An Autonomous Architecture that Protects the Right to Privacy

Alan R. Wagner

Department of Aerospace Engineering, Rock Ethics Institute, The Pennsylvania State University, University Park, PA 16802 USA
alan.r.wagner@psu.edu

Abstract

The advent and widespread adoption of wearable cameras and autonomous robots raises important issues related to privacy. The mobile cameras on these systems record and may re-transmit enormous amounts of video data that can then be used to identify, track, and characterize the behavior of the general populous. This paper presents a preliminary computational architecture designed to preserve specific types of privacy over a video stream by identifying categories of individuals, places, and things that require higher than normal privacy protection. This paper describes the architecture as a whole as well as preliminary results testing aspects of the system. Our intention is to implement and test the system on ground robots and small UAVs and demonstrate that the system can provide selective low-level masking or deletion of data requiring higher privacy protection.

Introduction

Networks of interconnected cameras currently provide constant surveillance over many metropolitan cities. In the near future, vision-based drones, robots and wearable cameras may expand this surveillance to rural locations and one's own home, places of worship, and even locations where privacy is considered sacrosanct, such as bathrooms and changing rooms. As the applications of robots and wearable cameras expand into our homes and begin to capture and record all aspects of daily living, we begin to approach a world in which all even bystanders are being constantly observed by various cameras wherever they go.

Recent examples of the vast expansion of surveillance capability are disturbing for those concerned with privacy. In 2016 an Ohio judge ruled that data collected by a man's pacemaker could be used as evidence that he committed arson (Moon, 2017). Previously, data from Fitbits and other wearable devices have been used as evidence (Watts, 2017). In another recent case, data collected by an Amazon Alexa device was used as evidence (Sauer, 2017). Hundreds of connected home devices, including appliances and televisions, now regularly collect data that can be used either as evidence or simply as a monitoring device by hackers or whomever can access the data. Comparatively little technical effort is being invested in ensuring personal privacy.

There are several reasons why protecting privacy is desirable. Device makers may be concerned with the theft of data from their devices (and the resulting lawsuits) but nevertheless view the risk of such theft as acceptable when comparing that risk with the limitations imposed by ensuring privacy. Encrypting data is one method commonly employed to protect privacy and security. Yet encryption brings its own set of limitations in that export restrictions are imposed on encryption technology, encryption is computational expensive process that can slow processors or processing, and the amount of data generated by devices such as wearable cameras makes encryption unfeasible. Another concern is that data is unprotected prior to encryption or after unencryption. Moreover, for staunch privacy advocates, a better solution is for the data to simply never have been collected in the first place.

The purpose of the paper is to present a computational architecture that can be used by artificial intelligence product developers and researchers which will alleviate privacy concerns. Our work focuses on privacy with respect to streaming video data. As mentioned, camera surveillance is rapidly expanding. Moreover, video can be used to for a variety of exceedingly intrusive purposes, such as detecting and characterizing a person's emotions, or constructing a visual network of their social contacts. Finally, as roboticists we are concerned that the large scale adoption of consumer robots might further restrict public privacy. The architecture presented is preliminary in that we present only working pieces and have yet to test a fully functional architecture at this time. This paper focuses more on the technical aspects of a privacy ensuring computational process than on the societal implications.

The section that follows summarizes the relatively few computational approaches to ensuring privacy. Next we

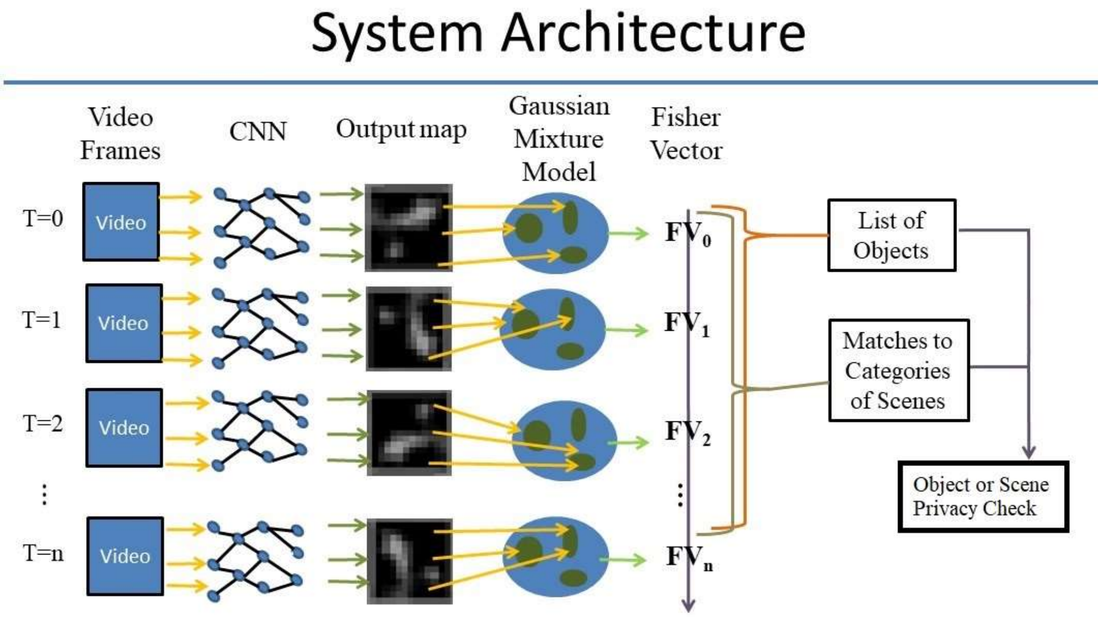

Figure 1. The figure depicts a privacy protecting architecture. The system takes streaming video on the left of the diagram, converts this input into lists of objects using a Convolutional Neural Network, and Gaussian Mixture Model. The output from this conversion encodes a list of objects in the video image. These objects are cross-checked with a list of objects requiring greater privacy. Collections of frame outputs are clustered into scene segments (not shown). These scene segments are then cross-checked against scenes requiring greater privacy.

present our architecture for ensuring privacy. We present preliminary results for some of the architecture's pieces and conclude with a discussion examining how the architecture might be incorporated by product developers and researchers.

An Architecture for Ensuring Privacy

There are some people, places, and things that a robot, wearable camera, or other device should not record. As discussed above, there are many different types of information that a device could record in violation of one's privacy. The work presented here will focus on camera data, although we believe that a similar system could likely also be developed for audio data. From a privacy perspective, camera data is important because it if allows one to monitor and track a person in an environment, identify relationships, capture a person's behaviors, monitor their emotions, and possibly even recreate conversations.

Our architecture focuses on three types of privacy: privacy for certain types of persons, places, and things. Privacy for certain types of individuals recognizes that particular categories of people, such as children, or for all people in certain roles, such as patients in a hospital, may need additional privacy guarantees above and beyond the general public. The architecture also recognizes that certain, particular places may require greater privacy for all people than other places. For instance, a wearable camera currently records images regardless of whether the person is in a public park or a public bathroom even though the privacy expectations in these two locations differ. Finally, the proposed system can vary the level of privacy in relation to objects identified in the environment. For example, amorous or provocative clothing, medicines in a bathroom, or private letters and banking information may all signal increased expectations of privacy. Figure 1 presents the overall system architecture. Overall, the proposed system is meant to ensure privacy rights for those people encountering robot based or wearable cameras either in public or in private.

In previous work we developed a computational process that allows one to create high-level representations from first-person video and to use these representations to evaluate the similarity of different scenes (Doshi *et al*., 2015). The process begins when a Convolutional Neural Network (CNN) is used to convert individual video frames into a set of 256 output maps (Figure 1). CNNs are a class of deep learning architectures often used for object recognition that are capable of recognizing objects within images

Previously we have used Caffe, an open-source framework for deep learning. Caffe includes several neural networks which have been pre-trained on the ImageNet dataset consisting of 1.2 million images capturing 1000 different categories of objects. AlexNet can recognize thousands of

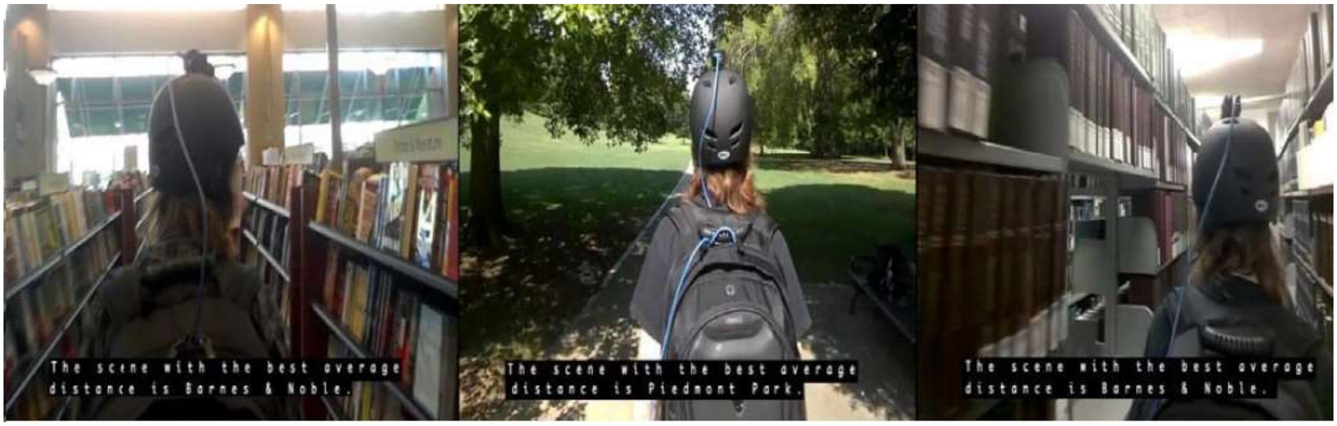
The scene with the best average distance is Barnes & Noble.
The scene with the best average distance is Piedmont Park.
The scene with the best average distance is Barnes & Noble.

Figure 2 A helmet-worn prototype is used to match the scene that the student is currently encountering to a set of stored scenes. For this prototype the system then stated the scene's label to the user. For the proposed system, the scene will be matched to predefined privacy settings.

different objects with a top-5 accuracy of over 90% and consists of a five convolutional layer architecture with three fully connected layers. Our research uses the output generated by the fifth convolutional layer as a higher-level representation roughly capturing object-level information and their spatial patterns in a scene.

Higher convolution layers are able to pick out object parts and then entire objects. Therefore, the output from the fifth convolutional layer of the network capture the identity, strength, and spatial distribution of object-level information throughout the image. In previous experiments, we have demonstrate that we can leverage this information to categorize the location of the video stream (the type of scene) or identify visible objects in the environment (Kira et al. 2016).

We then use Improved Fisher Vectors in conjunction with a Gaussian Mixture Model (GMM) pre-trained on a subset of the video data to produce a fixed length encoding which summarizes the strength of association between the set of output maps across multiple frames to the different modes in the GMM (Perronnin and Dance, 2007). Because the representation is based on object-level features and their positions in the images sampled from the environment, we have shown that our approach is robust enough to noise and blur that it can be used on video captured by a wearable camera.

We have shown empirically, both using preprocessed data and using live prototypes, that this process allows us to cluster visual scenes with respect to higher, more abstract concepts such as library, restaurant, or kitchen. The distance metric $D(S_i, S_j)$ can be used to match one's current visual scene to an existing scene category, previously encountered environments, or to generate a new scene category.

Privacy for Places

Our previous and ongoing work has demonstrated that the system can be used to recognize different categories of places. Using Google glasses we collected a large dataset from wide variety of locations in the Atlanta area. This information was then processed using the architecture in Figure 1. Later we tested the system using a different head mounted camera (Figure 2) at several locations. Some of the test locations were places where the data had been generated for the dataset. Most were not. Regardless of whether or not a location existed in the dataset, the system was able to generate a measure of similarity to previously encountered locations. We found that when this similarity metric was used to cluster video recordings unique categories of places were generated often related by purpose (Figure 3). For example, the system matches the library on the right of figure 2 to the Barnes and Nobles on the left. But when the person arrives at the coffee shop located in the middle of the library, the system categorizes the streaming video as a coffee shop. If the system's data captured within coffee shops is removed, the system recognizes the coffee shop as a type of restaurant (Figure 3). When streaming video of a picnic, the system notes that streaming video most closely matches a park and a restaurant. Our previous research shows that the computational framework can be used to categorize the incoming video stream in terms of places.

Once a scene is categorized, the category label can be cross-referenced with a list of high privacy areas. If the video is streaming from a high privacy location, then the entire video stream can be blurred. For example, if a drone is flown over a beach, video from the camera can be used with our system to identify the location as a beach and recognize that it is a higher privacy location. Privacy rules such as blurring the video feed, or blurring the faces and bodies of those in the video can then be enacted.

Privacy for Things

The presence of certain objects signals the need to greater privacy. For instance, medications, private letters and bills,

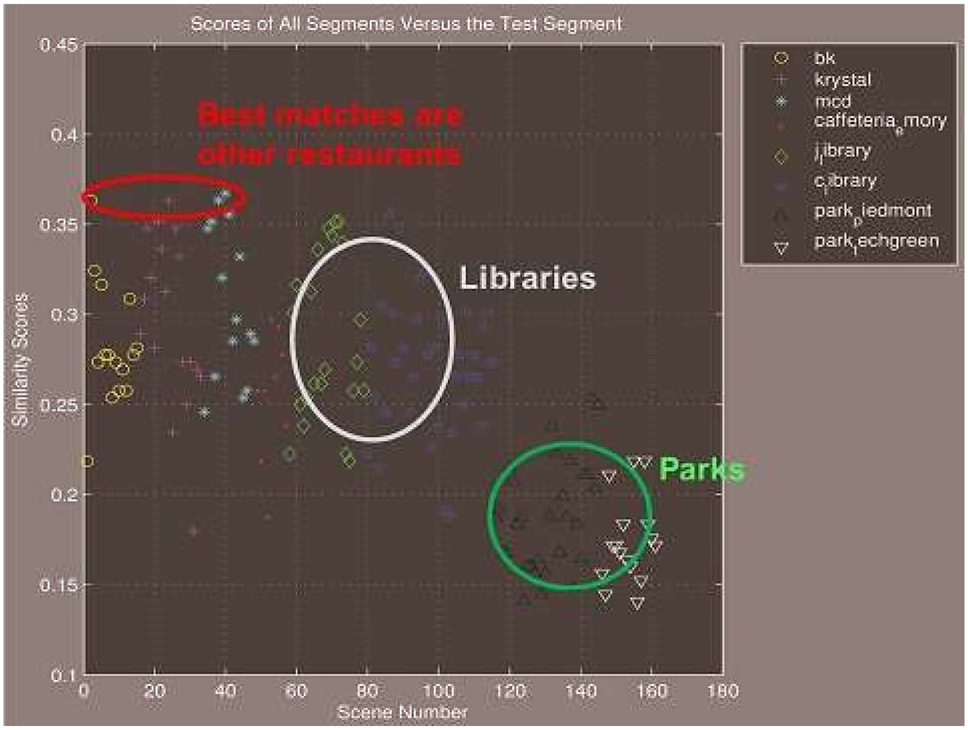

Figure 3. The graph above compares similarity scores for video segments from 8 scenes to segments taken from a McDonalds restaurant scene. The data depicts a spread across scenes with some segments matching multiple scenes. Nevertheless, the best match for each category of segments is a member of the same class. As indicated by the circle in red, the best matches to the McDonald's target are video segments from the other restaurant scenes. Some video segments taken from library scenes also match well to restaurant target. Finally, parks do not match well to restaurants. The individual data points correspond to different video segments.

objects related to intimacy, all signify the need for privacy. For a mobile camera on a robot or a wearable camera, encounters with these types of objects should signal the need for increased privacy.

Convolutional Neural Networks have the highest object recognition accuracy rates among computer vision classifiers (Razavian, 2014). Typically trained on large datasets, such as ImageNet containing millions of images of everyday objects, these classifiers recognize thousands objects.

Our proposed system leverages the incredible ability of Convolutional Neural Networks to recognize objects to identify a limited set of predefined, privacy signaling objects. Once such an object or several such objects are identified, a predefined privacy rule is enacted. This predefined rule may blur the object or the entire video stream depending on the circumstance. A full implementation of this idea may need to be trained on some or several high privacy items. Although ImageNet does contain images of medicines, a list of privacy inducing objects may need to be created, a dataset created, and neural network training performed.

Privacy for Persons

A person's age plays an important role with respect to privacy. We recently conducted online surveys asking adults about their privacy expectations with regard to drone flybys. Subjects strongly indicated that their privacy expectations were higher when they were in the presence of children. The results from these surveys are still being compiled but it is safe to say that age and gender impact one's expectations related to privacy.

Convolutional Neural Networks have been developed that can estimate a person's age or gender (Levi and Hassner, 2015). Recently published results by Levi and Hassner indicate an accuracy of 86.8 percent for gender classification and 50.7 percent for classifying an individual's age into one of eight categories. They note a 1-category off classification accuracy of 84.7 percent. Figure 4 depicts an example of off-the-shelf age and gender classification using convolutional neural networks.

We envision augmenting our architecture with an age pathway that identifies and categorizes faces in terms of age. This pathway could then use statically set privacy rules to blur out the faces of all individuals that are within specific recognition category. Alternatively, age and context could be used to selectively blur the faces of individuals within some age category and at a specific type of place or scene. Our recent surveys on privacy demonstrated that respondents were most concerned about privacy in relation to drone flyovers while they are at the beach with their children.

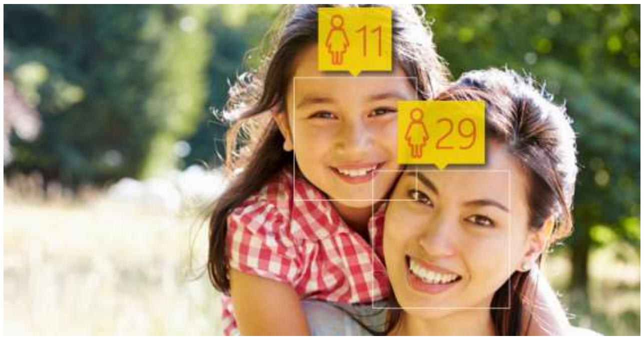

Figure 4 Convolutional Neural Networks have been trained to classify a person's gender and age. These classifiers can be used to recognize situations and people that may need greater system privacy protections.

Development and Limitations

We are currently in the process of developing and evaluating the system. As part of the development process several issue will need to be addressed.

One issue that arises is the need of the robot or camera wearer versus the need for privacy of the individual being observed. For a robot attempting to localize or using camera information to identify objects or people, blurring incoming data may limit the robot's usefulness or perhaps even endanger the people around it. One possible solution is to apply the privacy protection architecture after the robot has already used the video stream to situation awareness. Although this solves the immediate problem of interfering with the robot's perception it may compromise privacy and data protection, or at a minimum, open the door for hackers to intercept data before it reaches the privacy protection system.

For users of wearable cameras such as police officers or those with vision impairments, the tension between privacy and functionality is even more serious. Blurring out entire scenes or environments would quickly limit the usefulness of a wearable camera. Even blurring faces may generate problems for those that rely on the camera for social interaction.

In recognition of the tradeoff between privacy and usability, we are developing an interface that allows a person (perhaps the owner of a product or a user) to vary the privacy settings of the technology, perhaps contextualizing one's privacy concerns or, at least, having control over when and what the technology records and transmits. For example, for appliances an owner may want to set the maximal privacy settings ensuring that data is not stored or transmitted and that one's face is always blurred or removed. The user would also have control over the types of objects that are blurred when seen by the appliances camera. For a laptop camera, the user may want to alter the privacy settings on a daily or as needed basis. Even if most people choose simply to let the factory settings stand at least the user is given a choice as to whether and to what degree they prefer to protect their privacy.

Privacy Star

Ideally, product developers will create, test, implement, and adopt system architectures that guarantee privacy. The adoption of these types of architectures might be signaled by using an Energy Star like labeling program. Energy Star began as a voluntary labeling program designed to promote energy efficient products. Like Energy Star, Privacy Star would serve to identify and promote consumer products that protect user data and personal privacy. A certification process could be developed ensuring that a product meets predetermined privacy standards and the users have access and control over the privacy settings of their devices. Privacy Star does not necessarily need to be a government run entity. Like Energy Star it might begin as a grassroots effort.

Conclusion

This paper has presented a preliminary architecture for protecting user privacy. We have focused on ensuring privacy with respect to streaming video from a camera. Clearly there are many other aspects to protecting a user's data. We have specifically chosen to focus on streaming video for a number of important reasons. First, camera-based surveillance is expanding rapidly and, as noted in the introduction, is encroaching upon the privacy of our homes and perhaps even our bodies. Second, the possibility of large scale adoption and use of consumer robots will further infringe upon our right to privacy. Finally, unlike most other types of data collection, camera-based video data is broad in the amount, types, and uses of the information that is collected. Video can be used to gauge and classify a person's emotions, reconstruct what they said, or connect them to social networks to deduce their beliefs and principles. While the architecture we present is not without limitations, we feel that it is an important first step towards using artificial intelligence to protect a person's right to privacy.

References

Boyle, M., Edwards, C., Greenberg, S. 2000. The effects of filtered video on awareness and privacy. In: ACM Conference on Computer-supported cooperative work. pp. 1–10

Boyle, M., Neustaedter, C., Greenberg, S.: Privacy factors in video-based media spaces. In: Media Space 20 + Years of Mediated Life, pp. 97–122. CSCW (2009)

Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G. J., Durand, F., & Freeman, W. T. (2014). The visual microphone: Passive recovery of sound from video" ACM Transactions on Graphics (Proc. SIGGRAPH), 33(4), pages 79:1-79:10.

Doshi, J., Kira, Z., & Wagner, A. (2015). From deep learning to episodic memories: Creating categories of visual experiences. In *Proceedings of the Third Annual Conference on Advances in Cognitive Systems ACS* (p. 15).

Edgcomb, A., & Vahid, F. (2012). Privacy perception and fall detection accuracy for in-home video assistive monitoring with privacy enhancements. *ACM SIGHIT Record*, *2*(2), 6-15.

Hubers, A., Andrulis, E., Smart, William D., Scott, L., Stirrat, T., Tran, D., Zhang, R., Sowell, R., and Grimm, C. 2015. Video Manipulation Techniques for the Protection of Privacy in Remote Presence Systems. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (New York, NY, USA,59-60.

Kaminski, M. E., Rueben, M., Smart, W. D., & Grimm, C. M. (2016). Averting Robot Eyes. *Md. L. Rev.*, *76*, 983.

Kira, Z., Li, W., Allen, R., & Wagner, A. R. (2016) Leveraging Deep Learning for Spatio-Temporal Understanding of Everyday Environments, IJCAI Workshop on Deep Learning and Artificial Intelligence.

Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 34-42).

Moon, M., (2017, June 13). Judge allows pacemaker data to be used in arson trial. *Engadget.* Retrieved from https://www.engadget.com/2017/07/13/pacemaker-arson-trial-evidence/

Perronnin and Dance, 2007] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. 2007.

Razavian ,A.S., H. Azizpour , J. Sullivan , S. Carlsson. (2014) CNN features off-the-shelf: An astounding baseline for recogniton, CVPR 2014, DeepVision Workshop.

Sauer, Gerald (2017, February 28). A Murder Case test's Alexa's Devotion to your Privacy. *Wired.* Retrieved from https://www.wired.com/2017/02/murder-case-tests-alexas-devotion-privacy/

Watts, Amanda (2017, April 27). Cops use murdered woman's Fitbit to charge her husband. *CNN* Retrieved from http://www.cnn.com/2017/04/25/us/fitbit-womans-death-investigation-trnd/index.html