

WCAM: Smart Encoding for Wireless Surveillance

J. Meessen^a, C. Parisot^a, C. Le Barz^b, D. Nicholson^b and J.F. Delaigle^a

^a Multitel asbl, Mons, Belgium
jerome.meessen@multitel.be

^b Thales Communications, Colombes, France

ABSTRACT

In this paper, we present an integrated system for smart encoding in video surveillance. This system, developed within the European IST WCAM project, aims at defining an optimized JPEG 2000 codestream organization directly based on the semantic content of the video surveillance analysis module. The proposed system produces a fully compliant Motion JPEG 2000 stream that contains regions of interest (typically mobile objects) data in a separate layer than regions of less interest (e.g. static background). First the system performs a real-time unsupervised segmentation of mobiles in each frame of the video. The smart encoding module uses these regions of interest maps in order to construct a Motion JPEG 2000 codestream that allows an optimized rendering of the video surveillance stream in low bandwidth wireless applications, allocating more quality to mobiles than for the background. Our integrated system improves the coding representation of the video content without data overhead. It can also be used in applications requiring selective scrambling of regions of interest as well as for any other application dealing with regions of interest.

1. INTRODUCTION

Today, wireless transmission of video is one of the main issues of surveillance. Regarding the limited transmission bandwidth provided by this type of channel, video compression is then unavoidable, with the drawback of loss of quality after rendering the received data. One common approach addressing this issue is the rate adaptive coding, where the compression ratio is automatically modified based on feedback information about the available bandwidth. Modifying the global rate will affect instantaneously the whole quality of the transmitted data. However, in the case of video surveillance, this is sub-optimal for the end user. In this type of application, some objects, i.e. the foreground, are indeed much more important compared to their background containing less relevant information. The foreground objects must be visualized with a high quality after decoding, even though the channel bandwidth is narrow. As a consequence, content detection must be included in the video coding strategy in order to meet these requirements.

More than using only bandwidth information for adapting the video compression, this paper proposes a new approach based on smart coding. Here, real-time automatic scene analysis is exploited to extract the semantically relevant objects that will be favored during the video encoding process. This is achieved thanks to the region of interest (ROI) framework provided by the Motion JPEG 2000 standard.

This paper is organized as follows. Section 2 introduces the European IST WCAM project wherein this study has been realized, and, in section 3, we describe the chosen scene analysis method. The flexibility of the JPEG 2000 coding standard is briefly described in section 4 as well as the Motion JPEG 2000 streaming method implemented by the WCAM project. Region of interest coding is discussed in section 5 before the proposed smart coding strategy and its results are detailed in section 6. Section 7 concludes the paper.

2. WCAM SYSTEM DESCRIPTION

The objective of the WCAM project is to study, develop and validate a wireless, seamless and secured end-to-end networked audio-visual system [1]. This new project, started in January 2004, focuses on the technology convergence between video surveillance and multimedia content distribution over the Internet. Therefore, in this IST project, the video content is encoded in emerging content formats: Motion JPEG 2000 and MPEG-4 AVC/H.264, and transmitted

through Wireless LAN to different types of decoding platforms like PDA's and Set Top Boxes. While robust wireless transmission is taken into account, the video content will also be secured using a Digital Right Management system.

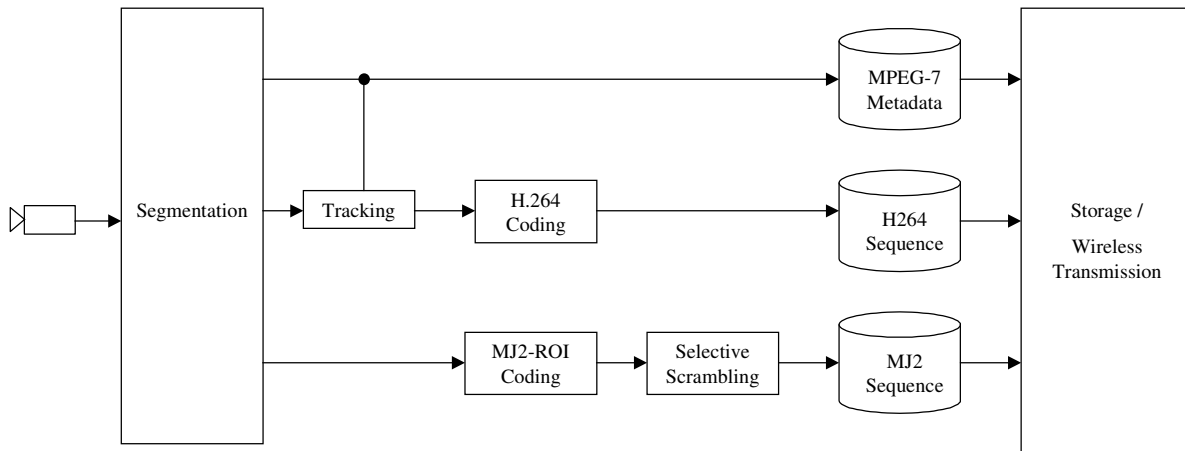


Figure 1. WCAM coding system architecture.

Figure 1 presents the smart coding architecture system developed by WCAM. The automatic scene analysis, i.e. segmentation and tracking, is linked to the video coding guaranteeing good quality for semantically relevant regions and keeping a low average data rate; enhancing the transmission errors concealment and allowing selective protection of the content.

The smart encoding consists in using spatial (segmentation) analysis to produce efficient JPEG 2000 encoding through bitrate allocation and regions of interest definition. From the original video sequence, the relevant objects, i.e. moving objects, are segmented. The segmentation results are used as input ROI masks for the JPEG 2000 encoding of the frame. Regarding the number of wavelet decompositions and code-block size, portions of the background can be included in the ROI, improving the contextual understanding of the scene by the end user. We are currently working on the optimization of the rates associated to the ROI and its background respectively. WCAM will also use image analysis in the same way for helping H.264 motion estimation and rate allocation and for favoring error concealment in the semantically relevant regions of the transmitted video frames. As an example, the temporal information of the tracking can as well be used for predictive encoding. Examples of data that can be exploited by the coding module (either Motion JPEG2000 or H.264) include the masks of the detected objects in each frame, an estimate of the correspondence between objects of two consecutive frames, an estimate of the objects motion in order to reduce the block-matching complexity of H.264.

Thanks to the segmentation results, the Motion JPEG 2000 coded data can also be selectively protected against transmission errors.

The coded video data and its associated MPEG-7 description are stored and/or transmitted through wireless network to the different users. An optional feature will be added for making images anonymous by masking (or scrambling) objects corresponding to persons. When the interpretation of objects is performed (i.e. by segmentation and tracking), the ROI-oriented multi-resolution approach of JPEG 2000 enables differentiated encoding of objects and levels within objects. One can imagine scrambling higher resolution coefficients, or simply removing them, when the image analysis module recognizes that the object is a person or a group of persons.

3. AUTOMATIC SCENE ANALYSIS

The goal of the scene analysis module is to detect and track regions of interest (ROI) of the video stream in order to both generate metadata describing the relevant events for the surveillance application and provide information for the coding module optimization. These data can be used either to optimize the coding process complexity but more than that, to optimize the video stream coding representation (e.g. providing more priority for regions of interest than for the background in wireless low bandwidth transmission conditions).

The automatic ROI extraction module of WCAM is based on a real-time statistical segmentation algorithm detailed hereafter. The algorithm is based on a mixture of Gaussians modeling of the luminance for each pixel of the background [11,12,13]. The main advantage of this technique is that it can automatically deal with backgrounds that have multiple states like cyclic states such as blinking lights, grass and trees moving in the wind, acquisition noise etc. Furthermore, the background model update is done in an unsupervised manner when the scene conditions are changing (lighting conditions in outdoor applications, increase or decrease of the background states number for each pixel independently...).

Let N be the maximum number of Gaussians for each pixel. At the beginning, each pixel mixture is composed of only one Gaussian with mean equal to its luminance in the first frame, standard deviation equal zero and frequency of appearance equal to one. For each new frame, the algorithm is the following:

If the current pixel luminance belongs to one of the Gaussians of the mixture (i.e. $\text{abs}(\text{luminance} - \text{mean}) / \text{standard_deviation} \leq \text{threshold}$), we check whether this Gaussian is one of the most frequent in the mixture. If the Gaussian appears often, the pixel is classified as part of the background, else it belongs to the foreground. Then, we update the Gaussian parameters (mean, standard deviation and frequency of appearance) with the current pixel luminance;

Else, if we have less than N Gaussians in the mixture, we initialize a new Gaussian for this pixel with its luminance for the mean, zero for the standard deviation and one for its frequency of appearance;

Else, if the mixture is composed of N Gaussians and the current pixel luminance has not been classified as background, we update the Gaussian having the lowest frequency of appearance.

Then, common algorithms such as erosion, dilatation, contour closing and labeling are used to get more accurate segmentation masks and high-level description of the objects shape and position in the scene.

Figure 2 shows how pixels are classified as foreground or background according to their luminance value and the current state of the background mixture of Gaussian model. Pixels are classified as background if and only if they belong to one of the most probable Gaussians of the mixture.

Figure 3 shows the difference between a classical frame differencing approach and the proposed one for efficient foreground/background pixel classification. One can observe that the frame differencing approach provides a wrong classification for some of the pixels while the proposed method provides a segmentation that takes into account the multi-modality of the sequence background.

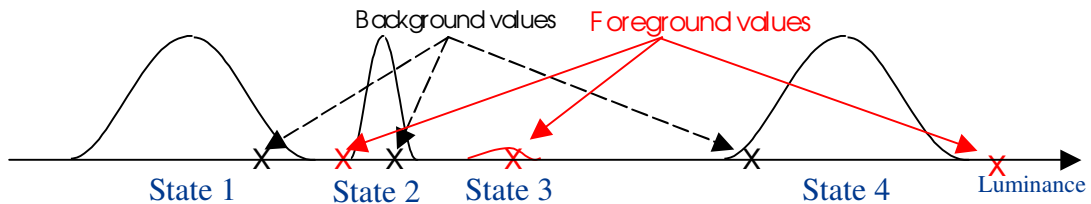


Figure 2. Foreground/Background pixel classification

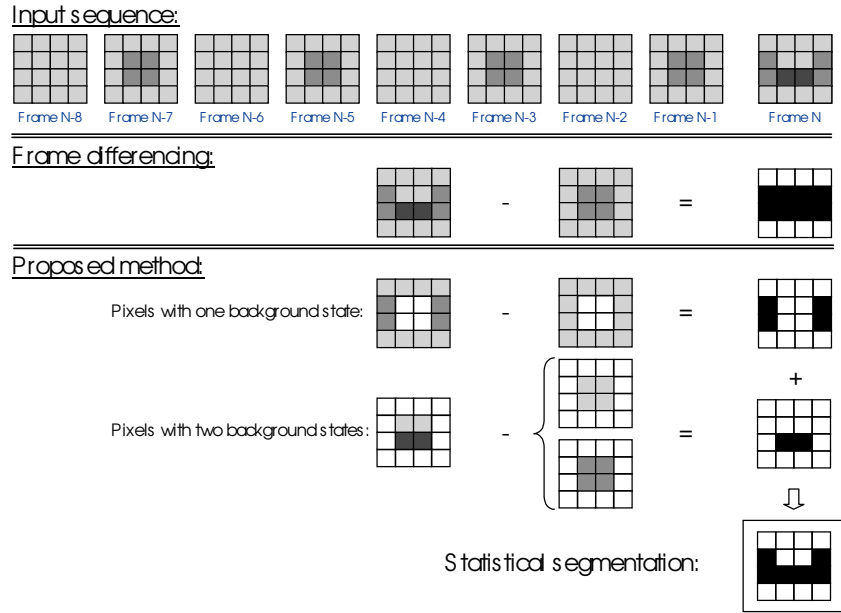


Figure 3. Sample input sequence of size 4x4 and its associated proposed segmentation compared to frame differencing.

The metadata resulting from the automatic scene analysis can then be of different semantic levels. Low level includes for example the size of the ROI, the medium level describes tracking information while the highest level of metadata contains scene description like ‘abandoned luggage’, ‘car on wrong lane’ or ‘alarm’ messages. Metadata storage is compliant with the MPEG-7 description scheme in order to be transmitted to users when necessary (alarm messages) or on demand. These metadata can also be stored for off line video content browsing [4] or retrieval though this is not directly addressed by our project.

4. USE OF JPEG 2000 FOR WIRELESS VIDEO STREAMING

JPEG 2000 [5][6][7] is the most recent from the international standards developed by the Joint Photographic Expert Group (JPEG). JPEG 2000 defines an image compression system that allows great flexibility not only for the compression of images but also for the access in the codestream. A key feature of JPEG2000 is the flexible bit stream representation of the images that allows to access to different representations of images using its scalability features (resolution, quality, position and image component) and the Region of interest (ROI) feature. The JPEG 2000 ROI coding [8] allows giving more quality to parts of the image of importance that can be of arbitrary shape thanks to the standard max-shift method.

A JPEG 2000 compressed image uses markers and marker segments to delimit and signal the compressed information, organised in headers (Main and Tile Parts) and packets. This modular organisation of the bitstream enables progressive data representation, such as quality progressive and resolution progressive. A JPEG 2000 codestream starts always by the Main Header followed by one or several Tile Part Headers, each of them followed by compressed data packets, and ends by an End Of Codestream (EOC), as shown in the following figure:



Figure 4: JPEG 2000 codestream structure

JPEG 2000 allows scalable decoding, i.e. partially decoding the compressed image at a desired bit-rate, image resolution, image region or color component. The following Figure 5 gives an example of scalable decoding in rate, which results in different quality images.

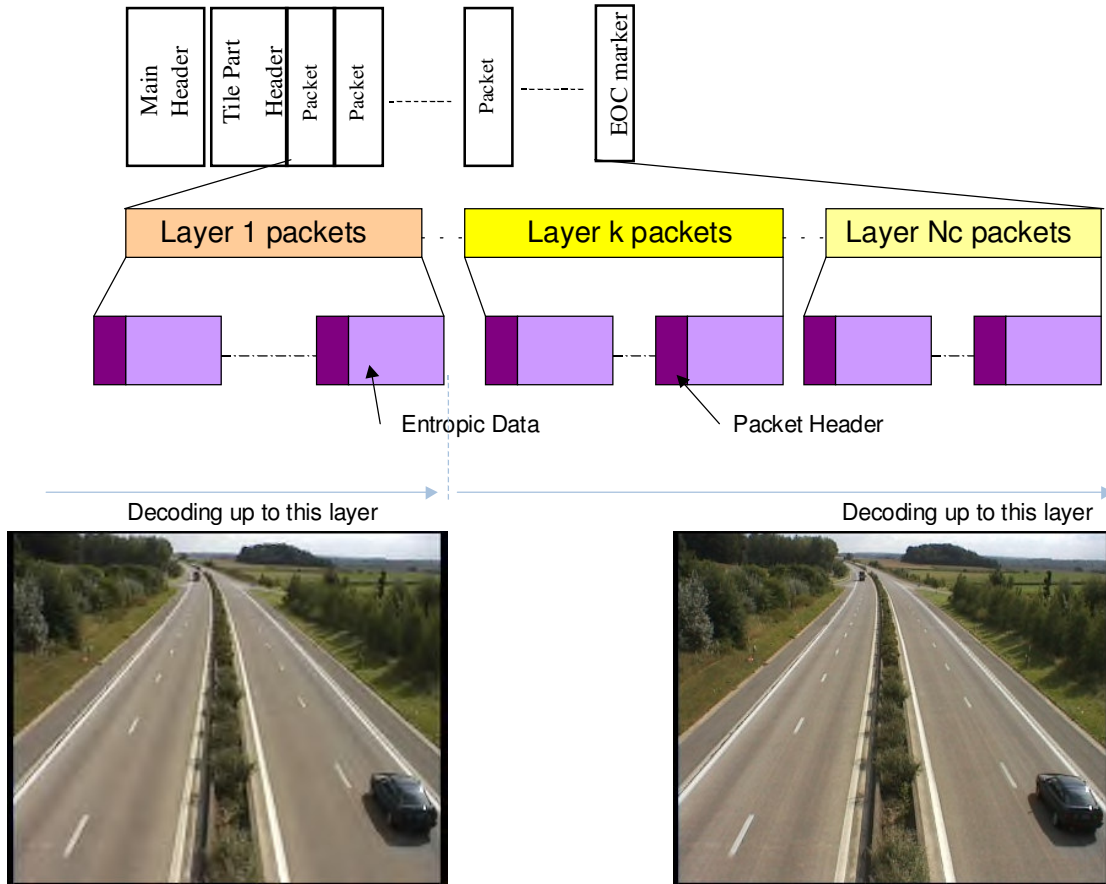


Figure 5: JPEG 2000 scalable decoding example

Part 3 of JPEG 2000, called MotionJPEG 2000 has defined a dedicated file format (MJ2) for storing video images compressed using JPEG 2000, as well as compliance points definitions (see Table 1) as well as test sequences and codestreams (see Table 2).

Parameter	Cpoint-0	Cpoint-1	Cpoint-2	Cpoint-3
Maximum decoded image size (WxH)	360x288	720x576	1920x1080	4096x3112
Maximum decoded components number	4	4	4	4
Maximum decoded bitdepth per component	8 bits	10 bits	12 bits	16 bits
Max Layer number	15	15	15	15
Max decoded resolution	3	4	5	5

levels				
--------	--	--	--	--

Table 1 : MotionJPEG 2000 compliance points

Test Sequence name	Resolution	Video Type	Number of Frames	Format	Bit depth (bits)
cp01.mj2	360x240	30P	150	YCC 4:2:0	8
cp02.mj2	320x240	24P	150	RGB 4:4:4	8
cp03.mj2	352x288	25P	200	Y 4:0:0	8
cp11.mj2	720x480	30P	150	YCC 4:2:0	8
cp12.mj2	640x480	24P	90	RGB 4:4:4	8
cp13.mj2	720x576	50I	200 (400 fields)	YCCA 4:2:0:4	8
cp21.mj2	1920x1080	30P	240	YCC 4:2:2	8
cp22.mj2	1920x1080	24P	90	RGB 4:4:4	12
cp23.mj2	1280x1024	15P	90	YCC 4:4:4	10
cp31.mj2	4096x3112	24P	80	RGB 4:4:4	16

Table 2 : MotionJPEG 2000 test sequences

Motion JPEG 2000 while providing more compression efficiency than MotionJPEG (see Figure 6) is suitable for applications where low delay, compressed frame independence, ability of directly extracting a still picture from video, higher bit-depth than 8 bits, other color spaces than YCbCr, scalability or selective compression (Region Of Interest) are required. Frames being compressed independently, the drawback towards video compression schemes like MPEG-4 AVC/H.264 is of course the need of higher bit-rates for the transmission.



MotionJPEG 2Mb/s

MotionJPEG 2000 2Mb/s

Figure 6: Comparison between MotionJPEG and MotionJPEG 2000 (352x288 25fps 4:2:0 2Mb/s)

It has to be noted that WCAM project provided the cp03 and cp13 test sequences, corresponding to the road surveillance video used in this document. These test sequences include ROI issued from an automatic scene analysis and segmentation as explained in section 3. This means that, often several ROIs are present on a frame, their shape is non rectangular and their positions vary from one frame to another. The segmented video has been also used as a fourth component for the cp13 test sequence. Demonstration sequences presented in this paper will be publicly available on the WCAM project's website [1].

IETF is currently defining how JPEG 2000 will have to be packetised using RTP [17], which will lead to interoperable video streaming of JPEG 2000 over IP. On the Internet, several % of packet loss is common and this value becomes worse in the context of wireless networks. To split JPEG 2000 video streams into RTP packets, efficient packetisation of the code stream has been studied to minimize problems in decoding due to missing parts of the JPEG 2000 codestream. The current IETF draft defines the JPEG 2000 RTP payload header, including some signaling, and the way the different parts of a JPEG 2000 codestream will be separated into RTP packets. The JPEG 2000 codestream is packetised by packetisation units, which are defined either as a JPEG 2000 main header, a tile-part header, or a JPEG 2000 data packet (see Figure 4 and Figure 5). RTP packets bounds are therefore aligned with JPEG 2000 data packets bounds, which allows to take advantage of JPEG2000 data packet loss resilience, thanks to the JPEG 2000 error resilience options. If an RTP packet is lost, the corresponding JPEG 2000 packet will be lost, and thanks to the Start of Packet header marker segment (SOP) [14], the decoder can be re-synchronized to the next JPEG 2000 data packet.

In very difficult wireless environments, an additional error correction mechanism can be necessary, and robustness toward packet loss can be also enhanced by using tools of JPEG 2000 part 11 (Wireless JPEG 2000) that allows particularly using Unequal Error Protection techniques [15][16]. A different error protection capacity can be then applied to different parts of a JPEG 2000 compressed image. When quality data order is used, each layer corresponds naturally to a decrement in sensitivity to errors. More importance can be also given to particular parts like Region of interest.

5. SMART REGIONS-OF-INTEREST CODING IN JPEG 2000

When encoding a region of interest (ROI) in JPEG 2000, an upshift of the wavelet coefficients bits corresponding to this region is realised, in such a way that they will be located above the maximum value of the background wavelet coefficients. The shape of the Region of interest is not described in the JPEG2000 codestream; only the value of the shift is transmitted. The upshift of the wavelet coefficients corresponds to a local increase of the dynamic range of these coefficients, and its value is determined by the maximum value of coefficients after the discrete wavelet transform.

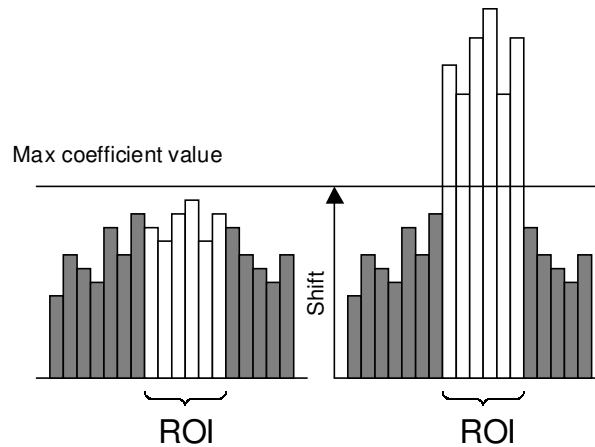


Figure 7: Region Of Interest max-shift method

JPEG 2000 proposed rate allocation method is based upon a Lagrangian rate/distortion optimisation. Due to their high dynamic range, all of the foreground wavelet coefficients (i.e. the ROI) are then prioritised.

One drawback of this method is that it is not possible to specify the rate to get a ROI and a background with a targeted quality, especially for high compression rate or narrow transmission bandwidth. For high compression rates, there can be cases where the background does even not exist anymore while still having a lossless compressed foreground.

Some methods that are not using the max-shift, but the flexible organisation of JPEG 2000 codestream, have been already proposed [10][9]. In these methods, some selected information from JPEG 2000 code-blocks corresponding to the Region of interest are selected for being included in one or several quality layers, and the remaining information corresponding to the background are placed in the following quality layer(s).

In the context of the WCAM project, both methods (max-shift and local allocation) have been implemented and compared. As a result, a third method combining the two approaches has been also studied which allows a more precise control of the rate and the prioritisation through quality layers, while keeping the spatial precision of the max-shift method.

From the original video sequence (Figure 3a), a segmentation mask is automatically generated (Figure 3b) using the analysis method described in section 3. This mask is used to define regions of interest that are coded with better quality than the background. Figure 3c shows the mobile regions of interests when no data from the background is included. Figure 3d shows the result of the ROI coding (Lossless ROI with low quality background).

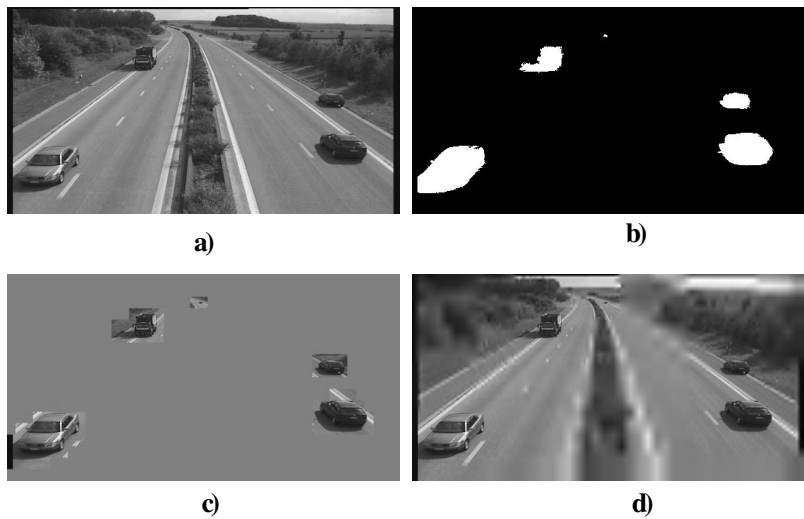


Figure 8: Smart Region of interest coding using segmentation mask

The ROI compressed data are located in a dedicated quality layer, while other compressed data (background) are located in the following quality layers. This allows taking advantage of scalable features of JPEG 2000. A video sequence can be then decoded/transmitted on a lower bit-rate than the rate chosen for the compression. Having used ROI for the important objects in the scene allows guarantying the quality of the content of the ROI, until the decoded data number is lower than the size of the compressed region of interest, as depicted in the following figure.

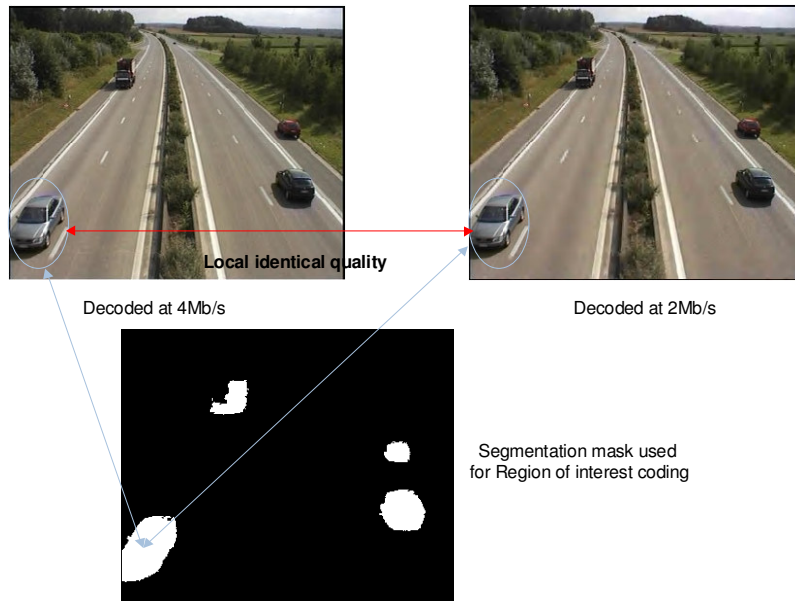


Figure 9: Scalable decoding of JPEG 2000 including Region of interest(s)

6. SMART CODING RESULTS

In the following results, two rate allocations have been performed: one for the background and another one for the ROI. This means that there are two layers, one containing mainly ROI information and the other one the background information.

In the case of local allocation, the limits of the ROI are matched on code-block limits (see Figure 10a). On one hand, in order to get good compression performance, the number of decomposition level as well as the code-block size must be high. But on the other hand, in order to obtain a spatially accurate region of interest, the minimum size for code-blocks as well as a limited number of decomposition levels must be used. The clear advantage of the local allocation is to allow a precise control of the rate allocated to foreground and background.

It is possible to improve this method, combining it with max-shift (see Figure 10b), which is called hereafter ‘max-shift with local allocation’. It consists in an up-shifting of the foreground wavelet coefficients as with the max-shift method, but followed by a code-block based local allocation. This allows getting the spatial precision of the max-shift while allowing the control of the foreground and background quality.

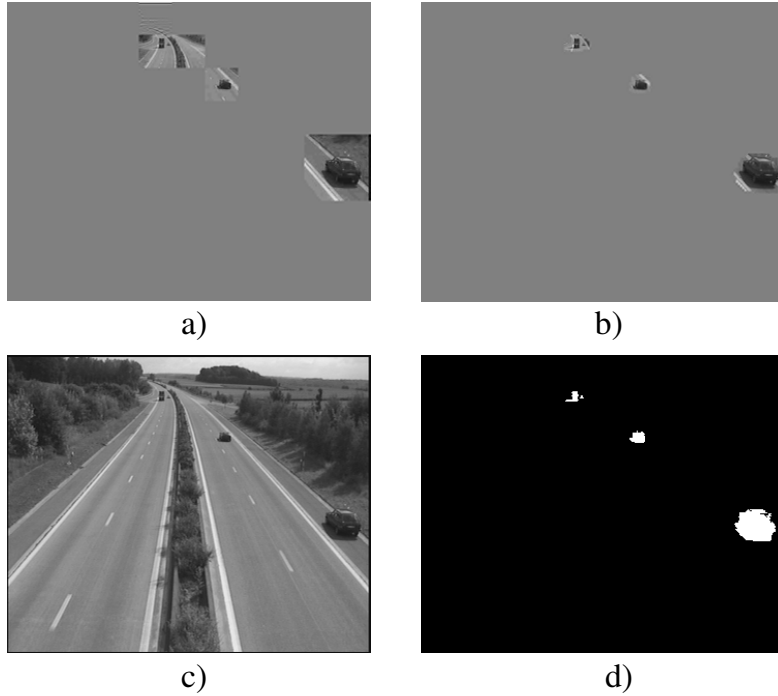


Figure 10 – Coded Region of interest (Decoding of the first quality layer only)
 a) Local rate allocation method, b) Local rate allocation method and max-shift,
 c) Original test image, d) Segmentation mask

In the following curves, max-shift method as proposed in JPEG 2000 standard, local allocation without and with max-shift have been compared to a JPEG 2000 encoding without region of interest. Figures are given separately for the complete image, the region of interest and the background.

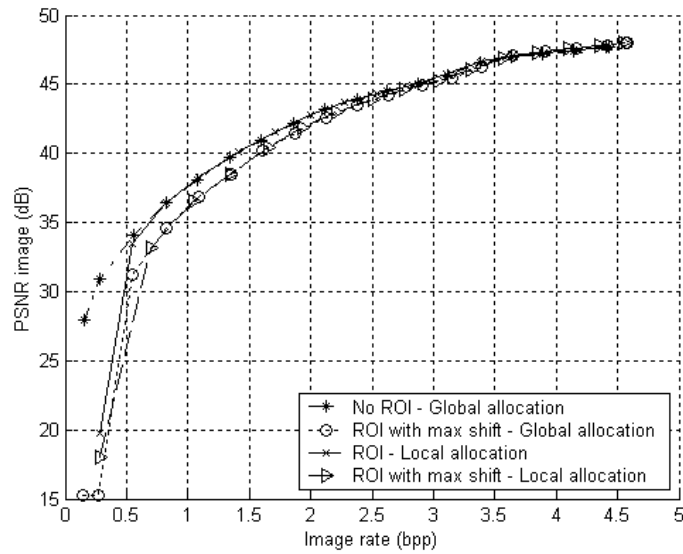


Figure 11: Total image compression efficiency

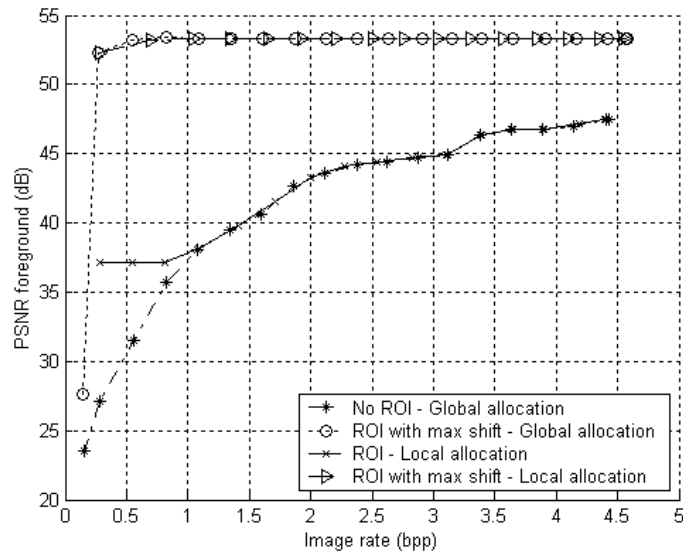


Figure 12: Region of interest compression efficiency

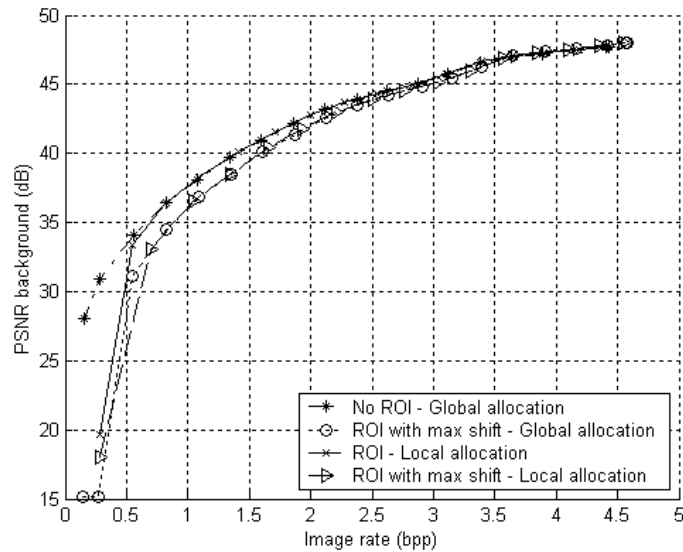


Figure 13: Background compression efficiency

It can be observed that the ROI being small regarding the total image, the performances obtained for the background and the overall image are very similar. As it can be seen, the max-shift allows keeping a very high quality for the segmented objects with a low impact on the background quality. Nevertheless, in the case of larger objects and with max-shift only, the performance of the background should be less. The proposed method combining max-shift and local allocation control, achieves the same performance as max-shift only, while giving control of the allowed rate for both foreground and background.

7. CONCLUSIONS

We have presented an integrated smart encoding system for wireless video surveillance producing fully compliant Motion JPEG 2000 streams. The unsupervised segmentation module automatically determines video surveillance objects of interest. Code-blocks that contain wavelet coefficients from regions of interest are placed in a JPEG 2000 layer different than the rest of the video. This codestream structure provides a trivial way for dynamically adapting regions of interest and background transmissions to channels bandwidth variations, e.g. in wireless conditions. Two methods have been tested for the rate allocation between foreground and background data, one of these integrating the standard max-shift step before the encoding of code-blocks corresponding to the foreground. They both allow flexible background rendering at high compression rates while the max-shift method proposed in JPEG 2000 allows background decoding only when the foreground has been totally decoded.

ACKNOWLEDGEMENT

This work has been funded by the EU commission in the scope of the FP6 IST-2003-507204 project WCAM "Wireless Cameras and Audio-Visual Seamless Networking". The video sequence used for the results presented in this document (speedway test sequence) have been generated by Université catholique de Louvain, in the context of the past ACTS project MODEST.

REFERENCES

1. IST-2003-507204 WCAM "Wireless Cameras and Audio-Visual Seamless Networking," *project website*: <http://www.ist-wcam.org>
2. E. Durucan and Touradj Ebrahimi, "Change Detection and Background Extraction by Linear Algebra," Special Issue of Proceeding of IEEE on Video Communications and Processing for Third Generation Surveillance Systems, November 2001.
3. ISO/IEC ISO/IEC 15 938-5, "Information Technology – Multimedia content description interface – Part 5: Multimedia description schemes", 2002.
4. J. Meessen, J.-F. Delaigle, L.-Q. Xu and B. Macq, "JPEG 2000 Based Scalable Summary for Understanding Long Video Surveillance Sequences", *to appear in proc. of SPIE Image and Video Communications and Processing (IVCP 05)*, San Jose, USA, January 2005.
5. ISO/IEC 15444-1/ IUT-T T.800, "JPEG2000 Image Coding System - Part 1 : Core Coding System", 2000.
6. M. Rabbani, R. Joshi, "An overview of the JPEG2000 still image compression standard", *Signal Processing Image Communication, Eurasip*, Volume17, No. 1, pp. 3-48, January 2002.
7. D. Santa-Cruz, T. Ebrahimi, "An analytical study of JPEG2000 functionalities", *ICIP2000*, Vancouver, Sept. 2000.
8. C.Christopoulos et al., "Efficient method for encoding regions of interest in the upcoming JPEG 2000 still image compression standard", in *IEEE Signal Processing Letters*, pp. 247-249, September 2000.
9. V. Sanchez, A. Basu, M. Mandal, "Prioritized Region Of Interest Coding in JPEG2000", *International Conference on Pattern Recognition (ICPR'04)*
10. A. Nguyen , V. Chandran , S. Sridharan, Robert Prandolini, "Progressive coding in JPEG2000 – Improving content recognition performance using ROIs and Importance Maps », *EUSIPCO'02*, Toulouse, France September 2002
11. 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.
12. K. Kim, T. Horprasert, D. Harwood, L. Davis, "Codebook-based Background Subtraction and Performance Evaluation Methodology", 2003.
13. 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*, 2005.
14. Moccagatta I., Soudagar S., Liang J., and Chen. H., "Error-Resilient Coding in JPEG-2000 and MPEG-4", *IEEE Journal on Selected Areas in Communications* , Vol. 18, No. 6, pp. 899-914, June 2000.

15. D. Nicholson, C. Lamy, X. Naturel, C. Poulliat, "JPEG 2000 backward compatible error protection with Reed-Solomon codes", *IEEE transaction on consumer electronic*, Nov 2003.
16. F. Frescura, M. Giorni, C. Feci, S. Cacopardi, "JPEG2000 and MJPEG2000 transmission in 802.11 Wireless Local Area Networks", *IEEE transaction on consumer electronic*, Nov 2003.
17. S. Futemma, A. Leung, E. Itakura, "RTP payload format for JPEG 2000 Video Streams Standard"; Internet -Draft, (draft-ietf-avt-rtp-jpeg2000-06.txt), October 25, 2004