MediaEval 2012: Scrambling Faces for Flexible Privacy Preservation using Image Background Self-similarities

Claudia Araimo
EURECOM
450 Route des Chappes
Biot, France
araimo@eurecom.fr

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

ABSTRACT

In this paper we present a system for preserving the privacy of individuals in video sequences by obfuscation of facial appearance, so as to render recorded people unrecognizable. Therefore we address the following two problems: automatic face detection and scrambling of face region. A multi-resolution access control scheme is used in order to decrease the visual quality of regions containing faces, according to the user access right. The proposed scrambling scheme is achieved by combining a self-VQ coding module and a bit masking algorithm.

Categories and Subject Descriptors

K.4.1 [**Computing Milieux**]: Computers and SocietyPublic Policy Issues[Privacy]

Keywords

Video surveillance, privacy, scrambling, access control

1. INTRODUCTION

The escalation in video camera deployment has led to a growing interest in protection of individual privacy. Obscuring faces is an effective way to preserve personal information, and several solutions have been developed in this direction; some few examples are: scrambling [1], face de-identification [3], complete anonymization [6]. Although preservation of sensitive data is of fundamental importance, we think that reversibility of the obscuration process should be guaranteed, so that authorized people can access to original data.

In this work a scheme for region-based obfuscation is proposed. This is achieved by combining coding and bit masking within a single algorithm [5]; the coding scheme exploits self-similarities in the frame: square partitions of the detected Region of Interest (ROI) are approximated by transformed versions of some other blocks extracted from the background. By partially hiding the coded values through bit masking, several gradual quality levels can be provided at the decoding stage. Such an approach allows a multi-resolution access control scheme, depending on user access rights.

2. SYSTEM DESCRIPTION

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

The proposed privacy-protection system consists of two steps: first, a face segmentation module, that extracts a bounding-box containing the detected face; second, an obfuscation module, that encrypts coded data corresponding to the extracted face regions.

2.1 Face segmentation

Two different algorithms have been used for the detection of faces: the well-known Viola-Jones frontal face detection algorithm, and a head detection algorithm. The first one has high degree of accuracy for near frontal face without occlusions, while it is not able to detect faces in the distance or occluded faces. For this reason we implemented a customized head detection algorithm. A foreground detection is achieved by subtracting and thresholding a background image estimation and the current frame. The bounding box of the head is obtained from the horizontal and vertical projections of the silhouette mask, by automatic selection of histogram steep rises.

2.2 Coding scheme

For each frame two images are taken: the face image, and the remaining image, that is obtained by replacing the detected ROI with a black box. The face image is encoded blockwise by using an adaptive codebook, which consists of affine transformed blocks taken from the remaining image. Hence the proposed coding system can be seen as a hybrid scheme combining fractal image compression [2] with mean-removed shape-gain vector quantization (MRSG-VQ) [4].

In MRSG-VQ an image block $R \in \mathbb{R}^n$ is approximated as $R = sD + o\mathbf{1}$, where s, o are scalar parameters, representing *scale* and *offset* of an intensity transformation, $\mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$ is a constant intensity block, and $D = (d_1, \ldots, d_n)^T$ is a shape block taken from the VQ codebook. The difference between VQ and fractal coding is that in the first one the codebook blocks are generated from a set of training images, while in the latter an image adaptive codebook is used, which is obtained from the original image itself.In the proposed hybrid coding scheme the block codebook consists of blocks taken from the remaining image, i.e. the original frame with black pixels within the face region.

The proposed encoder includes the following steps:

(i) *Image partitioning.* Face image is partitioned into nonoverlapping *range blocks* R of size $g \times g$. In our experiments we fixed $g = 8$ pixels.

(ii) *Domain codebook design.* The remaining image is partitioned into nonoverlapping $2g \times 2g$ square blocks.

Figure 1: Decoded image using no encryption (left), 6/8 masked bits (centre), 8/8 masked bits (right).

	Evening	Morning	Occlusions
ROI Accuracy	0.71	0.57	0.19
Body tracking	0.98	0.95	0.84
SSIM	0.95	0.92	0.92
PSNR	37.36	29.21	32.49

Table 1: Objective evaluation results

The domain pool is obtained applying the 8 canonical isometries to each block, and then pixel averaging by a factor of two. This produces the codebook of *domain blocks* D_i.

(iii) *Least squares search.* For each range block R the best approximation $R \approx sD + o\mathbf{1}$ is searched. Thus, for all the domain blocks D_i, the optimal scale and offset parameters are computed, by minimizing the least square approximation error: $E(D_i, R) = \min_{s,o} \|R - (sD + o\mathbf{1})\|$. The real optimal coefficients s, o are then quantized using a uniform scalar quantization. Finally the codebook block D_k which gives the smallest error $E(D_k, R)$ is selected. The output code for the current range block R consists in the index k of the most similar domain block D_k, and the indices s, o for the scale and offset coefficients.

Thus, the final output code is nothing other than a description of a blockwise image affine transformation. At the decoding step, an approximation of the original face image is obtained from its code and the remaining image, by computing the transformation defined during the coding stage.

2.3 Bit masking

Our access control is based on the control of the reconstruction process, that is obtained by partially hiding the value of s through bit masking. In our experiments we have quantized s using 8 bits. If each of these bits is readable, the decoder will reconstruct the face image using the highest resolution (Fig.1 left). On the other extreme, masking all 8 bits leads to an unreadable decoded image (Fig.1 right). Between these two configurations, masking a certain number of bits, starting from LSB up to MSB, leads to intermediate levels of visualization quality (Fig.1 centre).

3. RESULTS

The resulting obscured videos have been evaluated using both objective and subjective procedures.

The objective metrics compare each pair of original and obscured video in terms of: face detection accuracy, human body tracking accuracy, image quality metrics.

Table 1 provides the average objective evaluation values for the three categories of clips analyzed: evening (1), morning (2) and occlusions (3). As we can expect, the accuracy of the detection module heavily decreases in the case of people wearing occlusions, so a more robust face detection module is needed to overcome this kind of challenge. On the other hand performances of human tracking show that body structure and motion information are quite well preserved for all the videos, and similarly the visual impact of the obscuration does not change for the three categories.

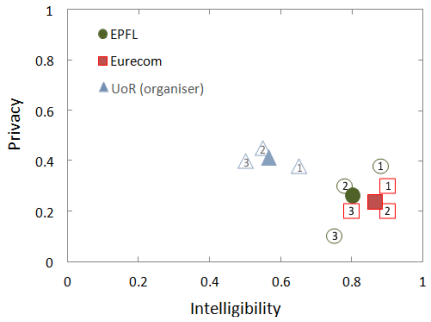

Figure 2: Subjective evaluation results

The subjective evaluation provides a measure of the preservation of privacy of individuals against the intelligibility of activities recorded. The aggregated scores are presented in the graph of Fig.2, for each of the three categories. As a whole, all the three categories show a low degree of privacy, and a quite high level of intelligibility. Anyway, the proposed scrambling provides a multi-resolution access control, that can be adjusted according to the system requirements. For all our experiments 6/8 bits were masked, therefore increasing this parameter can lead to a higher degree of obfuscation.

4. CONCLUSIONS

In this paper we propose a reversible region-based scrambling scheme that provides a multi-resolution access control facility. This access control allows us to calibrate the intelligibility-privacy tradeoff according to the system requirements. To encode face region we exploit self-similarities extracted from the background, so both the original code and background frame image are needed in order to decode the obsured face region.

5. REFERENCES

[1] F. Dufaux and T. Ebrahimi. Video surveillance using JPEG 2000. In *Proceedings of the SPIE*, volume 5588, pages 268–275, 2004.

[2] A. E. Jacquin. Image coding based on a fractal theory of iterated contractive image transformations. *Trans. Img. Proc.*, 1(1):18–30, Jan. 1992.

[3] E. M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Trans. on Knowl. and Data Eng.*, 17(2):232–243, Feb. 2005.

[4] T. Ramstad and S. Lepsoy. Block-based attractor coding: potential and comparison to vector quantization. In *Signals, Systems and Computers, Record of the 27th Asilomar Conference on*, nov 1993.

[5] S. Roche, J.-L. Dugelay, and R. Molva. Multi-resolution access control algorithm based on fractal coding, 1996.

[6] C. Velardo, C. Araimo, and J. Dugelay. Synthetic and privacy-preserving visualization of video sensor network outputs. In *Distributed Smart Cameras (ICDSC), 5th ACM/IEEE International Conference on*, 2011.