

Scene Analysis for Reducing Motion JPEG 2000 Video Surveillance Delivery Bandwidth and Complexity

Jérôme Meessen, Christophe Parisot, Xavier Desurmont and Jean-François Delaigle

Image Processing Department
Multitel A.S.B.L.
Mons, Belgium
jerome.meessen@multitel.be

Abstract— In this paper, we propose a new object-based video coding/transmission system using the emerging Motion JPEG 2000 standard [1] for the efficient storage and delivery of video surveillance over low bandwidth channels. Some recent papers deal with JPEG 2000 coding/transmission based on the Region Of Interest (ROI) feature and the multi-layer capability provided by this coding system [2][3]. Those approaches allow delivering more quality for mobile objects (or ROI) than for the background when bandwidth is too narrow for a sufficient video quality. The method proposed here provides the same features while significantly improving the average bitrate/quality ratio of delivered video when cameras are static. We transmit only ROIs of each frame as well as an automatic estimation of the background at a lower frame rate in two separate Motion JPEG 2000 streams. The frames are then reconstructed at the client side without the need of other external data. Our method provides both better video quality and reduced client CPU usage with negligible storage overhead. Video surveillance streams stored on the server are fully compliant with existing Motion JPEG 2000 decoders.

Keywords – *segmentation; object-based coding; Motion JPEG 2000; video surveillance.*

I. INTRODUCTION

In 2005 the market of digital video cameras should exceed the one of analog video cameras. Video surveillance systems now use IP networks and digital video coding standards to transmit the huge amount of video data from cameras or storage servers to control rooms and sometimes to mobile surveillance agents with wireless PDA interactivity. While widely tested codecs such as MJPEG and MPEG2 are the most used, state of the art MPEG4/H26L[4] and Motion JPEG 2000 (MJ2)[5] codecs are now integrated in new video surveillance devices and systems.

When they are compared to MPEG-based systems, the MJ2 video surveillance ones can take advantage of the JPEG 2000 unequalled amount of features [6]. This standard provides Regions Of Interest (ROIs) definition, as well as spatial, component, resolution and quality scalability [7][8]. The bitstream can be easily parsed and adapted in real-time in each of these scalabilities without the need of decoding frames. This enables the optimisation of video compression/transmission according to client decoding capabilities as well as the varying network load with a minimum impact on the storage server processing requirements. Furthermore, Motion JPEG 2000 provides state of the art compression efficiency that can be exploited in countries where inter-frame coding techniques are not recognized by courts as admissible evidences [9].

Figure 1 presents our approach for video surveillance delivery based on automatic background extraction and ROI segmentation followed by MJ2 coding. Our system manages two Motion JPEG 2000 streams stored by the video surveillance data server: the first one contains the MJ2 encoded video sequence with the ROIs and the background of each frame, separated into two quality layers. The second stream only contains the estimated average background either at a low fixed frame-rate or when it changes. Typically, we assume that it is never useful to update the background more than once every 4 seconds.

In section II, we present the automatic background extraction/video segmentation technique we use for unsupervised ROI detection. Section III describes the video delivery architecture at both the server and the client sides and discusses system complexity. Section IV deals briefly with network/client dynamic adaptation of the bitrate and our experimental results are shown in section V before the conclusion in section VI.

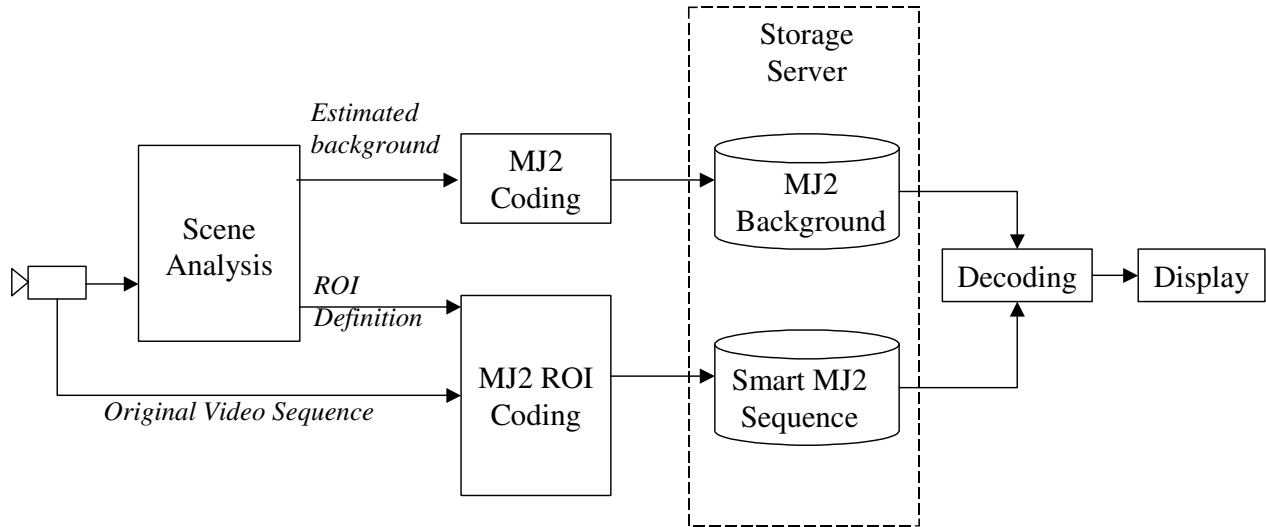


Figure 1: Proposed architecture for ROI based smart video surveillance delivery.

II. FOREGROUND/BACGROUND AUTOMATIC SEPARATION

The role of the scene analysis is to detect and track ROIs of the video stream in order to provide information to the smart encoder. It automatically computes an estimate of the scene background based on past frames and extracts mobile objects. Final ROIs are mobile objects matching the decision characteristics (size, position, texture...) and/or the behaviour of controlled elements (people, cars...).

The analysis is based on a real-time statistical segmentation algorithm using a mixture of Gaussians modelling for the background luminance of each pixel [10] [11]. The main advantage is that it automatically supports backgrounds having multiple states like blinking lights, grass and trees moving in the wind, acquisition noise etc. Furthermore, the background model update is done in an unsupervised way when the scene conditions are changing.

Figure 2 shows the mixture of Gaussians for one pixel at a given time. The new pixel luminance, i.e. in the new frame, is compared to the current mixture. If this luminance belongs to one of the most probable Gaussians, the pixel is then classified as background and the relevant Gaussian parameters (i.e. mean, variance, frequency) are updated. Otherwise, the pixel is classified as foreground and the parameters of the less probable Gaussian are updated according to this new luminance value. At the beginning, a new Gaussian is initialised each time a pixel is classified as foreground until the maximum number of Gaussians is reached.

At any time, a current estimated background frame can be constructed. It just requires getting the mean of the most probable Gaussian for each pixel. Such estimated background frames are less sensitive to noise than the original frames. In the proposed system, they are generated either at a fixed low frame-rate or only when major background changes are detected, so as to optimise the video surveillance delivery.

III. OBJECT-BASED MOTION JPEG 2000 ENCODING AND DECODING

We propose to use two compliant Motion JPEG 2000 (MJ2) streams. Both are encoded exploiting the analysis presented in the previous section. These streams are encoded and saved with high quality since they could be used further as evidences if necessary. As described in section 0, data delivered to the control room or wireless PDA clients consist in a truncation of both MJ2 streams.

A. Encoding

The first stream is obtained by MJ2 encoding the original video sequence using ROIs corresponding to the relevant objects detected in the frames. We use the segmentation results as the ROI definition for each frame. Different methods exist for coding ROIs in JPEG 2000: shifting the wavelet samples [6], re-organizing the codestream data as done in [2] or combining both of these approaches as in our previous work [3]. In our case, we choose to organize the codestream of each frame with different quality layers. For the sake of simplicity, we consider here only two layers with the first one dedicated to ROI data. The method can be easily extended to n layers for ROIs and m layers for remaining data, n and m being fixed parameters of the system.

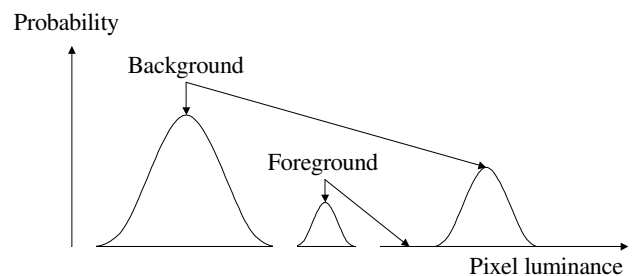


Figure 2: Statistical background modelling of a pixel using three Gaussians.

As an example, we suggest the first layer to include a lossless representation of ROIs. Decoding only this layer will then result in lossless visualizing the ROIs. Truncating the codestream within this layer allows us to adapt the decoded ROI quality to bandwidth constraints. The second layer allows recovering the full original video surveillance sequence if necessary.

The second MJ2 stream is created by MJ2 encoding the background estimated by the scene analysis. Let us suppose here that this stream contains one single quality layer including the lossless compressed background. This background estimation must be updated rarely, i.e. either on demand or when necessary, e.g. because of a significant illumination change. We assume the background to be encoded at most once every 4 seconds.

Assuming that both MJ2 streams are coded with the same quality, the overhead then equals to add/encode less than one background frame data every hundred sequence frames at 25 fps. The precise storage overhead ratio is:

$$\frac{\text{background_frame_rate}}{\text{sequence_frame_rate}} \% \quad (1)$$

Therefore, in typical background update rates, the storage/encoding overhead will never exceed 1%.

B. Decoding

As described in [13], the powerful scalability of JPEG 2000 allows merging images in the coded domain. In particular, it is easy to replace an image region by data from another image, as far as the coding parameters, i.e. number of wavelet decompositions, tiles size, code-blocks size etc., are identical for both images.

We exploit this feature at the video receiver side to merge the ROIs stream with the low rate background stream. In practice, for each received background frame, we perform the arithmetic decoding of the wavelet samples without applying the inverse DWT. When decoding the ROIs stream, if a code-block has not been decoded, i.e. no data were received for this region in the sub-band, we simply copy the background wavelet samples located at the same code-block position, as depicted in Figure 3.

Assuming that the background extraction/segmentation is correct, the wavelet coefficients of the estimated background frames are very similar to the actual background coefficients in the current frame. Therefore, the method doesn't produce undesirable artefacts at code-blocks borders. Furthermore, the segmentation mask propagation used to determine which wavelet coefficients belong to the ROI in each sub-band [12] ensures that ROI pixels reconstruction only depends on data of the ROI stream. This guarantees a perfect JPEG 2000 integrity of the reconstructed ROI pixels.

To evaluate the transmitted bitrate and decoding complexity, we assume that both the ROI and the background must have the same quality. In the worst case, ROIs always cover the whole frames. Thus, a background frame has been unnecessarily transmitted and decoded. The bitrate/decoding

overhead is then of one frame at each background update, which is identical to the storage overhead ratio (1).

Therefore, in typical background update rates, both the transmission bitrate and decoding complexity overhead are less than 1%.

On the contrary, if no ROIs are detected in the frames, only the background bitrate and decoding complexity must be taken into account, i.e. only one background frame to be transmitted and decoded for a set of sequence frames. The bitrate/complexity reduction factor reaches:

$$\frac{\text{sequence_frame_rate}}{\text{background_frame_rate}} \quad (2)$$

As an example, this represents a bitrate and decoding complexity reduction factor of 100 for a 25 fps sequence with background update rate of 0.25 fps. When there are very few and very small ROIs in the sequence, the complexity is approximately reduced to the one of the IDWT which represents a minor part of the whole decoding complexity [14].

In most cases, the detected ROIs are not numerous and/or are small compared to the frame size. The gain in complexity is then remarkably high, in terms of transmission rate and decoding complexity. In the worst case, the overhead is negligible.

IV. TERMINAL ADAPTIVE VIDEO DELIVERY

Based on the available bandwidth, a bitrate must be allocated for both the objects of interest (ROIs) and background streams. One can equivalently choose the maximum bitrate for each MJ2 stream or fix the ratio $\text{background stream rate} / \text{ROI stream rate}$. Let us assume that the ROIs and the background have comparable textures. If the bitrate target is identical for an individual background frame and an individual ROI data frame, ROIs quality vary from the background quality to lossless. With this very basic strategy, we ensure ROIs to have a better quality than the background. The smaller the ROIs size, the better their quality compared to background.

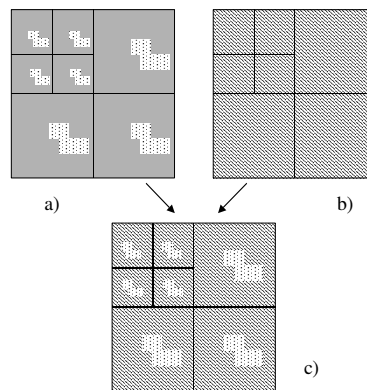


Figure 3: Fusion of background (b) and ROIs (a) in the wavelet/code-block domain.

V. RESULTS

The proposed system has been tested on the MJ2 *speedway* CIF reference sequence at 25 fps. A bitrate of 634 kbps has been allocated to the MJ2 ROI stream and 26 kbps for the MJ2 background stream. Figure 4 presents the results for frame 197. The PSNR we obtain for the whole 200 frames sequence is 34.7 dB while it is 30.6 dB for standard MJ2 coding. Thus, we get a gain of 4.1 dB for the sequence and 4.6 dB for frame 197. Notice that traditional MJ2 ROI coding methods act on the relative quality of the foreground with respect to the background while our system allows optimising both at the same time. It is worth noting that the processing cost of the automatic segmentation is small compared to the JPEG 2000 encoding. It requires only 15% of the CPU on a Dell desktop 3GHz machine for this sequence. Moreover, the results of the automatic scene analysis can be directly exploited for intelligent video surveillance.

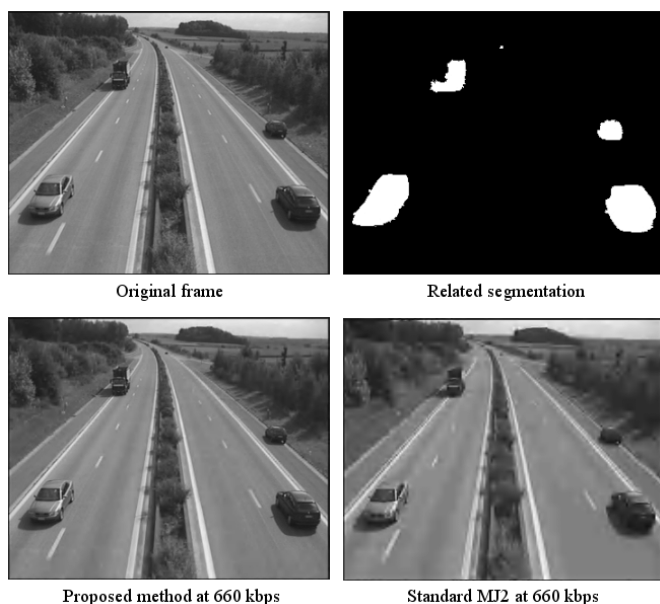


Figure 4: Comparison of the proposed approach with respect to standard MJ2 coding. Respective PSNRs are 34.4 dB and 29.8 dB.

VI. CONCLUSION

We proposed a new delivery strategy for video surveillance exploiting the powerful scalability of the Motion JPEG 2000 standard. The presented solution consists in distinguishing the objects of interests from their background through unsupervised statistical scene analysis. Both the ROI and the background are encoded in separate Motion JPEG 2000 streams for delivery. The receiving client will merge the frames in the wavelet/code-block domain when necessary. Since the background must be transmitted only when important state variations are detected or on demand, this method

strongly reduces the average transmitted bitrate while preserving a high visual quality after decoding.

Moreover, we have shown how the processing overhead in terms of encoding and storage is negligible. The maximum overhead that could exist in the worst case (ROIs always covering the whole frames) is also negligible in terms of bandwidth usage and decoding CPU requirements. When compared to traditional MJ2 video surveillance delivery, our method presents remarkable bandwidth and decoding complexity reduction since there are very few video surveillance applications where the full video is crowded whatever the daytime is.

ACKNOWLEDGMENT

This work has been funded by the EU commission in the scope of the FP6 IST-2003-507204 project WCAM [15].

REFERENCES

- [1]. ISO/IEC 15444-3 "JPEG 2000 part 3: Motion JPEG 2000"
- [2]. V. Sanchez, A. Basu, M. Mandal, "Prioritized region of interest coding in JPEG2000," IEEE trans. on Circuits and Systems for Video Technology, vol. 14(9), pp. 1149-1155, September 2004.
- [3]. J. Meessen, C. Parisot, C. Le Barz, D. Nicholson and J.-F. Delaigle, "WCAM: smart encoding for wireless surveillance," Proc. of SPIE Image and Video Communications and Processing (IVCP), San Jose, USA, January 2005.
- [4]. Provideo surveillance: http://www.provideo.com.tw/index_S.htm
- [5]. Sanyo VCC - WB 2000, <http://www.sanyocctv.com>.
- [6]. ISO/IEC 15444-1/ IUT-T T.800, "JPEG 2000 Image Coding System - Part 1 : core coding system," 2000.
- [7]. M. Rabbani and R. Joshi, "An overview of the JPEG2000 still image compression standard," Signal Processing Image Communication, Eurasp, vol.17(1), pp. 3-48, January 2002.
- [8]. D. Santa-Cruz, T. Ebrahimi, "An analytical study of JPEG2000 functionalities," proc. of IEEE International Conference on Image Processing (ICIP), Vancouver, September 2000.
- [9]. Avid Technology, Forensic Video Decisions, 15th May, 2001.
- [10]. C. Stauffer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.
- [11]. X. Desurmont, C. Chaudy, A. Bastide, C. Parisot, J.F. Delaigle, B. Macq, "Image analysis architectures and techniques for intelligent systems," IEE proc. on Vision, Image and Signal Processing, Special issue on Intelligent Distributed Surveillance Systems, vol. 152(2), pp. 224-231, April 2005.
- [12]. C.Christopoulos, J. Askelöf and M. Larsson, "Efficient method for encoding ROI in the upcoming JPEG 2000 still image compression standard," IEEE Signal Processing Letters, vol. 7(9), pp. 247-249, September 2000.
- [13]. R. Rosenbaum and D. Taubman, "Merging images in the JPEG 2000 domain," Proc. of Visualization, Imaging and Image Processing (VIIP), vol. 1, Benalmadena, Spain, pp. 249-252, September 2003.
- [14]. C. Le Barz, R. Elmostadi and D. Nicholson., "Real time implementation of JPEG 2000," proc. of European Signal Processing Conference (EUSIPCO), Toulouse, France, September 2002.
- [15]. FP6 IST-2003-507204 WCAM, "Wireless Cameras and Audio-Visual Seamless Networking," <http://www.ist-wcam.org>