Adaptive Cartooning for Privacy Protection in Camera Networks

Ádám Erdélyi[1], Tibor Barát[2], Patrick Valet*[1], Thomas Winkler[1] and Bernhard Rinner[1]

[1]Institute of Networked and Embedded Systems and Lakeside Labs, Alpen-Adria-Universität
Lakeside Park B02b, 9020 Klagenfurt, Austria {firstname.lastname}@aau.at
[2]Institute of Informatics, University of Szeged
Árpád tér 2, 6720 Szeged, Hungary barat.tibor@stud.u-szeged.hu

Abstract

Visual privacy in video-based applications such as surveillance, assisted living or home monitoring is a highly active research topic. It is critical to protect the privacy of monitored people without severely limiting the utility of the system. We present a resource-aware cartooning privacy protection filter which converts raw images into abstracted frames where the privacy revealing details are removed. Cartooning can be applied either to entire images or pre-selected sensitive regions of interest. We provide an adaptation mechanism to our cartooning technique where the operator can easily change the filter intensity. The feasibility of this new approach is demonstrated by its deployment to real-world embedded smart cameras. We evaluate privacy protection and utility of cartooning with the PEViD data set and compare it with the two widely-used privacy filters: blurring and pixelation.

1. Introduction

The widespread deployment of video surveillance cameras [21] also increases concerns about privacy and data security [26, 31]. On-board processing capabilities of modern smart camera systems [22] allow to integrate privacy protection directly into the camera and provide distributed processing opportunities [10]. Various methods have been developed to protect privacy based on the prior identification of sensitive regions of interest (ROI) [1, 3, 6, 8, 13, 19, 20, 32] such as human faces. An ideal algorithm preserves privacy while behavioural information remains perceptible. The trade-off between privacy protection and the intelligibility of the resulting video, and hence the utility of a camera system, is a critical aspect.

*pvalet@edu.uni-klu.ac.at

Relying on the identification of ROIs prior to privacy protection limits the real-world applicability of many proposed solutions. The limited performance of any image-based ROI detector is critical in the context of privacy protection. Even a single mis-detection in a frame can seriously violate the privacy of a captured person, because the revealed identity will spread across time and space (i.e., to the whole length of the video footage and over to other cameras in case of multi-camera system). To avoid the dependency on the detector performance, privacy protection can be applied globally to the entire image [6].

In this paper we introduce a cartooning technique for privacy protection. Cartooning converts raw and photo-realistic images into abstracted frames where the privacy revealing details such as facial features, structures and landmarks are removed. Cartooning maintains a high utility of the videos, since behavioural information is still readily perceivable. Naturally, the effect of privacy protection and utility depends on the concrete settings of the cartooning filter. An ultimate objective would be that the cameras autonomously select the proper setting for a specific scene and the given privacy requirements. As a step towards this objective, we provide an adaptation mechanism to our cartooning technique where the operator can easily change the filter intensity. In this paper we focus on global cartooning which is applied to the entire image instead of pre-selected ROIs and can be implemented on-board of cameras.

Our contribution involves three major aspects. First, we introduce a resource-aware cartooning technique for global privacy protection. Second, we demonstrate the applicability and feasibility of this technique on different hardware platforms. Third, we evaluate privacy protection and utility of cartooning with the PEViD data-set [16] and compare it with blurring and pixelation — two widely-used privacy filter methods.

The rest of this paper is structured as follows. Section 2 summarises related work, and Section 3 presents our car-

tooning filter. Section 4 discusses the evaluation results including a comparison with two standard privacy filters. Finally, Section 5 concludes with an outlook to future work.

2. Related Work

Researchers have developed various methods for privacy protection in videos. These methods basically rely on image processing algorithms such as scrambling by JPEG-masking [19], in-painting [8], pixelation [13], blanking [3], replacement with silhouettes [32], blurring [20], warping or morphing [17]. The vast majority of these methods are used to protect only human faces and therefore assume that these ROIs can be reliably detected. However, protecting only primary privacy channels is insufficient in numerous scenarios [24]. Obscuring secondary (implicit) privacy threatening areas in videos, such as clothes and carried items, is typically required. At the MediaEval 2013 workshop [4], the participants proposed and evaluated several primary and secondary protection approaches including shape- and colour-aware segmentation, cartooning, adaptive edge detection, pixelation, silhouettes, replacing pixels based on the global minimum of surface spectral reflectance, pseudo-random pixel scrambling, warping and Discrete Cosine Transform based scrambling. Our cartooning algorithm was ranked top based on the total average scores for intelligibility, privacy and appropriateness [1].

In contrast to the object-based methods as evaluated in the recent workshop, global methods (e.g., [6]) are not reliant on inaccurate object detectors and thus will not miss any sensitive area of the image frame. Beyond detector independence, global techniques are easily applicable to moving (PTZ) cameras. Even if the global approach is detector-independent, the use of detectors can further increase its performance in terms of privacy protection by applying stronger filters on more sensitive areas of image frames. In terms of computational complexity the pipeline of the global approach typically consists of simpler building blocks, however significantly more pixels have to be processed. A particular challenge for global methods is to maintain a high utility of the video [25].

The level of visual privacy is considered sufficient when monitored people cannot be recognized or identified in the captured videos. A sufficient utility level is maintained when people and even faces can be detected and behavioural information can be perceived — despite the fact that persons can not be identified. The level of privacy protection and utility can be assessed by subjective and objective methods. Subjective methods are expensive but quite common in this area and include techniques such as questionnaires and user studies [6,15,23,27]. Objective methods are based on mathematical models [24] or automated methods [12,18] such as various image metrics or object detection and recognition algorithms from the field of computer vision. SSIM (Structural Similarity index) and PSNR (Peak Signal-to-Noise Ratio) [11] are often used for appropriateness evaluation. For utility evaluation, object detection is typically performed by using the Viola-Jones algorithm [29] or HOG-based (Histogram of Oriented Gradients) detectors [9]. Face recognition, with the purpose of evaluating privacy protection, can be based on PCA (Principal Component Analysis) [28], LDA (Linear Discriminant Analysis) [5] and LBP (Local Binary Patterns) [2]. of what these metrics are used for can be found in Section 4.

3. Privacy Protection by Adaptive Cartooning

The basic idea of cartooning is to convert raw and photo-realistic images into abstracted versions which still provide high utility but preserve some privacy. The objective is thus to generate "cartoons" which allow to recognise behaviours but hinder the identification of persons in the scene. The two key techniques for cartooning are colour filtering and edge enhancements, i.e., smoothing areas with moderate colour variations to single coloured areas and to enhance important areas with emphasized edges.

In this work we apply cartooning globally to entire frames. Thereby, we avoid privacy breaches due to unreliable region of interest detectors. Furthermore, since the required level of privacy protection highly depends on the application context, we introduce a simple mechanism to change the cartooning effect and can thus adapt the level of privacy protection as required.

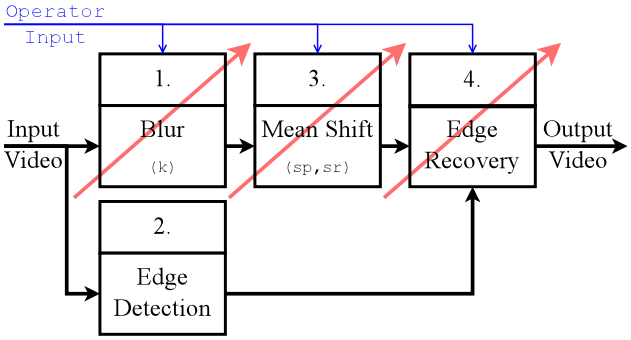


Figure 1: Image processing pipeline for cartooning.

Figure 1 depicts the image processing pipeline of our cartooning approach. Each frame is passed through the following four stages:

1. Initial blurring is performed to reduce noise and fine grained details in the input image.

2. A Sobel edge detector is applied on the input image.

3. A Mean Shift filter [7] is applied as colour filter on the blurred image.

4. The filtered image is superimposed by a bitwise weighted copy of the original image using the gradient mask from the Sobel edge detector. This strengthens the cartooning effect and makes the image less blurry.

The cartooning effect is adapted primarily via the kernel size for blurring (k), the Mean Shift spatial radius (sp) and the Mean Shift colour radius (sr). The following equations define the filter adaption within the range $i = (0..100]$, where i is the filter intensity. Intensity $i = 0$ means no filtering (i.e., output = input).

$$k_i = \lfloor i \cdot k_{orig}/50 \rceil|_{k_{orig}=k_{50}=7,(\forall k_i>1)} \quad (1)$$

$$sp_i = \lfloor i \cdot sp_{orig}/50 \rceil|_{sp_{orig}=sp_{50}=20} \quad (2)$$

$$sr_i = \lfloor i \cdot sr_{orig}/50 \rceil|_{sr_{orig}=sr_{50}=40} \quad (3)$$

Furthermore, the intensity of edge recovery is decreased towards $i = 100$ in order to avoid too many artefacts.

$$E_i = \begin{cases} \frac{E}{i/25}, & \text{if } i > 50. \\ E, & \text{otherwise.} \end{cases} \quad (4)$$

where E is the initially detected edge mask. The parameters k_{orig}, sp_{orig} and sr_{orig} are manually fine tuned for intensity $i = 50$ and other intensity levels are aligned around it proportionally.

Figures 2a to 2c illustrate the effect of different strengths of the cartooning filter and compares them also to blurring (2d to 2f) and pixelation (2g to 2i) which are widely used for privacy protection. From a visual assessment it can be seen that cartooning achieves visually appealing results which maintain higher intelligibility than blurring and especially pixelation. In Section 4 we are going to present results of a systematic evaluation of the privacy protection vs. intelligibility of our cartooning technique.

4. Evaluation

The evaluation of our cartooning technique focuses on two aspects — the tradeoff between privacy protection and utility as well as the runtime performance on state-of-the-art hardware. Privacy protection is measured by the recognition rate of standard face recognition methods, i.e., PCA [28], LDA [5] and LBP [2]. Naturally, lower recognition rates indicate a higher privacy protection. The utility of the filtered video is assessed based on the structural similarity (SSIM) and the peak signal-to-noise ratio (PSNR) w.r.t. the unmodified video frames. The performance of the standard Viola-Jones face detector is used as an additional parameter for the utility. In Section 4.1 a detailed discussion of the used evaluation methodology is presented followed by a discussion of the results in Section 4.2. Section 4.3 summarises the results of the runtime measurements.

(a) 10% Cartooning (b) 50% Cartooning (c) 100% Cartooning

(d) 10% Blurring (e) 50% Blurring (f) 100% Blurring

(g) 10% Pixelation (h) 50% Pixelation (i) 100% Pixelation

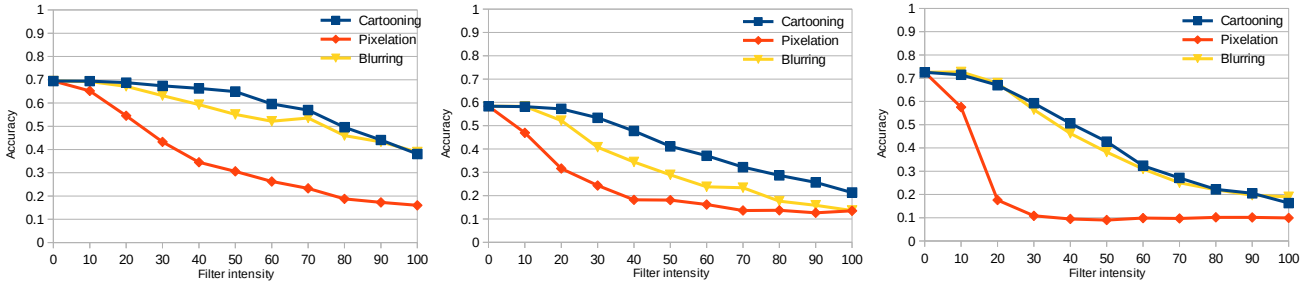
Figure 2: Comparison of globally cartooned, blurred and pixelated frames at different filter intensities.

4.1. Privacy vs. Utility Evaluation Method

Our evaluation is based on 20 full HD videos from the PEViD data-set [16]. In these video clips people perform various actions such as leaving a building, dropping a bag or fighting. Each clip consists of 400 frames and includes annotations for all persons, faces and objects of interest. These annotations serve as ground-truth for our evaluations, i.e., to determine the recognition and detection rates, respectively. All videos of the PEViD data-set have been globally filtered using cartooning, blurring and pixelation with intensity values ranging from 0% to 100%. In order to establish some correspondence among the different filter methods, we apply the same kernel sizes for the same intensity values, i.e., the kernel size for blurring and pixelation is equal to the Mean Shift kernel size sp for cartooning.[1] 50% intensity refers to a kernel size of 20×20 pixels while other intensity levels are calculated proportionally using Equation (2).

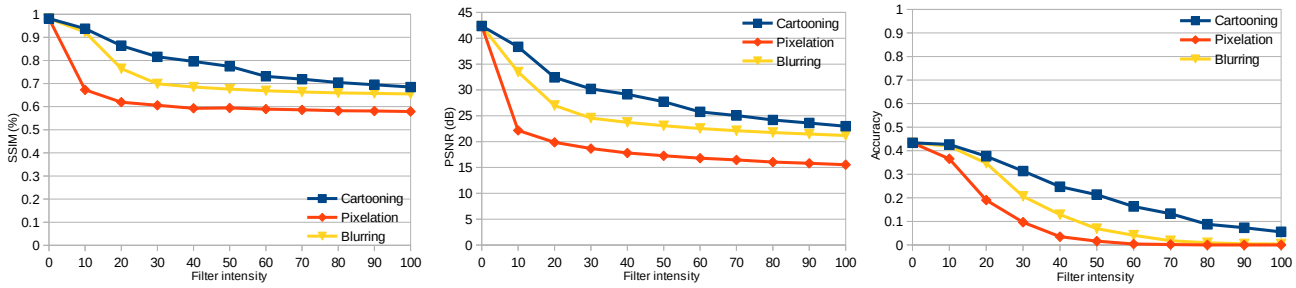
The presented recognition and detection rates in Section 4.2 are the average values for all frames of the 20 video sequences where at least one valid face appears. A face is considered valid if the coordinates of both eyes are available in the ground-truth data. This is necessary to rescale (based on the distance between the eyes) and realign the faces (by transforming the eyes above the reference points)

[1] Note that same kernel size does not result in the same utility or privacy protection due to the different amount of information abstraction for cartooning, blurring and pixelation.



(a) PCA Face Recognition (b) LDA Face Recognition (c) LBP Face Recognition

Figure 3: PCA, LDA and LBP face detection for cartooned, blurred and pixelated frames at different filter intensities.



(a) Structural Similarity Index (b) Peak Signal to Noise Ratio (c) Viola Jones Face Detection

Figure 4: SSIM, PSNR and Viola Jones for cartooned, blurred and pixelated frames at different filter intensities.

before training the recognizers. For face detection, a true positive is counted if

- the top left corners of the detected (x_{det}, y_{det}) and the annotated (x_{ann}, y_{ann}) bounding boxes are not further from each other by more than 25% of the width and height (w_{ann}, h_{ann}) of the annotated bounding box along the x and y axes respectively (i.e., $x_{ann} - w_{ann}/4 \leq x_{det} \leq x_{ann} + w_{ann}/4$ and $y_{ann} - h_{ann}/4 \leq y_{det} \leq y_{ann} + h_{ann}/4$),
- the size of the detected bounding box does not differ from the size of the annotated bounding box by more than 50% of its size (i.e., $|w_{det} - w_{ann}| \leq w_{ann}/2$ and $|h_{det} - h_{ann}| \leq h_{ann}/2$).

We use the following metrics for privacy evaluation:

1. The recognition rate of PCA-based face recognition (`EigenFaceRecognizer()` from OpenCV).
2. The recognition rate of LDA-based face recognition (`FisherFaceRecognizer()` from OpenCV).
3. The recognition rate of LBP-based face recognition (`LBPHFaceRecognizer()` from OpenCV).

And the following list contains the metrics for evaluating utility:

4. The detection rate of the built-in Viola-Jones face detector from OpenCV (`CascadeClassifier::detectMultiScale`).
5. The structural similarity (SSIM) index.
6. The peak signal-to-noise ratio (PSNR).

We trained all face recognizers (metric 1, 2 and 3) with the 1627 valid faces from the data-set with an overall population size of 12 persons. For face detection (metric 4), a generally trained Viola-Jones detector was used based on the standard "haarcascade_frontalface_default.xml" file from OpenCV.

4.2. Privacy vs. Utility Evaluation Results

The results of our experiments with three different face-recognizers are summarized in Figure 3. Taking into account the general accuracy level of these face-recognizers it turns out that the PCA-based face-recognizer (Figure 3a) performs better than the others. Since it is the strongest recognizer and hence threatening privacy the most, we want to focus on this primarily. Figure 4 depicts the achieved utility of the three filtering methods. As can be clearly seen, cartooning outperforms the other two methods in both the global utility metrics SSIM and PSNR as well as the object-based metric (face detection rate). On the other hand, pixelation achieves the lowest utility which was expected due to its strong data abstraction. According to Figure 3a cartooning is fairly close to blurring with regard to privacy. However, it is quite far from utility based on the charts of Figure 4. If we choose a different basis for filter intensity alignment and thus slightly lower the utility level until it is still better than the other two methods (i.e., cartooning curves

Device	CPU	Memory	OS	Execution Time
Desktop	Intel Core i7 (3770) Quad Core + HT, 3.4 GHz	16 GB DDR3 1333 MHz	Ubuntu 13.10	77 ms
Desktop with CUDA (GeForce GTX 560Ti)	Intel Core i7 (3770) Quad Core + HT, 3.4 GHz	16 GB DDR3 1333 MHz	Ubuntu 13.10	20 ms
Laptop	Intel Core i5 (3320M) Dual Core + HT, 2.6 GHz	8 GB DDR3 1600 MHz	Ubuntu 13.10	190 ms
Odroid-U2	ARM Cortex-A9 (Exynos 4412) Quad Core, 1.7 GHz	2 GB LPDDR2 800 MHz	Ubuntu 13.09	400 ms
Pandaboard ES	ARM Cortex-A9 (TI OMAP 4460) Dual Core, 1.2 GHz	1 GB LPDDR2 400 MHz	Ubuntu 12.04	1600 ms

Table 1: Average execution times over 2000 frames for the cartooing algorithm (cp. Fig. 1) for different platforms. The input resolution is 320×240. Multi-threading is used to utilize all available CPU cores of the individual platforms.

will be shifted to the left in Figure 4), we can achieve higher privacy levels with cartooning than with blurring (since a shift will occur also in Figure 3). Pixelation causes a huge loss of visual information and therefore it provides high privacy levels but performs quite bad in terms of utility. This makes it inadequate for surveillance purposes.

4.3. Execution Time Evaluation

Our cartooning filter is implemented in C++ using OpenCV [14]. The most important OpenCV functions in our processing pipeline are `blur()`, `Sobel()`, `pyrMeanShiftFiltering()` and its GPU counterpart `gpu::meanShiftFiltering()`. The Mean Shift filter is configured to use maximum 2 levels of the Gaussian pyramid and the termination criteria is set to 2 iterations (with $\epsilon = 1$). 95% of the processing time is spent on `pyrMeanShiftFiltering()`, therefore this is the bottleneck of the processing pipeline.

An important aspect is the potential applicability of the cartooning on embedded camera devices. In Table 1 we provide the specification of the different platforms used for our measurements together with the achieved filter execution times for a single frame. Our cartooning implementation is designed to utilize all CPU cores provided by the platforms. The given execution times are averages for 2000 frames. The input resolution is 320×240.

In its current version the cartooning achieves about 2.5 fps on a state-of-the-art embedded devices such as the Odroid-U2. On a standard PC we achieve about 13 fps. With a GPU implementation of mean shift using the CUDA cores, the frame rate is boosted to about 50 fps. With new embedded platforms and embedded GPUs this is a promising direction for increasing substantially the frame rates and making cartooning feasible for embedded smart cameras.

5. Conclusion and Future Work

We presented a globally applied privacy filter based on cartooning that can achieve an acceptable level of privacy protection while maintaining a good utility level. Our filter provides adaptive functionality so that the balance between privacy and utility can be freely adjusted and matched to arbitrary surveillance scenarios. Furthermore, the applicability of such a filter on real-world embedded devices was also demonstrated. Ongoing work involves the performance enhancements of our cartooning technique [] as well as optimization for embedded, GPU-enabled smart camera devices. Our preliminary results show that our algorithms strongly benefit from moving parts of the computation from the CPU to the GPU.

In order to further enhance the performance of our method in terms of privacy, extra image filters could be applied at sensitive areas of the video frames such as an extra blurring on faces or the re-colouring of certain regions.

A totally automated evaluation framework can be used to do the adaptation of filter intensity to arbitrary scenarios. We are planning to extend and further develop our current evaluation framework in order to create a prototype of such an autonomously adaptive privacy preserving video filter.

Acknowledgment

This work was performed in the project *TrustEYE: Trustworthy Sensing and Cooperation in Visual Sensor Networks* [30]. It was funded by the European Regional Development Fund (ERDF) and the Carinthian Economic Promotion Fund (KWF) under grant KWF-3520/23312/35521.

References

[1] Á. Erdélyi, T. Winkler, and B. Rinner. Serious Fun: Cartoon-

ing for Privacy Protection. In *Proceedings of the MediaEval Multimedia Benchmark Workshop*, page 2, 2013.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[3] A. J. Aved and K. A. Hua. A general framework for managing and processing live video data with privacy protection. *Multimedia systems*, 18(2):123–143, 2012.

[4] A. Badii, M. Einig, and T. Piatrik. Overview of the MediaEval 2013 Visual Privacy Task. In *Proceedings of the MediaEval Multimedia Benchmark Workshop*, page 2, 2013.

[5] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[6] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *Proceedings of the Conference on Computer Supported Cooperative Work*, pages 1–10, 2000.

[7] Y. Cheng. Mean shift, Mode Seeking, and Clustering. *Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[8] S.-C. Cheung, M. Venkatesh, J. Paruchuri, J. Zhao, and T. Nguyen. Protecting and managing privacy information in video surveillance systems. *Protecting Privacy in Video Surveillance*, pages 11–33, 2009.

[9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[10] B. Dieber, L. Esterle, and B. Rinner. Distributed resource-aware task assignment for complex monitoring scenarios in visual sensor networks. In *Proceedings of the International Conference on Distributed Smart Cameras*, pages 1–6, 2012.

[11] F. Dufaux. Video scrambling for privacy protection in video surveillance: recent results and validation framework. In *Proceedings of SPIE*, volume 8063, page 14, 2011.

[12] F. Dufaux and T. Ebrahimi. A framework for the validation of privacy protection solutions in video surveillance. In *Proceedings of the International Conference on Multimedia and Expo*, pages 66–71, 2010.

[13] B.-J. Han, H. Jeong, and Y.-J. Won. The privacy protection framework for biometric information in network based cctv environment. In *Proceedings of the Conference on Open Systems*, pages 86–90, 2011.

[14] itseez. OpenCV – Open Source Computer Vision. `http://opencv.org`, 2013. Last accessed: June 2014.

[15] P. Korshunov, C. Araimo, F. Simone, C. Velardo, J. Dugelay, and T. Ebrahimi. Subjective study of privacy filters in video surveillance. In *Proceedings of the International Workshop on Multimedia Signal Processing*, pages 378–382, 2012.

[16] P. Korshunov and T. Ebrahimi. PEViD: Privacy Evaluation Video Dataset. In *Proceedings of SPIE*, volume 8856, page 9, 2013.

[17] P. Korshunov and T. Ebrahimi. Using face morphing to protect privacy. In *Proceedings of the 10th International Conference on Advanced Video and Signal Based Surveillance*, pages 208–213, 2013.

[18] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi. Framework for objective evaluation of privacy filters. In *Proceedings of SPIE*, volume 8856, page 12, 2013.

[19] K. Martin and K. N. Plataniotis. Privacy protected surveillance using secure visual object coding. *Transactions on Circuits and Systems for Video Technology*, 18(8):1152–1162, 2008.

[20] A. Martinez-Balleste, H. A. Rashwan, D. Puig, and A. P. Fullana. Towards a trustworthy privacy in pervasive video surveillance systems. In *Proceedings of the Pervasive Computing and Communications Workshops*, pages 914–919, 2012.

[21] F. Porikli, F. Brémond, S. L. Dockstader, J. Ferryman, A. Hoogs, B. C. Lovell, S. Pankanti, B. Rinner, P. Tu, and P. L. Venetianer. Video Surveillance: Past, Present, and Now the Future. *IEEE Signal Processing Magazine*, 30(3):190–198, 2013.

[22] B. Rinner and W. Wolf. An introduction to distributed smart cameras. *Proceedings of the IEEE*, 96(10):1565–1575, 2008.

[23] M. Saini, P. Atrey, S. Mehrotra, and M. Kankanhalli. Anonymous surveillance. In *Proceedings of the International Conference on Multimedia and Expo*, pages 1–6, 2011.

[24] M. Saini, P. Atrey, S. Mehrotra, and M. Kankanhalli. W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications*, pages 1–24, 2012.

[25] M. Saini, P. K. Atrey, S. Mehrotra, and M. Kankanhalli. Adaptive transformation for robust privacy protection in video surveillance. *Advances in Multimedia*, 2012:4, 2012.

[26] C. Slobogin. Public privacy: Camera surveillance of public places and the right to anonymity. *Mississippi Law Journal*, 72(1):213–300, 2002.

[27] S. Tansuriyavong and S.-i. Hanaki. Privacy protection by concealing persons in circumstantial video image. In *Proceedings of the Workshop on Perceptive User Interfaces*, pages 1–4, 2001.

[28] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[29] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[30] T. Winkler, Á. Erdélyi, and B. Rinner. TrustEYE – Trustworthy Sensing and Cooperation in Visual Sensor Networks. `http://trusteye.aau.at`, 2012. Last accessed: June 2014.

[31] T. Winkler and B. Rinner. Security and Privacy Protection in Visual Sensor Networks: A Survey. *ACM Computing Surveys*, 47(1):42, 2014. (in print).

[32] C. Zhang, Y. Tian, and E. Capezuti. Privacy preserving automatic fall detection for elderly using rgbd cameras. In *Proceedings of the International Conference on Computers Helping People with Special Needs*, pages 625–633, 2012.