Motion Reference Image JPEG 2000: Road Surveillance Application with Wireless Device

Théodore Totozafiny[a], Olivier Patrouix[b], Franck Luthon[a], Jean-Marc Coutellier[c]

[a]Computer Science Lab, University of Pau, Château Neuf, Place Paul Bert, 64100 Bayonne, France;
[b]Laboratory for Industrial Process and Services, Technopole Izarbel, 64210 Bidart, France;
[c]Magys, Technopole Izarbel, 64210 Bidart, France

ABSTRACT

This paper deals with a new codec based on the JPEG 2000 standard that will use a market hardware codec in order to build a road surveillance device. The developed coder consists in 4 processing steps, namely construction of a reference image, foreground extraction (ROI mask), encoding with JPEG 2000 and transmission through a wireless device. A first order recursive filter is used to build a reference image that corresponds to the background image and the updated reference image is computed according to a mixture of Gaussians model. The system builds a reference image and transmits it towards a decoder through the GSM network. After the initialization phase, the reference image is updated automatically according to a Gaussian mixture model, and when the ROI can be considered as null, a piece of the updated background image is sent. We perform motion detection in order to extract a binary mask. The motion mask gives the region of interest for the system. The current image and the motion mask are coded using the ROI option of JPEG 2000 codec with a very low bit rate and transmitted towards the decoder. The complete scheme is implemented and it reaches the expected performances. We also showed how the local background image is built and updated at each frame. We presented the strategy in order to update smoothly the remote background image. The implementation runs at 5-8 frames per second on a 1.8 GHz AMD processor for 320x240 color images.

Keywords: JPEG 2000, ROI, Reference image, Video segmentation, Background substraction, Motion detection

1. INTRODUCTION

Our study addresses the problem of road surveillance for safety purposes with an imaging device. Magys company* is currently using an intelligent camera which includes the camera itself, the frame grabber, and the JPEG codec. The image file is sent to the remote station using a wireless communication system based on a GSM modem. The base station uses a hardware or software JPEG decoder. By now, the performance is about 1 image every 8 seconds due to the low bandwidth and the limited ratio between quality and compression in JPEG. Our goal is to provide a new device to the customer that allows at least the same quality but respecting the following criteria: video rate of 1 image per second and use only a market ready codec. In order to cope with these restrictions, we choose to develop our system on a JPEG 2000 codec because of its robustness to transmission errors, its high level of compression and the fact that it implements Region Of Interest (ROI). JPEG 2000 is the new ISO/ITU-T still image coding standard developed by the Joint Photographic Experts Group (JPEG). JPEG 2000 is designed to provide numerous capabilities within a unified system. It supports various types of still images, multi-components, color, gray-level images and it supports also images that have different characteristics[4, 12, 14] .

Further author information (send correspondence to O.P.):
O.P.: E-mail: o.patrouix@estia.fr, Telephone: +33 (0)5 59 43 84 10
T.T.: E-mail: t.totozafiny@estia.fr, Telephone: +33 (0)5 59 43 84 96
F.L.: E-mail: Franck.Luthon@univ-pau.fr, Telephone: +33 (0)5 59 57 43 44
J-M.C.: E-mail: jm.coutellier@magsys.net, Telephone: +33 (0)5 59 43 85 05

*Magys is a SME involved in developing video devices mainly for video surveillance.

We have developed a Motion Reference Image codec based on the JPEG 2000 standard (MRIJ2K). First, the system builds a reference image and transmits it towards a decoder through the GSM network. After the initialization phase, the reference image is updated automatically according to a Gaussian mixture model, and when the ROI can be considered as null, a piece of the updated background image is sent. Second, we perform motion detection in order to extract a binary mask. The motion mask gives the region of interest for the system. The current image, containing only data linked with the mask, is coded using the ROI option of JPEG 2000 codec with a very low bit rate and transmitted towards the decoder. We focus on our motion detection method based on background suppression. By this approach, the background model is updated and computed frame by frame: moving objects in the scene are detected by the difference between the background model and the current image.

This paper is organized as follows: we describe our MRIJ2K approach in section 2 and the initialization phase is given in section 3. In section 4 and 5, we deal with the motion detection algorithm and the updating of the reference image respectively. Finally experimental results and conclusions are given in the section 6 and 7.

2. OUR MRIJ2K APPROACH

In our approach, the local device is the one that is directly connected to the camera and the remote one is the ground station which receives the encoded images. In the industrial context, some points need to be taken into account:

- The small amount of memory and limited CPU power on the final industrial product,
- The small bandwidth on the GSM network (9600 bauds),
- The fact that wireless transmission is prone to packet errors.

In order to obtain a very low bit rate with MPEG-4, most of the frames need to be bidirectional ones which increase the computation time for encoding and decoding. Moreover, some Intra-coded frames may be lost at the reception. Thus, JPEG 2000 standard is chosen for our development, instead of MPEG.

Fig.1 shows a codec motion reference image based on the JPEG 2000 codec. We describe our system which has been developed assuming that the mobile device is static during the functioning (typically, a patrolman puts the device on the side of the road). This device is mainly based on a camera, a CPU and a wireless transmission device. Our codec is based on 4 processing steps:

1. An initialization phase where a reference image is built (i.e. the background image) and transmitted towards the decoder which runs on a ground station,
2. The automatic ROI construction including motion detection and image processing,
3. The JPEG 2000 encoding using ROI, based on standard specification: Maxshift technique,
4. The image transmission with a GSM modem.

The next section describes the functionning of each step.

3. THE INITIALIZATION STEP

3.1. Constructing the reference image

In a static camera context, the reference image extraction of the scene is a complex task and several assumptions are made. Background modelling is at the heart of any background substraction algorithm, and there are several methods in the computer vision literature. In,[6] the authors evaluate the advantages and drawbacks of some techniques. Existing methods for background modelling may be classified as either predictive or non-predictive. Predictive methods model the scene as a time series and develop a dynamic model to recover the current input based on past observations. Non-predictive methods build a probabilistic representation of the observations at a particular pixel.[2,8] Common techniques used to extract the reference image are the following:

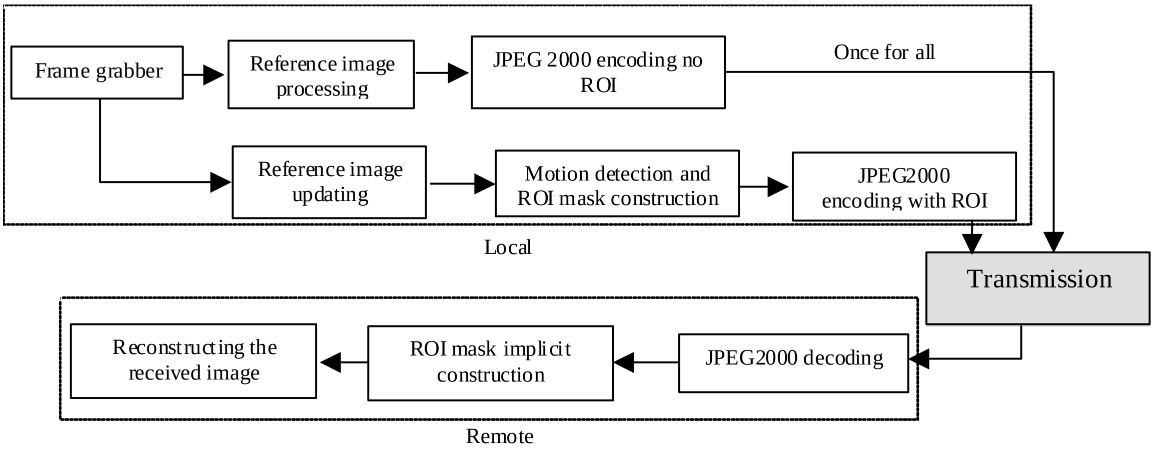
Frame grabber
Reference image processing
JPEG 2000 encoding no ROI
Once for all
Reference image updating
Motion detection and ROI mask construction
JPEG2000 encoding with ROI
Local
Transmission
Reconstructing the received image
ROI mask implicit construction
JPEG2000 decoding
Remote

Figure 1. Illustration of our MRIJ2K approach

- **Median filter:** the reference image estimate is defined to be the median at each pixel location of all the frames in the buffer.
- **Frame differencing:** uses the video frame at time $(t-1)$ to build the background model for the frame at time t.
- **Kalman filter:** a recursive technique for linear dynamic systems under Gaussian noise.
- **Mixture of Gaussians:** a method that tracks multiple Gaussian distributions simultaneously.

Considering the constraint of our final application (lack of memory and embedded system working on a PC104 format board), we use the following first order recursive filter:

$$I_{ref}(p,t+1) = \alpha_p I(p,t) + (1-\alpha_p) I_{ref}(p,t) \tag{1}$$

where $I_{ref}(p,t)$ and $I(p,t)$ are the intensity values of pixel p in the reference image and in the current image at time t respectively. $\alpha_p \in [0,1]$ is the learning rate that gives the training speed, and p is the pixel location in the image.

Of course when the object moves slowly in the scene, it will be included in the constructed reference image. The reference image must be updated from time to time to solve this problem.

3.2. Learning rate filter

Since the reference image is the static part of the scene, the value of α_p for all pixels belonging to a moving object must be 0. In order to determine α_p we have to know all pixels that belong to the background. When p is a background picture element, the α_p value must be in the $]0,1]$ interval, otherwise α_p is 0. The binary stability charts of the three consecutive images $I(t-2)$, $I(t-1)$ and $I(t)$ are used in order to know the object states in the scene. To compute the binary stability chart, we use the edge information and the noise of the acquisition camera is attenuated by an average filter (3×3). The Canny edge detector is used to calculate the gradient of each frame in the sequence.

$$d_1 = |I_g(p,t-2) - I_g(p,t-1)| \tag{2}$$

$$d_2 = |I_g(p,t-1) - I_g(p,t)| \tag{3}$$

where I_g is the spatial gradient of the processed image in the sequence.

A logical AND between d_1 and d_2 followed by an entropic thresholding indicates whether an object is moving or not. Thus, when the condition $((d_1 < \lambda) \,\&\, (d_2 < \lambda))$ is true, the object is a background element. λ can be obtained according to an entropy power threshold selection method.[7]

4. MOTION DETECTION AND ROI MASK GENERATION

Motion detection is a binary labelling problem. It consists in attributing to each pixel p of on image I at time t one of the two following label values:

$$e_p = \begin{cases} 1 & \text{if } p \in \text{moving object} \\ 0 & \text{if } p \in \text{static background} \end{cases} \tag{4}$$

Assuming static camera and quasi constant illumination of the secne, motion information is closely related to temporal changes of the intensity function $I\,(p,t)$ and to changes between the reference image $I_{ref}(p,t)$ and the current image. We use 2 observations in order to carry out the binary labelling. For every pixel, we calculate at pixel p and time t:

1. the difference between the reference image and the current image:

$$O_{RD}(p,t) = |I(p,t) - I_{ref}(p,t)| \tag{5}$$

2. the consecutive frame difference:

$$O_{FD}(p,t) = |I(p,t) - I(p,t-1)| \tag{6}$$

In the first observation, in order to carry out motion detection, most commonly existing techniques check whether the input pixel is significantly different from the background such as $O_{RD} > T_f$ and use edge information from $I(p,t)$ and $I_{ref}(p,t)$ to compute the absolute difference. The moving object threshold T_f is determined experimentally or according to an automatic entropy-based threshold selection. Then, the obtained binary mask is improved by mathematical morphology.

4.1. The Maxshift method in the JPEG 2000 standard

JPEG 2000 Part 1 has adopted a specific implementation of the scaling based ROI approach. The algorithm is called Maxshift method.[1, 10, 12] The method consists in scaling down the background coefficients. In the Maxshift method, the ROI mask is generated in the wavelet domain. All wavelet coefficients that belong to the background are examined and all bit-planes of the background coefficients are shifted down by s bits. The presence of ROI is signalled to the decoder by a marker segment and the value of s is written into the codestream. Thus, with the Maxshift method, no extra information about the shape of the ROI is required for the decoder. But the encoding of the ROI and the background coefficients are completely disjoint processes, *i.e.* the ROI needs to be completely decoded before any background information is reconstructed.

4.2. Compression with JPEG 2000

The ROI automatically extracted by motion detection contains only the useful data. The pixels not belonging to the ROI are removed (set to black color) in order to minimize the data to be sent *i.e.* in the spatial domain when $I_{roi}(p,t)$ value is 0, then the $I(p,t)$ value is set to 0, where $I(p,t)$ is the pixel value of the current image and $I_{roi}(p,t)$ is the mask of the ROI binary image obtained by motion detection. The current image is compressed using a ROI option with a very low bit rate in order to obtain a 9600 bits image size (which will give a rate of 1 img/s with GSM). Then, the compressed current image is transmitted towards the decoder.

4.3. Building the received image in the decoder

The decoder receives a transmitted image and checks whether it is a reference image or a current image. Then the decoder decodes the image and builds implicitly a binary mask. Due to the Maxshift technique, according to,[5] the decision of whether or not a coefficient belongs to the background is taken by comparing the number of decoded bits of current coefficient with M_b, the nominal maximum number of magnitude bit-planes in subband b. This number is computed by the following formula:

$$M_b = G + \epsilon_b - 1 \tag{7}$$

where G is the number of coding guard bits and ϵ_b is the exponent appearing in the definition of the subband's quantization step.

The resulting image is built with three images, namely reference image, current image received and mask built. A pixel substitution is used to build the final image.

5. UPDATING THE REFERENCE IMAGE

Background maintenance is an important problem for many computer vision applications.[13] In visual surveillance applications that deal with outdoor scenes, the background of the scene contains many non-static phenomenae such as illumination variation, tree branches and leaves, object occlusions, shadows, objects being introduced or removed from the scene. In our approach, two background image updatings are needed. The first one uses a high frequency (each frame) and is done locally (at the encoder level) because the updated image is used in the ROI computation. The second uses a lower frequency (as soon as possible when ROI is empty) and is proceeded on the remote computer (at the decoder level).

5.1. Local updating

Due to this kind of background variation, the pixel intensity values vary significantly over time. In computer vision literature, several methods were proposed to solve these problems. The background model must be adapted to take into account the background variation. In order to update the reference image, we use an adaptive mixture of Gaussians initially developed by Stauffer and Grimson.[3, 9, 11] Each pixel will be modelled with K Gaussian distributions. The probability $P(x_t)$ that a certain pixel has intensity x_t at time t is given by:

$$P(x_t) = \sum_{j=1}^{K} \frac{\omega_j}{(2\pi)^{\frac{d}{2}} \left|\sum_j\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t-\mu_j)^T \sum_j^{-1}(x_t-\mu_j)} \tag{8}$$

where ω_j is the weight, μ_j is the mean and $\sum_j = \sigma_j^2 I$ is the covariance for the j^{th} distribution. The K distributions are classified based on ω_j/σ_j^2 ratio, and the first B distributions are used as a model of the background of the scene.

$$B = \arg\min_b \left(\frac{\sum_{j=1}^{b} \omega_j}{\sum_{j=1}^{K} \omega_j} > T \right) \tag{9}$$

T is the fraction of the total weight given to the background model and K is a small number between 3 to 5. The parameters of the matched components are updated as follows:

$$\omega_{j,t} = (1-\alpha)\omega_{j,t-1} + \alpha \tag{10}$$

$$\mu_{j,t} = (1-\rho)\mu_{j,t-1} + \rho x_t \tag{11}$$

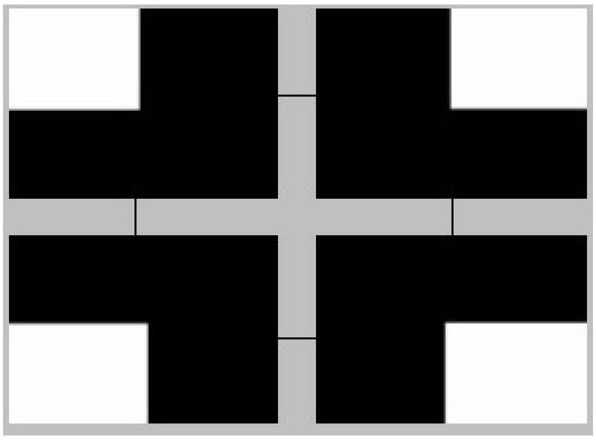

Figure 2. The updated reference image strategy

$$\sigma^2_{j,t} = (1-\rho)\sigma^2_{j,t-1} + \rho(x_t - \mu_{j,t})^2 \tag{12}$$

where α is the learning rate with $0 \leq \alpha \leq 1$. ρ is a learning parameter that can be defined according to[9]:

$$\rho \approx \frac{\alpha}{\omega_{j,t}} \tag{13}$$

For unmatched components, parameters μ and σ remain the same, but ω is updated as follows:

$$\omega_{j,t} = (1-\alpha)\omega_{j,t-1} \tag{14}$$

The updated reference image is given by the matched background model. This image is stored on the local device and is used to upgrade the remote reference image.

5.2. Remote updating

Our strategy in order to update the remote reference image is the following. When we detect no moving object in the scene according to the binary mask ROI image, we use the bandwidth to update the reference image for the decoder. We choose to transmit towards the decoder a piece of the actual background image. In this case, we create an ROI mask based on a rectangular shape. A compressed quarter of the reference image is obtained. Then, we send it to the decoder. This strategy is applied to each quarter of the reference image (Fig.2).

The process of updating is achieved according to a clockwise loop. In order to separate the received images whether it is an updated reference or a moving object image, a flag needs to be added. The updating ratio is not controlled but it is not a problem in our application context because the background dynamics is very slow.

6. EXPERIMENTAL RESULTS

For the JPEG 2000 encoding, we used both the Kakadu codec SDK. Our goal is to obtain a 9600 bits image file in order to reach the image rate of 1 img/s with the GSM. We use the ROI in the compression process because it leads to a smaller size for the image file but the quality of the moving region is still good enough for our purpose (videosurveillance).

At the ground station, the image reconstruction is done using the background image and the current JPEG 2000 image. While decompressing the current image, we are able to retrieve the ROI mask which contains the spatial information using the method described in Sec.4.3. The current image is computed by copying the

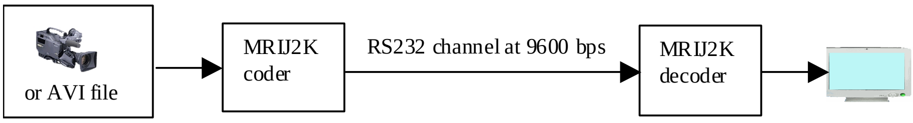

Figure 3. Experimental set-up.

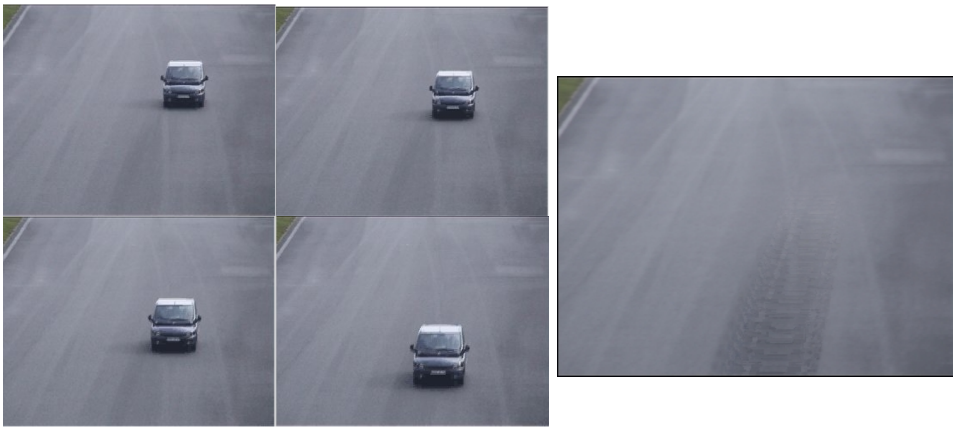

Figure 4. Left: image sequence. Right: background image.

background image and replacing the background pixel by the current pixel if it belongs to the ROI. We have developed an application using Visual C++ for the main frame, DirectShow and OpenCV for our custom image processing filters.

For our first experiments, we decide to use 2 PCs with Windows operating system. The first transmission tests are done using a serial port (RS232 restricted to $9600bps$) between the 2 PCs. Our experimental set-up is described on Fig.3.

6.1. Background extraction

In our experimentation, the training frame number is set to 30. The tests have been done with live video using a webcam or recorded video sequences of typical highway scenes. Fig.4 represents on the left a snapshot of the sequence, and on the right the computed background image.

In order to check the quality of the background reference image, a frame without any moving object is grabbed manually in the sequence and the PSNR is evaluated. The value is $33.87dB$ which leads to an average quality level. This level is in accordance to our context. The quality is linked to the number of frames and also to the tuning of the first order parameters according to typical speed of the moving objects in the scene. So the quality can be improved.

Then, the reference image is compressed to a JPEG 2000 file at a moderate bit rate of $0.4bpp$, and sent once for all towards the decoder.

6.2. Motion ROI and current image coding

Fig.5 represents the current frame on the left, the ROI mask in the middle, and the image of the moving regions on the right. The ROI mask is obtained by computing the difference between the current frame and the background image. Then, the obtained image is thresholded using an experimental value. Morphological filters (opening and closing) are applied in order to remove isolated pixel (white pixels in black region) and also to fill holes in a moving region (black pixels in white zone). The current moving object image is compressed using the ROI mask with a bit rate of $0.128bpp$.

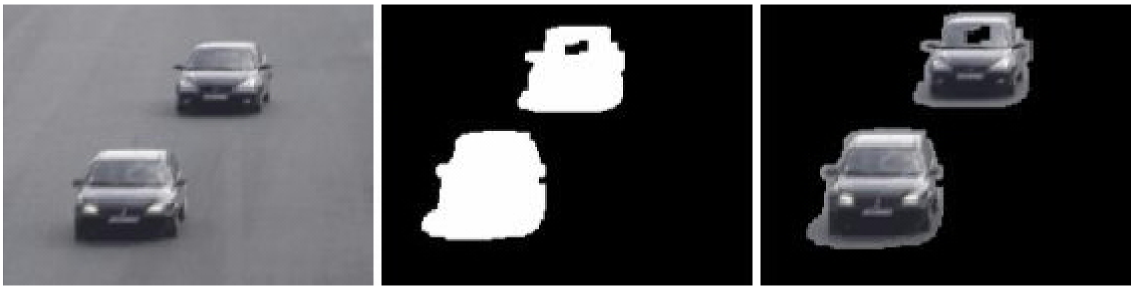

Figure 5. From left to right: current image, ROI mask, image to be compressed.

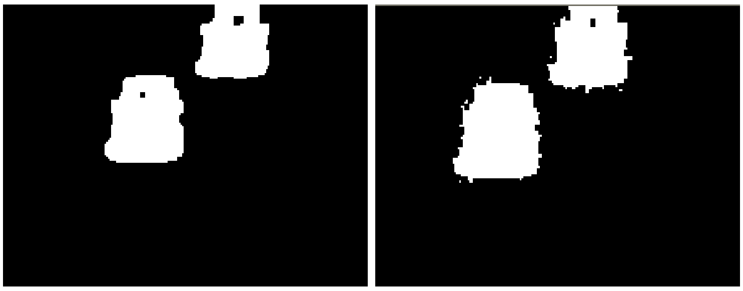

Figure 6. From left to right: Original ROI mask, reconstructing ROI mask.

6.3. Implicit ROI construction

Fig.6 represents the ROI mask computed while the remote PC is decompressing the JPEG 2000 file on the right side and the original on the left. In order to check the quality of our reconstructed mask, the difference between the original ROI mask and the computed one is evaluated. The number of different pixels is about 200 over 76800 pixels. The level of quality of the implicit ROI mask is considered good enough for our application.

6.4. Remote current image processing

Fig.7 shows on top the current image and ROI mask on the local device, and at the bottom the remote image and the implicit ROI mask. In order to check the quality of the remote current image, the PSNR between remote image and initial current image is computed. The value is $34.41dB$ which leads to an average quality level.

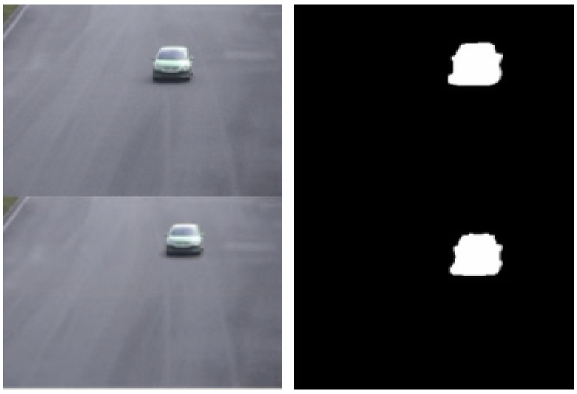

Figure 7. Above: local results. Below: remote results.

7. CONCLUSIONS

A video coding approach has been developed in the context of road surveillance. The complete scheme is implemented and it reaches the expected performances. We also showed how the local background image is built and updated at each frame. We presented the strategy in order to update smoothly the remote background image. The implementation of the MRIJ2K runs at 5-8 frames per second on a 1.8 GHz AMD processor for 320×240 color images. The transmission, using a restricted serial port, has been tested and the rate of 1 frame per second has been reached with GSM. Note that by using GPRS or future UMTS, one might expect a video rate of 25 img/s.

As for future extensions, we are trying to build more accurate motion detection. We are planning to implement the entropy power technique for thresholding in order to automatically tune the filter parameter. The combination of two differences, one between current frame and background reference image and the other one between current frame and past frame, is also studied in our development for the ROI improvement.

Normally according to the Maxshift method in JPEG 2000 standard, the ROI information should be completely found while the JPEG 2000 file is decompressed. In our first implementation, *i.e.* modification of the SDK package, the reconstructed ROI is slightly different. When this problem will be fixed, the reconstructed image at the ground station will be improved.

Preliminary experiments show that our strategy for updating the remote background reference image can lead to more fluid video sequence transmission through the network.

Finally, the GSM modem device needs to be implemented in our software package. Our results are obtained using a restricted serial port so the final system performances should be close to the present ones.

REFERENCES

1. A. P. Bradley and F. W. M. Stentiford. JPEG 2000 and region of interest coding. DICTA, Melbourne, Australia, 2002.
2. S-C. S. Cheung and C. Kamath. Robust techniques for background subtraction in urban trafic video. In WA SPIE, editor, *Image Processing*, volume 5308 of *VCIP*, pages 881–892, San Jose, California Bellingham, 2004.
3. R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. International Conference on Image Analysis and Processing, Palermo, Italy, 2001.
4. T. Fukuhara and al. Motion-JPEG2000 standardization and target market. In *Image Processing*, ICIP, Vancouver, Canada, 2000.
5. R. Grosbois, D. Santa-Cruz, and T. Ebrahimi. New approach to JPEG 2000 compliant region of interest coding. volume 4472 of *46th annual meeting, Applications of Digital Image Processing XXIV*, pages 267–275, San Diego, CA, USA, 2001.
6. D. Gutchess, M. Trajkovicz, E. Cohen-Solal, D. Lyons, and A. K. Jain. Background model initialization algorithm for video surveillance. volume 1 of *ICCV*, page 733, Vancouver, BC, Canada, 2001.
7. F. Luthon, M. Liévin, and F. Faux. On the use of entropy power for threshold selection. *Signal Processing*, 84:1789–1804, 2004.
8. A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Image Processing*, volume 2 of *CVPR*, Washington, DC, USA, 2004.
9. P. W. Power and J. A. Schoonees. Understanding background mixture models for foreground segmentation. Proceedings Image and Vision Computing, pages 267–271, Auckland, New Zealand, 2002.
10. M. Rabbani and R. Joshi. *An overview of the JPEG 2000 still image compression standard.* Kluwer academic publishers, The Netherlands, 2002.
11. C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. volume 2 of *CVPR*, pages 246–252, 1999.
12. D. S. Taubman and M. W. Marcellin. *JPEG 2000 Image Compression fundamentals standard and practice.* Kluwer academic publishers, Netherlands, 2002.

13. D. Wang, T. Feng, H-Y. Shum, and S. Ma. A novel probability model for background maintenance and subtraction. International Conference on Vision Interface, Calgary, Canada, 2002.
14. W. Yu, R. Qiu, and J. Fritts. Advantages of Motion-JPEG2000 in video processing. In *Electronic Imaging*, Fritts SPIE Photonics West, San Jose, CA, 2002.