ROBUST HUMAN FACE HIDING ENSURING PRIVACY

Isabel Martínez-Ponte, Xavier Desurmont, Jerome Meessen and Jean-François Delaigle

{martinisa, desurmont, jerome.meessen, delaigle}@multitel.be
Multitel A.S.B.L, Avenue Copernic 1, B-7000 Mons, Belgium.

ABSTRACT

Nowadays, video surveillance of people must ensure privacy. In this paper, we propose a seamless solution to that problem by masking faces in video sequences, which keeps people anonymous. The system consists of two modules. First, an analysis module identifies and follows regions of interest (ROI) where faces are detected. Second, a Motion JPEG 2000 encoding module compresses the frames keeping the ROI in a separate data layer, so that the lossless rendering of human faces can be restricted.

The analysis module combines two complementary methods: face detection, to locate faces in the image, and tracking, to follow them seamlessly along the time. The fusion of these two methods increases robustness: once a face has already been detected in a frame, tracking may locate it in the consecutive frames, even where a face detection algorithm would not. In addition, detection of faces prevents tracking from losing its targets.

The encoding module downshifts the JPEG 2000 data corresponding to the identified ROI to the lowest quality layer of the codestream. When the transmission bandwidth is limited, the human faces are then decoded with a lower visual quality, up to invisibility when required.

1. INTRODUCTION

As the number of video surveillance systems increases, ensuring privacy gets more and more importance. The solution we present is based on face masking and consists of two steps: first, an analysis module locates Regions Of Interest (ROI) where faces are detected; second, the JPEG 2000 frame encoding module isolates these ROI at the end of the codestream, ensuring poor visual quality when the bandwidth is limited and allowing a restricted decoding of the human faces. The proposed work flow is shown on Figure 1.

Locating faces in images is a complex problem, because of several factors: occlusions, variable face orientation, facial expressions, presence of facial features such as beards or moustaches... An example of face detection used for video surveillance is described in [4] where scrambling is used to guarantee privacy. However, using face detection only is not reliable enough. In order to ensure robustness along a video sequence, our analysis module combines two complementary methods: face detection and tracking. Another cooperative approach between spatial detection and temporal tracking is described in [1].

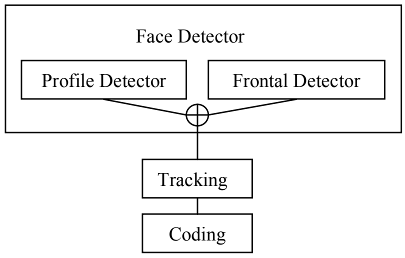

Figure 1. System configuration

JPEG 2000, the new standard for image compression, offers a new flexibility to the coded image [6][12]. In particular, JPEG 2000 allows differentiating regions of interest from their background, by providing them with higher or lower visual quality. Different approaches have been proposed for coding JPEG 2000 ROI [14][2][13]. In this paper, we rely on a codestream organization driven by the ROI identification. The ROI related code-blocks are isolated in the lowest quality layer, which has two consequences: first, when transmission bandwidth is limited, the ROI will be rendered with poor visual quality and, second, the full decoding of the ROI can be restricted to particular users. Compared to the ROI protection used in [4], our system does not need any scrambling or encryption tool. Note that here, the ROI must be hidden rather than coded with higher quality as in traditional definition of ROI. However, this does not affect the compliance of the proposed system with the Motion JPEG 2000 standard.

The paper is organized as follows: in section 2 we explain the face detection and tracking algorithms used. The encoding strategy is described in section 3. Section 4 concludes our work.

2. ROBUST FACE DETECTION

2.1. Overview

Many techniques have been proposed for face detection [16]. The face detection method we use follows an appearance-based approach and is based on the implementation from OpenCV [11] of the classifier presented in [15] and [9]. This system consists of a cascade of boosted classifiers working with Haar-like features, i.e. rectangular features which are a reminiscent of Haar Basis functions and can be computed in any scale and location in constant time. This kind of classifiers shows very good performances, see section 2.2, but is very specific. A single cascade classifier is indeed overstrained to accommodate all in-class variability. So, classifiers trained for frontal faces can not be used to detect profile faces since the Haar features associated are not the same. Several solutions were proposed in the literature to extend the previous approach to multi-view faces. In [7] a decision tree is trained to determine the viewpoint class before applying the appropriate detector on that viewpoint. In [8], a FloatBost learning method with pyramid architecture is used to deal with different kind of rotations faces may present.

2.2. Algorithm evaluation

One profile face detector and four frontal face detectors are available in OpenCV. The frontal face detectors are the following:

1. stump-based 24x24 discrete adaboost,
2. stump-based 20x20 gentle adaboost,
3. tree-based 20x20 gentle adaboost,
4. stump-based 20x20 gentle adaboost (a tree of stage classifiers instead of a cascade).

We have evaluated these detectors, by carrying performance tests on a set of 160 images with 540 frontal faces. Some of these images are obtained from [3] while the others are extracted from own video surveillance sequences. Two qualitative error rates (detection rate [5] and absolute false alarm rate) are used to compare the detectors.

cascade	detection rate	abs. false alarm rate
1	0.86	108
2	0.87	60
3	0.90	67
4	0.83	31

Table 1 Comparison of OpenCV frontal face detectors

As shown on Table 1 frontal faces were generally well located when the area was not too small and no in-plane rotations occurred. Some faces were detected twice and several false alarms occurred. As expected, profile faces were not detected.

The performance of the first detector is a little lower than with the other detectors. We can see that, as shown on [9], gentle adaboost outperforms discrete adaboost and an input pattern size of 20x20 provides better results. The last three classifiers show a compromise between detection rate and false alarm rate. We can appreciate that, according to [9], tree slightly outperforms stumps as weak classifiers since stumps do not allow learning dependencies between features. So, we decided to use cascade 3 in our system.

The profile face detector located most of the profile faces that were not detected with the previous classifiers as well as some frontal faces. However, some profile faces remained not detected. This is due to the great variability of the profile faces in terms of in-plane and out-of-plane rotations.

2.3. Proposed Face Detection Approach

In our system (see Figure 1), we propose to use in parallel two of the cascades available in OpenCV, one for frontal faces (cascade 3) and one for profile faces, in order to handle multi-view. Combining a frontal and a profile face detector increases detection rate, but also false detection rate as expected. Tracking will allow us to reject most of the remaining false alarms and to follow the detected faces along the time even if they are not detected in all the frames.

2.4. Tracking

Regarding the tracking algorithm, two different approaches were tested: ‘bottom-up’, where the tracking algorithm tries to match along the time the faces located by the detection algorithm, and ‘top-down’, where tracks are initialised by the detection algorithm and the tracking tries to locate along the time certain features of the detected ROI.

Results obtained with the bottom-up approach were not satisfactory because the detected faces were temporally too distant from each other to be matched. The top-down approach illustrated on Figure 2 is then preferred. Once a face is detected, we calculate its features and estimate its position in the following frame by first order prediction. From this estimation, we search its new position and update the associated features.

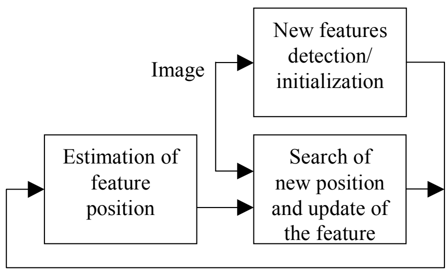

Figure 2. Top-down approach for face tracking

2.5. Merging Face Detection and Tracking

The detection algorithm works along in time with the tracking algorithm in order to detect the ROI and certify they actually correspond to human faces.

The same face is usually detected several times. However, these detections must not lead to new tracks, but have rather to be considered as new occurrences of the same object. The tracking algorithm merges them if the distance between their associated features is inferior to an established threshold.

Detected objects whose position did not change for a fixed number of frames were discarded. Indeed, in the applications of interest, people do not remain immobile for a long time. This constraint reduces false detections.

When false detections occur for moving objects, these "false faces" are initially tracked by the algorithm. We have decided to give an initial belief ($b_{t0}=\beta$) to the detected face which decreases with time:

$$(1)\begin{cases} b_{t+1} = \max(b_t + \delta, \gamma_1) \text{ if face is detected} \\ b_{t+1} = \alpha b_t, \alpha < 1 \text{ if face is not detected and } b_t < \gamma_1 \end{cases}$$

where b_t represents the evolution of the belief along the time. If $b_{t+1} < \gamma_2$, the face is discarded. Figure 3 illustrates the behaviour of the proposed algorithm. We consider two faces that are detected at time $t=t0$ and $t=t1$. The first one is again detected at $t=t2$, $t=t3$, $t=t4$ and $t=t5$, while the second is not re-detected and so it is considered to be a false detection. Tracks become darker when their belief is smaller than a third threshold, γ_3, in order to show that the detection may be false. The values of γ_1 and γ_2 were set to 1 and 0.5 respectively and then the other parameters were set experimentally.

The results obtained by combining face detection and tracking are significantly better than the ones obtained using face detection only. On the one hand, tracking locates faces in frames where the detection algorithm would not. On the other hand, face detection finds new faces and prevents tracking from losing its targets. As a consequence, detection rate increases while false detection rate decreases.

3. JPEG 2000 frame encoding

The JPEG 2000 image compression framework allows defining quality layers. Each of them contains data from pre-defined spatial elements, which are corresponding to a entire number of JPEG 2000 code-blocks in the wavelet domain. So, in the following, we refer to code-blocks as the data chunks associated to spatial regions. Usually, given a target bit rate, a quality layer consists of the code-blocks contributions maximizing the PSNR quality improvement of the decoded image when this layer is added to the preceding layers, see e.g. Figure 4. However, the code-blocks contribution selection can also be achieved based on the definition of regions of interest.

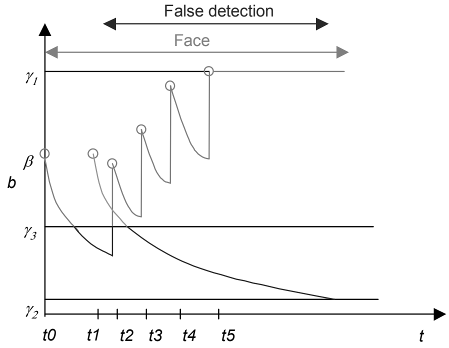

Figure 3. Behaviour of the detection and tracking algorithm when the object detected is a face and when it is a false detection.

In section 2, we have shown how the analysis module identifies human faces that must be hidden through encoding. So, we decided to isolate the data from the ROI in a separate quality layer as shown on Figure 5. This particular layer is lower than every other quality layer in the coded frame. As a consequence, if the viewer does not decode this last layer, the human faces will be invisible as on Figure 6. Privacy is then ensured. Moreover, in case of limited bandwidth, the ROI will be rendered with a poorer quality compared to the background, see e.g. Figure 7. Let's note that this particular codestream ordering can be achieved online, through transcoding.

The spatial precision of the downshifted ROI depends on the code-block size. The larger the code-blocks, the larger the hidden ROI, but the higher the compression ratio.

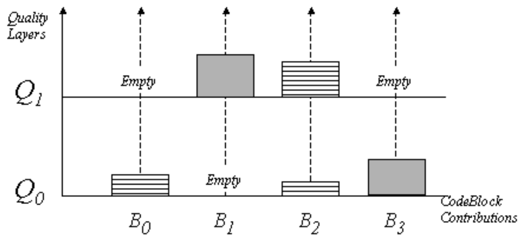

Figure 4. JPEG 2000 Quality layers

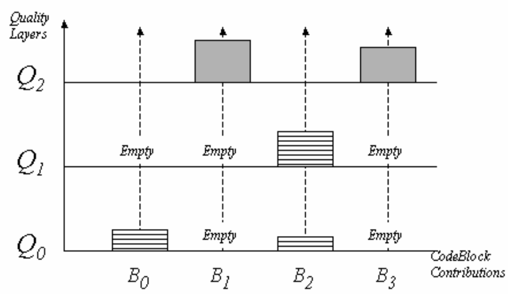

Figure 5. JPEG 2000 quality layers with isolated ROI

Figure 6. Human faces hiding: only the first JPEG 2000 quality layer is decoded.

Figure 7. Human Face blurring: the second quality layer is partially decoded.

4. CONCLUSIONS

We have presented a robust system to mask faces in video surveillance sequences in order to ensure privacy.

First, the analysis module locates faces and follows them seamlessly along the time by the combination of face detection and tracking. These two methods complement each other to increase robustness: once a face has already been detected in a frame, tracking may loate it in frames where a face detection algorithm would not. In addition, the detection method prevents tracking from loosing its targets. The results obtained by the analysis module are satisfactory, especially in sequences were we can distinguish frontal up-right faces as the detection algorithm locates them accurately, which allows a good tracking afterwards.

Second, a Motion JPEG 2000 encoding module isolates, within each frame, the data corresponding to the human faces in the lowest quality layer of the codestream. This ensures poor visual quality for these ROI in lossy compression, up to invisibility if required. The produced video stream is fully compliant with the Motion JPEG 2000 standard.

In the near future, the analysis module will be improved to detect faces in a wider variety of poses.

9. ACKNOWLEDGEMENT

This work has been partially funded by the European Commission, within the scope of IST WCAM project (IST-507204.)

10. REFERENCES

[1] J.R. Casas, A. Puig, P. Puig, "Spatial-temporal Video Analysis for Improved Pedestrian Detection", *5th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS-04)*, Lisbon, Portugal , April 2004.

[2] C. Christopoulos, J. Askelf, M. Larsson, "Efficient method for encoding regions of interest in the upcoming JPEG 2000 still image compression standard", *IEEE Signal Processing Letters*, pp. 247-249, September 2000.

[3] CMU/VASC image database http://vasc.ri.cmu.edu/idb/html/face/index.html

[4] F. Dufaux, T. Ebrahimi, Emitall SA, "Smart video Surveillance System Preserving Privacy", *Proc. of SPIE - Image and Video Communication and Processing*, San Jose, USA, January 2005.

[5] T. Ellis, "Performance Metrics and Methods for Tracking in Surveillance", *Proc. of the Third IEEE Workshop on PETS*, pp. 26-31, June 2002.

[6] ISO/IEC 15444-1/ IUT-T T.800, "JPEG2000 Image Coding System - Part 1: Core Coding System", 2000.

[7] M. Jones and P. Viola, "Fast Multi-view Face Detection", *IEEE Conf. on Computer Vision and Pattern Recognition*, June 2003.

[8] S. Li, L. Zhu, Z. Zhang and H. Zhang, "FloatBoost Learning and Statistical Face Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, n. 9, September 2004.

[9] R. Lienhart, A. Kuranov and V. Pisarevsky, "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection", *Proc. of 25th German Pattern Recognition Symposium*, September 2003.

[10] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection", *IEEE ICIP 2002*, vol. 1, pp. 900-903, September 2002.

[11] Open Computer Vision Library http:/sourceforge.net/projects/opencvlibrary/

[12] M. Rabbani, R. Joshi, "An overview of the JPEG2000 still image compression standard", *Signal Processing Image Communication*, Eurasip, vol. 17, n. 1, pp. 3-48, January 2002.

[13] V. Sanchez, A. Basu and M. Mandal, "Prioritized Region of Interest Coding in JPEG2000", *IEEE Trans. on CSVT*, vol. 14 (9), pp. 1149-1155, September 2004.

[14] D. Santa-Cruz, T. Ebrahimi, "An analytical study of JPEG2000 functionalities", *ICIP2000*, Vancouver, September 2000.

[15] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.

[16] M. Yang, D. Kriegman and N. Ahuja, "Detecting faces in images: A survey", *IEEE Transactions on Pattern analysis and Machine intelligence*, vol. 24, n. 1, pp. 34-58, January 2002.