MediaEval 2012 Visual Privacy Task: Applying Transform-domain Scrambling to Automatically Detected Faces

Pavel Korshunov
Multimedia Signal Processing Group, EPFL
CH-1015 Lausanne, Switzerland
pavel.korshunov@epfl.ch

Aleksei Triastcyn
Multimedia Signal Processing Group, EPFL
CH-1015 Lausanne, Switzerland
aleksey.tryastsyn@epfl.ch

Touradj Ebrahimi
Multimedia Signal Processing Group, EPFL
CH-1015 Lausanne, Switzerland
touradj.ebrahimi@epfl.ch

ABSTRACT

In this paper, we describe our approach and discuss evaluation results for the MediaEval 2012 Visual Privacy task. The goal of the task is to obscure faces of people visible in provided surveillance clips to preserve their personal privacy. We also additionally assume, although it is not explicitly stated in the task description, that the privacy protection should be done in an automated way and the applied privacy tool should be reversible and prone to attacks. We use a combination of a face detection algorithm and transform-domain scrambling technique, which pseudo-randomly scrambles bits during encoding, that was applied to the detected face regions. The evaluations of the resulted automated privacy protection tool showed that inaccuracies of the face detection algorithm affected both objective and subjective results. An interesting finding is however that scrambling, while being non-distractive to the evaluating subjects, appeared significantly irritating with score of 0.8, but only for 'evening' subset of the dataset.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*video analysis, representations, data structures, and transforms*

Keywords

Privacy protection, video surveillance, scrambling, face detection.

1. INTRODUCTION

The Visual Privacy tasks focuses on the problem of privacy protection in video surveillance, which is gaining more and more attention from research community and concerned public. Wide and growing adoption of video surveillance systems, their intrusive nature into personal space, and general social aspect of such systems present several challenges to the tools aiming on preserving privacy. The main challenges, specified in Visual Privacy task this year, include obfuscation of faces and supporting accessories (i.e., scarf, cap, or glasses) of people in a given surveillance scenario and making such obfuscation visually pleasing, i.e., socially acceptable. Many existing tools, such as blurring, pixelization, or masking, for protection of personal privacy in image and video do not respond well to the defined challenges. Therefore, we participated in this challenge to start creating a comprehensive and socially acceptable tool for protection of visual privacy.

In addition to the challenges specified in the Visual Privacy task description, we assume that a privacy protection tool should be fully automated, without any manual pre-processing. Since the goal of the task is to obscure faces of the persons in the video, we employed a face detection algorithm from the OpenCV library[1].

Also, we assume that obfuscation of faces in a video should be done in a reversible manner, so the original version of the video can be recovered if necessary, as opposed to such irreversible and simple techniques like blurring or pixelization. Several reversible approaches to obfuscation of regions in video exist, such as encryption of selected regions [1], complete removal of sensitive regions from the video [5] (encoding and storing them separately), transmitting only the information necessary to perform the surveillance task at hand with information about sensitive regions superimposed on the main stream as separately encrypted and encoded visual objects [3]; other notable techniques also include anonymization [4].

In this challenge, however, we use a transform-domain scrambling technique [2], where pixels in the sensitive regions are scrambled in a pseudo-random way based on a secret key. The scrambling is done during encoding process by randomly flipping significant bits of DCT components of the video frames (assuming MPEG-4 encoding). Unscrambling can be performed by flipping back the bits of DCT components during decoding process, provided the pseudo-random algorithm is seeded with the same key. Since the scrambling is done on DCT components and not pixels directly, the resulted scrambled video is highly tolerant to later modifications, alterations, and security attacks. Even after such alterations, the scrambled video can be unscrambled with very little loss in visual quality, which makes this technique highly attractive to use in practical applications, such as video surveillance.

2. PRIVACY PROTECTION TOOL

We built an automated visual privacy protection tool using Visual C++. The tool essentially combines OpenCV face detection algorithm with scrambling algorithm that is based on MPEG-4 encoding standard software.

2.1 Detecting Faces

Since videos in the provided dataset were very different in terms of lighting (shot during evening and morning) and visibility of the faces (faces of various sizes, occlusions from caps, scarfs, and sunglasses), tuning face detection algorithm from OpenCV was a chal-

[1]http://opencv.org/

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

lenge. Based on the training videos, we had come up with the following tuning parameters to achieve the best tradeoff between the number of false and true positives. We set to 5 the minimum number of neighbour rectangles combined into object. The pruning flag of the algorithm was set to CANNY. The size of the scanning window was decreased to 10×10 pixels from the default 20×20 pixels, because many videos in the dataset had very small faces (as people walked further away from the camera), so the smaller search window would help to detect more faces.

Despite the tuning efforts, the face detection under performed, especially, on videos from the 'evening' subset and on videos with occlusions (scarf, cap, and sunglasses).

2.2 Scrambling Detected Regions

We run the face detection algorithm on each frame for each input video. Resulted bounding boxes around faces are saved into the binary video mask file. Since scrambling is implemented as a component of the MPEG-4 standard encoder (software for one of the pre-final standard drafts), we set the corresponding configuration files for the scrambled encoding for each video, which can have a different frame rate and resolution. Our scrambling technique randomly changes bits of 63 DCT coefficients for every macro block in a given video frame. The final output of the scrambling is an MPEG-4 video.

3. EVALUATION RESULTS

Results obtained with the objective evaluation tool provided by the organisers demonstrated the weaknesses of the OpenCV face detection algorithm. Accuracy of the detection is reported as 0.24 on average, with almost no detections for videos with occlusions. One possible reason for such detection results is the over-conservative choice of values for tuning parameters of the algorithm. The PSNR and SSIM metrics for scrambled videos are 28.26 and 0.92 respectively, which are moderately high, especially since videos were re-encoded with older standard draft, which encodes less efficiently than the currently used MPEG-4 encoder.

The subjective evaluations are summarised in Figure 1 and Figure 2. Ideally, we would like to have high privacy scores (visual privacy is preserved) and high intelligibility scores (surveillance task is not obstructed by privacy protection), but, in reality, there is a tradeoff between privacy and intelligibility. Since no face detection algorithm has 100% of the detection accuracy, in some video frames, faces remained undetected and hence unscrambled. As the whole videos were played back to the subjects during evaluations, those unscrambled faces led to low subjective privacy scores (see data for EPFL in Figure 1).

Although face detection had a significant impact on the with privacy and intelligibility scores, irritation and distraction (Figure 2) should only be affected by the privacy protection technique itself. Interestingly, our scrambling approach shows significantly high irritation score, but only for videos from the 'evening' subset. A possible explanation could be that evening videos after scrambling appeared more unnatural or 'scary' to the subjects due to a poorer lighting. But since there is no such questionary data available, this hypothesis should be explored further by studying the irritation effect of scrambling on people.

4. CONCLUSION AND FUTURE WORK

The main focus of our approach, when participating in a Visual Privacy task, was on combining together in a single tool an automated face detection and reversible scrambling technique that is prone to attacks. Evaluation results show that a better face detection algorithm is necessary, while irregularity in the irritation scores suggest the need to develop a more comprehensive tests for irritation and distraction subjective factors and investigate these effects of scrambling technique further.

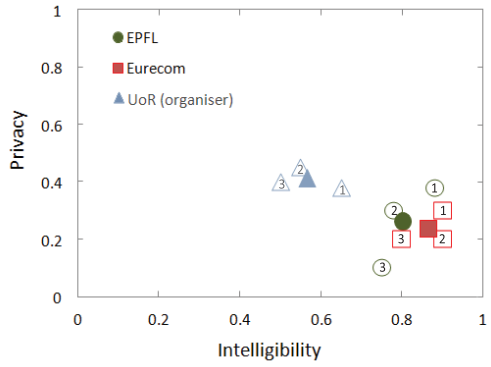

Figure 1: Privacy vs. intelligibility for different participants.

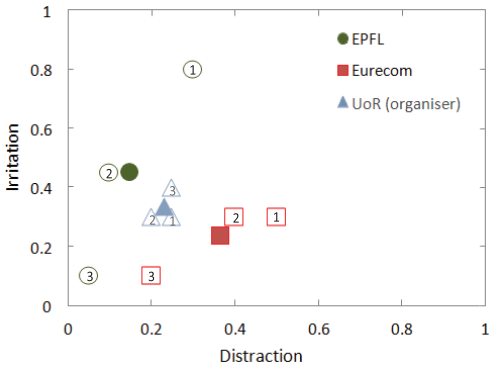

Figure 2: Irritation vs. distraction for different participants.

5. ACKNOWLEDGMENTS

This work was conducted in the framework of the EC funded Network of Excellence VideoSense.

6. REFERENCES

[1] T. Boult. PICO: Privacy through invertible cryptographic obscuration. In *IEEE Workshop on Computer Vision for Interactive and Intelligent Environments*, Nov 2005.

[2] F. Dufaux and T. Ebrahimi. Video surveillance using jpeg 2000. In *SPIE Proc. Applications of Digital Image Processing XXVII*, Denver, CO, Aug 2004.

[3] F. Dufaux and T. Ebrahimi. Recent advances in mpeg-7 cameras. In A. G. Tescher, editor, *Applications of Digital Image Processing XXIX*, volume 6312. SPIE, 2006.

[4] C. Velardo, C. Araimo, and J.-L. Dugelay. Synthetic and privacy-preserving visualization of video sensor network outputs. In *5th ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC'11)*, pages 1–5, Ghent, Belgium, Aug 2011.

[5] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *Proceedings of the ACM international conference on Multimedia*, pages 48–55, NY, USA, 2004.