# JPEG 2000 vs. JPEG in MPEG Encoding

**V.G. Ruiz, M.F. López, I. García and E.M.T. Hendrix** *

Dept. Computer Architecture and Electronics
University of Almería.
04120 Almería. Spain.
E-mail: vruiz@ual.es, mflopez@ace.ual.es, inma@ace.ual.es, eligius.hendrix@wur.nl
Tel: +34 950 015711

**Abstract**
The MPEG-2 standard is the most used codec for video compression. This paper investigates the consequence of replacing the JPEG core system of the MPEG-2 video codec by a progressive image codec, specifically JPEG 2000, so creating a video codec (VC) that can match any desired bit rate. The quality of compressed video sequences between MPEG-2 and VC is compared. Results show that our codec improves the quality of the decompressed images.

**Keywords**
MPEG-2, JPEG 2000, bit rate control.

## 1 Introduction

Digital video coding is one of the most important applications in the area of signal processing and telecommunications, because the amount of data that a digitized video signal produces is too large to be stored or transmitted efficiently without compression. The main features of a video coding system are: minimal image degradation, less computational resources requirements at the decoder than at the coder (asymmetry) and temporal and spatial (SNR and resolution) scalability at decoding time. Moreover, depending on the application for which the codec is used (Internet streaming, real-time transmission, digital storage, video-conference, etc.), other interesting properties may be: exact bit-rate control, resilience to errors, random access to individual pictures and minimal delay coding.

The MPEG (Moving Picture Experts Group) standards are the state of the art in video coding [1]. They exploit temporal and spatial redundancies of the image sequence to create a compressed bit-stream that represents the original sequence without a great visual degradation. MPEG-2 [2] is based on a motion estimation (ME) technique [3] and on the JPEG (Joint Photographic Experts Group) coding system [4]. ME allows MPEG-2 to reduce the temporal redundancy of the sequence of images and JPEG reduces the spatial redundancy in every residual single image. Recently, a new and powerful standard called JPEG 2000 [5] has appeared for still image coding. It is based on the discrete wavelet transform (DWT) instead of the discrete

*This work was supported by the Ministry of Education of Spain (CICYT TIC2002-00228).

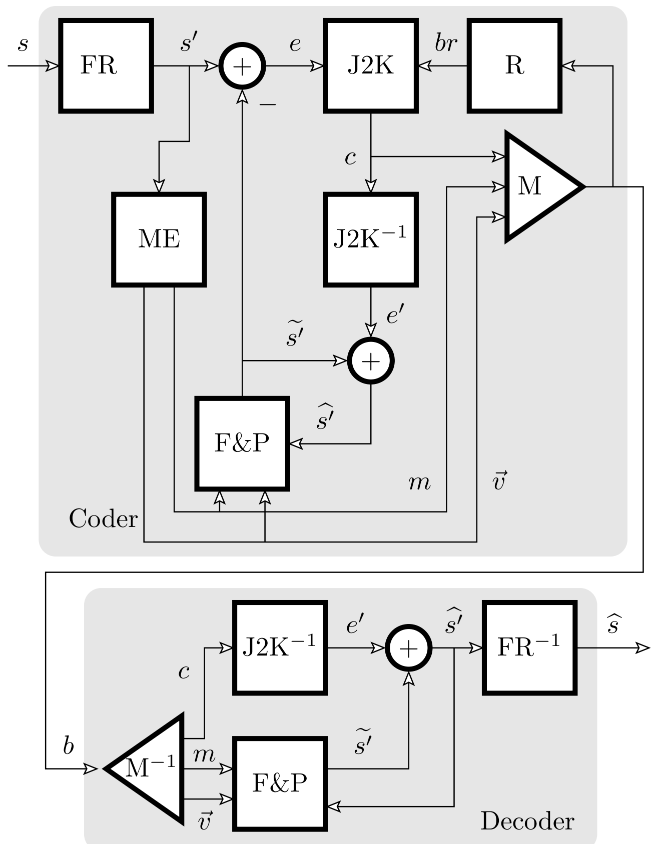


Figure 1: The proposed video codec (VC).

cosine transform (DCT) which is used by JPEG. This feature allows JPEG 2000 to increase the compression ratios and to obtain the exact bit rate more easily.

In this paper, the idea behind the proposed video codec (VC) is to replace the JPEG compression engine by a progressive image compressor such as JPEG 2000. VC inherits from MPEG-2 the codec asymmetry and the temporal random access. JPEG 2000 provides a very good bit rate control and robustness to bit-errors. This is essential essential in frameworks like Internet, where the packet transmission imposes specific bit rates.

The rest of the document is organized as follows. Section 2 describes VC, Section 3 presents experimental results and in Section 4 contains the main conclusions and the future work.

## 2 The Proposed Video Codec (VC)

Our codec fits into the framework of the MPEG-2 standard [2]. It is a hybrid system in which the JPEG 2000 coder (J2K for short) replaces the DCT, Q (quantization), $Q^{-1}$ (inverse quantization), $DCT^{-1}$ (inverse DCT) and VLC (variable length coding) modules in the MPEG-2 coder. The JPEG 2000 decoder ($J2K^{-1}$) replaces the VLD (variable length decoding), $Q^{-1}$ and $DCT^{-1}$ modules in the MPEG-2 decoder (see Figure 1). A sequence of images $s$ is processed GOP by GOP (group of pictures). The FR (frame re-order) module performs the image reordering necessary in each GOP leading to the re-ordered image sequence $s'$. VC uses the same block-based motion estimation (ME) module and predictive feedback loop as MPEG-2 to reduce the temporal redundancy in $s$. This is done by using the sequence $\widehat{s'}$ that the decoder has at the decoding time to generate $\widetilde{s'}$. Thus, for every image in $s'$ a prediction image in $\widetilde{s'}$ is built by the F&P (frame-store and predictor) module and a sequence of error images $e$ is computed. The error sequence is sent to the J2K module which generates the sequence $c$ of compressed residual images at the desired bit-rate (br). The bit-rate regulator (R) determines the size of every image in $c$. It exploits the progressive feature of JPEG 2000 to control the bit-rate accurately. Finally, the sequence $c$, data for the decision modes $m$ and the motion vectors $\vec{v}$ are multiplexed by M, generating the bit-stream $b$. From the MPEG-2 layer description [2], the DCT data (the block layer), the quantizer data (at the macro-block layer) and the slice layer, are substituted by data of J2K.

At the decoder, the $\text{J2K}^{-1}$ module restores the error images $e'$ and the F&P module creates the prediction images $\widetilde{s'}$. With this information, $\widehat{s'}$ is reconstructed. In the end, the $\text{FR}^{-1}$ module puts the images in the correct order.

We have selected the JPEG 2000 compressor because is progressive. This kind of codecs produce a bit stream that can be truncated at any point and a approximated full-resolution of the original image can be restored. This property is very suitable for our intentions because we can match any desired bit rate simply truncating the J2K's bit stream ($c$).

## 3 Results

Numerical and visual evaluations of VC are reported and a comparison to the TM5 implementation [6] of the MPEG-2 standard is given. In our experiments, the ME module and the bit allocation for every frame in the GOP in the VC and in MPEG-2 were identical. We have used the peak signal to noise ratio (PSNR) measure, defined (in dB) for each image of the sequence as

$$\text{PSNR[dB]} = 10 \log \frac{255^2}{\frac{1}{N}\sum_{i=1}^{N}(s[i] - \widehat{s}[i])^2} \quad (1)$$

for 8 bpp images, where $N$ is the number of points in an image, and $s[i]$ and $\widehat{s}[i]$ are points of the original and decompressed images, respectively. The test sequences are *akiyo*, *foreman* and *mobile*. They have $352 \times 288$ points per frame (CIF), 300 frames per sequence, and 30 frames/second.

Depending of the sequence, the amount of movement if different. Thus, *Akiyo* is a static talking-head, without zooms or camera pam motions. *Mobile* has slow moving objects and the camera is zooming in, and *foreman* is an extreme case, displaying a periods of time in which the camera makes very complex non-linear pam motions. Two different GOP sizes (6 and 12) have been tested. We used three different bit-rates: 0.6 Mbps, 1.2 Mbps and 2.6 Mbps.[1]. These bit rates are low compared with the normal working point of the MPEG-2 standard, but if we want to measure better the visual quality of the reconstructions we need to use them.

Numerical results are summarized in Table 1 and graphically illustrated in Figure 3 and show that VC always outperforms MPEG-2, and in *akiyo*, the PSNR value of VC for 0.6 Mbps is better than the PSNR value of MPEG-2 for 2.6 Mbps. These numerical differences were measured in the RGB domain.

For the visual comparison, we have selected one frame of each sequence (GOP size = 12 and bit-rate = 0.6 Mbps). Images are displayed in Figure 2. It can be seen that the blocking artifacts produced by the block transform at low bit-rates are eliminated using VC, although macroblocks

[1]1.2 Mbps and 2.6 Mbps are the bit-rate for the video stream in a typical 1.5 Mbps and 3.0 Mbps MPEG-2 stream, respectively.

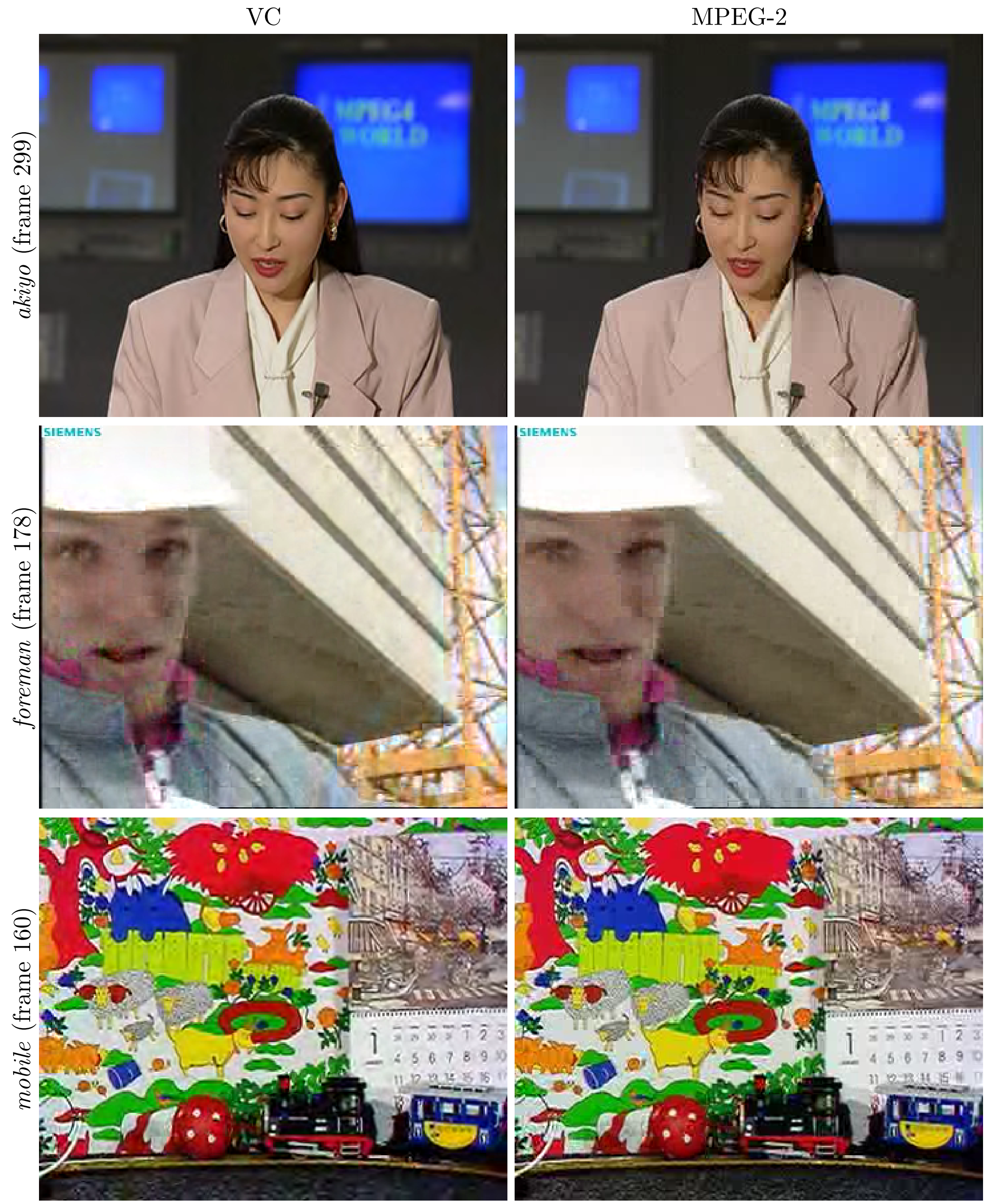


Figure 2: A visual comparison of MPEG-2 and VC. GOP size = 12, and bit-rate = 0.6 Mbps.

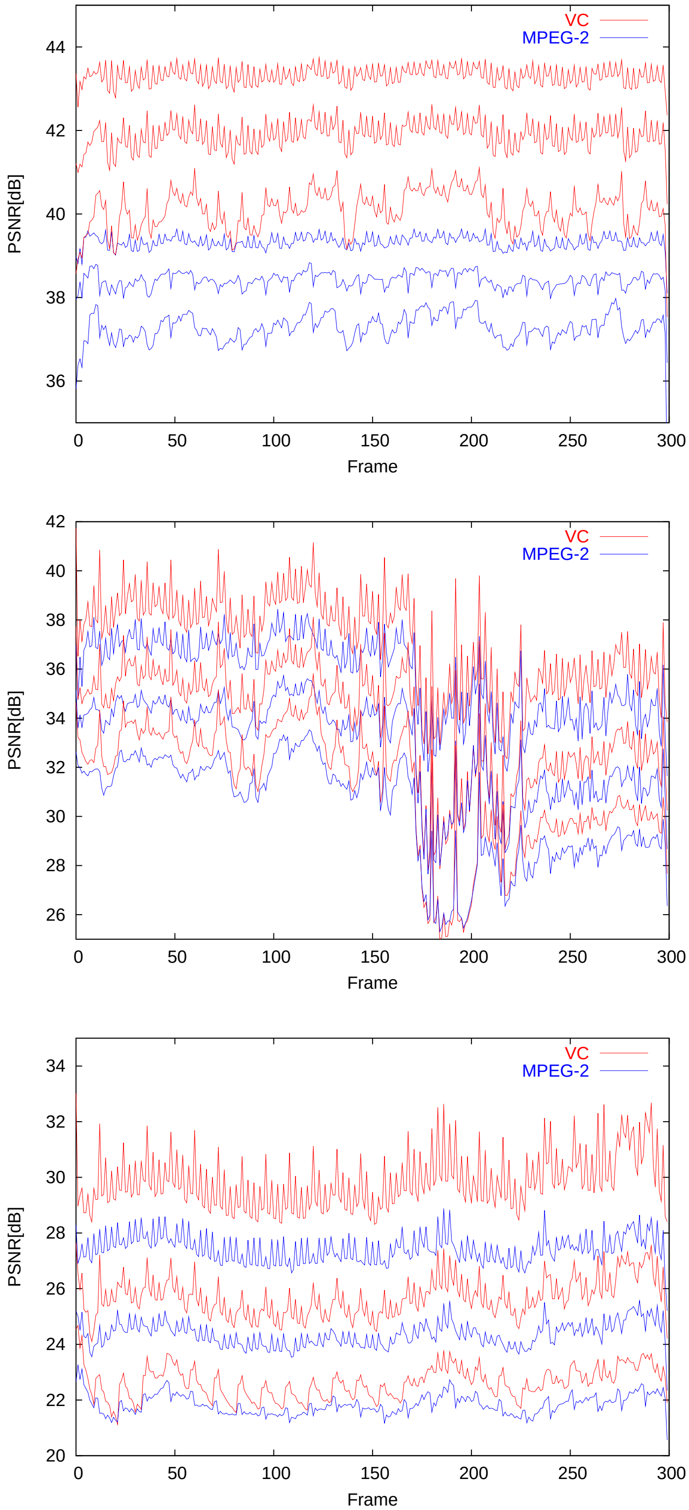


Figure 3: Numerical results for 0.6 (bottom curve), 1.2 and 2.6 Mbps (upper curve). GOP size = 12. From top to bottom: *akiyo*, *foreman* and *mobile*.

Table 1: Average values of PSNR[dB].

GOP size = 12

| | *akiyo* | | *foreman* | | *mobile* | |
|---|---|---|---|---|---|---|
| Mbps | MPEG-2 | VC | MPEG-2 | VC | MPEG-2 | VC |
| 0.6 | 37.2 | **40.0** | 30.2 | **31.2** | 21.8 | **22.6** |
| 1.2 | 38.4 | **41.8** | 32.8 | **33.9** | 24.3 | **25.6** |
| 2.6 | 39.3 | **43.3** | 35.7 | **37.2** | 27.4 | **29.8** |

GOP size = 6

| | *akiyo* | | *foreman* | | *mobile* | |
|---|---|---|---|---|---|---|
| Mbps | MPEG-2 | VC | MPEG-2 | VC | MPEG-2 | VC |
| 0.6 | 36.3 | **39.3** | 29.8 | **30.9** | 20.7 | **21.3** |
| 1.2 | 38.0 | **41.2** | 32.4 | **33.8** | 23.6 | **25.1** |
| 2.6 | 39.1 | **43.1** | 35.4 | **37.1** | 26.9 | **29.5** |

can still be distinguished. This effect is produced by the use of non-overlaped macroblocks in the motion compensation.

## 4 Conclusions

We have tested the efficiency of a video compressor based on the use of a progressive image codec. Our codec fits into the framework of the MPEG-2 standard where the DCT core system is replaced by the JPEG 2000 codec. The rest of the modules remain unchanged. Our main conclusions are that VC outperforms MPEG-2 (numerically spoken) in every case tested and produces better quality images.

VC has also a more exact bit rate control that MPEG-2. In our experiments we have selected the same bit rate that the MPEG-2 encoder to compare the reconstructions, but we can chose any other bit rate. This freedom at the encoding time is other of the main advantages of our VC.

## 5 Future Work

There are other several characteristics of the JPEG 2000 bit stream that have no been explored in this paper and that can be used to produce new features, like the scalability, in the compressed video stream. For example, the SNR progressive representation of $e$ (see Figure 1) could be used to create a compressed bit stream with SNR scalability. There is no necessity of use multiple layers. We could truncate the compressed bit stream to produce a optimal reconstruction of the original video sequence. This is very useful in the transmission of compressed sequences in the Internet, because most of links of this kind of networks have a capacity that is variable and for this reason, unknown at encoding time.

## References


[1] T. Sikora, "MPEG Digital Video-Coding Standards," *IEEE Signal Processing Magazine*, vol. 14, no. 5, pp. 82–100, September 1997.

[2] International Organisation for Standardisation (ISO/IEC JTC1 SC29 WG11), *International Standard ISO/IEC 13818*, 1996.

[3] J.S. Baras, Y.-Q. Zhang, and S. Zafar, "Predictive Block-matching Motion Estimation Schemes for Video Compression (Digest)," in *IEEE International Conference on Consumer Electronics*, June 1991, pp. 300–301.

[4] G.K. Wallace, "The JPEG Still Picture Compression Standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–40, 1991.

[5] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 Still Image Compression Standard," *IEEE Signal Processing Magazine*, pp. 36–58, September 2001.

[6] International Organisation for Standardisation (ISO/IEC JTC1 SC29 WG11), *Test Model 5*, 1993.