MediaEval 2012 Visual Privacy Task: Privacy and Intelligibility through Pixellation and Edge Detection

Atta Badii
University of Reading
ISR Research Laboratory, School of Systems Engineering
United Kingdom
atta.badii@reading.ac.uk

Mathieu Einig
University of Reading
ISR Research Laboratory, School of Systems Engineering
United Kingdom
m.l.einig@reading.ac.uk

ABSTRACT
In this paper, we describe a method that can protect privacy in video footage by automatically obscuring faces. In order to retain some level of intelligibility, a specific filter has been added for showing some facial features without revealing the person's identity.

Keywords
Privacy, privacy protection, video analytics, image processing, filtering, face detection, tracking

1. INTRODUCTION
The ubiquity of surveillance combined with the new technological advances in terms of Video Analytics is creating an important need for mitigation technologies that could protect people's privacy. However, basic filters such as blurring have been shown to be inadequate for protecting privacy [1], and it is therefore crucial to develop new methods that can offer a good balance between privacy and intelligibility. Our participation to the MediaEval Visual Privacy Task [2] aims at addressing this issue.

2. PROPOSED APPROACH
The challenge is therefore to find a filter that can reliably be applied automatically to video footage that will protect privacy as much as possible, but at the same time, also try to retain some level of intelligibility in order to ensure that the videos can still be used for surveillance purposes. The following sections explain our approach.

2.1 Face detection
The specific data set of the task [3] makes the face detection challenging as the subjects can be wearing different accessories which hide their face such as sunglasses, hats and scarves. Furthermore, videos were recorded in different lighting conditions (morning and evening) meaning that an automatic system must be able to cope with a wide range of scenarios including hard shadows, saturated and over exposed images. Our approach therefore relies on a fast face detector using Local Binary Patterns (LBP) features [4] and on the Histogram of Oriented Gradient (HOG) detector [5] trained for detecting upper bodies (shoulders and head).

The LBP detector will fail for small or occluded faces, unlike the HOG detector which will fail only when the subject's face is close to the borders of the frame, making the two detectors complementary. However, the HOG detector cannot discriminate between the front and back of the person. Therefore, when only the HOG detects a face, it is unclear whether the person is facing the camera and should be anonymised or not. We use a Hidden Markov Model to classify the track and size of the detected bounding box into three categories: moving closer, static, moving away. If the subject is detected as moving away, then no anonymisation is performed as it is likely that the subject is turning their back to the camera. This technique is not perfect as people could potentially be moving backwards – although this does not occur in the data set – but it has the advantage of working regardless of the environment, and the person's skin or hair colour and accessories.

Faces are tracked, and when lost, their position is extrapolated from their previous positions for a few frames, before being discarded if they do not reappear. This allows a better protection of privacy in case of difficult scenes where faces are not easily detected because of occlusions or poor lighting.

2.2 Masking
We combine two different masks for achieving the two goals of this challenge which are anonymity and intelligibility.

The anonymity filter uses a standard pixellation technique that reduces the image to 12x12 pixels, making the identification of a person or even the detection of a face extremely difficult.

Figure 1. Input image (left) and result of the anonymity filter (right)

The intelligibility filter results from the use of a Sobel filter on the saturation component of the face in HSV colour space. The saturation channel has been chosen because it features good contrast under all lighting conditions, and preserves the edges of the eyes and mouth.

These edges alone are not sufficient to get information about the person's identity, but allow a good detection of changes in facial expression, which could be useful for surveillance purposes.

The edges are thresholded and attenuated by modulating their pixel value by the inverse of the squared distance to the centre, keeping only the strongest central ones.

MediaEval 2012 Workshop, October 4-5, 2012, Pisa, Italy

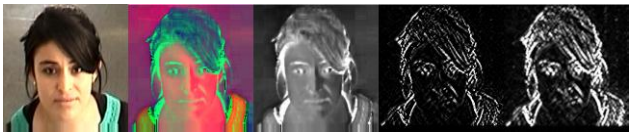

Figure 2. Different stages of the intelligibility filter (from left to right): original image, HSV image, saturation component, Sobel filter, and final image after morphological filters

The results of the anonymity and intelligibility filters are then blended into the original image with a radial attenuation in order to avoid having very visible and irritating edges around the filtered areas of the video.

Figure 3. Final obscured face

3. RESULTS

Our technique has been evaluated both objectively through the use of computer vision and image processing algorithm, and subjectively through a questionnaire. The objective evaluation relies on 5 different criteria:

- Accuracy: the overlap of detected faces with respect to the ground truth.
- Anonymity: the ratio of faces that were automatically detected after the filter was applied.
- Intelligibility: the ratio of detected pedestrians from the original and filtered videos.
- SSIM and PSNR: visual similarity between the original and filtered videos.

The objective results achieved by our method on 12 different videos are as follows:

	Accuracy	Anonymity	Intelligibility	SSIM	PSNR
Mean	0.50	1.00	0.93	0.96	35.80
Std Dev	0.19	0.00	0.06	0.02	1.07

The relatively poor face detection accuracy can be explained by the fact that even though our face detector tends to be less sensitive to occluding accessories (such as scarves, caps or sunglasses), it also tends to recognise some parts of the background as faces. The anonymity score was evaluated by running the standard OpenCV face detector on the filtered image. Our methods achieved a perfect anonymity score, which does not mean that people are completely unrecognisable, but that it is impossible to perform automated face detection and recognition on our output videos. Our high intelligibility score indicates that other algorithms such as the Histogram of Oriented Gradient for pedestrian detection are not influenced by our filter and automatic processing of our videos are still possible. The SSIM and PSNR scores show that the output image remains quite similar to the input.

The subjective evaluation was carried out through a questionnaire asking 20 participants to recognise people, or guess the gender and ethnicity of the subjects shown in the video as well as the items they could be wearing. Participants were also asked questions regarding their feelings about the results by seeing it from the side of both the subject and CCTV operator.

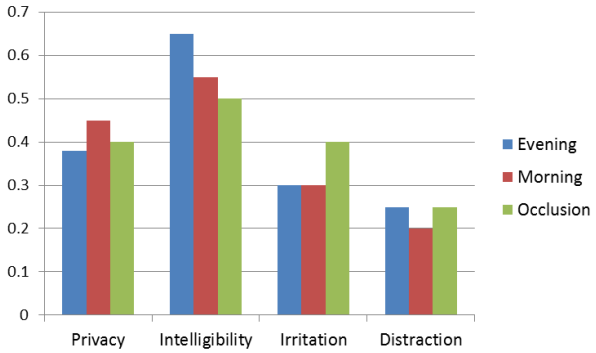

Figure 4. Subjective evaluation results for each subset

The subjective evaluation shows that the privacy and intelligibility scores are balanced, but that the effect is distracting and irritating. The results also indicate that our results are very consistent and are not significantly affected by the lighting conditions or occluding accessories.

4. CONCLUSION

We have presented a privacy protection filter relying on face detection that balances privacy and intelligibility. The face detection relies on two widely used object detectors combined together with a Hidden Markov Model for filtering out some false positives. Privacy is achieved through pixellation, and intelligibility through edge detection. This method has been validated through both objective and subjective evaluation. Our results show that this method prevents automatic face detection and recognition, but does not disrupt other video analytics algorithms such as pedestrian detection.

5. ACKNOWLEDGMENTS

This work was supported by the European Commission under contracts FP7-261743 VideoSense.

6. REFERENCES

[1] P. Korshunov, C. Araimo, F. De Simone, C. Velardo, J. L. Dugelay, and T. Ebrahimi, "Evaluation of visual privacy filters impact on video surveillance intelligibility," in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, 2012, pp. 150–151.

[2] A. Badii, T. Piatrik, M. Einig, and C. Lallah, "Overview of MediaEval 2012 Visual Privacy Task," *MediaEval 2012 Workshop, Pisa, Italy*, Sep. 2012.

[3] C. Velardo, C. Araimo, and J. L. Dugelay, "Synthetic and privacy-preserving visualization of video sensor network outputs," in *Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on*, 2011, pp. 1–5.

[4] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893.