Negotiating Privacy Preferences in Video Surveillance Systems

Anonymous AVSS submission for Double Blind Review

Paper ID 55

Abstract

We propose a novel privacy aware video surveillance system. The proposed system encodes privacy preferences using P3P-APPEL framework that was first proposed for managing data privacy on the web. To this end, we have proposed extensions to P3P-APPEL to make it suitable for video surveillance applications. A noteworthy feature of the proposed system is its ability to interact with individuals present in the scene. Users with appropriate security credentials have access to one of three privacy settings: L0 (no privacy), L1 (face blur), and L2 (full body blur). User can thus choose the level of privacy (or surveillance) they are comfortable with. This is an extremely desirable capability that shifts the relationship between those who are observed and those who operate video surveillance systems.

1. Introduction

Recent advances in image processing, camera hardware, and wireless communication have led to an increase in the rate at which new video surveillance systems are being installed in private and public places, such as airports, train stations, banks, restaurants, schools, hospital, homes, etc. The law enforcement community is increasingly relying on video surveillance for crime prevention and community safety. Video footage captured through surveillance cameras is routinely used to identify suspects and as evidence in the courts. Video data collected by cameras present in the scene has many applications outside of the surveillance domain. Most notable among these are smart environments where video data can be used to identify occupants and analyze activities taking place in these environments. The ability to "see" enables smart environments to respond to the needs of its occupants.

Aforementioned benefits notwithstanding, pervasive video surveillance raises serious privacy concerns. Experts agree that video surveillance undermines our "right to anonymity." Video surveillance augmented with biometric technology (e.g., face recognition) raises even more privacy concerns. Who is collecting information about us? How this information is being used? What information is being collected? Who has access to this information? What is the retention policy for the collected information? These privacy concerns have discouraged the adoption of video surveillance technologies in non-security related domains, such as smart homes, hospitals, assisted living facilities for the elderly, smart meeting rooms, etc. A timely challenge for computer vision researchers is to develop video surveillance systems that include the security features of traditional systems and that have built-in *privacy protection* capabilities.

This paper presents a Privacy Aware Surveillance System (PASS) that is capable of enforcing user-specific privacy policies. PASS differs from existing privacy-aware video surveillance systems in an important way—it supports user interaction 1) to collect privacy preferences and 2) to get user consent to video data collection. Users can use gestures to indicate their desired privacy settings to the system. PASS employs visual analysis to locate individuals present in the scene and hide their identities by applying one or more privacy preserving filters. For example, PASS is able to locate and blur the faces of individuals present in the scene. It is also able to obscure the entire person; thereby, hiding their gender, age, height, size, ethnicity, etc. While a number of video surveillance systems implement privacy protection strategies, such as blurring individuals to preserve their anonymity, PASS is unique in its ability to interact with a user to collect his privacy preference and to get his consent about video data collection. Fig. 1 illustrates how PASS can respond to user requests about privacy levels and how PASS is able to ask for a person's consent about video data collection.[1]

The following scenario illustrates our vision of how PASS might operate in a real setting. Persons A and B are meeting in a room that is under video surveillance. Both of them decide that they do not want to be recorded, so they use gestures to tell the surveillance system to stop recording. The video system responds to their request for privacy and alter the video stream to obscure their identities.[2]

[1] We envision that user consent negotiations takes place at an access station (privacy decision point).

[2] It is possible to still recover the raw footage with proper authorization.

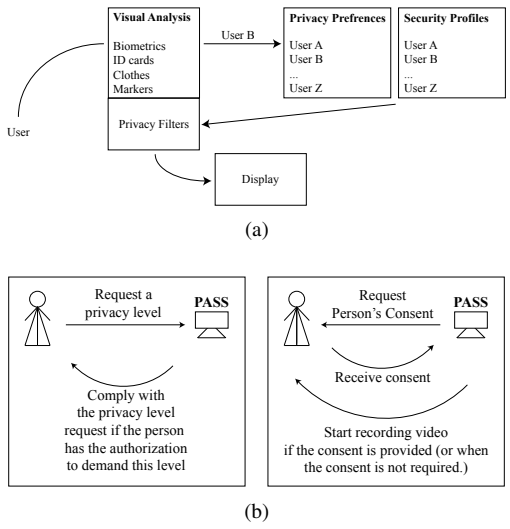


Figure 1: (a) PASS relies upon computer vision technology to match an individual against its privacy preferences and security profile stored in the system. A user with proper credentials is able to request one of the available privacy settings. (b) Left: PASS is able respond to a user's need for privacy. Right: PASS is also able to acquire an individuals consent about video data collection.

Shortly thereafter person C enters the room. The three decide to record the conversation and instruct PASS to discontinue privacy protection filters. Since C is not an employ of this company, PASS requires C's consent to record this meeting. C provides this consent at a designated access point. Depending upon their respective roles and security credentials, individuals have access to different levels of privacy protection. For example, person A may only invoke level 1 privacy that includes face blurring; while person B can invoke both level 1 and 2 privacy, which includes both face and body blurring.

The Office of the Privacy Commissioner of Canada has issued a number of guidelines regarding the installation and use of video surveillance systems operated by law enforcement agencies, private organizations, and owners of private properties [1]. For example, the public must be informed via clearly marked signage that they are under video surveillance. Currently, these guidelines do not mention whether user consent is necessary to collect video data despite the fact that it is mentioned under Personal Information Protection and Electronics Document Act (PIPEDA). To the best of our knowledge, none of the existing camera systems acquire user consent for video surveillance. PASS is a first step in that direction.

The rest of the paper is organized as follows. Sec. 2 presents an overview of the work done by previous researchers in the area of privacy in video surveillance systems. In Sec. 3, we explain the computer vision algorithms needed to realize PASS. Sec. 4 discusses how PASS manages privacy preferences of various users. Results are presented in Sec. 5, and we conclude the paper with conclusions and future directions in Sec. 6.

2. Relevant Literature

Cavallaro stresses the need to tackle privacy issues due to the widespread use of video surveillance [2]. The American Civil Liberties Union (ACLU) has outlined a number of concerns about video surveillance. ACLU maintains that data collected through Closed-Circuit Television (CCTV) cameras have routinely been misused by those with access to the data. In the United Kingdom the video privacy falls under the Data Protection Act of 1998 [3]. Most countries follow the Organization of Economic Cooperation and Development (OECD) guidelines for privacy protection of personal data, including video data. There is obviously a need to develop video surveillance systems that implement legal controls need to preserve the privacy of individuals, while at the same time allowing video surveillance to be an effective tool against criminal activity.

Computer vision technologies can be used to develop camera networks that can uphold privacy policies and regulations [4, 5]. Pedestrian detection and tracking routines can identify individuals present in the scene and obscure them to hide their identities. The operator can still see the scene and know how many people are present in the scene without knowing the identities of those people [6]. An activity recognition technique can reveal an individual if it detects an anomalous behavior.

Schiff *et al*. develop a video surveillance system capable of obscuring the faces of individuals present in the scene [7]. Individuals who do not want to be identified wear a visual marker, which allows the video surveillance system to locate the face of the individual and obscure it with an ellipse, while allowing observation of his or her actions in full detail. This allows the operator to observe the activities taking place in the scene without knowing the identities of the people present. Newton *et al*. propose de-identification of facial features to preserve the privacy of individuals present in the scene [8].

Sony patented a privacy mode for camcorders that replaces the skin color of individuals so as to avoid race-based discrimination [9]. [10] patented a system capable of obscuring a privacy region in a pan-tilt-zoom camera. [11] develops a system that is able to locate and obscure people in a video, thereby preventing statistical inferences from the video. Chattopadhyay and Boult developed a privacy preserving smart camera, called *PrivacyCam* [5]. PrivacyCam uses on-board digital signal processor to locate and encrypt human faces in the image. The original image can be recovered given the correct decryption key.

Senior *et al*. study the problems associated with human monitoring of video surveillance feeds and stress the need for systems that can carry out observation tasks au-

	Room A: All individuals	Room A: Individuals wearing Blue clothes	Room A: Individuals wearing Red clothes	Corridor: All individuals	Corridor: Individuals wearing Blue clothes	Corridor: Individuals wearing Red clothes	Room B: All individuals	Room B: Individuals wearing Blue clothes	Room B: Individuals wearing Red clothes
L0 (no privacy)	X	X	X	X	X	X	X	X	X
L1 (head blurring)		X	X	X	X	X			
L2 (body blurring)			X		X	X			

Figure 2: PASS supports user-specific and location-aware privacy policies. In Room A, individuals wearing blue clothes has access to L1 privacy and those in red clothes has access to L2 privacy. Other individuals has no access to the privacy features present in PASS. In the corridor, however, individuals wearing either red or blue has access to L2 privacy. Other users have access to L1 privacy. Similarly, no one is authorized to access the privacy features in Room B.

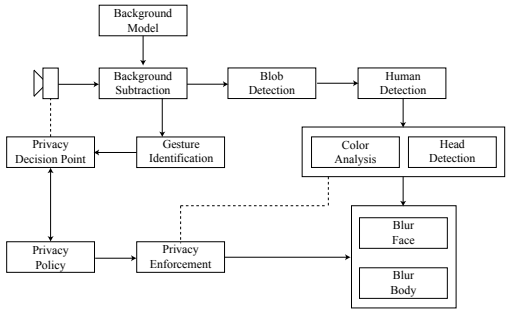
tonomously [12]. Such systems, they maintain, reduce the breach of privacy associated with video surveillance systems monitored by human monitors. The propose a privacy aware video surveillance system that uses encryption and access control lists to restrict the access to raw video footage. Tansuriyavong and Hanaki employ face recognition technologies to find candidate individuals in a video and replace their faces with their IDs [13].

Zhang *et al*. propose a secure watermarking technique to generate privacy protected video from the raw video footage [14]. Hudson and Smith study the question of privacy vs. situational awareness in video surveillance systems [15]. Qureshi proposes decomposing raw video into object video streams, which can latter be recombined to render privacy enabled video feed [16].

3. System Overview

PASS employs video analysis to locate and identify individuals present in the scene (Fig. 3). Individuals interact with PASS and specify their preferences via gestures. As shown in Fig. 2, pass supports both user-specific and location-specific privacy levels, i.e., different users may have access to different privacy levels based upon their credentials and their current location. User are mapped to their privacy preferences and credentials by leveraging computer vision technologies (Fig. 1). Currently, PASS implements three privacy levels: L0 (no privacy), L1 (head blurring), and L2 (body blurring). In the current implementation PASS maps an individual to his privacy preferences and credentials through color analysis. Individuals wearing red shirts have access to the highest level of privacy; while those in blue clothes can only access L1 privacy. Other individuals are deemed unknown. We imagine that in a real implementation, PASS would rely upon more sophisticated biometric analysis to map individuals to their privacy settings and security credentials.

The performance of PASS is ultimately tied to the capabilities of the vision pipeline that is responsible for analyzing raw video data. PASS vision module is responsible for both tracking individuals present in the scene and recognizing hand gestures to support user interactions. We now describe these two aspects of the vision module.



(a)

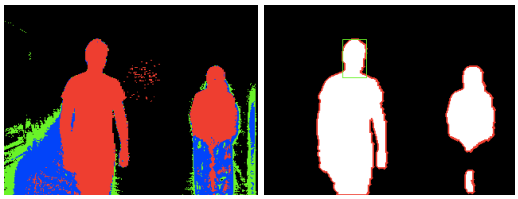
Figure 3: PASS relies upon well-understood computer vision technologies, such as pedestrian identification and tracking, to implement user-specific and location-aware privacy policies. Users can indicate their privacy preferences to PASS via gestures.

3.1. Background Learning

During an initial training phase, when no pedestrians are visible, each camera learns a background model of the scene. We model the chromaticity of each pixel in RGB space. Specifically, the mean and variance of brightness and chromaticity distortion is stored for each pixel. This information is then used to classify each pixel as either belonging to the background or foreground (Fig. 4(a)). Background subtraction step involves comparing the current frame against the learnt background model and constructing a (in general, noisy) foreground mask. In a real system, we would also need a mechanism to update the background model to account for changes in the background. It is straightforward to incorporate this capability into our background model. We employ automatic thresholding to classify each pixel as foreground or background. We refer the reader to [17] for more details.

3.2. Blob Detection and Pedestrian Tracking

The foreground mask obtained through background subtraction is cleaned up through connected component analysis and blobs representing foreground objects are extracted



(a) (b)

Figure 4: (a) Foreground pixels are identified via background subtraction. Each pixel is classified as either belonging to the background (black) or the foreground (red, blue, or green). Foreground pixels are further classified into normal pixels (red) or shadows (blue)/highlights (green).(b) Blob silhouette shape analysis identifies the head location. The green rectangle indicates the head location.

(Fig. 4(b)). In our case, each blob represents one or more pedestrians. Pedestrian signatures encode pedestrian color distribution as 2D histograms in Hue-Saturation (HS) colorspace. We have empirically selected 32 bins along the two (Hue-Saturation) dimensions. Tracking is performed by setting up a bipartite graph matching problem as suggested in [18]. The optimal solution to the matching problem resolves pedestrian identities across multiple frames. We refer the reader to [18] for more details. Pedestrian tracker assigns each blob to one or more pedestrians.

The tracker maintains a list of pedestrians that are currently being tracked. In each frame, each pedestrian is either matched to a blob (using pedestrian signature matching) or to the background. The tracker is robust to short-duration occlusions.

3.3. Color Analysis

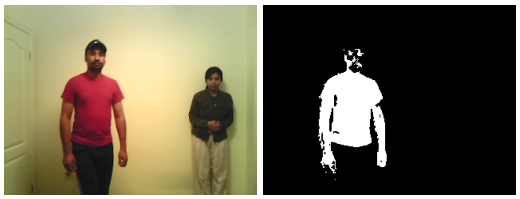
Currently the security credentials of each pedestrian is determined by the color of his clothes. We expect that biometric analysis will be used to identify a pedestrian in a more realistic setting. We follow the steps below to determine if an individual is wearing a red shirt:

1. Apply median filters to RGB channels
2. Compute $M = R - (G + B)$, where R, G, and B represent the red, green, and blue channels, respectively
3. Threshold M; white pixels represent red pixels
4. Compute the ratio of red pixels to the total number of pixels in a blob

Similarly, we can determine if a person is wearing a blue shirt.

3.4. Head Localization

Our strategy for head localization is premised upon two observations: 1) it is highly likely that the head is adjacent



(a) (b)

Figure 5: Blob color analysis determines if a person is wearing a red or a blue shirt. In our current implementation security credentials of a person is determined by the color of his shirt. (a) raw frame. (b) red shirt classifier has detected that the individual is wearing a red shirt.

(a) (b) (c) (d)

Figure 6: PASS supports low-effort user interaction via gesture recognition. (a) An open hand gesture grabs tells PASS that the user is about the indicate his desired privacy setting to PASS. (b) A fist gesture instructs PASS to terminate privacy protection filters. (c) An index finger gesture requests L1 privacy. (d) A V gesture requests a L2 privacy.

to the silhouette boundary and 2) the relative location of the head is constrained by the posture. Furthermore, we observe that in an upright position, head silhouette is roughly Ω-shaped. Head localization is then performed via silhouette location and shape analysis for each blob (Green rectangle in Fig. 4(b)).

3.5. Gesture Recognition

An individual interacts with PASS via gesture recognition. We have implemented a skin classifier that finds pixels belonging to a hand in an image. An alternate would be to use background subtraction to identify pixels belonging to hand. We found that background subtraction did not perform as well as we had hoped. We compute the contour for the blob representing the hand of an individual. Peaks along the contour represent fingers. We refer the reader to [19] for further details about hand detection and gesture recognition. Currently, the system can recognize the gestures shown in Fig. 6.

4. Privacy Preferences in PASS

PASS' privacy framework is inspired by the Privacy Preference Project (P3P)—a widely adopted framework for data privacy management in the World Wide Web. The pri-

AVSS #55

(a) (b)

Figure 7: Pass is able to blur the entire body (a) or just he face (b) of the selected individual. The level of blurring is user controlled. It is also straightforward to perform in-place encryption instead of the näive blurring demonstrated above.

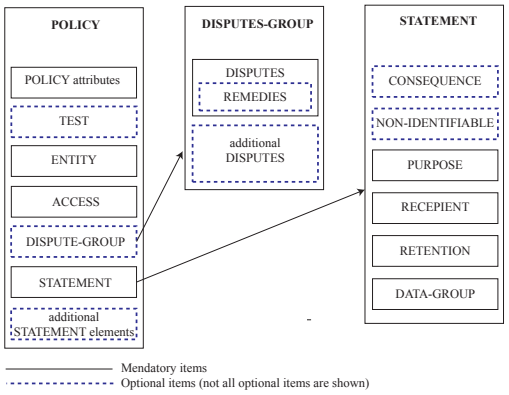


Figure 8: Structure of a P3P XML policy file

mary goal of P3P is to enable websites express their privacy policy in a standard, machine-readable XML format that the software agents such as web browsers can retrieved automatically and interpret easily. A P3P policy file would typically contain seven segments as shown in Fig. 8:

- Policy attributes is a mandatory segment that contains information about policy location(s).
- Test is an optional segment that contains information about the status of the policy (test or active).
- Entity is a mandatory segment that contains information about the legal entity responsible for the privacy policy.
- Access is a mandatory segment that encodes user access to their own data.
- Dispute-Group is an optional segment that handles disputes between conflicting policy requirements.
- Statement is a mandatory segment describing privacy practices related to data storage, retrieval, etc. Statement is further subdivided into the following segments:
  - Consequence: an optional segment containing human readable information about the policy.
  - Non-Identifiable: an optional segment that informs the agent that no data is being collected.
  - Purpose: a mandatory segment providing a reason for collecting the data.
  - Recipient: a mandatory segment containing information about the recipient of the data.
  - Retention: a mandatory segment outlining the retention policy.
  - Data-Group: a mandatory segment describing the data collected by the web site.

A P3P Preference Exchange Language (APPEL) can be used to define a list of privacy preference rules. These rules can then be used by user agents, e.g., web browsers, to make decisions about the privacy policies of P3P enabled websites. APPEL rules are matched against a P3P policy in the order in which they appear. Each rule consists of the two parts: 1) behavior/output states the action to be taken if the rule evaluates to be true and 2) body/input specifies a pattern that needs to be matched against the privacy policy. Currently, three types of actions are supported. A *request* action is taken if the website's policy conforms to user's preferences. A *block* action is taken when the website's privacy policy does not match with user's preferences. Lastly, a *limited* action is taken place when the privacy policy of the website is only partially acceptable to the user.

4.1. Communicating Privacy Preference in PASS

A noteworthy feature of PASS is the protocol that communicates the system privacy policy and the user privacy preferences between the surveillance system and the monitored user. While P3P and APPEL has been widely adopted over the Internet, we find them lacking for our purposes. Specifically, the existing P3P-APPEL model does not support negotiation and prescribe "take-it-or-leave-it" interactions between a web server and its client. Equally importantly P3P-APPEL model only includes a description about the entity collecting the personal information. It does not describe the entity who owns the collected information, thus prohibiting a binding contract about data usage between the individual and the data gathering entity.

We propose the following three extensions to P3P-APPEL model in order to render it suitable for use within video surveillance systems:

- The P3P-APPEL model must support the visual description of the monitored individual. E.g., a biometric face or gait signature or some other visual characteristic that can be extracted from video data. As mentioned earlier, presently we identify a person by the color of his clothes. In our toy universe individuals belong to three categories: 1) those with red shirts, 2) those wearing blue shirts, and 3) all others.

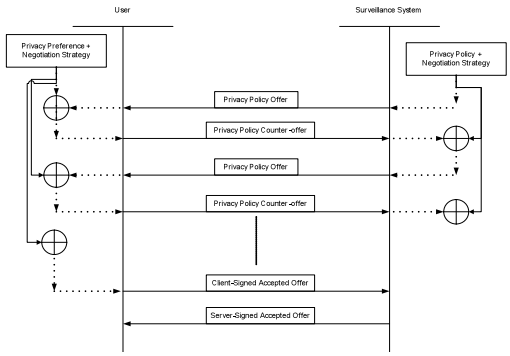


Figure 9: Proposed extension to P3P-APPEL allows privacy policy negotiation between an individual and the video surveillance system operator.

- Privacy policy must store the signatures of both the individual under surveillance and the entity operating the video surveillance system. The presence of these signatures mark the commitment of both parties to privacy policy. Currently, this feature is only implemented for known individuals, i.e., those wearing either red or blue shirts.

- P3P-APPEL model should also incorporate a negotiation strategy that allows an individual under surveillance and the entity that operates the video surveillance system to arrive at a mutually agreeable privacy policy. This feature is currently not implemented. Fig. 9 suggests a scheme for implementing such a protocol. We imagine that an access station that supports two-way communication between the individual and the visual surveillance system will be used for this purpose (see privacy decision point in Fig. 3).

5. Results

Our preliminary results appear promising. Fig. 10 shows a room with an individual wearing a red shirt. As seen in Fig. 10(a), PASS is providing L1 privacy to the individual. The person performs a palm gesture (Fig. 10(b)) and requests PASS to discontinue privacy protection (Fig. 10(c)). PASS reduces the privacy level to L0. Fig. 10(d) shows the individual again performing a palm gesture. Next, person requests privacy level L1. PASS complies and hides persons face (Fig. 10(e)). After some time the individual again grabs PASS attention and instruct it to increase the privacy level to L2 (Fig. 10(f-g)). PASS complies and obscures the entire person.

Our gesture recognition system work fairly well. This is in part due the fact that we are employing very simple gesture vocabulary. Each gesture is easily differentiated. A limitation of the current approach is that the individual must face the camera when making the gesture, since our system can only recognize gestures when viewed from the front. Also, the hand must be kept away from the body when making a gesture. We use skin colors to segment the hand, so the gesture recognition system gets confused when the hand is close to face. Gesture recognition performs poorly when the person is far from the camera. Similarly, gesture recognition assumes that the entire individual is visible in the video.

6. Conclusions and Future Work

We propose a novel privacy aware surveillance system that supports user-specific privacy policies. The system encode privacy within the P3P-APPEL model that was proposed for managing data privacy on the Internet. We have proposed extensions to P3P-APPEL model to render it suitable for describing privacy preferences and assist negotiations about these preference within the context of video surveillance.

A noteworthy feature of the proposed system is its ability to interact with individuals under observation. Users can describe their privacy preferences via gestures. To the best of our knowledge no other video surveillance system support such interactions. We believe that this is a crucial capability as we build more sophisticated video surveillance systems. This capability is especially useful for video surveillance systems for smart homes, elderly care, office buildings, etc. We imagine that this capability will change how we perceive video surveillance and how we deal with cameras that surround us.

We are currently working on incorporating biometric analysis routines—face detection, gait analysis, etc.—into PASS.

References

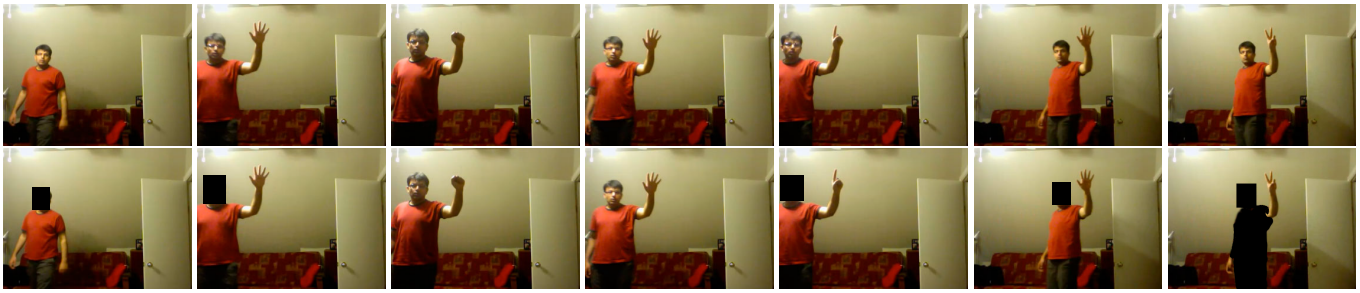
[1] "Office of The Privacy Commissioner of Canada." [Online]. Available: http://www.priv.gc.ca 2

[2] A. Cavallaro, "Privacy in Video Surveillance [In the Spotlight]," pp. 166–168, March 2007. 2

[3] "Privacy International." [Online]. Available: http://www.privacyinternational.org 2

[4] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, C. F. Shu, and M. Lu, "Enabling video privacy through computer vision," *IEEE Transactions on Security and Privacy*, vol. 3, no. 3, pp. 50–57, 2005. 2

[5] A. Chattopadhyay and T. E. Boult, "PrivacyCam: a Privacy Preserving Camera Using uCLinux on the Blackfin DSP," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*, Minneapolis, MN, June 2007, pp. 1–8. 2



(a) (b) (c) (d) (e) (f) (g)

Figure 10: Video sequence showing an individual interacting with PASS using gestures. Top row shows the raw video captured by the camera and the bottom row shows the privacy-enabled video.

[6] D. Chen, Y. Chang, R. Yan, and J. Yang, "Tools for Protecting the Privacy of Specific Individuals in Video," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 1–10, 2007. [Online]. Available: http://www.hindawi.com/journals/asp/2007/075427.abs.html 2

[7] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: Detecting visual markers in real-time to address privacy concerns," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, San Diego, CA, November 2007, pp. 971–978. 2

[8] E. M. Newton, L. Sweeney, and B. Malin, "Preserving Privacy by De-Identifying Face Images," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005. 2

[9] A. M. Berger, "Privacy Mode for Acquisition Cameras and Camcorders," US patent 6,067,399 to Sony Corp., Patent and Trademark Office, 2000. 2

[10] J. Wada, K. Wakiyama, H. Kogane, and N. Takada, "Monitor Camera System and Method of Displaying Pictures from Monitor Camera Thereof," European patent EP 1 081 955 A3 to Matsushita Electric Industrial, European Patent Office, 2001. 2

[11] J. Fan, H. Luo, M.-S. Hacid, and E. Bertino, "A novel approach for privacy-preserving video sharing," in *Proc. 14th ACM international conference on Information and knowledge management (CIKM05)*. New York, NY, USA: ACM, November 2005, pp. 609–616. 2

[12] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-l. Tian, and A. Ekin, "Blinkering Surveillance : Enabling Video Privacy Through Computer Vision," NY, pp. 1–9, 2003. 3

[13] S. Tansuriyavong and S.-i. Hanaki, "Privacy protection by concealing persons in circumstantial video image," in *Proceedings of the 2001 workshop on Percetive user interfaces (PUI01)*. Orlando, Florida: ACM Press, November 2001, pp. 1–4. [Online]. Available: http://portal.acm.org/citation.cfm?doid=971478.971519 3

[14] W. Zhang, S. Cheung, and M. Chen, "Hiding privacy information in video surveillance system," in *Proceedings of the IEEE International Conference on Image Processing (ICIP05)*. Genova: IEEE, September 2005, pp. 868–871. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1530530 3

[15] S. E. Hudson and I. Smith, "Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, Boston, November 1996, pp. 248–257. 3

[16] F. Z. Qureshi, "Object-Video Streams for Preserving Privacy in Video Surveillance," in *Proc. 6th International Conference on Advanced Video and Signal Based Surveillance (AVSS 09)*, Genovo, Italy, September 2009, pp. 1–8. 3

[17] T. Hoprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. 7th IEEE International Conference on Computer Vision, Frame Rate Workshop (ICCV99)*, Kerhyra, September 1999, pp. 1–19. 3

[18] H.-T. Chen, H.-H. Lin, and T.-L. Liu, "Multi-object tracking using dynamical graph matching," in *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR01)*, vol. 2, Hawaii, December 2001, pp. 210–217. 4

AVSS #55

[19] S. Malik, "Real-time Hand Tracking and Finger Tracking for Interaction," Toronto, pp. 1–21, 2003. 4