to recognize activities always requires more data, which in this case are egocentric images typically collected in real-world settings. These images must be captured, collected and reviewed for annotations, a process that typically takes place in centralized cloud-based computing systems, i.e., not locally on the device, and most likely not performed by the individual who wore the camera. Even in the most privacy-protecting scenario, when a classifier has been built and is packaged as part of a self-contained "closed" system that does not expose images externally, there will always be a need or opportunity to improve the performance of said classifier. By intentionally degrading captured photographs before processing them during training and run-time operation, our approach adds a new layer of privacy protection that would not be present otherwise.

6.2 Image Degradation Reversibility

The proposed method hinges on the intentional degradation of egocentric photos. It is critically important that the photos remain in their degraded state so that increased privacy can be guaranteed. Recently, several CNN-based approaches for addressing image degradation and in particular image deblurring, have been proposed. However, these works were aimed at the removal of a specific type of non-intentional degradation due to the imaging process itself. For example [50] addressed the problem of removing image degradation due to saturation, camera noise, and compression artifacts. To model focal blur they used, as it commonly done, a disk kernel of size 7. This kind of degradation is not comparable to the severe, intentional blurring (with kernel sizes ranging from 11 to 45) used in our approach. The work of Sun et al. aimed at removing motion blur due to the camera motion or very fast moving target [40]. Their proposed approach relies on optical flow estimation, which does not perform well when the frame-rate of the camera is too low (less than a few pictures per second).

Instead of focusing on the task of image deblurring, Vasilevij investigated the effect of fine-tuning a CNN including blurred images into the training set, with the goal of improving the performances of CNN based object recognition on blurred image [47]. Their study demonstrated that, under these conditions, standard CNN architectures are able to produce internal representations that are invariant to blur. However, in their experiments they used very small kernel sizes ranging from $k=2$ to $k=8$.

To conclude, it is not possible to fully reverse the degradation we intentionally applied to the egocentric images due to the large kernel sizes we used. From a mathematical point of view, if the exact size of the kernel used for blurring and the type of blurring are known, the transformation is reversible. However, in practice, even knowing these variables, numerical errors and border effect do not allows a perfect reconstruction. Since the effects of these errors increases with the size of the filter, blurring transformations with kernel sizes above $k = 8$ are not reversible.

6.3 Limitations

While our studies led to encouraging results, it is important to highlight their limitations. As mentioned in Section 3.2, the participants surveyed in the crowd-sourced study represented a population skewed towards a younger demographic who reported being familiar with technology. Consequently, our survey findings might not apply to older adults, for example. Also, even though we evaluated our computational method with 84,078 egocentric images captured in real-world conditions, these images reflect activities of only a small number of individuals. Finally, our approach was not designed nor is expected to completely eliminate privacy concerns through image blurring. Instead, our aim was to show the rate at which a simple, scalable method reduces the privacy exposure of bystanders and at what recognition performance cost.

7 CONCLUSIONS

Attempts to address privacy concerns in wearable camera photographs have mostly focused on selectively filtering sensitive image regions from photographs. In this paper, and informed by a crowd-sourced study with 640 participants, we have taken a different approach to the problem: focus on intentionally blurring whole images and using them in a deep learning based activity recognition approach. Experimental results performed on two large annotated datasets, have demonstrated that it is possible to apply intentional degradation to the full image (with kernel size *k=11* or above) without leading to a significant performance drop in terms of activity recognition accuracy. To our knowledge, our work represents the first quantitative study investigating the trade-off between perceived bystander privacy protection and activity recognition performance with egocentric photos. By releasing our FPPWild-4 dataset of 39,176 annotated egocentric photos across 17 activity classes, we encourage the research community to build upon and extend our work.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Maura Turolla of Telecom Italia for providing specifications about the application scenario.

The work is supported by the National Natural Science Foundation of China under Grant No.: 61273304_a and Young Scientists' Support Program (http://www.nnsf.cn/youngscientists).

REFERENCES

[1] Gregory D Abowd, Gillian R Hayes, Giovanni Iachello, Julie A Kientz, Shwetak N Patel, Molly M Stevens, and Khai N Truong. 2005. Prototypes and paratypes: Designing mobile and ubiquitous computing applications. *IEEE pervasive computing* 4, 4 (2005), 67–73.
[2] Anne Adams. 2000. Multimedia information changes the whole privacy ballgame. In *Proceedings of the tenth conference on Computers, freedom and privacy: challenging the assumptions.* ACM, 25–32.
[3] Victoria Bellotti and Abigail Sellen. 1993. Design for Privacy in Ubiquitous Computing Environments.(1993). (1993).
[4] Andrew Besmer and Heather Richter Lipford. 2008. Privacy perceptions of photo sharing in facebook. In *Proc. SOUPS*.
[5] Michael Boyle, Christopher Edwards, and Saul Greenberg. 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.
[6] Alejandro Cartas, Mariella Dimiccoli, and Petia Radeva. 2017. Batch-based activity recognition from Egocentric Photo-Streams. In *To appear in the proceedings of the International Conference on Computer Vision (ICCV), Workshop on Egocentric Perception, Interaction and Computing*. IEEE.
[7] Alejandro Cartas, Juan Marín, Petia Radeva, and Mariella Dimiccoli. 2017. Recognizing Activities of Daily Living from Egocentric Images. In *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 87–95.
[8] Daniel Castro, Steven Hickson, Vinay Bettadapura, Edison Thomaz, Gregory Abowd, Henrik Christensen, and Irfan Essa. 2015. Predicting daily activities from egocentric images using deep learning. In *proceedings of the 2015 ACM International symposium on Wearable Computers.* ACM, 75–82.

[9] Fanny Coudert, Denis Butin, and Daniel Le Métayer. 2015. Body-worn cameras for police accountability: Opportunities and risks. *Computer law & security review* 31, 6 (2015), 749–762.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 248–255.

[11] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2377–2386.

[12] Alireza Fathi, Yin Li, and James M Rehg. 2012. Learning to recognize daily actions using gaze. In *European Conference on Computer Vision*. Springer, 314–327.

[13] Alireza Fathi, Xiaofeng Ren, and James M Rehg. 2011. Learning to recognize objects in egocentric activities. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On*. IEEE, 3281–3288.

[14] Md Sadek Ferdous, Soumyadeb Chowdhury, and Joemon M Jose. 2017. Analysing privacy in visual lifelogging. *Pervasive and Mobile Computing* (2017).

[15] Batya Friedman, Peter H Kahn Jr, Jennifer Hagman, Rachel L Severson, and Brian Gill. 2006. The watcher and the watched: Social judgments about privacy in a public place. *Human-Computer Interaction* 21, 2 (2006), 235–272.

[16] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. 2009. Large-scale privacy protection in google street view. In *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2373–2380.

[17] Iván González-Díaz, Vincent Buso, Jenny Benois-Pineau, Guillaume Bourmaud, Gaelle Usseglio, Rémi Mégret, Yann Gaestel, and Jean-François Dartigues. 2015. Recognition of Instrumental Activities of Daily Living in Egocentric Video for Activity Monitoring of Patients with Dementia. In *Health Monitoring and Personalized Feedback using Multimedia Data*. Springer, 161–178.

[18] Mark S Granovetter. 1973. The strength of weak ties. *American journal of sociology* (1973), 1360–1380.

[19] Ralph Gross, Latanya Sweeney, Fernando De la Torre, and Simon Baker. 2006. Model-based face de-identification. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*. IEEE, 161–161.

[20] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: a retrospective memory aid. In *UbiComp'06: Proceedings of the 8th international conference on Ubiquitous Computing*. Springer-Verlag.

[21] Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. 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.

[22] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 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.

[23] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 675–678.

[24] Paul Kelly, Simon J Marshall, Hannah Badland, Jacqueline Kerr, Melody Oliver, Aiden R Doherty, and Charlie Foster. 2013. An ethical framework for automated, wearable cameras in health behavior research. *American journal of preventive medicine* 44, 3 (March 2013), 314–319.

[25] J Kerr, S J Marshall, S Godbole, J. Chen, and A Legge. 2013. Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings. (2013).

[26] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. 2016. Enhancing lifelogging privacy by detecting screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4309–4314.

[27] Pavel Korshunov, Claudia Araimo, Francesca De Simone, Carmelo Velardo, J-L Dugelay, and Touradj Ebrahimi. 2012. Subjective study of privacy filters in video surveillance. In *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*. Ieee, 378–382.

[28] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[29] Yin Li, Zhefan Ye, and James M Rehg. 2015. Delving into egocentric actions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 287–295.

[30] Gabriela Marcu, Anind K Dey, and Sara Kiesler. 2012. Parent-driven use of wearable cameras for autism support: A field study with families. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012), 401–410.

[31] Michael Massimi, Khai Truong, David Dearman, and Gillian Hayes. 2010. Understanding recording technologies in everyday life. *IEEE Pervasive Computing* 9, 3 (2010), 64–71.

[32] David H Nguyen, Aurora Bedford, Alexander Gerard Bretana, and Gillian R Hayes. 2011. Situating the concern for information privacy through an empirical study of responses to video recording. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems. ACM, 3207–3216.
[33] David H Nguyen, Gabriela Marcu, Gillian R Hayes, Khai N Truong, James Scott, Marc Langheinrich, and Christof Roduner. 2009. Encountering SenseCam: personal recording technologies in everyday life. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 165–174.
[34] Gabriel Oliveira-Barra, Mariella Dimiccoli, and Petia Radeva. 2017. Leveraging Activity Indexing for Egocentric Image Retrieval. In *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 295–303.
[35] Hamed Pirsiavash and Deva Ramanan. 2012. Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2847–2854.
[36] Moo-Ryong Ra, Ramesh Govindan, and Antonio Ortega. 2013. P3: Toward Privacy-Preserving Photo Sharing.. In *NSDI*. 515–528.
[37] Michael S Ryoo, Brandon Rothrock, Charles Fleming, and Hyun Jong Yang. 2016. Privacy-Preserving Human Activity Recognition from Extreme Low Resolution. *arXiv preprint arXiv:1604.03196* (2016).
[38] Krishna Kumar Singh, Kayvon Fatahalian, and Alexei A Efros. 2016. Krishnacam: Using a longitudinal, single-person, egocentric dataset for scene understanding tasks. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 1–9.
[39] Sibo Song, Vijay Chandrasekhar, Bappaditya Mandal, Liyuan Li, Joo-Hwee Lim, Giduthuri Sateesh Babu, Phyo Phyo San, and Ngai-Man Cheung. 2016. Multimodal multi-stream deep learning for egocentric activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 24–31.
[40] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. 2015. Learning a Convolutional Neural Network for Non-Uniform Motion Blur Removal. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[41] M Sun, J D Fernstrom, W Jia, S A Hackworth, N Yao, Y Li, C Li, M H Fernstrom, and R J Sclabassi. 2010. A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association* 110, 1 (2010), 45.
[42] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1–9.
[43] Yasuhiro Tanaka, Akihisa Kodate, Yu Ichifuji, and Noboru Sonehara. 2015. Relationship between Willingness to Share Photos and Preferred Level of Photo Blurring for Privacy Protection. In *Proceedings of the ASE BigData & SocialInformatics 2015*. ACM, 33.
[44] Robert Templeman, Mohammed Korayem, David Crandall, and Apu Kapadia. 2014. PlaceAvoider: Steering first-person cameras away from sensitive spaces. (2014).
[45] Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, and Gregory D Abowd. 2013. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 739–748.
[46] Edison Thomaz, Aman Parnami, Irfan A Essa, and Gregory D Abowd. 2013. Feasibility of identifying eating moments from first-person images leveraging human computation. *SenseCam* (2013), 26–33.
[47] Igor Vasiljevic, Ayan Chakrabarti, and Gregory Shakhnarovich. 2016. Examining the Impact of Blur on Recognition by Convolutional Networks. *arXiv preprint arXiv:1611.05760* (2016).
[48] M Weiser. 2002. The computer for the 21st Century. *Pervasive Computing, IEEE* 99, 1 (2002), 19–25.
[49] Rose Wiles, Jon Prosser, Anna Bagnoli, Andrew Clark, Katherine Davies, Sally Holland, and Emma Renold. 2008. Visual ethics: Ethical issues in visual research. (2008).
[50] Li Xu, Jimmy SJ Ren, Ce Liu, and Jiaya Jia. 2014. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*. 1790–1798.
[51] Hong Zhang, Lu Li, Wenyan Jia, John D Fernstrom, Robert J Sclabassi, and Mingui Sun. 2010. Recognizing physical activity from ego-motion of a camera. *Proceedings of IEEE EMBS* 2010 (2010), 5569–5572.

Received May 2017; revised August 2017; accepted October 2017