ROBUST VIDEO TRANSMISSION USING H.264 AND REAL-VALUED BCH FRAMES

A. Ouled Zaid, M. Kieffer, C. M. Lee and P. Duhamel

LSS – CNRS – SUPELEC – Paris-Sud University
Plateau de Moulon, 91192 Gif sur Yvette, France

ABSTRACT

With the increasing use of multimedia technologies, video coding requires higher performances as well as new features. To address this need, the latest ITU-T video coding standard, H.264, has been developed. The motion compensation algorithm is the key of the improvements over H.263+. This paper proposes an H264-based algorithm which does not require the motion vectors to be transmitted. It extends previous results obtained in the H263+ context. The main purpose of this work is to check whether our procedure of reestimating the motion vectors at the receiver is compatible with the initial H264 scheme efficiency. It is shown that this property is obtained at some cost in terms of bit rate for comparable PSNRs, but that our new H264-based scheme has about the same performance level as the plain H263+ coder. This is essentially a preliminary work, which shows that the most sensitive part of the bit stream can be removed, thus allowing new robust coders to be developed[1].

1. INTRODUCTION

The new H.264 standard [1] achieves considerably higher coding efficiency compared to older standards, such as, *e*.g., H263+ [2]. These performances are partly due to the adoption in Inter-picture coding of more refined motion models and sophisticated coding modes. A macroblock (MB) can be split in various subareas to each of which is assigned a single motion vector (MV). Multiple reference-picture motion compensation is also available. These predictive coding tools contribute to the overall performances of H.264 but at the cost of a reduced robustness. Indeed, Inter picture coding encodes only the motion compensation residuals and the MVs. Any error affecting the MVs will propagate until non-predictively coded information appears in the bitstream. This results in substantial, and sometimes catastrophic deterioration of the reconstructed video signal. Consequently, research on video coder robustness is facing major new challenges.

The problem of balancing the tradeoff between compression efficiency and robustness has received much attention. For example, the design of soft decoding of VLC codes [3] was developed taking into account the structure of the bitstream generated by H.263+. In [4], UEP (Unequal Error Protection) was developed by partitioning the bitstream into different classes of priority. While state-of-the-art approaches improve the robustness of video coders, they share the common limitation of being still very sensitive to MV transmission errors.

Recently, a new approach avoiding any motion vector transmission was developed for the H.263+ coding scheme [5]. This approach is based on BCH frame expansion. This expansion introduces structured redundancy, by imposing some specific property to the MB of the images to be encoded. This property has to be at least partly retrieved on the reconstructed images at decoder side. It is thus possible to reestimate the MVs by ensuring that the reconstructed images have recovered the imposed property. Consequently, MVs transmission prove to be useless, since they may be reestimated at decoder side. Here, it is shown that the same type of technique may be applied to the H.264 coding chain, in order again to avoid transmission of the MVs. However, contrary to the modified H.263+ coding scheme, we take advantage of the adaptive MB coding modes to select Intra mode at the MB level whenever the MV reestimation process is inefficient.

This paper first gives a brief overview of the BCH frame expansion. In Section 3, the H.264 coding scheme is briefly described and more attention is accorded to MV estimation method. Section 4 deals with the modified H.264 coding scheme where frame expansion is incorporated. Experimental results are then given in Section 5, followed by some conclusions and perspectives in Section 6.

2. BCH FRAME EXPANSION

A BCH frame expansion as introduced in [6] is a linear operator $\mathbf{F}^{\text{BCH}}_{(n,k)}$ which expands vectors of $\mathbf{R}^k$ into vectors of $\mathbf{R}^n$ such as

$$\mathbf{c}_{(n)} = \mathbf{W}_{(n)}\mathbf{P}_{(n,k)}(\mathcal{A})\mathbf{W}^{-1}_{(k)}\mathbf{i}_{(k)} = \mathbf{F}^{\text{BCH}}_{(n,k)}\mathbf{i}_{(k)}$$

where $\mathbf{W}_{(n)}$ is a DFT (or DCT) transformation matrix and $\mathbf{P}_{(n,k)}(\mathcal{A})$ is a matrix that pads properly $n-k$ zeros. These

[1]This work is partly supported by RNRT VIP.

zeros are inserted at $n-k$ locations specified in the set $\mathcal{A}$. Assume that an expanded word $\mathbf{c}_{(n)}$ is transmitted over a noisy channel and that $\mathbf{r}_{(n)}$ is received. To determine whether $\mathbf{r}_{(n)}$ has been transmitted without any error, it sufficies to check whether

$$\mathbf{s}_{(n-k)}(\mathbf{r}_{(n)}) = \mathbf{R}_{(n-k,n)}(\mathbf{W}_{(n)})^{-1}\mathbf{r}_{(n)} = \mathbf{H}_{(n-k,n)}\mathbf{r}_{(n)}$$

is zero. Here $\mathbf{H}_{(n-k,n)}$ plays thus the role of a parity-check matrix, and $\mathbf{s}_{(n-k)}$ that of a syndrome. In the case of a bi-dimensional signal such as images of a video sequence, the redundancy has to be introduced by a product frame expansion applied on each $\mathbf{I}_{(k,k)}$ block of the pictures which constitutes the initial video sequence. The result will be expanded blocks $\mathbf{M}_{(n,n)}$ of size $n \times n$. The syndrome is now a matrix given by

$$\begin{aligned}\mathbf{S}_{(n,n)}(\mathbf{M}_{(n,n)}) = \mathbf{W}_{(n)}^{-1}\mathbf{M}_{(n,n)}(\mathbf{W}_{(n)}^{-1})^{\mathbf{T}} \\ -\mathbf{F}_{(n,k)}^{\text{BCH}}\mathbf{M}_{(n,n)}(\mathbf{F}_{(n,k)}^{\text{BCH}})^{\mathrm{T}}\end{aligned} \quad (1)$$

$n\times n-k\times k$ of the syndrome components will take non-zero values whenever $\mathbf{M}_{(n,n)}$ is affected by some transmission errors.

Note that in the case of Fourier BCH frames, the block padding with $n-k$ zeros should respect the Hermitian symmetry in order to keep the coded signal in the field of the real numbers.

The following section recalls the basics of the H.264 coding process with the MV estimation concept. This is necessary to understand our approach, since we must ensure that the decoder has the same MV estimates as the encoder.

3. H.264 CODER AND MV ESTIMATION

3.1. Encoder (forward path)

An input picture is processed at the MB level. Each MB $\mathbf{M}_{(n,n)}$ is encoded in Intra or Inter mode. In both cases, a predicted MB $\mathbf{X}_{(n,n)}$ is evaluated. In Intra mode, $\mathbf{X}_{(n,n)}$ is formed using the previously encoded MBs in the same picture. In Inter mode, $\mathbf{X}_{(n,n)}$ is formed by motion-compensated prediction from reference picture(s).

The prediction $\mathbf{X}_{(n,n)}$ is subtracted from the current MB $\mathbf{M}_{(n,n)}$ to produce a residual texture MB $\mathbf{T}_{(n,n)}$, that is transformed, quantized and entropy encoded.

3.2. Encoder (reconstruction path)

After re-scaling and inverse transform, an estimate of the texture MB $\tilde{\mathbf{T}}_{(n,n)}$ is obtained at decoder side. It corresponds to a distorted version of $\mathbf{T}_{(n,n)}$. The prediction MB $\mathbf{X}_{(n,n)}$ is added to $\tilde{\mathbf{T}}_{(n,n)}$ to obtain the reconstructed MB $\tilde{\mathbf{M}}_{(n,n)}$. Reconstructed reference picture is finally created from a series of MBs $\tilde{\mathbf{M}}_{(n,n)}$.

3.3. Finding the best MV

MV estimation is performed for motion compensated prediction. H.264 coder supports motion compensation block sizes ranging from 16×16 to 4×4. A separate motion vector is then required for each partition. In our study, to simplify the MV estimation process, we limit ourselves to 16×16 MB partition. Considering the high cost of transmitting motion parameters, a MV $\mathbf{m}$ is coded differentially versus a MV predictor $\mathbf{p}$ [1]. For each block or MB the MV is determined by integer-pixel position followed by sub-pixel refinement. The integer-pixel motion search as well as the sub-pixel refinement returns the MV that minimizes

$$J(\mathbf{m},\lambda) = SAD(\mathbf{M}_{(n,n)}, \tilde{\mathbf{M}}_{(n,n)}(\mathbf{m})) + \lambda \cdot R(\mathbf{m}\text{-}\mathbf{p}) \quad (2)$$

with $\mathbf{m} = (m_x, m_y)^T$ being the MV, $\mathbf{p} = (\mathbf{p}_x, \mathbf{p}_y)^{\mathrm{T}}$ being the prediction for the MV, and λ being the Lagrange multiplier. The rate term $R(\mathbf{m}\text{-}\mathbf{p})$ corresponds to the number of bits assigned to the motion information.

4. MODIFIED H.264 CODER

We now discuss the MV robustness technique which was recently applied to the H.263+ coding scheme and integrate this approach to the H.264 coder. Our objective is here to take advantage of the H.264 architecture, to balance the tradeoff between compression efficiency and robustness.

The adopted approach consists of procedures to select the MVs with the goal of avoiding the motion information transmission.

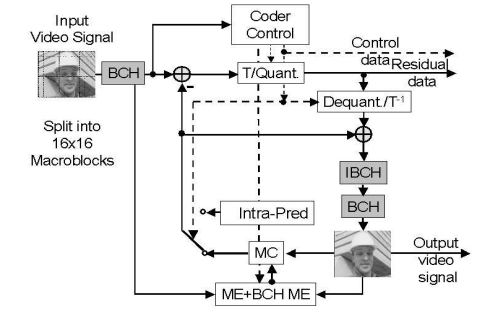
The modified version of H.264 coding/decoding schemes are respectively illustrated in Fig. 1 (a) and Fig. 1 (b). The first step consists in introducing redundancy (based on BCH frames) before entering the encoder. In the figure, the expansion and synthesis on a BCH frame is done through the BCH and IBCH blocks. In order to avoid some problems, and to check the feasability of this new scheme, we limit ourselves to half-pixel precision to make the MV reestimation. Current work is undertaken in order to avoid this restriction, and to obtain better performances. However, the results reported here already allow to evaluate the cost (in terms of bitrate) of avoiding the MV transmisison.

4.1. MV reestimation based on BCH frames

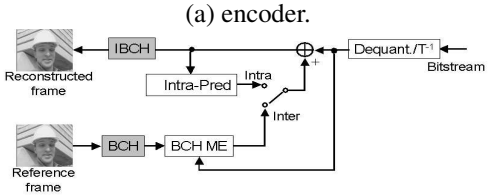
Given a MB $\mathbf{M}_{(n,n)}$ to be coded, the MV reestimation serves to determine the MV $\hat{\mathbf{m}} = (m_x, m_y)^{\mathbf{T}}$ that minimizes the Frobenius norm $||\mathbf{S}_{(n,n)}||_{\mathrm{F}}$ of the syndrome associated to the reconstructed MB $\tilde{\mathbf{M}}_{(n,n)}$ which is the estimate of $\mathbf{M}_{(n,n)}$.

$$\tilde{\mathbf{M}}_{(n,n)}(\mathbf{m}) = \tilde{\mathbf{T}}_{(n,n)} + \mathbf{X}_{(n,n)}(\mathbf{m})$$

At the decoder side $\mathbf{X}_{(n,n)}(\mathbf{m})$ is extracted from a search area in the previously reconstructed reference picture. Assume that the search range $\mathrm{N}_{(l,l)}$ in the reference picture is



(a) encoder.



(b) Decoder.

Fig. 1. H.264 coding/decoding scheme without MV transmission.

available, the MV $\hat{\mathbf{m}}$ can be reestimated based on the same BCH frame expansion principle. In fact, it is clearly seen that, in the absence of quantization noise, the MV estimate $\hat{\mathbf{m}}_{\text{BCH}}$ of $\hat{\mathbf{m}}$ can be obtained from the minimum of the objective function $J_{\text{BCH}}(\cdot,\cdot)$:

$$J_{\text{BCH}}(\mathbf{m}, \tilde{\mathbf{T}}_{(n,n)}(\hat{\mathbf{m}})) = \quad (3)$$
$$||\mathbf{S}_{(n,n)}(\tilde{\mathbf{T}}_{(n,n)}(\hat{\mathbf{m}}) + \mathbf{X}_{(n,n)}(\mathbf{m}))||_{\text{F}}$$

For nonpathological images, the property that has been imposed in $\mathbf{M}_{(n,n)}$ by frame expansion, will be retrieved.

4.2. Robust MV estimation

Due to quantization noise in the texture, there is no guarantee that $\hat{\mathbf{m}}$ is the argument of the global minimum of $J_{\text{BCH}}(\mathbf{m})$. A loss of agreement can occur between $\hat{\mathbf{m}}$ and the reestimated MV at coder side. To solve this problem, the classical MV estimation is replaced by a robust one. The principle is that the decoder is simulated at the encoder, so that the decoder estimate will be chosen at the emitter side. Let $\ell = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_N\}$ be a MV list generated by the original H.264 MV estimation process. This list satisfy

$$J(\mathbf{m}_1, \lambda) \leq J(\mathbf{m}_2, \lambda) \leq \ldots \leq J(\mathbf{m}_N, \lambda).$$

In the robust MV estimation process, the $\tilde{\mathbf{T}}_{(n,n)}(\mathbf{m}_1)$ residual data to be transmitted is replaced by $\mathbf{T}_{(n,n)}(\mathbf{m}_{\underline{k}})$ which satisfies

$$\underline{k} = \min\{k \mid \mathbf{m}_k = \arg\min_{\mathbf{m}} J_{\text{BCH}}(\mathbf{m}, \tilde{\mathbf{T}}_{(n,n)}(\mathbf{m}_k))\},$$

In some situations, no such $\underline{k}$ may exist. In such case, the Intra MB coding mode is automatically selected. Moreover, even when such a match is found, it may happen that the compression cost exceeds that given by Intra MB coding mode. In this case, there is no interest in realizing this motion compensation and again the Intra MB coding mode is applied.

5. EXPERIMENTAL SETUP AND RESULTS

5.1. The experimental conditions

All of the above concepts were introduced within version JM 7.0 of the H.264 reference software. For our simulations we have selected the first 101 pictures of QCIF resolution sequence Foreman. The spatial resolution of each frame is 176 pixels by 144 lines. The first picture is Intra coded, while all others are Inter coded. The UVLC entropy coder was used for all our tests, a search range of 8 and half-pixel MV precision. To evaluate the performance of our MV removed approach, we use the PSNR as a function of the resulting bitrates. Three BCH frames have been put at work: a Fourier-based BCH frame $\mathbf{F}^{\text{BCH-F}}_{(16,15)}$ with $\mathcal{A} = \{8\}$ and two DCT-based BCH frames $\mathbf{F}^{\text{BCH-D}}_{(16,15)}$ with $\mathcal{A} = \{4\}$ and $\mathcal{A} = \{12\}$.

5.2. Tuning our method

Fig. 2 depicts the PSNR-Rate curves using the above cited video coding methods and modified H.263+ coder with robust reconstruction of MV using a DCT-based BCH frame $\text{F}^{\text{BCH-D}}_{(16,15)}$ with $\mathcal{A} = \{8\}$. First concentrate on the comparison of the results obtained with our various BCH codes. All redundancies are similar, hence the results can be straightforwardly compared. Among all BCH codes, the DCT with the fourth coefficient forced to zero is the worst one. This is due to the fact that the $\mathcal{A}$ is chosen in low frequencies, hence blocking effects and texture problems can appear in the image which is fed into the encoder. The only choice for such a redundancy location with DFT is 8, since it is the only transform value which is real-valued. It is seen to perform better than DCT(4), but worse than DCT(12) which is the best choice. This is explained as follows : there is a tradeoff in the choice of this location : If very high frequencies are chosen, the initial content of the texture is very small at this frequency, and the criterion based on the norm of the syndrome is not very discriminant. However, when going to lower frequencies, the MV estimation lacks efficiency. This is why, in the DCT case, the best choice is in the mid-high frequencies (remember that, for Fourier, 8 is the highest possible frequency).

5.3. Comparison with other cases

Note that, for comparison purposes, all methods here have been restricted to a MV estimation with a pixel accuracy. The robust procedure, applied to H264 is seen to obtain a large improvement over the same procedure applied to H263+ (3 to 5 dB). This is attributed to the possibility of using the Intra mode whenever this mode is more efficient than

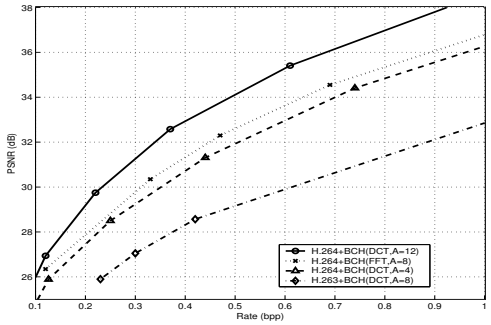


Fig. 2. PSNR/Rate curves using our method : BCH((FFT, A=8),(DCT, A=4), (DCT, A=12)) and robust H.263+ coder.

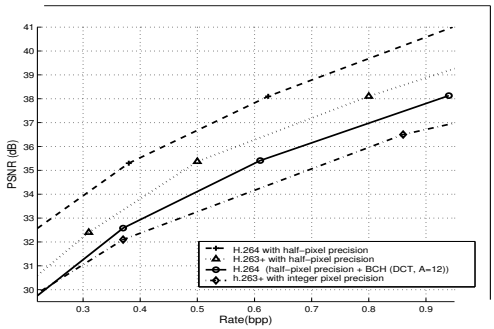


Fig. 3. PSNR/Rate curves using our method BCH(DCT, A=12), H.264 and H.263+ with half-pixel precision coders.

Motion Prediction from a rather non adapted region (for the purpose of using the Froebenius norm of the syndrome) as a criterion. Our results in the case of DCT-based BCH frames with $\mathcal{A} = \{12\}$ have been compared to those obtained using the same H.264 and H.263+ coders with halp-pixel precision (without BCH frame expansion). Fig. 3 shows that the performance decreases a lot : about 2.5 dB compared to the plain H264 scheme and 1 dB compared to the plain H263+ scheme. This fact has several origins : (i) the size of the original pictures was extended before BCH expansion (this was due for a honest comparison). (ii) the expansion introduces some redundancies to the video sequence. (iii) Inter MB coding mode is replaced by the Intra coding mode if no possible MV can be found that can be reconstructed at decoder side. However, the robustness provided by Intra coding may be costly in bit rate. Thus, the scheme of switching between Intra coding and Inter coding, so as to achieve the right balance between compression efficiency and robustness, is very important in our method. The final comparison may be more relevant : It turns out that our robust scheme based on H.264 with half-pixel precision is almost as efficient as the plain H263+ encoder with integer-pixel MV precision in the case of low bitrates and it performs better in higher bitrates. We are thus able to obtain a video encoder which does not require any MV transmission, with H363 like performance. Therefore, this algorithm seems to be a good basis for obtaining robust video codecs, if combined with other robustification tools, such as joint source/channel decoding using source semantics [3].

6. CONCLUSION

It has been recognized for a long time that MV was very sensitive to any transmission error. Thus, proposing algorithms avoiding this bottleneck deserves consideration. The performance that we obtained can clearly be improved, but we have shown that a BCH frame expansion combined with H264 had the potential of avoiding any MV transmission, at the cost of H263-like performances.
Furthermore, an analysis of the several factors that contribute significantly to the increase of bitrate have been detected, and will be improved in future work. First, the rate distortion could eventually be more adapted to this new situation. Then (and mainly) the performance of our algorithm can be substantially improved by taking into account a quarter-pixel motion compensation. This will help in improving the nominal source encoder. Finally, technique has to be combined with soft source or source/channel decoding in order to obtain textures with a good quality even in severely distorted channels.
It shoud be noted that the generated bitstream does not contain any motion information, the remainder is identical to the original H.264 bitstream. As a consequence, the decoder is no more in conformity with the standard.

7. REFERENCES

[1] ITU-T Recommendation, "Advanced Video Coding," FINAL Committee Draft, Document JVT-E022 11496-10, H.264/ISO/IEC, September 2002.

[2] ITU-T Recommendation, "Video Coding for Low Bitrate Communication," Tech. Rep., H.263 Version 2, 1998.

[3] H. Nguyen and P. Duhamel, "Compressed image and video redundency for joint source-channel decoding," *Golobecom*, 2003.

[4] U. Horn, B. Girod, and B. Belzer, "Scalable video coding for multimedia applications and robust transmission over wireless channels," in 7^{th} *Int. Workshop on Packet Video*, March 1996.

[5] C. M. Lee, M. Kieffer, and P. Duhamel, "Robust Motion Vectors and Texture Transmission for the H263 Video encoder family," in *Proc. of PCS*, Saint–Malo, 2003.

[6] P. G. Casazza, "The art of frame theory," *Taiwanese J. Math.*, vol. 4, no. 2, pp. 129–202, 2000.