

MOTION-COMPENSATED OVERSAMPLED FILTERBANKS FOR ROBUST VIDEO CODING

J. C. Chiang, M. Kieffer and P. Duhamel

LSS – CNRS – Supélec – Université Paris-Sud
Plateau de Moulon, 91192 Gif-sur-Yvette, France

ABSTRACT

During the past years, many efforts are devoted for efficient coding of video. However, this high coding efficiency results in an increased sensitivity to transmission errors. This is the motivation for developing techniques allowing altogether to transmit video sequences efficiently and to be robust to the transmission errors. This paper proposes a joint source-channel coding scheme using motion-compensated oversampled filterbanks (OFBs). Previous work for robust transmission using OFBs have concentrated on still images. Here, OFBs are combined with motion compensation techniques. The redundancy introduced by the overcomplete representation is employed for error correction of the motion compensated frames. Experimentations have shown that performances are acceptable even when the compressed stream is transmitted through a binary symmetric channel with crossover probability of 10^{-2} .

1. INTRODUCTION

With the rapid advance of wireless communication, the need to transmit video over mixed internet and wireless channels is increasing. As a result, reliable transmission over wireless links has gained growing interest. The main difficulty is to mitigate the effect of both packet loss and channel transmission errors. Many solutions have been proposed, but the problem remains still open.

For standard video coding schemes, such as *MPEG-4* or *H264*, hybrid coding techniques are employed to remove the temporal and spatial redundancy in the video frames. A block-based motion-compensation (MC) is realized between consecutive frames and a 2-D DCT is performed in the spatial domain after motion compensation. Finally, entropy coding is realized. When transmitted over a noisy channel, the use of such DPCM feedback loop causes channel errors to propagate in consecutive frames until the next refresh frame is received. This error propagation comes in addition with the one due to the entropy coding. In order

This work has been partly supported by the VIP RNRT project. Thanks are also due to J. Woods, who kindly provided the initial $2D + t$ coding program

to increase the robustness of these hybrid coders, several techniques were proposed: forward error correction, automatic repeat request, reversible variable length codes, multiple description coding, etc. Joint source-channel coding (JSCC) methods have also been proposed [1, 2] to facilitate the trade-off between source and channel coding efficiency.

Alternative $t + 2D$ MC subband coding schemes have been proposed [3, 4]. In such coders, a MC temporal filtering (MCTF) is performed along the motion trajectories before realizing spatial filtering. In the MC-EZBC video coder proposed by [5] a sub pixel invertible MCTF is realized. Then, a wavelet subband decomposition is employed on the motion-compensated frames. Finally, the resulting spatio-temporal subband coefficients are encoded by EZBC [6]. A very efficient compression can be achieved using this scheme, but is obtained at the cost of a high sensitivity to channel noise. Classically, robustness is improved by the use of multiple description coding [7]. In [8], the robustness is achieved by temporal oversampling of a motion-compensated lifting scheme. However, both schemes are robust to packet losses, but less to random transmission errors, that are likely to arise in the wireless links.

This paper proposes a $t + 2D$ video coder based on oversampled filterbanks (OFBs) [9] which provide a redundant description of the signal. This redundancy has been put at work to correct transmission errors in a noisy environment [10, 11] when applied to still images [12]. The basic idea of this paper is to analyze the Motion Compensated frames of the video sequence using an OFB in order to be able to correct transmission errors, as was the case in a still image transmission context.

Section 2, briefly recalls the structure of a $t + 2D$ video coding scheme before presenting the proposed scheme. Section 3 recalls that OFBs can be seen as joint-source channel coders. For this purpose, a channel model is defined in Section 3.1, then Section 3.2 presents the error correction scheme. Finally, Section 4 illustrates the performance of the proposed technique by some simulations.

2. MOTION-COMPENSATED $T + 2D$ VIDEO CODING

In $t + 2D$ video coding, efficient compression is achieved by removing the temporal redundancy between the consecutive frames of a group of frames (GOF) and the spatial redundancy in the resulting MC frames. Unlike hybrid coding, where previously decoded frames are used as reference for the current frame, MCTF is performed along the motion trajectory. Temporal low and high frequency frames are generated for each two or more consecutive input frames by filtering. The short-term dependency between these consecutive frames are removed in such a way. Furthermore, with a pyramid decomposition structure, the long-term dependency in a GOF is removed by further filtering the low frequency frames.

2.1. Original scheme

Figure 1 presents the general architecture of a $t + 2D$ video coding system. For every GOF, the motion vectors are estimated first. MCTF is performed recursively for a given filtering level. Then the spatial analysis is applied on the MC frames. Then, quantization and entropy coding is realized, after bitrate allocation between the motion field and the motion compensated frames to produce the final bitstream that is sent to channel. To increase robustness against transmission errors, usually channel codes are incorporated.

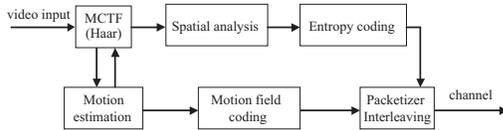


Fig. 1. Original $t + 2D$ coding scheme

2.2. Proposed scheme

Figure 2 shows the proposed $t + 2D$ video coding scheme based on OFBs. As in the scheme presented in Section 2.1, two parts may be distinguished, namely MCTF and filtering of the MC frames. For the MC, the techniques developed by Chen [5] is employed. The motion vectors (MVs) are coded using lossless DPCM and adaptive arithmetic code. The JSCC scheme described in [12] is applied on these MC frames (redundancy is introduced spatially). After filtering using an OFB, quantization is performed. No entropy coding is realized, in order to prevent propagation of channel errors in the bitstream. The bit assignment and the choice between scalar quantizer and pyramid vector quantizer (PVQ) is realized for every subband, based on the Shoham-Gersho algorithm. Finally, the bitstream is generated by interleaving the output of quantized subband signals and coded motion vectors. Headers and motion vector are particularly sensitive to transmission errors. This paper assumes that they are transmitted and received without any error (they are assumed to

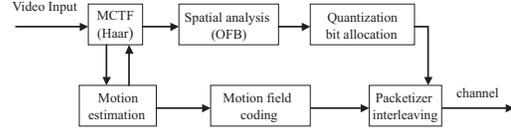


Fig. 2. MC OFB-based video coding scheme

be strongly channel coded). It will be shown in Section 3 that OFBs may be used to correct MC texture errors. At decoder side, a synthesis OFB is employed in conjunction with error correction to decode the MC frames. Inverse temporal filtering is then realized on each GOF to obtain again decoded frames.

3. ROBUSTNESS TO TRANSMISSION ERRORS

The robustness of the proposed video coding scheme is based on OFBs. Here, subband transformation and redundancy introduction are combined in a single step using OFBs. In OFBs, the output signal is an overcomplete representation of the input signal. This redundancy has been employed in [10] to correct part of the noise introduced in the subbands by the communication scheme represented by a Gaussian Bernoulli-Gaussian (GBG) channel model, *i.e.*, a channel introducing a noise consisting of a mixture of Gaussian background noise and Bernoulli-Gaussian (BG) impulse noise. The GBG channel model characteristics will be first briefly recalled before sketching the error correction technique using OFBs and the GBG channel model.

3.1. GBG joint channel model

In separate source and channel coding, the channel is the part of the communication scheme that is between the output of the channel coder and the input of the channel decoder. Similarly, the joint channel is situated between the output of the analysis OFB and the input of the synthesis OFB. The joint channel model gathers the quantization error introduced by source coding and the errors remaining after channel decoding. In [14], such a memoryless communication channel described as the sum of Gaussian plus BG noises has already been studied. Under these assumptions, for every subband generated by the OFB, the relation between $y(n)$ and $\tilde{y}(n)$, the input and output of the joint channel can be written as

$$\tilde{y}(n) = y(n) + a(n) + b(n), \quad (1)$$

where $b(n)$ is some Gaussian noise (representing quantization errors) and $a(n)$ is an impulse noise (corresponding to uncorrected channel errors). The Gaussian noise is assumed to be of zero mean and variance σ_g^2 , while the impulse noise is modeled as BG $a(n) = \xi(n)b'(n)$, where $\xi(n)$ stands for a Bernoulli process, an i.i.d. sequence of zeros and ones with $\text{prob}(\xi(n) = 1) = p$, and $b'(n)$ represents a Gaussian

noise with zero mean and variance σ_i^2 , such that $\sigma_i^2 \gg \sigma_g^2$. The probability density function (pdf) of the channel noise $c(n) = a(n) + b(n)$ can be expressed as

$$p(c) = (1 - p) G(c, 0, \sigma_g^2) + p G(c, 0, (\sigma_g^2 + \sigma_i^2)), \quad (2)$$

with $G(c, m, \sigma^2)$ denoting a Gaussian pdf having mean m and variance σ^2 . A GBG channel is thus characterized by there 3 parameters σ_g^2 , p and σ_i^2 .

If L quantized subbands are sent over a binary symmetric channel with crossover probability p_b , in fact, L such GBG channel models have to be built, depending on the number of bits per sample b assigned to the subband and the characteristics of the quantization that has been used. The parameter σ_g^2 corresponds to quantization noise only and is easily derived from the distortion introduced by quantization. The sample error probability p and the variance of the impulse noise σ_i^2 depend on b and on the quantizer outputs when scalar quantization is considered. For PVQ or product-code PVQ, the estimation of these 3 parameters is intricate, but can be performed. The dimension D of the quantized vectors has also an impact on p and σ_i^2 , see [15] for more details.

3.2. Impulse correction scheme

Figure 3 summarizes the impulse error detection and correction algorithm proposed in [10]. This scheme is implemented before the synthesis stage of the OFB.

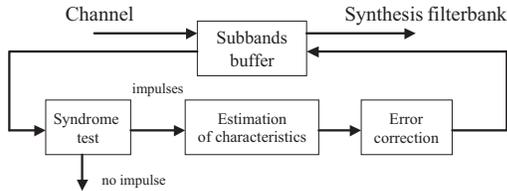


Fig. 3. Error correction scheme

The algorithms consists of two steps. First, an hypothesis test is used to determine whether impulse errors are present at a given time instant ℓ and optionally to estimate their number. For computational simplicity, this test is based on the norm of samples of the syndrome taken over a sliding window of length $N_V + 1$, where N_V is the order of the parity-check matrix associated to a given analysis OFB. Once ℓ is determined, it is possible to obtain maximum *a posteriori* estimates of the subband k and the amplitude a of the impulse error. For more details, see [11].

4. SIMULATION RESULTS

The original $t + 2D$ program was kindly provided by Dr. John Woods, and our version with improved robustness is organized as follows: Each GOF contains 8 frames. A 3-level pyramid MCTF is realized with Haar filters. During this filtering, if the number of unconnected pixels be-

Scheme tuning	Condition	PSNR (dB)		
		Y	U	V
Reference, $p_b = 0$	no imp.	35.94	41.83	44.19
MC-OFB, $p_b = 0$	no imp.	35.49	41.90	42..23
Reference, $p_b = 0.01$	with imp.	Crash		
MC-OFB, $p_b = 0.01$	no imp.	34.13	41.76	42.05
MC-OFB, $p_b = 0.01$	no corr.	26.23	38.61	39.23
MC-OFB, $p_b = 0.01$	with corr.	30.60	41.40	41.66

Table 1. Simulation results. The reference scheme is an MC-EZBC source coder followed by a rate 1/2 convolutional code

tween two consecutive frames is larger than a given threshold, MCTF is replaced by motion-compensated prediction. The obtained motion vectors are encoded by the lossless DPCM and adaptive arithmetic code before being strongly channel coded. Then, the MC frames are spatially filtered by an evenly-stacked DFT modulated OFB, with oversampling ratio 8/6 and real prototype filter order of 47 [16]. All video components Y, U and V are filtered separately with the same OFB. The quantization step involves SQ or PVQ, depending on the result of a bit-rate optimization taking the channel effects into account. When PVQ is used, the vector components are taken from the same subband, at the largest possible distance. It should be noted that PVQ is used here to allow smaller bit rates than SQ; this is absolutely required for a good optimization). Finally, the data is sent over a binary symmetric channel with crossover probability p_b .

Bit assignment is realized based on Shoham-Gersho algorithm. Rates correspond to scalar quantization rates or several different dimensional (4,8,16,32 and 64) PVQ rates and distortions correspond to the distortions introduced by the quantization *and the errors introduced by the channel* after error correction. For a given global rate, the bit assignment varies according to the crossover probability in the channel.

The `foreman.sif` sequence is encoded at 30 fps with a bitrate of 600 kbps (rate on the channel). Table 1 compares the performances of the proposed technique to those of a reference scheme consisting of an MC-EZBC followed by a rate 1/2 convolutional code protecting the EZBC encoded data. In both cases, the MV part is assumed to be transmitted and received without any error while the MC frames will be corrupted by channel errors. Without any transmission error, the optimization resulting from Shoham-Gersho algorithm indicates that the first subband for Y, U and V should be quantized by a SQ and the remaining subbands by a 64-D PVQ. However when the transmission error is non zero, the assignment changes and lower dimensional PVQ are employed to obtain a trade-off between error correction efficiency (which decreases with the dimension of the PVQ) and image quality.

Table 1 shows that when tuned without error, the perfor-

mances of our compression algorithms are sufficient for a practical use, despite the inherent bitrate increase due to the oversampling. In the case of $p_b = 0.01$, the decoders associated with MC-EZBC followed by a convolutional code are not able to reproduce any frame due to the residual channel errors which were not corrected by channel decoder. However, in our proposed scheme, although the performances decrease with 1 % of bit errors (which is a very large percentage for a transmission system!) but are reasonable. This number is the target for the error correction algorithm, which cannot perform better. It has to be noted that, even if the error correction process is not activated, the scheme that has been tuned for robustness is quite robust to transmission errors, since there is no error propagation, and since the tuning takes into account channel induced errors. The performances are improved when the error correction scheme is put at work, as can be seen in the last line of Table 1. Better performances could certainly be obtained by a better tuning of the decoding algorithms, but, it can be seen that in this example, the error correction algorithm improves the luminance by 4.3 dB.

The performance of this error-correcting algorithm is illustrated on Figure 4 displaying the same frame, before and after correction.

5. CONCLUSIONS

In this paper, $t + 2D$ video coder based on MC OFBs has been proposed. OFBs provide a JSCC tool that enables robust video transmission. The structured redundancy provided by OFBs is used to correct transmission errors. First experimental results have been given in this paper. More detailed comparisons with other $t + 2D$ coding scheme have to be performed in a noisy context.

6. REFERENCES

- [1] R. E. Van Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated and joint source-channel coding," *Proc. IEEE*, vol. 87, no. 10, pp. 1734–1750, 1999.
- [2] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE transactions on Image Processing*, vol. 11, no. 9, pp. 1043–1052, 2002.
- [3] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transaction on Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.
- [4] S.-J. Choi and John W. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.
- [5] P. Chen, "Fully scalable subband/wavelet coding," *PhD thesis*, 2003, Rensselaer Polytechnic Institute.
- [6] S.-T. Hsiang, "Highly scalable subband/wavelet image and video coding," *PhD thesis*, 2002, Rensselaer Polytechnic Institute.
- [7] I.V. Bajic and J.W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. on Image Processing*, vol. 12, pp. 1211–1225, 2003.
- [8] C. Tiller, . B. Pesquet-Popesc, and M. Van der Schaar, "Multiple description scalable video coding," *EU-SIPCO*, 2004.
- [9] Z. Cvetković and M. Vetterli, "Oversampled filter banks," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1245–1255, 1998.
- [10] J. C. Chiang, M. Kieffer, and P. Duhamel, "Oversampled filterbanks seen as channel codes : impulse noise correction," in *Proceedings of ICASSP*, 2003, pp. 249–252.
- [11] F. Labeau, J. C. Chiang, M. Kieffer, P. Duhamel, L. Vandendorpe, and B. Marq, "Oversampled filter banks as error correction codes: Theory and impulse noise correction," *submitted to IEEE Transactions on Signal Processing*, Jan. 2004.
- [12] J. C. Chiang, M. Kieffer, and P. Duhamel, "Robust image transmission using oversampled filterbanks," in *Proc. Eusipco*, 2004.
- [13] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.
- [14] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Transactions on Communication*, vol. 44, no. 2, pp. 145–147, 1996.
- [15] A. C. Hung, E. K. Tsern, and T. H. Meng, "Error-resilient pyramid vector quantization for image compression," *IEEE Trans. Image Processing*, vol. 7, no. 10, pp. 1373–1386, 1998.
- [16] M. Harteneck, S. Weiss, and R. W. Stewart, "Design of near perfect reconstruction oversampled filter bnaks for subnd adaptive filters," *IEEE Transactions on Circuit and Systems-II*, vol. 46, no. 8, pp. 1081–1085, August 1999.



Fig. 4. Reconstructed frame (a) without correction, (b) with correction