Shape and Color-aware Privacy Protection

Andrea Melle
EURECOM
450 Route Des Chappes
Biot, France
andrea.melle@eurecom.fr

Jean-Luc Dugelay
EURECOM
450 Route Des Chappes
Biot, France
jean-luc.dugelay@eurecom.fr

ABSTRACT

We introduce a novel content-independent filter to protect privacy sensitive Regions Of Interest (ROI) in video surveillance sequences. An abstracted version of the original image is rendered such as the general appearance of shapes and colors is preserved, while obfuscating fine details carrying personal visual information. We use shapes and colors-aware, temporally coherent segmentation algorithm, combined with a color quantization and patch rendering step.

1. INTRODUCTION

The increasing adoption of video surveillance systems has led to a growing research interest in privacy protection methods. A review of principles for privacy protection in video surveillance can be found in [9], while an evaluation of several existing protection filters is reported in [3]. One persistent challenge in privacy protection remains to find the correct balance between obfuscation of personal visual information, intelligibility of the source and pleasantness.

Non-photorealistic rendering techniques described in the literature achieve artistic effects such as tooning, painting, or sketching. For example, in [10] the authors propose a video abstraction pipeline based on bilateral filter and color quantization, and subjectively evaluate both visual pleasantness and intelligibility, coming to the conclusion that abstracted images favor general content understanding. The use of segmentation to obtain a pixelized result resembling *pixel art* has been proposed in [4]. However, this method applied to privacy protection would carry the same drawbacks of the commonly adopted pixelization filter [7].

We propose a new privacy protection filter inspired by results in image abstraction and non-photorealistic rendering fields. Our method is based on a boundaries and regions-aware segmentation algorithm, combined with a color quantization and patch rendering step, which transforms the original privacy sensitive ROI in a stylized and simplified version. While the general appearance of shapes and colors is preserved, to allow for people and actions detection tasks, identification details, such as faces and clothes traits, are obfuscated to render identification impossible.

2. PROPOSED APPROACH

Given a video sequence together with bounding boxes

defining the privacy sensitive ROIs, our algorithm proceeds in three steps. First, a segmentation algorithm divides the image in boundaries-aware patches. Second, the image is abstracted by replacing the pixels in each patch with a single color chosen from a palette. Finally, the abstracted image is rendered on top of the original frame to produce the final output. If additional region annotations or background subtraction maps are available, the final result can be further refined by binary masking. Figure 1 shows an example of original and filtered frame.

The algorithm allows adaptation to the desired strength of privacy protection. By varying the number of patches, either globally or independently in certain regions, we can obtain different levels of abstraction. Our C++ implementation takes on average less than 0.5 seconds/frame for segmentation and about 0.3 seconds/frame for color quantization and rendering.

2.1 Segmentation

The intuition behind our privacy protection algorithm is to render an abstracted version of the image by replacing patches of pixels with a single color chosen from a palette. To preserve intelligibility and visual pleasantness, we aim for a region and boundary-aware process. Accordingly, we adopt a segmentation procedure which divides the image in a user-specified number N of arbitrarily shaped patches, maximizing both their spatial and color consistency. A good review of patch, or *superpixel*, segmentation methods, together with the original description of the algorithm we adopted in our work ($SLIC$) can be found in [1].

The SLIC segmentation algorithm is based on *K-means* clustering [8] performed in a $5D$ space which includes both spatial coordinates (x, y) and color values in the perceptually uniform (L, a, b) space. While the original formulation of SLIC works best for still images, when applied to video sequences jittery artifacts appear, due to temporal inconsistencies in color and shape of patches over several frames. Therefore, we adopt an extension of the algorithm which enforces temporal consistency by including the temporal dimension t in the clustering distance metric: a video can be represented as a 3D volume by stacking up its frames over the time dimension, and therefore segmented in supervoxels. A combined distance metric is obtained as a linear combination of the two L^2 norms on space-time (x, y, t) coordinates and color (L, a, b) values:

$$D = d_{Lab} + c \frac{R \times C}{N} d_{xyt} \qquad (1)$$

Where N is the desired number of patches, (R, C) are the

height and width of the ROI and c is a *compactness* parameter balancing the trade-off between spatial proximity and color similarity of the resulting clusters.

The output of the segmentation algorithm is a segmentation label map, where each patch is identified by a unique label. When additional annotation corresponding to specific regions, such as a face, is available, we enforce a higher level of privacy protection by merging all the patches substantially overlapping with such region, to ensure proper obfuscation of shape and color details.

2.2 Color quantization and patch rendering

We keep a palette of a small fixed number of colors (e.g. 8) progressively updated from the upcoming frames as following: we first compute the average color for each patch and subsequently build the palette with a *K-medoids* quantization [5] over all the color occurrences at the current and previous most recent $n = 5$ frames. Each patch is then filled by the closest color in the palette. The resulting filtered image still resembles the original one in the general shape and color appearance, but the fine details are destroyed.

2.3 Masking

To make the result visually more appealing and avoid filtering nonsensitive regions, we crop the abstract image with a foreground mask, inferred from the annotations and background subtraction maps, when available. Very sensitive regions such as face and skin are represented with an ellipse in the mask, to enforce maximum protection. The final frame is computed as:

$$I_{out} = I_a \cap [\mathcal{S}(L, M_f) \geq T] \oplus I_{in} \cap [\mathcal{S}(L, M_f) < T] \quad (2)$$

where I_{out} is the final rendered image, I_a the abstract image, I_{in} the input image, L the segmentation labels map and M_f the foreground mask. $\mathcal{S}$ is a support operator which counts the number of foreground pixels for each given patch label. In such way, each patch is either fully rendered abstracted or fully rendered original in the final image.

(a)   (b)

Figure 1: Example of filtered frame (b) vs. original (a) .

3. RESULTS

We applied the proposed method on selected sequences from the PEViD dataset [6]. Evaluation has been performed according to the MediaEval 2013 Visual Privacy Task guidelines, as described in great details in [2]. Table 1 reports our scores, together with the average score of all participants to the challenge.

	Objective		Subjective	
	Score	Average	Score	Average
Intelligibility	0.563	0.502	0.728	0.656
Privacy	0.576	0.665	0.607	0.684
Appropriateness	0.385	0.56	0.514	0.492

Table 1: Results: our method and average of all participants.

4. CONCLUSIONS

In this paper we have proposed a novel privacy filter based on a region-aware segmentation algorithm combined with a color quantization and abstract rendering step. The result is a stylized image where the general intelligibility of shape and color is preserved, but the fine details of visual features are destroyed.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. SuÌĹsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, 2012.

[2] A. Badii, M. Einig, and T. Piatrik. Overview of the mediaeval 2013 visual privacy task. *MediaaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain*, 2013.

[3] F. Dufaux and T. Ebrahimi. A Framework for the Validation of Privacy Protection Solutions in Video Surveillance. In *Proc. of IEEE International Conference on Multimedia & Expo*, IEEE International Conference on Multimedia and Expo. Ieee Service Center, 445 Hoes Lane, Po Box 1331, Piscataway, Nj 08855-1331 Usa, 2010. special session on "Privacy-aware Multimedia Surveillance".

[4] T. Gerstner, D. DeCarlo, M. Alexa, A. Finkelstein, Y. Gingold, and A. Nealen. Pixelated image abstraction. In *Proceedings of the International Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, June 2012.

[5] L. Kaufman and P. Rousseeuw. Clustering by means of medoids. 1987.

[6] P. Korshunov and T. Ebrahimi. Pevid: privacy evaluation video dataset. 2013.

[7] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi. A framework for objective evaluation of privacy filters in video surveillance. In *Proceedings of SPIE Volume 8856*, volume 8856, 2013.

[8] S. Lloyd. Least squares quantization in pcm. *IEEE Trans. Inf. Theor.*, 28(2):129–137, Sept. 2006.

[9] A. Senior. Privacy protection in a video surveillance system. In A. Senior, editor, *Protecting Privacy in Video Surveillance*, pages 35–47. Springer London, 2009.

[10] H. Winnemöller, S. C. Olsen, and B. Gooch. Real-time video abstraction. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 1221–1226, New York, NY, USA, 2006. ACM.