Color and R.O.I. with JPEG2000 for Wireless Videosurveillance

Franck Luthon, Brice Beaumesnil

COLOR AND R.O.I. WITH JPEG2000 FOR WIRELESS VIDEOSURVEILLANCE

Franck Luthon, Brice Beaumesnil

University of Pau and Adour
LIUPPA Lab
IUT Informatique, Château Neuf, Place Paul Bert, 64100 Bayonne, FRANCE

ABSTRACT

For a videosurveillance application based on wireless transmission of images acquired with a static camera, we investigate two features within JPEG2000, namely the choice of a nonlinear color transform in order to improve color rendering at very high compression ratios, and the unsupervised extraction of regions of interest (ROI) by motion detection.

1. INTRODUCTION

JPEG2000 is much more than a new compression tool for replacing JPEG [1]. Indeed it is a flexible and open framework for the representation of still images [2]. The standard is based on the use of a discret wavelet transform (DWT) instead of DCT, and an arithematic bit-plane coding (EBCOT) instead of Huffman coding (Fig. 1). The new features and

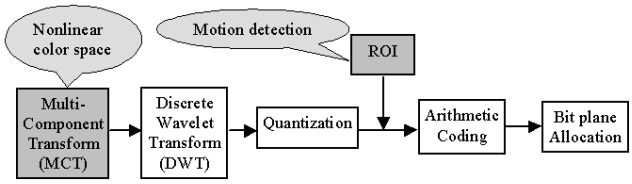
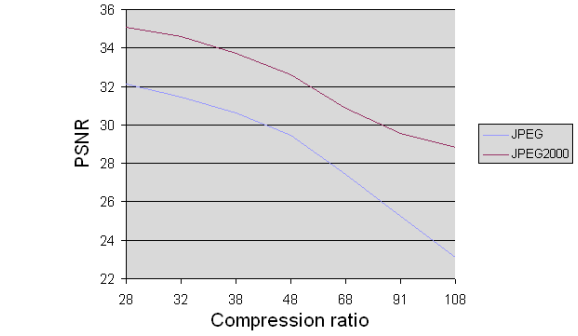


Fig. 1. JPEG2000 block-diagram (in gray: our contribution)

induced improvements are: superior performance in terms of compression ratio (typ. 1:60 instead of 1:30 for comparable visual quality, see Fig. 2 for rate-distortion curve); lossless or lossy compression modes; scalability in resolution, in quality or spatial; progressive transmission and reconstruction of images; robustness w.r.t. errors for mobile applications at very low bitrate; definition of regions of interest (ROI); open architecture (choice of multi-component color transform MCT, choice of DWT); flexible file format.

In the context of road surveillance based on the use of a portable video module with acquisition board at 8 img/s and GSM transmission at 9600 bauds, a high compresssion

Thanks to MAGYS S.A., Bidart, and ENSEIRB, Bordeaux, France, for providing images. Thanks to ANVAR and IRA for funding.



(a) JPEG (1:90) (b) JP2k (1:90;Yuv; Dwt9/7)

Fig. 2. JPEG2000 versus JPEG: gain $\times 2$ in compression.

ratio is needed to achieve a transmission rate of about 1 img/s which is convenient for such an application (highway patrolmans transmitting information to distant control center). However, the current prototype using JPEG allows a compression ratio of about 1:30, which yields a too low transmission rate (about 8sec/img for an image size of 420×255). On the other hand, MPEG4 may not be used because the reconstruction of B and P coded images might become impossible: the transmitted video rate is too low for allowing a correct motion estimation at the decoder side (not all acquired images are transmitted); even if the transmission rate were high enough, packet errors which occur with GSM would lead to lose some Intra-coded images required for proper decompression. Therefore, we choose JPEG2000 with two main objectives in mind : to get an acceptable quality (color rendering) at very high compression ratios ($>$1:90), by using an original nonlinear color transfom; to gain another factor in the compression ratio (typ. $\times 2$) by taking advantage of the ROI: since the application

addresses the case of a static camera (the patrolman stops and puts his camera module at a given place, e.g. spot of an accident or traffic jam), we propose to extract the ROI by motion detection.

2. NONLINEAR COLOR TRANSFORM

JPEG2000 handles multi-component images. From the three basic color components RGB, the Multi-Component Transform (MCT) allows to change the color space in order to satisfy to two criteria : a better decorrelation of color components for efficient compression, a better correspondance to the human visual system in order to decrease the information loss due to quantization (better rendering). Two linear transforms YUV and YCrCb are basically proposed in the standard. The well known YCrCb color space is defined by the matrix relationship : $[Y\ Cr\ Cb]^t = T.[R\ G\ B]^t$, where t denotes transpose and:

$$T = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5 & -0.41869 & -0.08131 \\ -0.16875 & -0.33126 & 0.5 \end{bmatrix} \quad (1)$$

Here, we propose to use an original nonlinear transform, called the LUX transform, that already proved to be efficient for color segmentation [3]. This color transform originates both from biology [4] and mathematics [5]. The idea of introducing a logarithmic non linearity is in adequation with the human visual system (Fig. 3): the cone transduction function may be described by a loglike function, while the action of horizontal and bipolar cells (weighted average and weighted difference resp.) may be modelled by a linear matrix like T. From the mathematical viewpoint, the diagram below helps understand how we build the LUX color space by composition of three functions:

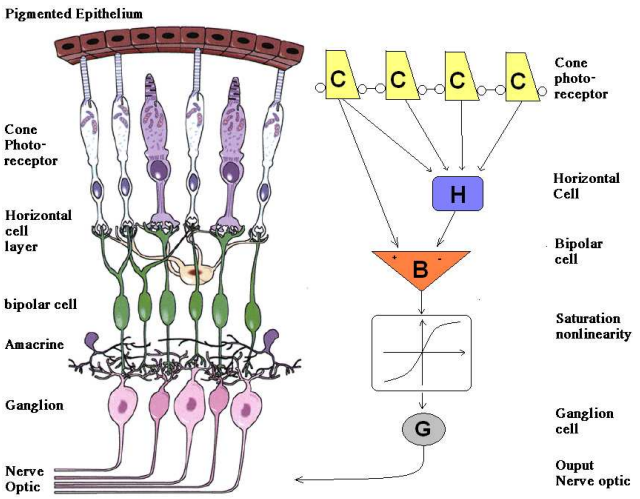


Fig. 3. Biological analogy.

$$(R,G,B) \xrightarrow{\Phi^{-1}\circ T\circ\Phi} (L,U,X) \quad (2)$$

where Φ is an isomorphism that provides a normalized log transform whereas its inverse Φ^{-1} is a normalized exponential function (for details, see the theory of Logarithmic Image Processing model [5]). From the RGB image, the basic LUX transform is then defined as:

$$\begin{aligned} L &= (R+1)^{t_{11}}(G+1)^{t_{12}}(B+1)^{t_{13}} \\ U &= (R+1)^{t_{21}}(G+1)^{t_{22}}(B+1)^{t_{23}} \\ X &= (R+1)^{t_{31}}(G+1)^{t_{32}}(B+1)^{t_{33}} \end{aligned} \quad (3)$$

where t_{ij} are coefficients of matrix T.

Now, we adapt it specifically for compression purpose (requirements of bounded dynamic range and inverse transform). The inverse transform of (3) is given by:

$$\begin{aligned} R &= L^{a_{11}}U^{a_{12}}X^{a_{13}} - 1 \\ G &= L^{a_{21}}U^{a_{22}}X^{a_{23}} - 1 \\ B &= L^{a_{31}}U^{a_{32}}X^{a_{33}} - 1 \end{aligned} \quad (4)$$

where a_{ij} are the coefficients of inverse matrix $A = T^{-1}$. In fact, any matrix corresponding to one of the various TV standards like YIQ, NTSC, PAL, YUV might be used. The best results for LUX were obtained with a matrix T_{lux} that is a mix of YCrCb and YIQ: i.e. matrix T with the last row replaced by $[-0.202\ \ 0.5\ \ -0.298]$. Fig. 4 illustrates the tests carried out on the typical image of Fig. 6a.

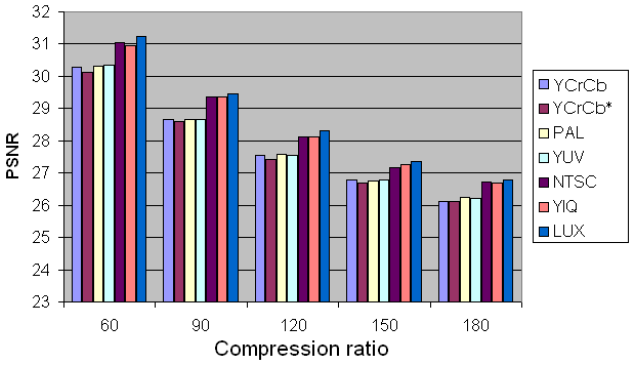


Fig. 4. Various color matrices tested within LUX transform

Before applying the DWT, we have to stick to the numerical format in the compression tool, and we want to use all the dynamics available for color pictures coded with 24 bits. Therfore, we normalize each component in (3) by adjusting its dynamic range depending on the image content: $x = 255(X - X_{min})/(X_{max} - X_{min})$.

Fig. 5 shows the improvement of color rendering at very high compression ratio (1:90) on a typical image of road

surveillance. The uncompresssed BMP image is shown on Fig. 6a. Whereas JPEG yields inacceptable distortion (blocking artifacts, Fig. 2a), JPEG2000 still gives good result with YUV and Daubechies 9/7 wavelet transform (Fig. 2b). However, one can see (Fig. 5, top) that the red lights and orange warning strips at the back of the truck have lost their color. The nonlinear LUX transform proposed here yields a better rendering of original colors (Fig. 5, bottom).

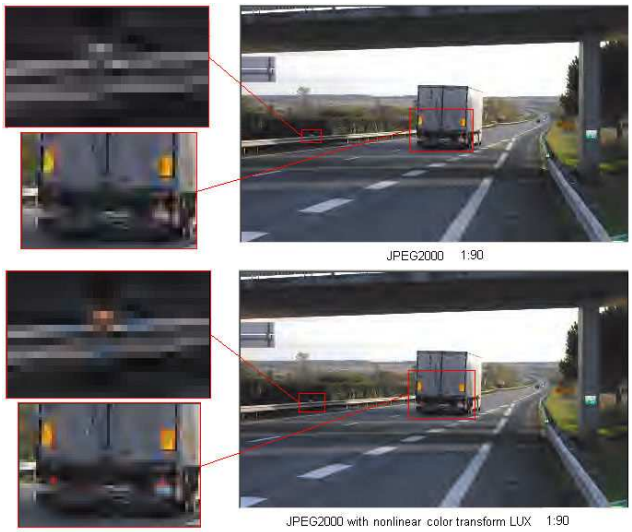


Fig. 5. Compression with JPEG2000 (ratio 1:90 ; Dwt9/7): comparison between YCrCb (top) and LUX (bottom)

3. ROI FROM MOTION DETECTION

Another new feature of JPEG2000 is the processing of ROI, which allows to have different compression ratios for different parts of the image, the regions of interest defined by the user being encoded with more precision than the rest of the image (background). Fig. 6b illustrates this feature (in the case of a rectangular ROI placed manually). The coding of ROI is done by the Maxshift technique, which takes place between quantization and arithmatic coding (cf. Fig. 1). It consists in applying a binary shift to the right to coefficients related to the *background* (BG), while coefficients of the ROI are coded with the most significant bits (Fig. 6c). The main advantage of this technique is that no spatial information about the ROI (e.g. coordinates) needs to be explicitely transmitted to the decoder. Only the position of the shifting point is transmitted. Then, the various wavelet coefficients are entropy-coded with the most significant bit-planes first. This method enables to visualize the ROI first, in the case of progressive transmission with progressive reconstruction of the image. But it can also be used in order to increase the compression ratio of images where only some regions are relevant for the end-user. To this end, we propose to

(a) original image. b) compression with ROI (1:200)

c) Maxshift technique applied on wavelet coefficients.

Fig. 6. Principle of ROI.

automatically extract the ROI by motion detection. Motion detection aims at labelling each pixel $s = (x, y)$ of image at time t in order to get a binary map of temporal changes. In the case of a static camera, a low-cost motion detection algorithm is easily implemented (computation of temporal observation, thresholding, mathematical morphology). The observation o_s is the absolute value of the intensity difference between two consecutive time instants. After adequate thresholding (threshold θ), one gets the binary labels l_s:

$$o_s = |I_t(s) - I_{t-1}(s)| \quad (5)$$

$$l_s = \begin{cases} \text{“1”} & \text{if } s \in \text{mobile area (ROI)} \\ \text{“0”} & \text{if } s \in \text{static background (BG)} \end{cases} \quad (6)$$

All mobile-labeled pixels ($o_s > \theta$) are candidates for the ROI. Since observation is noisy (camera and quantization noise), a post-processing like erosion-dilation or opening-closing is applied on the binary map to fill in the holes in the moving masks and to erase isolated points due to noise, in order to get connected areas whose union will constitute the ROI. The ROI extracted that way is not a window with regular shape applied on the image (like a rectangle or an ellipse), but it is a set of regions with arbitrary shape corresponding to the masks of moving objects in the scene (Fig. 7). The use of ROI with high compression ratio (Fig. 6b) proves that data outside the ROI carry little information. Therefore, BG pixels may be discarded (irrelevant), leaving more place for relevant information (Fig. 7, right). If only mobile pixels (ROI) are coded and transmitted (without taking into account BG pixels which will be reconstruted provided some reference image of the background is available at the decoder), it allows to gain greatly in compression ratio (1:240), so that the transmission rate is

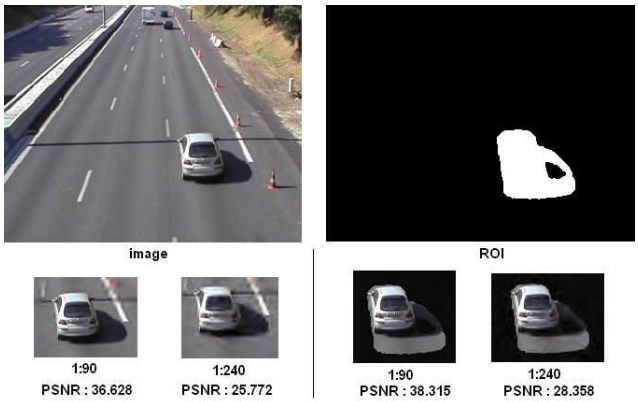


Fig. 7. ROI extracted by motion detection. Comparison of compression performance with or without background data.

significantly increased (1 img/s), while preserving the ROI with good quality (PSNR around 30dB).

This approach leads to our proposal of a Motion-JPEG 2000 codec scheme for very low bitrate transmission, working in the restricted case of a static camera. The idea is the following : a reference image of the background is first provided to the decoder, then updated at a low video rate like every 16 images (or a partial updating of the reference image is done, like 1/16 of the image size at each time), in order to refresh BG pixels on reception display. Other images are highly compressed using only ROI pixels extracted by motion detection. It enables to handle a basic motion information (binary motion detection) that is intra-coded in JPEG2000 (thanks to the ROI), contrarily to MPEG (cf. B and P images). Therefore, it works even if the video rate at the decoder is very low, since the decoding does not rely on previous frames (that might have been acquired and compressed at the coder side, but not been transmitted towards the decoder). The extraction of ROI by motion detection is a significant part of the codec: the better the detection, the better the transmitted images. Another important part is the reference updating process: currently, the updating of the reference image at a lower rate is the weak part our codec: between two updates, there are ROI border artifacts (Fig. 8).

Fig. 8. ROI border artifacts (compression 1:240)

4. DISCUSSION

The LUX transform, adapted for the purpose of compression within JPEG2000 standard, proved to be efficient on various test images corresponding to road surveillance application, but also on other benchmark images (like mandril and clown) that have much more color contrast. Nevertheless, further tests are required to validate the approach. The use of ICC profile seems to be well suited for implementing our choice about a nonlinear MCT. The connection between the choice of a nonlinear color transform and the choice of quantization parameters must be investigated (also the choice of visual weighting factors in each sub-band: here, we discarded the standard weights since they are adapted to a linear, but not to a nonlinear color space).

For the road surveillance application, JPEG2000 robustness to errors will also be tested (and compared to JPEG that fails for packet errors). Currently, the GPRS protocole is being implemented in the propotype (instead of GSM). A gain of $\times 5$ is expected in the transmission rate, in addition to the improvements resulting from our use of JPEG200. So that the application might work at 5img/sec in a near future, with reasonable visual quality for videosurveillance. With UMTS, the rate of 25 img/s would even be achievable. For this work on JPEG2000, we tested both JasPer (C implementation) and Kakadu codec (C++) in order to compare processing speed and computation costs which are important issues for hardware implementation on a PC104 board.

5. REFERENCES

[1] D.S. Taubman and M.W. Marcellin, *JPEG 2000 Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, The Netherlands, 2001.

[2] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Processing: Image Communication*, vol. 17, pp. 3–48, 2002, http://www.jpeg.org/JPEG2000.htm.

[3] M. Liévin and F. Luthon, "Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video," *IEEE Trans. on Image Processing*, vol. 13, no. 1, pp. 63–71, Jan. 2004.

[4] S. Shah and M.D. Levine, "Visual information processing in primate cone pathways-Part I: A model," *IEEE Trans. on Systems, Man and Cybernetics-B*, vol. 26, no. 2, pp. 259–274, Apr. 1996.

[5] M. Jourlin and J.C. Pinoli, "Image dynamic range enhancement and stabilization in the context of the logarithmic image processing model," *Signal Processing*, vol. 41, pp. 225–237, 1995.