Reliable Transmission of Motion JPEG 2000 Codestream by using Motion Estimation

Abdou Khadre DIOP, Sidi Mohamed FARSSI, Khaly TALL, Idy DIOP
Laboratory of Medical Imaging & BioInformatics
Department of Computer Science
Ecole Supérieure Polytechnique
Cheikh Anta Diop University
Dakar, SENEGAL
djeylani2001@yahoo.fr

ABSTRACT: *The Motion JPEG 2000 (Part 3 of the JPEG 2000 standard) has a higher performance compared to traditional video compression standards [1], however the problem encountered in the transmission of Motion JPEG 2000 files is the loss of packets in the codestream Motion JPEG2000 files. Conditional Replenishment has been proposed [2], but controlling the movement is a major difficulty due to decomposition algorithms used by the DWT (Discrete Wavelet Transform).*

In this paper, we propose the use of the method based on the principle of motion estimation to reliably reconstruct the lost packets in the transmission of Motion JPEG2000 codestream with a lot of movement. An objective and subjective assessment of the proposed method was performed in terms of measurement of PSNR (Peak Signal to Noise Peak) and SSIM (structural similarity).

Keywords: Motion JPEG 2000, Conditional Replenishment, Motion Estimation, Motion Compensation, Order of Progression, PSNR, SSIM

Received: 2 December 2012, Revised 8 January 2013, Accepted 12 January 2013

1. Introduction

A video signal is 3D (2D + t) with temporal and spatial correlations. Temporal correlations are mainly related to the motion of objects over time. Spatial correlations are related to the particular contours of objects and reasons for textures.

While traditional video coders like H.264 or MPEG4 predictive coders are based on a motion compensation loop and encoding 2D prediction residuals by the DCT (Discrete Cosine Transform), Motion JPEG 2000 Part 3 of the JPEG 2000 standard that define the use of the JPEG 2000 standard for the treatment of video sequences is a concatenation of images in JPEG 2000 encoded using the DWT. And controlling the movement in video compression using the DWT is a major difficulty due to decomposition algorithms used by the DWT [3].

Faced with the transmission of the codestream Motion JPEG 2000 file represented as a sequence of packets, there is packet loss in the codestream of these files. Therefore we can get a bad quality in the reconstruction of the video file transmitted especially if the sequence has a lot of movement.

To alleviate the consequences of these transmission errors, we use the method of motion estimation, a method well known in the inter-frame coding of video sequences using traditional video compression, to reconstruct reliably lost packets in the codestream of Motion JPEG 2000.

To describe our work, the paper presents in the first part features of Motion JPEG 2000 and Motion JPEG 2000 packetization to locate the various packages in the codestream video file. The second part presents the method of motion estimation and motion compensation: method used to reconstruct lost packets. The final section presents the results obtained in terms of objective and subjective qualities. Finally we present the conclusion.

2. Motion JPEG 2000 Features

Motion JPEG 2000 is defined as Part 3 of the JPEG 2000 standard. It is an extension of the JPEG 2000 for the treatment of video sequences. The principle is that each frame of video is encoded in JPEG 2000. Figure 1 shows a Motion JPEG 2000 format MJ2 or MJP2 represented as a concatenation of images in JPEG 2000: JP2 or J2K.

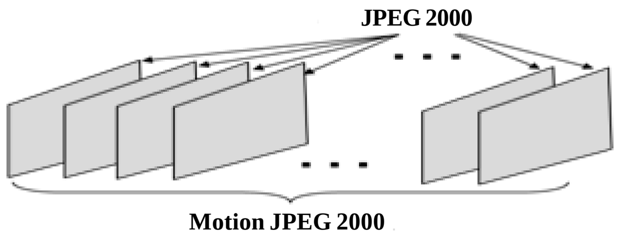

Figure 1. Composition of a Motion JPEG 2000

The allocation algorithm flow of the JPEG 2000 standard allows creating packages like all other allocation algorithms flow. Thus each packet is identified by four parameters: a quality level, a level of resolution, a component and a precinct [5].

The packetization is used to add to each packet codestream following a header containing all the coding parameters and in order of inclusion depending on the type of growth desired (e.g. by resolution, by quality level). Thus, it is possible to randomly have access to the codestream of each frame. So, we can determine the information in each packet (Figure 2).

3. Transmission of Motion JPEG 2000 Codestream

As a transmission technology is not 100% reliable, the transmission of MJ2 file can result in loss of packets in the codestream file. To reconstruct lost packets conditional replenishment is proposed [2]. In their work, they have extended the original conditional replenishment framework by increasing the number of references candidates to approximate the spatial zone to transmit. This proposed method permits a flexible transmission if the spatial zone presents no movement.

We propose the method of motion estimation and motion compensation used in the inter-frame coding of video sequences in MPEG4 or H.264 for the recovery of lost packets during transmission if the sequences present a lot of movement [6], [7]. The motion estimation and motion compensation techniques address the problem of estimating the movement between two frames: $f(x, y, t_1)$, which be called the anchor frame, and $f(x, y, t_2)$, which will be referred to as the target frame.

Motion representation methods vary in terms of the size of the region of support, which can vary among the Global model, the Pixel-based model, the Block-based model or the Region-based model. The Region-based model, used in the MPEG-4 standard and a region of support have been chosen, the problem shifts to the estimation of the model parameters. One of the most popular motion estimation criteria is based on the DFD (displaced frame difference). It consists in computing the difference in intensity between every point in the anchor frame and the corresponding point in the target frame [3]. Mathematically, the objective function can be written as:

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

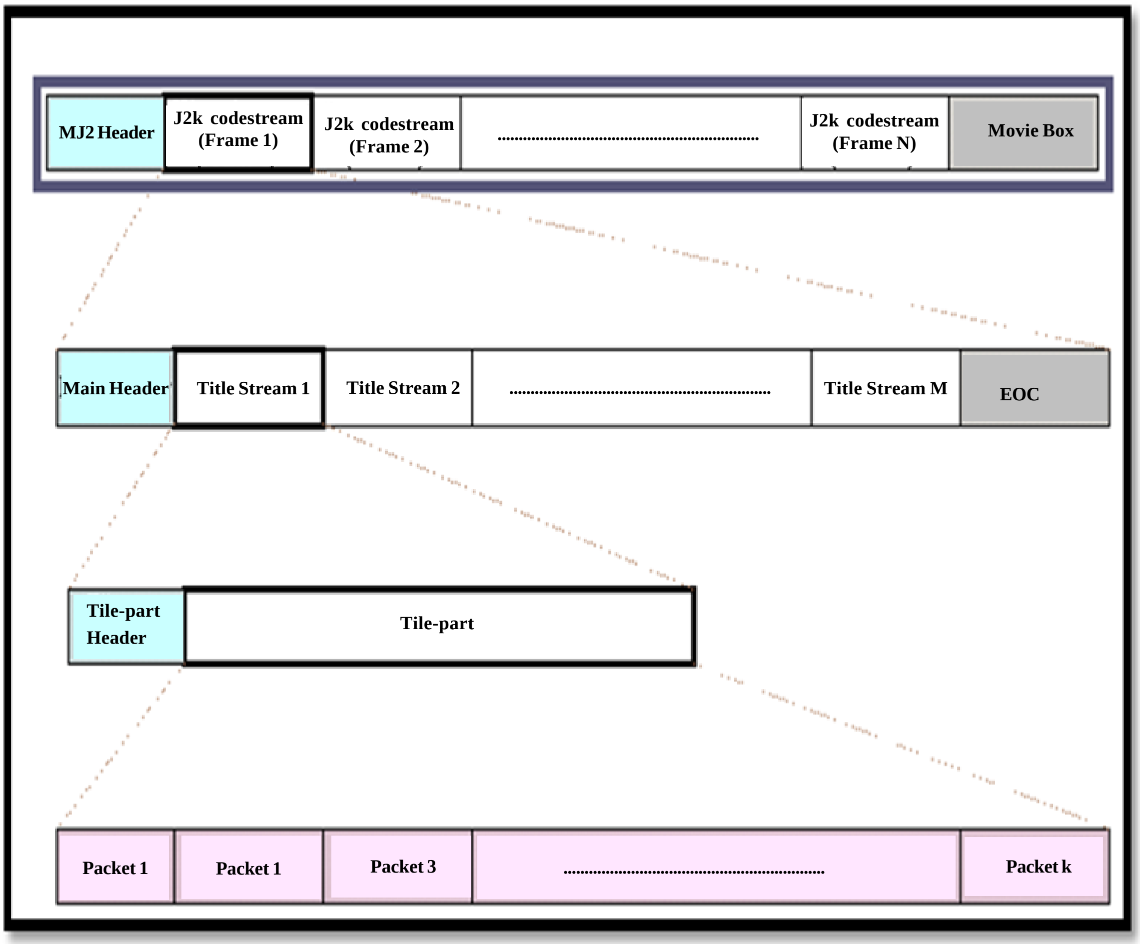

Figure 2. Structure of the Motion JPEG 2000 file

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

Using (1) we propose the following algorithm applied to two successive images can be shown in Figure 3.

Between these two images, the current image named target frame and the previous image, anchor image. If we lose packets in the codestream of the target image, conditional replenishment method is used to recover packets of anchor image for placement in the target image. After packets are placed in the target image, motion detector implemented from the equation Displaced Frames Difference (DFD) can locate the part of the image with movement. As well, movement is detected in a region of support at which we apply our method to reconstruct reliably lost packets.

In our algorithm, the compression format video sequence MJ2, the extracting images from a MJ2 to J2K file and the wrapping J2K images in a MJ2 file are obtained using the OPENJPEG[1] platform.

4. Results

The transmission method has been tested exhaustively, but we present the results on "*news.mj2*", a CIF video sequence (288 * 352 * 3 pixels per frame, 300 frames), transmitted at 25 fps. This sequence is characterized by a quasi-stationary movement of the camera. Figure 4 shows the target and the anchor frame. We can see the movement with the dancer in the background. Regarding the JPEG 2000 compression parameters, the sequence has been encoded with 1 quality layer, 6 levels of resolution, 3 components and 1 precinct so the codestream of each frame is constituted by 1 * 6 * 3 * 1 = 18 packets.

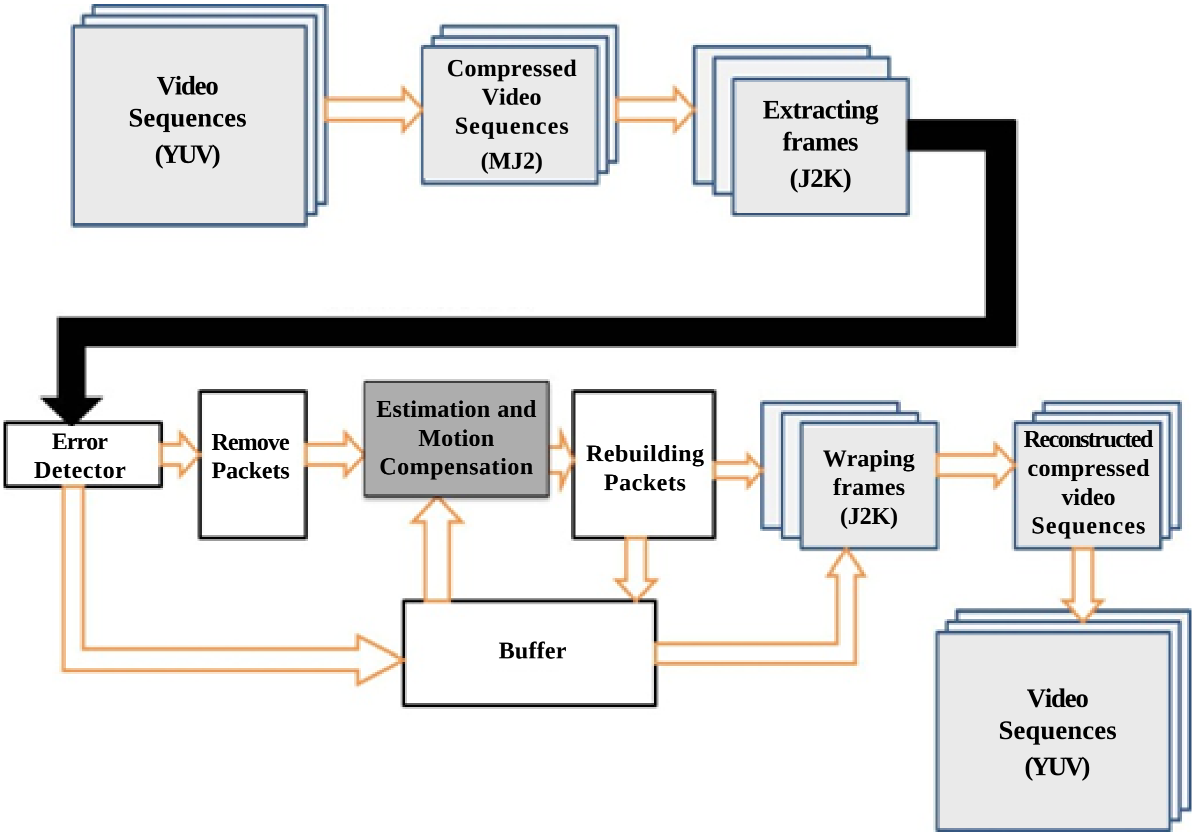

Figure 3. Algorithm using motion estimation and compensation method

Figure 4. Target frame (a) and anchor frame (b)

To transmit the codestream of these compressed video sequences obtained in format MJ2, it is possible to lose packets. If the lost packets are in the region of support where movement is detected, our method can reconstruct the lost packets in the target frame by using the anchor frame.

Figure 5 compares the PSNR at different bit-rates of the motion estimation and motion compensation method and the conditional replenishment method. PSNR_recRC is the PSNR value of the target frame reconstructed by conditional replenishment and PSNR_recEC is the PSNR value of the target frame reconstructed by motion estimation and motion compensation.

Equation (2) gives the PSNR between the original target image and the reconstructed target image.

$$PSNR = 10.\, Log_{10}\left(\frac{(2^{nb} - 1)}{MSE}\right) \tag{2}$$

The MSE (Mean Square Error) is obtained by (3)

$$MSE = \frac{1}{M.N} \sum_{n=1}^{N} \sum_{m=1}^{M} (X_{orig}(n,m) - X_{rec}(n,m))^2 \quad (3)$$

The PSNR value provides an objective assessment of the target frame, so we deduce that the transmission of Motion JPEG 2000 by using motion estimation is more reliable than the transmission of Motion JPEG 2000 by conditional replenishment. However it is possible to reconstructed images with good PSNR value unless they have a good perspective of the human observer. Thus it is necessary to resort to other method using the human observer.

Wang [4] 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):

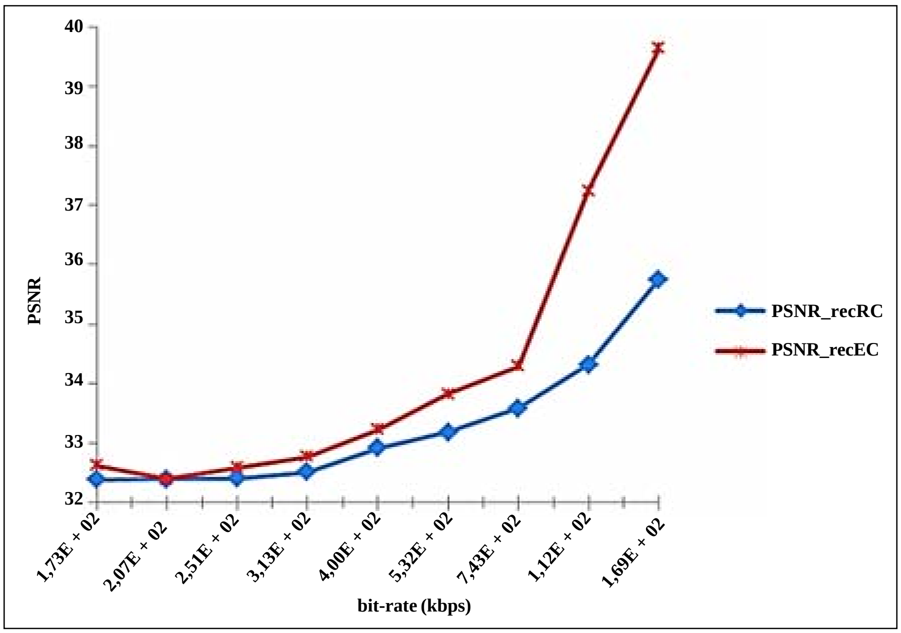

Figure 5. PSNR of target frame reconstructed by conditional replenishment and by motion estimation and motion compensation

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

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

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

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

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

Figure 6 compares the mean SSIM at different bit-rates of the motion estimation and motion compensation method and the conditional replenishment method. MSSIM_recRC is the mean SSIM value of the reconstructed target frame by conditional replenishment and MSSIM_recEC is the mean SSIM value of the reconstructed target frame by motion estimation and motion compensation.

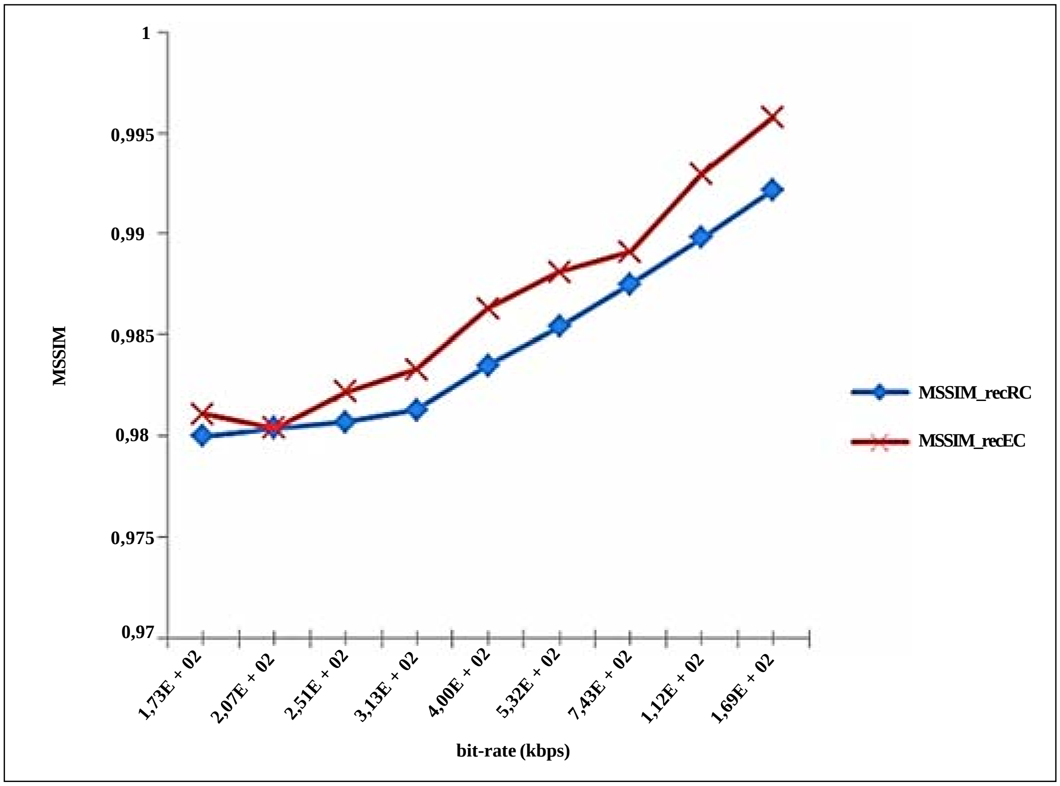

Figure 6. Mean SSIM of target frame reconstructed by conditional replenishment and by motion estimation and motion compensation

Figure 7. Region of support with movement

Figure 8. The reconstructed target frame by conditional replenishment

These objective and subjective assessments highlight the transmission reliability of the Motion JPEG 2000 codestream by using motion estimation and motion compensation method.

Figure 8, Figure 9 and Figure10 show respectively the region of support with movement, the reconstructed target frame by conditional replenishment and the reconstructed target frame by estimation and motion compensation.

5. Conclusion

In this work, we investigated the use of estimation and motion compensation to transmit reliably Motion JPEG 2000 file.

Figure 9. The reconstructed target frame by motion estimation and motion compensation

Although it is well known in the inter-frame coding sequences MPEG-4 and H.264, the method allows reconstructing lost packets in a target frame by using the anchor frame. To evaluate this method, we consider the PSNR expression for an objective assessment and the mean of the similarity structure for a subjective assessment. These results highlight the transmission reliability of the Motion JPEG 2000 codestream by using motion estimation and motion compensation.

As storage and transmission video surveillance content in JPEG 2000 are deployed of integrated solutions, these results encourage considering the use of motion estimation and motion compensation to reconstruct lost packets in video surveillance.

References

[1] Detlev Marpe, Valeri George, Hans L. Cycon, Kai U. Barthel. (2004). Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode, Wavelet Applications in Industrial Processing, Providence, RI, USA, 5266 of Proceedings of SPIE, p. 129-137.

[2] François-Olivier Devaux, Jérôme Meessen, Christophe Parisot, Jean-François Delaigle, Benoît Macq & Christophe De Vleeschouwer. (2007). A flexible video transmission system based on JPEG 2000 conditional Replenishment with multiples references, *International Conference on Acoustics, Speech, and Signal Processing* (ICASSP).

[3] Marques, O. (2011). Pratical Image and Video Processing Using MATLAB, Wiley IEEE, Florida Atlantic University.

[4] Wang, Z., Li, Q. (2011). Information content weighting for perceptual image quality assessment, *IEEE Transactions on Image Processing*, 20 (5) 1185-1198, May.

[5] Majid Rabbani, Rajan Joshi. (2002). An overview of the JPEG 2000 still image compression standard, *Signal Processing: Image Communication*, 17, p. 3-48.

[6] Christian Grava. (2003). Compensation de mouvement par réseaux neuronaux cellulaires : Application en imagerie médicale, PhD Institut National des Sciences Appliquées de Lyon, Formation doctorale: Images et Systèmes.

[7] Patrice Brault. (2005). Estimation de mouvement et segmentation d'images. Partie I : Estimation de mouvement par ondelettes spatio-temporelles adaptées au mouvement, PhD Université Paris-Sud XI, Ecole Doctorale Sciences et Technologies de l'Information des Télécommunications et des Systèmes.