

Performances of the estimation and motion compensation for the reconstruction of motion areas in a sequence video Motion JPEG 2000

Abdou Khadre DIOP *, K. TALL *, S. M. FARSSI *, I. DIOP *

* Laboratory of Medical Imaging & BioInformatics, Department of Computer Science,
“Ecole Supérieure Polytechnique”, Cheikh Anta Diop University
Dakar, SENEGAL

djeylani2001@yahoo.fr

Abstract— In a very noisy environment, the transmission of video sequences Motion JPEG 2000 is not entirely reliable since attending packet loss in the codestream the movie Motion JPEG 2000. For a reconstruction of the lost packets, the method of conditional replenishment has been proposed. However, with this method, the effects of blocks appear in motion areas making the quality of the video poor in these areas.

In this paper, we propose the method of estimation and motion compensation for optimal recovery of lost packets in motion areas. Thus, the application implementation in Matlab can see firsthand the disappearance of the effects of blocks in the transmission of video sequence Motion JPEG 2000 in a very noisy environment.

Keywords— Motion JPEG 2000, estimation and motion compensation, motion areas, effects of blocks, conditional replenishment.

I. INTRODUCTION

The standard JPEG compression can compress lossy video sequence using the discrete cosine transform (DCT). Processing the successive images DCT is performed on pictures previously partitioned into blocks of 8×8 pixels. So the effects of blocks appear in the movie playback compressed with JPEG video compression standard. However, the compression standard JPEG 2000, proposed for replacing the JPEG standard uses the discrete wavelet transform (DWT) in the compression of video sequences JPEG 2000. And for very high compression ratios, the occurrence of the effects of blocks is not observed when reading.

However, in a noisy environment, if packets are lost in motion areas of a video sequence Motion JPEG 2000, the effects of blocks appear when reading this video.

To reduce the effects of blocks related to packet loss, the method of conditional replenishment was proposed [1]. However, this method lacks very high reliability in the motion of areas video. We proposed to adapt the method of estimation and motion compensation to reconstruct lost packets [2], [3]. So we propose in this paper this last method to improve the

robustness to transmission of video sequence Motion JPEG 2000 by removing the effects of blocks.

To describe our work, the article presents in the first part features of Motion JPEG 2000. The second part introduces the concept of conditional replenishment used to reduce the effects of blocks recorded in the transmission video sequence Motion JPEG 2000 in a very noisy environment. The third section describes how the method of estimation and motion compensation improves the robustness of the transmission of video sequence Motion JPEG 2000. Finally the fourth part presents the results in terms of objective and subjective quality assessment.

II. THE MOTION JPEG 2000

Motion JPEG 2000 was standardized in 2002 as part of the JPEG 2000 ISO/IEC standard with later improvements. Motion JPEG 2000 is defined as the part 3 of the JPEG 2000 standard. It is an extension of the JPEG 2000 for processing video standard. The principle is that each frame of the video is encoded in the JPEG 2000 format. A Motion JPEG 2000 MJ2 or MJ2 format is a concatenation of images in JPEG 2000 format JP2 or J2K.

The allocation algorithm flow of the JPEG 2000 standard allows creating packages like all other algorithms flow. The packages are created depending on the number of resolution at the discrete wavelet transform (DWT), the number of precincts, the number of components in the image and the number of layers of quality defined when compressing an image. This organization of the codestream permit to decompress spatial zone, a layer of quality, a resolution or a component of the image or video sequence considering the order of progression which is the order in which the packets appear in the codestream.

Thus packet loss in the Motion JPEG 2000 codestream can lead to bad reconstruction of scalable specifications of an image especially in motion areas of the image. The method of conditional replenishment has been proposed to reconstruct the lost packets during transmission.

III. THE CONDITIONAL REPLENISHMENT

The conditional replenishment was introduced by Mounts [4] in 1969 in the early years of digital video coding. It is a method of compressing inter-frame video using temporal correlation for reducing the video transmission rate. This method works by detecting and sending only the changing parts of the image and by using the receiver video data from the previous frame to the part that does not change. The compression rate which can be obtained by this technique depends a large extent on rate of change within the image. In this context, the images to be transmitted are divided into spatial zones that are approached by a reference available to the decoder or supplemented by new data. Indeed, if the available reference and spatial zone to be transmitted are similar, then the number available will probably be used to estimate the spatial zone. Else new data will likely transmitted to reconstruct the spatial zone, because it differs significantly from the reference available.

The conditional replenishment coding in traditional compression schemes such as MPEG and H.264 consists in encode only the pixels whose luminance changed from one frame to the next and to transmit two informations: the address of the pixel whose luminance has changed and the value of luminance. Thus the encoder conditional replenishment distinguished two cases: motion areas where no information is transmitted and motionless areas where temporal information quantized and coded is transmitted.

In Motion JPEG 2000, the method of conditional replenishment is proposed to reconstruct the lost packets during transmission based on the codestream of the anchor image following the loss of packets in the codestream of the target image. However rebuilding motion area is not sharp with this method because of the effects of blocks appeared in the transmission of codestream of the video sequence rendering bad the quality of transmission. Figure 1 illustrates the effects of blocks.



Figure 1. Original image on left and Effects of blocks on right

In this context, the results show that the reconstruction of lost packets with this method has limitations. To optimize the reliability of the transmission of video sequences with better reconstruction of lost packets, the method of estimation and motion compensation is proposed for the reconstruction of lost packets in a motion area.

IV. ESTIMATION AND MOTION COMPENSATION

Detection and motion estimation in temporal sequences is a fundamental problem in the field of processing and analysis of

image sequences. The purpose is to detect and estimate the motion of objects in a video as a dense vector field [5].

The motion estimation techniques address to the problems of detection and motion estimation between two images $f(x,y,t_1)$ known as the anchor image and $f(x,y,t_2)$ called the target image. Motion estimation is called backward if $t_2 > t_1$ and forward motion estimation where $t_1 > t_2$ [6], [7].

A motion vector $\mathbf{d}(\mathbf{X}) = [d_x, d_y]^T$ represents the displacement of the particular point $\mathbf{X} = (x_i, y_i)$ between t_1 and t_2 . These concepts are illustrated in Figure 2.

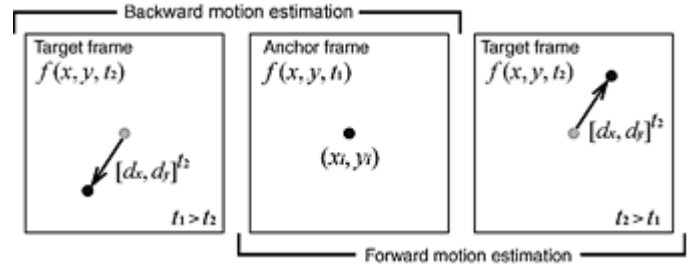


Figure 2. Anchor and target frames in forward and backward motion estimation

Order to get the estimated motion, several types of methods exists. They are classified into three categories: method of block matching, statistical methods and differential methods.

The differential motion estimation methods are based on temporal and spatial gradients of luminous intensity of the pixels. The principle of these methods rests on a strong assumption, which is that of the conservation of the luminous intensity of a pixel along the motion path. The criterion for motion estimation with these methods is based on the equation of the Displaced-Frame Difference (DFD) which consists in determining the difference in intensity at each point between the target image and the anchor image. Mathematically, the equation of the DFD is given by (1):

$$E_{DFD}(d) = \sum_{X \in R} (f_2(X + d) - f_1(X))^p \quad (1)$$

Where $f_1(X)$ is the pixel value of X at time t_1 , $f_2(X + d)$ is the pixel value at time t_2 , d is the estimated displacement and R the motion detection zone.

In this work the differential methods are used to implement the estimation of motion in Motion JPEG 2000. The equation of the DFD will detect motion in a motion area of the anchor image and optionally to reconstruct the lost packets in the target image.

As with the implementation of the method of conditional replenishment, the method of estimation and motion compensation is proposed to reconstruct the lost packets during transmission based on the codestream of the anchor image after the following algorithm (Figure 3).

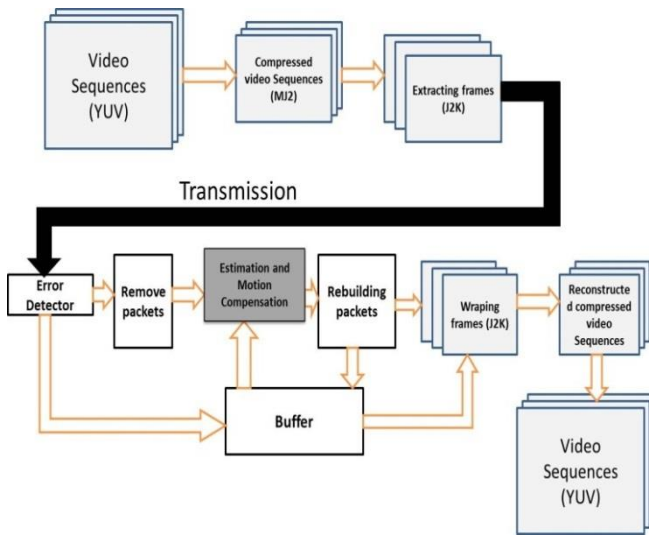


Figure 3. Algorithm using motion estimation and compensation method

The implemented motion estimation procedures are applied to the target and anchor images to detect motion area. If packets are lost in a motion area of the target image, these procedures will be used to estimate the motion in the anchor image and reconstruct the packets of the target image from the same anchor image by motion compensation.

The estimate of motion in the anchor image can be performed by estimating motion forwards or backwards. Assessment criteria described for determining the best reconstructed image by motion estimation in forward or backward.

V. RESULTS

We consider a video sequence Motion JPEG 2000 of 80 images transmitted in a very noisy environment. Assuming the images are noisy successively in motion areas, we reconstruct the lost packets in these areas by conditional replenishment then by estimation and motion compensation. First we see firsthand the quality of transmission is poor in the detected motion areas where noise has been applied. The effects of blocks are noted in the transmission of the video sequence. Then the effects of blocks are reduced in motion areas reconstructed by conditional replenishment making poor transmission quality in motion areas. Finally we note a marked improvement of effects of blocks in the video reconstructed by estimation and motion compensation. Figure 4 and Figure 5 illustrate the difference between the methods of reconstruction of lost packets in the codestream of a video sequence Motion JPEG 2000 by conditional replenishment and estimation and motion compensation. To assess these methods, we use objective and subjective assessment of the quality methods.



Figure 4. Reconstructed image by conditional replenishment



Figure 5. Reconstructed image by estimation and motion compensation

For an objective assessment of the quality of transmission, the PSNR (Peak Signal-to-Noise Ratio) is a metric that is used to assess the fidelity of an image relative to the reference. It is a criterion that treats the image as a purely digital signal. It is based on the mean square error MSE (Mean Square Error) connected between the degraded image and the reference in (2):

$$MSE = \frac{1}{M \cdot N} \sum_{n=1}^N \sum_{m=1}^M (X_{origine}(n, m) - X_{compressée}(n, m))^2 \quad (2)$$

We deduce the peak signal to noise ratio by (3):

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{MSE} \right) \quad (3)$$

Figure 6 shows the evolution of PSNR between the Motion JPEG 2000 video sequence transmitted with noise, the reconstituted by conditional replenishment and the reconstituted by estimation and motion compensation based on the number of images contained in the Motion video sequence JPEG 2000. In reading about the figure we can deduce that the method of estimation and motion compensation improves the robustness to transmission video sequence Motion JPEG 2000 face the loss of packets in motion areas of the sequence.

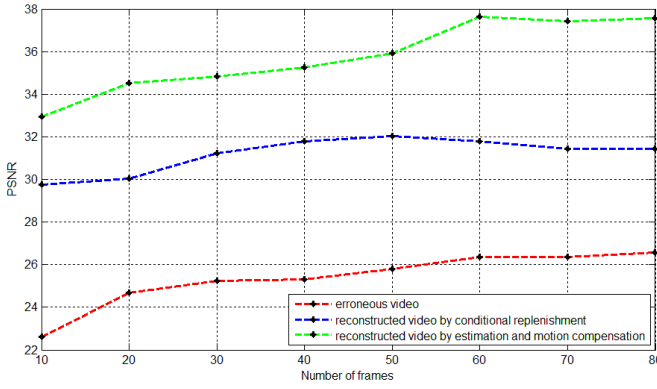


Figure 6. PSNR in terms of number of frames

For a subjective assessment of the quality of transmission, Wang [8]-[10] proposed a system of quality estimation based on the measurement of similarity structure (SSIM). The system consists of measuring the similarity in three comparisons: luminance, contrast and structure. The measurement of similarity structure is defined in (4):

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i; \text{ average intensity} \quad (5)$$

$$\sigma_x = \left(\frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2}; \text{ standard deviation} \quad (6)$$

$$\sigma_{xy} = COV(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y); \text{ covariance} \quad (7)$$

$$C_1 = (0.01 * 255)^2; C_2 = C_3 = (0.03 * 255)^2;$$

Figure 7 shows the evolution of the SSIM between the Motion JPEG 2000 video sequence transmitted with noise, the reconstituted by conditional replenishment and the reconstituted by estimation and motion compensation based on the number of images contained in the Motion video sequence JPEG 2000. In reading about the figure we can deduce that the method of estimation and motion compensation improves the robustness to transmission video sequence Motion JPEG 2000 face the loss of packets in motion areas of the sequence.

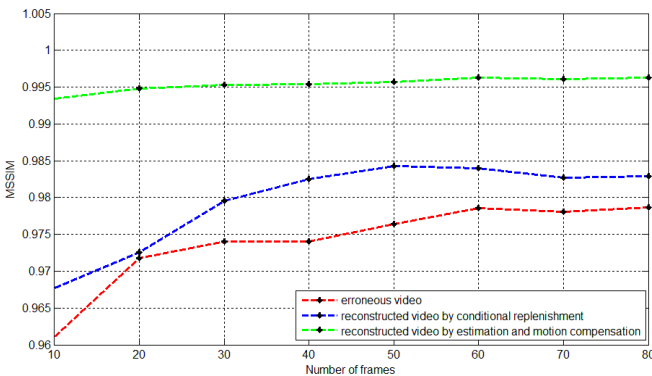


Figure 7. MSSIM in terms of number of frames

VI. CONCLUSION

In this work, we investigated the use of estimation and motion compensation to transmit reliably Motion JPEG 2000 file. Although it is well known in the inter-frame coding sequences MPEG-4 and H.264, the method allows reconstructing lost packets in motion areas of a target frame by using the anchor frame. The results of subjective and objective evaluation can be concluded that the method of estimation and motion compensation can be used to ensure a good quality of video transmission.

As storage and transmission video surveillance content in JPEG 2000 are deployed of integrated solutions, these results encourage considering the use of estimation and motion compensation to reconstruct lost packets in video surveillance. Likewise this method can be used in the field of medical imaging for the detection information.

REFERENCES

- [1] F.O. Devaux, J. Meessen, C. Parisot, J.F. Delaigle, B. Macq & C. Vleeschouwer, "A flexible video transmission system based on JPEG 2000 conditional Replenishment with multiples references", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 2007.
- [2] A.K. Diop, S.M. Farssi, K. Tall, I. Diop : "Reliable transmission of Motion JPEG 2000 by using motion estimation", *IEEE ICSPT 2013 International Conference on Signal Processing and Telecommunication. Journal of Intelligent Computing*, Volume 4, Issue: 1, March 2013, pp 23-29.
- [3] A.K. Diop, K. Tall, I. Diop, S.M. Farssi : "Adaptation of the method of estimation and motion compensation in the transmission of the video sequence in JPEG 2000" *International Journal of Distributed and Parallel Systems (IJDPs)* Vol.3, N°6, November 2012, pp 159-167.
- [4] F.W. Mounts. "A video encoding system with conditional picture-element replenishment". *Bell Systems Technical Journal*, 48(7):2545-2554, 1969.
- [5] V. Garcia. "Estimation de mouvement subpixelique par blocs adaptée à la couleur avec modèle de mouvement", *rapport de stage DEA Image Vision*, Septembre 2004.
- [6] Oge Marques, "Practical Image and Video Processing Using MATLAB" Chapter 22 Digital video processing techniques and applications 561-589. *John Wiley & Sons, Inc., Publication*, 2010.
- [7] Mohammad A. Khan et Mark J. T. Smith. Handbook of Image & Video Processing, "chapter 5.3 Fundamentals of Vector Quantization", pages 485-493. *Academic Press, al bovik edition*, 2000.
- [8] Z. Wang, A.C. Bovik et L. Lu. "Why is image quality assessment so difficult?", *IEEE International Conference On Acoustics, Speech, and Signal Processing*, 4:3313-3316, 2002
- [9] Z. Wang, A.C. Bovik, H. R. Sheikh et E. P. Simoncelli. "Image quality assessment: from error visibility to structural similarity", *Image Processing, IEEE Transactions on* 13(4):600-612, 2004.
- [10] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment", *IEEE Transactions on Image Processing*, vol.20, N°5, pp. 1185-1198, May 2011.



Author

Abdou Khadre DIOP was born the 10th of July 1979 in Pikine, Dakar, Senegal. He received the degree in Telecommunication and Electronics Engineering from the Gaston Berger University of Saint-Louis (Senegal) in 2005.

In 2012, he was employed as ATER at the Department of Technology in Information Communication of the Alioune Diop University of Bambey. In 2008, he joined the Laboratory of Biomedical Imaging & Bioinformatics for a joint PhD thesis. His current research areas cover transmission of the video sequence in JPEG 2000.