

## **IST WCAM PROJECT: SMART AND SECURE VIDEO CODING BASED ON CONTENT DETECTION**

J. MEESEN, C. PARISOT AND J-F. DELAIGLE

*Multitel asbl  
Avenue Copernic,1  
7000 Mons, Belgium  
E-mail: Jerome.meessen@multitel.be*

C. LEBARZ AND D. NICHOLSON

*Thales Communication  
Boulevard de Valmy  
Colombes, France  
E-mail: Didier.nicholson@fr.thalesgroup.com*

This paper proposes an integrated solution for smart delivery of video surveillance data. The system is developed within the IST project WCAM "Wireless Cameras and Audio-Visual Seamless Networking" which is presented as well [1]. The main feature of our system is it includes smart video coding based on automatic scene analysis and understanding. Specifically, the segmentation results are used for encoding regions of interest (ROI) in Motion JPEG 2000 guarantying good quality for the semantically relevant objects while keeping a low average data rate. Evaluation of different strategies for JPEG 2000 ROI encoding based on segmentation is presented. We propose a ROI coding strategy optimising the overall quality of the frames while keeping the average data rate low enough for wireless video transmission.

### **1 Introduction**

One critical issue when transmitting video surveillance data is to guaranty acceptable visual quality for the semantically relevant objects or events while keeping the wireless conditions into account, i.e. restricted bandwidth and transmission errors. Therefore, WCAM IST European project has linked image analysis tools to the video encoding. This guaranties optimal quality on the objects of interest that may appear in the video. Since usually only mobile objects are important in video surveillance, the WCAM system will include image analysis tools such as segmentation and object tracking for both vehicles and people. As an example, change detection based on frame difference [2] and/or reference images and segmentation using shape, colour or texture analysis will output active frames and region of interest (ROI) that need to be encoded with better quality and described by means of metadata, in the MPEG-7 philosophy [3].

## **2 IST WCAM objectives**

The goal of the WCAM project is to study, develop and validate a wireless, seamless and secured end-to-end networked audio-visual system. 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 (DRM). Privacy issues are also addressed by selective protection of sensitive frame regions.

## **3 Automatic Scene Analysis**

The goal of the scene analysis module is to detect and track regions of interest 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. 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 estimation of the correspondence between objects of two consecutive frames and an estimation of the objects motion in order to reduce the block-matching complexity of H.264. These data can be used to optimize the coding process complexity but more than that, to optimize the video stream coding representation.

Regions of interest are found with a real-time statistical segmentation module. The algorithm is based on a mixture of Gaussians modeling of the luminance for each pixel of the background [10,11,12]. The main advantage of this technique is that it can automatically deal with backgrounds that have multiple states (cyclic states such as blinking lights, grass and trees moving in the wind, acquisition noise...). 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...). 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.

The metadata resulting from the automatic scene analysis are stored in compliance with the MPEG-7 description scheme in order to be transmitted to the user when necessary (alarm messages) or on demand. These metadata can also be stored for off line video content browsing or retrieval though this is not directly addressed by our project.

#### **4 Smart Regions-of-Interest Coding in JPEG2000**

JPEG 2000 [4][5][6] 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 to the codestream. A key aspect of JPEG2000 is the flexible bit stream structure of the compressed images. This flexibility allows to access to different representations of an image following its scalability features (resolution, quality, position and image component) and the region of interest (ROI) feature. The JPEG 2000 ROI coding [7] allows to give more quality to important parts of the image. These privileged areas can be of arbitrary shape thanks if we used the so called max-shift method.

When encoding a ROI with the max-shift method, an upshift of the wavelet coefficients bits corresponding to this region is realised, in such a way these coefficients will be located above the maximum value of the background wavelet coefficients. The shape of the ROI 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.

The JPEG 2000 proposed rate allocation method is based upon a Lagrangian rate/distorsion optimisation. Due to their high dynamic range, all of the wavelet coefficients of the foreground (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 ratio. For high compression ratios, it can happen that the background is even not present anymore while still having a lossless compressed foreground.

Some methods that are not using the max-shift, but a flexible organisation of JPEG 2000 codestream, have been already proposed, see [9][8]. In these methods, some selected information from JPEG 2000 codeblocks corresponding to the ROI are selected for being included in one or several first quality layers, while the remaining information corresponding to the background is 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, for the ROI, the spatial precision of the max-shift method.

From the original video sequence (Figure 1a), a segmentation mask is automatically generated (Figure 1b). This mask is used to define ROI, which are coded with better quality than the background. Figure 1c shows the mobile ROI when no data from the background is included. Figure 1d shows the result of the ROI coding (Lossless ROI within low quality background).

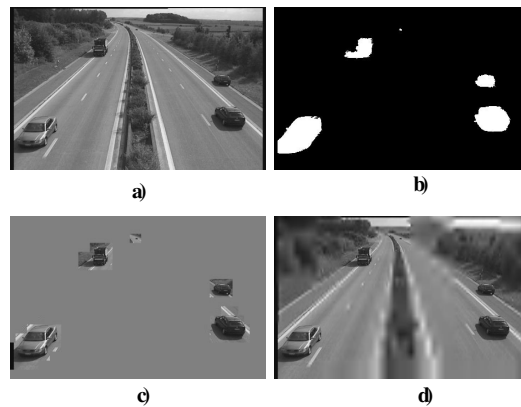


Figure 1: Smart Region of Interest coding using segmentation mask

## 5 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 quality layers, the first containing mainly ROI information and the second one, the background information.

In the case of local allocation, the limits of the ROI are matched on code-block limits (see Figure 2a). 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 precise ROI, the minimum size for code-block 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 by combining it with max-shift (see Figure 2b). This allows getting the spatial precision of the max-shift while allowing the control of the foreground and background quality. In the latter, we call this combined method max-shift with local allocation.

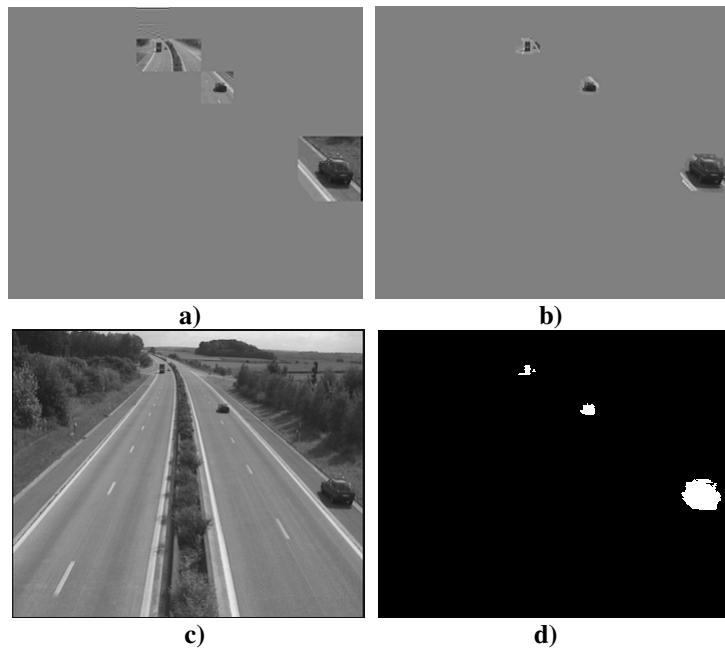


Figure 2 – Coded Region Of Interest (Decoding of the first quality layer only)  
a) Local rate allocation method, b) Local rate allocation method and max-shift,  
c) Original test image, d) Segmentation mask

In the following curves, max-shift method, as proposed in the JPEG 2000 standard, as well as local allocation without and with max-shift have been compared to a JPEG 2000 encoding without ROI. Results are given separately for the complete image (Figure 3), the ROI (Figure 4) and the background.

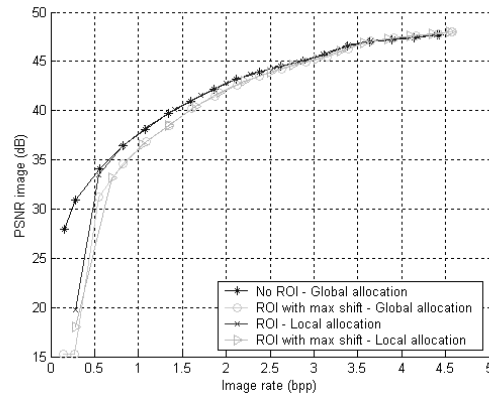


Figure 3: Total image compression efficiency

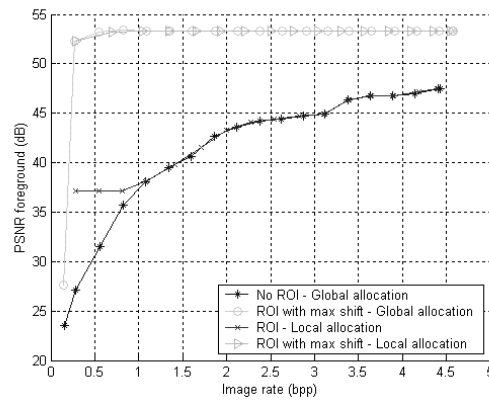


Figure 4: Region Of Interest compression efficiency

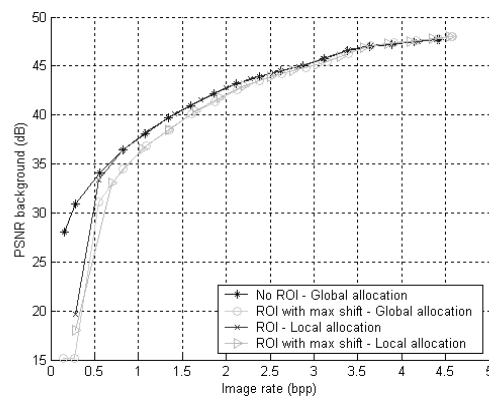


Figure 5: Figure 6: Background compression efficiency

It can be observed that as the ROI is small compared to the total image size, the performances obtained for the background and the overall image are very similar. We see also that max-shift allows to keep a very high quality for the segmented objects with a low impact on the background quality. Nevertheless with larger objects, and using max-shift only, the quality of the background is lower than with other methods. The proposed method combining max-shift and local allocation control, achieves the same performance as max-shift only, while allowing rate control for both foreground and background.

WCAM has also provided to the JPEG standardization group new video sequences for the Motion JPEG 2000 conformance tests. These sequences include ROI's issued from an automatic scene analysis. Demonstration sequences presented in this paper will be publicly available on the WCAM project's website [1].

## **6 Conclusions and future work**

We presented the scope and current results of the IST WCAM project. Study on JPEG 2000 ROI coding achieved within this project has been described. We have shown that the standard max shift method combined with a local rate allocation for the background improves the visual quality of the decoded frames while keeping the average data rate low enough for efficient wireless transmission.

Future works include optimising the coding and transmission of H.264 video data as well as addressing security and privacy issues based on content analysis.

## **Acknowledgments**

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 the 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. ISO/IEC 15444-1/ IUT-T T.800, “*JPEG2000 Image Coding System - Part 1: Core Coding System*”, 2000.
5. M. Rabbani, R. Joshi, “An overview of the JPEG2000 still image compression standard”, *Signal Processing Image Communication*, *Eurasip*, **17**(1), pp. 3-48, January 2002.
6. D. Santa-Cruz, T. Ebrahimi, “An analytical study of JPEG2000 functionalities”, *Proc. of IEEE International Conference of Image Processing (ICIP 2000)*, Vancouver, Sept. 2000.
7. 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.
8. V. Sanchez, A. Basu, M. Mandal, “Prioritized Region Of Interest Coding in JPEG2000”, *International Conference on Pattern Recognition (ICPR'04)*, Cambridge, UK, Aug. 2004.
9. A. Nguyen , V. Chandran , S. Sridharan, Robert Prandolini, “Progressive coding in JPEG2000 – Improving content recognition performance using ROIs and Importance Maps », *Proc. of European Conference on Signal Processing (EUSIPCO'02)*, Toulouse, France September 2002
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*, **2**, pp. 246-252, June 1999.
11. K. Kim, T. Horprasert, D. Harwood, L. Davis, “Codebook-based Background Subtraction and Performance Evaluation Methodology”, 2003.
12. X. Desurmont, C. Chaudy, A. Bastide, J.F. Delaigle, B. Macq, “A Seamless Modular Image Analysis Architecture for Surveillance Systems “, *Proc. of IEEE Intelligent Distributed Surveillance Systems*, Feb. 2004.