Statistical Reconstruction for Predictive Video Coding

Catarina Brites[1], Vitor Gomes[2], João Ascenso[3], Fernando Pereira[4]

Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001 Lisbon, Portugal
[1,3,4]{catarina.brites, joao.ascenso, fp}@lx.it.pt
[2]vitorgomes@gmail.com

***Abstract*— Substantial rate-distortion (RD) gains have been achieved in video coding standards by increasing the encoder complexity while maintaining the decoder complexity the lowest possible. On the other hand, the alternative distributed video coding (DVC) approach proposes to exploit the video redundancy mostly at the decoder side, keeping the encoder as simple as possible. One of the most characteristic DVC tools is the statistical reconstruction of the DCT coefficients, which plays a similar role to the inverse scalar quantization (ISQ) in predictive codecs. The main objective of this paper is to propose a statistical reconstruction approach for predictive coding (notably the H.264/AVC standard) as a substitute to ISQ, thus creating a coding architecture with a mix of predictive and distributed coding tools. Experimental results show that the proposed statistical reconstruction solution allows achieving Bjontegaard bitrate savings up to 2.4% regarding the ISQ based H.264/AVC High profile codec.**

***Index Terms*—H.264/AVC, predictive video coding, distributed video coding, inverse scalar quantization, statistical reconstruction**

I. Introduction

As for the previous video coding standards, the H.264/AVC standard [1] adopts a predictive coding paradigm combining temporal prediction with a spatial transform, quantization and entropy coding to achieve good rate-distortion (RD) performance. The complexity associated to this process lies mostly at the encoder side, keeping the decoder as simple as possible. An alternative video coding approach, the so-called *distributed video coding* (DVC), aims at lowering the complexity associated to the video encoders by exploiting the correlation of video signals mainly at the decoder side [2]. While many well know tools used in the predictive video coding paradigm have been integrated in DVC solutions, it became clear at some stage that also some novel tools typically used in DVC, such as the *statistical reconstruction*, could be integrated in predictive video coding. These tools may be used with the expectation of improving the predictive video coding RD performance, eventually also changing the asymmetric balance between encoder and decoder complexity.

In predictive video coding, the decoded frame is obtained by summing the Intra/Inter predicted frame (obtained with the modes/motion vectors sent by the encoder) with the residual frame (obtained by entropy decoding followed by inverse scalar quantization and inverse DCT of the residual). However, it is possible to employ a (DVC based) statistical reconstruction as a substitute of the inverse scalar quantization (ISQ) assuming that the side information (SI) is the Intra/Inter predicted frame and characterizing the correlation between the frame to be coded and the SI by a statistical model. In the context of the DVC paradigm, SI is an estimate of the original frame (source) to code, made at the decoder, based on already decoded information [2] and its quality as a major impact on the final RD performance. The same happens with the predictive coding SI, this means the prediction frame, which again has a major impact on the final RD performance. While in DVC the SI is only available at the decoder, in predictive coding the prediction SI is available at both sides.

In this context, the main goal of this paper is to propose a statistical reconstruction approach for predictive coding, notably in the context of the H.264/AVC standard, as a replacement to the usual (normative) ISQ approach, thus creating a video coding architecture where a mix of predictive and distributed coding tools coexists. The proposed statistical reconstruction solution adopts a Laplacian correlation model for the discrete cosine transform (DCT) residuals and estimates the model parameter best fitting the DCT residuals. This approach exploits the correlation noise statistics between the source and the prediction taken as the SI, thus improving the reconstruction process of the quantized DCT coefficients, and thus the final RD performance.

This paper is organized as follows: Section II briefly reviews the relevant aspects of the H.264/AVC quantization and DVC reconstruction processes. Next, Section III presents the proposed video codec architecture while Section IV describes in detail the novel video coding tools. Experimental results are presented and analysed in Section V. Finally, Section VI concludes this paper including also possible future work.

II. Background Review

This section briefly presents some relevant background information on the H.264/AVC quantization and DVC reconstruction processes.

A. H.264/AVC Quantization

The regular (non-normative) quantization process used in the H.264/AVC reference software [3] quantizes the residual DCT coefficients $R(u,v)$ according to (1) and (2), generating a block of quantized residual DCT coefficients, $R_q(u,v)$.

$$|R_q(u,v)| = (|R(u,v)|A(Q_M,u,v) + f2^{15+Q_E}) \gg (15+Q_E) \quad (1)$$

$$\text{sign}\{R_q(u,v)\} = \text{sign}\{R(u,v)\} \quad (2)$$

In (1), $Q_M = QP \bmod 6$, $Q_E = QP/6$, where QP is the quantization parameter, (u,v) stands for the DCT coefficient position within the block, $A(Q_M,u,v)$ is a H.264/AVC tabled value associated to the quantization operation, which depends on the QP value and DCT coefficient position, and f is the parameter controlling the

978-1-4799-6139-9/14/$31.00 ©2014 IEEE

quantization bin width around zero (the so-called *dead-zone*); typically, f is 1/3 for Intra blocks and 1/6 for Inter blocks [3]. In (1), $\gg$ represents a binary shift right, which is equivalent to a division operation in integer arithmetic. In (2), $sign\{x\}$ stands for the signal operator.

The ISQ process has to recover a specific value to represent the decoded quantization bin. In H.264/AVC, the ISQ of the residual DCT coefficients is obtained from

$$R_r(u,v) = \{[R_q(u,v)B(Q_M,u,v)] \ll Q_E + 2^3\} \gg 4 \qquad (3)$$

where $B(Q_M,u,v)$ is a H.264/AVC tabled value associated to the ISQ operation, which depends on the QP value and DCT coefficient position.

B. DVC Statistical Reconstruction

Again, the DVC statistical reconstruction has the target to recover a specific value to represent the decoded quantization bin. However, the DVC statistical reconstruction approach [2] exploits the correlation model between the original source and the SI to minimize the mean-squared error (MSE) of the reconstructed samples given the decoded quantization bin q' as in (4)

$$X_r = E[x|q',Y] = \frac{\int_l^u x\, p_{x|Y}(x|Y)\, dx}{\int_l^u p(x|Y)\, dx} \qquad (4)$$

where X_r is the reconstructed DCT coefficient, Y is the corresponding SI DCT coefficient, $E[.]$ is the expectation operator and l and u represent the lower and upper bounds of q', respectively. In (4), the conditional probability density function $p_{x|Y}(.)$ models the residual statistics between corresponding coefficients of the original and SI frames; as usual in the DVC literature, $p_{x|Y}(.)$ is assumed to be a Laplacian distribution [2].

III. Predictive Video Codec with Statistical Reconstruction: Architecture

The high-level encoder and decoder architectures of the proposed predictive video codec with statistical reconstruction are illustrated in Fig. 1 and Fig. 2. The proposed codec architecture is based on a standard H.264/AVC codec design where the highlighted modules correspond to the major contributions of this paper briefly described next:

- ***Optimal Transform, Scaling and Quantization:*** This step takes the prediction P (which is playing the role of the SI in a DVC codec) and brings it to the DCT domain while quantizing it with QP = 0; the quantization process is applied as in the regular H.264/AVC codec, which means the transform and quantization processes cannot be separated.
- ***Residual Statistical Modelling:*** This step estimates the Laplacian model parameter best fitting the received DCT residuals; this is very different from a DVC codec where no residuals are ever received and thus they have to be estimated at the decoder.
- ***Statistical Reconstruction:*** Finally, the source DCT coefficients are statistically reconstructed given the transformed P and the statistical model previously obtained. Note that, in the proposed codec, the reconstructed source DCT coefficients still have to be fed to the Scaling and Inverse Transform module while in the regular H.264/AVC codec that module acts over the residual DCT coefficients, R_r.

Since this is a predictive codec where the encoder and decoder have to be always in perfect prediction synchronism, all modules inserted at the decoder have to be replicated at the encoder to obtain the same reference (decoded) frames and guarantee the mentioned synchronism. In the next section, the novel modules are described in detail.

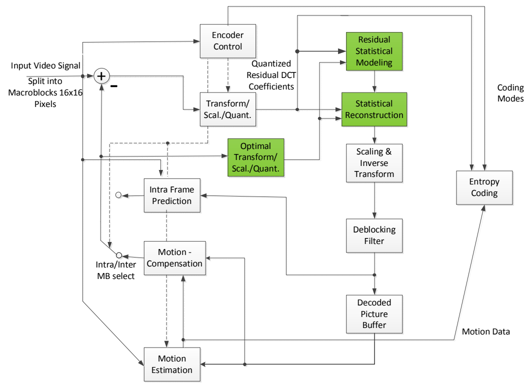

Fig. 1 High-level encoder architecture of the proposed video coding solution.

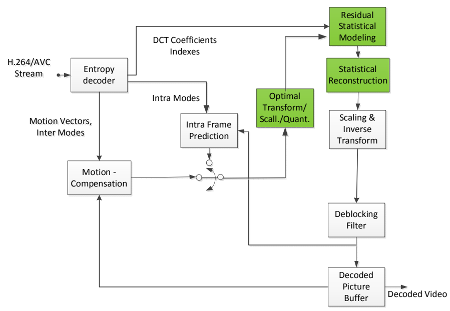

Fig. 2 High-level decoder architecture of the proposed video coding solution.

IV. Predictive Video Codec with Statistical Reconstruction: Novel Coding Tools

This section describes the operational details of the three modules associated to the novel reconstruction.

A. Optimal Transform, Scaling and Quantization

This module aims at converting the pixel based prediction (created with the encoder available and decoder received coding modes and motion vectors) to the DCT domain while quantizing it with QP = 0, i.e. the finest quantization possible, thus the attribute 'optimal'. This 'ideally' quantized DCT domain prediction avoids any quantization error, thus allowing to obtain the best transform domain 'SI', while reusing the transform, scaling and quantization operations performed in a regular H.264/AVC codec. In this context, each 4×4 luma block within a prediction macroblock (MB) is first transformed using a 4×4 integer DCT transform. After, the prediction DCT coefficients are scaled and quantized according to (1) and (2), generating a 4×4 block of quantized prediction DCT coefficients, P_q; $P_q(u,v)$ is also known as *level* or *quantization bin*. The ISQ of the prediction DCT coefficients is obtained from (3) with R_q replaced by P_q.

B. Residual Statistical Modelling

As usual in predictive video coding, a Laplacian distribution as in (5) is used to model the residual DCT coefficients distribution, R [4]; in (5), α stands for the Laplacian distribution parameter.

$$p(R) = \frac{\alpha}{2}\exp(-\alpha|R|) \tag{5}$$

The α parameter can be estimated using the maximum likelihood method [5] as in (6) where N represents the number of coefficients at the given DCT band for the full frame and R_k is the kth DCT coefficient value of that band:

$$\hat{\alpha}_{ML} = \frac{N}{\sum_{k=1}^{N}|R_k|}. \tag{6}$$

From (6), different α values are obtained for each residual DCT band, but the same value for the same band in all the blocks of the full frame. However, to better adapt to the changing statistics within a DCT band, this means along space/frame, it is proposed here to (decoder) estimate the α parameter at the DCT coefficient level, i.e. each residual DCT coefficient will have a different α value associated to it. In this case, adopting a finer modelling does not have any rate implications as the α parameter values are decoder estimated. Thus, the proposed α parameter estimation is performed at the DCT coefficient level for each coefficient within each 4×4 (luma) block in a MB as follows:

- **Residual DCT coefficients ISQ** – First, the residual DCT coefficients are inverse quantized according to (3).
- **DCT coefficients statistics accumulation** – Then, the sum of the absolute values of $R_r(u,v)$ for each DCT band (u,v) over the N_{prev_blocks} 4×4 blocks encoded/decoded so far (within a MB) is computed according to

$$V(u,v) = \sum_{k=1}^{N_{prev\_blocks}} |R_{r,k}(u,v)|. \tag{7}$$

In (7), N_{prev_blocks} can vary between 1 and 16. While it would be possible to consider all the previously coded/decoded MBs, e.g. in the frame or sequence, in the α estimation process, this would likely contemplate R_k values whose order of magnitude is quite different from the one in the current MB (thus outlier residuals). In statistical model parameter(s) estimation processes, outliers are in general responsible for lowering the statistical model accuracy and, therefore, should be avoided. For this reason, the α estimation technique proposed here considers only the data meanwhile available for the current MB.

- **α parameter estimation** – Finally, the α parameter for the (u,v) DCT coefficient is obtained from

$$\hat{\alpha}(u,v) = \frac{N_{prev\_blocks}}{V(u,v)}. \tag{8}$$

Although sharing some similarity with (6), the α parameter estimation approach proposed in (8) allows a finer adaptation to the residual changing statistics within a band as each DCT coefficient has a specific α parameter associated to it.

As the proposed statistical reconstruction is expected to improve the predictive video coding reconstructed quality, it makes sense to apply it as the MB coding process evolves to allow creating better predictions. Note that when Intra 4×4 prediction is used, prior reconstructed samples in adjacent blocks are used to create the prediction for a given 4×4 block. Thus, to perform the DCT coefficients (statistical) reconstruction as the coding process evolves, it is necessary to update $V(u,v)$ as the MB coding process takes place.

The $\hat{\alpha}(u,v)$ values obtained will be then used by the statistical reconstruction module to obtain the reconstructed DCT coefficients, as it will be seen in the next section.

C. Statistical Reconstruction

Finally, the statistical reconstruction module includes two main steps, the *DCT coefficients reconstruction bin bounds computation*, and the *DCT coefficients reconstruction*, which will be described in detail in the following.

1) *DCT Coefficients Reconstruction Bin Bounds Computation*: The (source, not residuals) DCT coefficients reconstruction bin bounds are obtained from the residual DCT coefficients reconstruction bin bounds and the prediction DCT coefficients, since this data is available at both the encoder and decoder sides. For this purpose, the following steps are performed:

- **Residual DCT coefficients reconstruction bin width computation** – First, the width of the residual DCT coefficients reconstruction bin, i.e. the quantization step size $\Delta(u,v)$, is computed. Although the quantization step size can be computed from the QP sent in the bitstream, the resulting value will not be scaled to the order of magnitude of the DCT coefficients for which the reconstruction bin bounds have to be computed; recall that quantization is performed over scaled DCT coefficients (see Fig. 1). Thus, it is proposed here to compute $\Delta(u,v)$ as the difference between the inverse quantized values of any two adjacent residual DCT coefficients bins for a given band; note that the ISQ operation is performed as in (3).
- **Residual DCT coefficients reconstruction bin bounds computation** – The lower (L_R) and upper (U_R) bounds of the bin where the residual DCT coefficient is reconstructed are obtained from (9)-(11). Those equations express the fact that the standard H.264/AVC inverse scalar quantization recovers the residual DCT coefficients at a distance f×Δ of the bin lower bound and the dead-zone width is equal to 2Δ(1–f) [3].

$$\begin{cases} U_R(u,v) = R_r(u,v) + \frac{f2^{15+Q_E}}{2^{15+Q_E}}\Delta \\ L_R(u,v) = U_R(u,v) - \Delta \end{cases}, \quad sign(bin) = -1 \tag{9}$$

$$\begin{cases} U_R(u,v) = 2\Delta \times (1 - f2^{15+Q_E})/2 \\ L_R(u,v) = -U_R(u,v) \end{cases}, \quad sign(bin) = 0 \tag{10}$$

$$\begin{cases} L_R(u,v) = R_r(u,v) - \frac{f2^{15+Q_E}}{2^{15+Q_E}}\Delta \\ U_R(u,v) = L_R(u,v) + \Delta \end{cases}, \quad sign(bin) = 1 \tag{11}$$

- **DCT coefficients reconstruction bin bounds computation –** Finally, the lower (L) and upper (U) bounds of the bin where the DCT coefficient will be reconstructed are obtained from (12). Basically, by shifting the residual DCT coefficients reconstruction bin bounds (previously computed) according to the reconstructed prediction value $P_r(u,v)$, the bin bounds where the (input video) DCT coefficient should be reconstructed can be found.

$$\begin{cases} L(u,v) = L_R(u,v) + P_r(u,v) \\ U(u,v) = U_R(u,v) + P_r(u,v) \end{cases} \tag{12}$$

2) *DCT Coefficients Reconstruction:* Finally, the

reconstructed DCT coefficients, $X_r(u,v)$, are obtained from

$$X_r = \begin{cases} L + \dfrac{1}{\alpha} + \dfrac{\Delta}{1 - e^{\alpha\Delta}} & , \ P_r < L \\ P_r + \dfrac{\left(\gamma + \frac{1}{\alpha}\right) e^{-\alpha\gamma} - \left(\delta + \frac{1}{\alpha}\right) e^{\alpha\delta}}{2 - \left(e^{-\alpha\gamma} - e^{\alpha\delta}\right)} & , \ P_r \in [L, U[\\ U - \dfrac{1}{\alpha} - \dfrac{\Delta}{1 - e^{\alpha\Delta}} & , \ P_r \geq U \end{cases} \quad (13)$$

which corresponds to the closed form derived from the statistical reconstruction solution presented in Section II.B used to avoid numerical computation of the integrals. In (13), the indices (u,v) are omitted just to make it less dense. In (13), $\delta = U(u,v) - P_r\,(u,v)$ and $\gamma = P_r\,(u,v) - L(u,v)$. As it can be observed from (13), the (statistically) reconstructed DCT coefficient value $X_r(u,v)$ depends on the prediction P_r (SI) location. Since the reconstructed value obtained from (13) corresponds to the minimum MSE estimate of the original DCT coefficient $X(u,v)$, the statistical reconstruction approach should allow improving the decoded video quality and the overall video codec RD performance.

V. Performance Assessment

To evaluate the proposed video coding solution, the novel tools described in Section IV have been integrated in the H.264/AVC reference software (JM 18.3).

A. Test Conditions

The detailed test conditions are summarized in Table I and correspond to representative conditions. Fig. 3 shows the first frame of each test video sequence. The compression efficiency gains are measured with Bjontegaard deltas for the bitrate (BD-Rate) and PSNR (BD-PSNR) [6]. The H.264/AVC High profile (with normative ISQ) is adopted as performance benchmark.

(a) (b) (c) (d)

Fig. 3 First frame of the selected video sequences: (a) *Night*, (b) *City*, (c) *Big Ships*, and (d) *Shuttle Start*.

TABLE I
Test Conditions

Sequences	*Night*, *City*, *Shuttle Start* and *Big Ships*
Resolution	1280×720@60Hz
No. of Frames	151
GOP size	15
GOP structure	IBBPBBP…
QPI (QP_P=QP_I+1; QP_B=QP_I+2)	10, 12, 14, 18, 24, 30, 36, 42, 48
Transform Size	4×4
No. of Reference Frames	4
Other Encoder Tools	RDO On, CABAC enabled

B. RD Performance Results and Analysis

Table II shows the RD performance of the proposed predictive video coding solution with statistical reconstruction versus the adopted benchmark. From Table II, it can be concluded that the proposed video coding solution outperforms the reference software H.264/AVC High profile solution for all the test sequences, with bitrate savings up to 2.4%. Higher coding gains are observed for the more detailed sequences, such as *Night* and *City*, and higher bitrates (i.e. lower quantization step sizes), where it is more difficult for the rigid H.264/AVC ISQ function, which recovers the residual DCT coefficients always at a fixed distance from the quantization bin lower bound, to minimize the (quantization) distortion. In fact, the exploitation of the correlation statistics between the source and the prediction (SI) is more beneficial for distortion minimization purposes, thus allowing to increase the reconstructed PSNR for a similar bitrate.

TABLE II
Bjontegaard Metric Results for H.264/AVC + Statistical Reconstruction vs H.264/AVC

Seq.	**H.264/AVC (Reference codec)**		**H.264/AVC+Stat. Rec. (Proposed codec)**		**Bjontegaard Metric**	
	Rate [kbps]	**PSNR [dB]**	**Rate [kbps]**	**PSNR [dB]**	**BD-PSNR [dB]**	**BD-Rate [%]**
Night	116141.65	49.30	116129.83	49.83	0.14	-2.38
	42993.94	42.75	43123.83	42.94		
	14671.87	38.61	14713.41	38.67		
	2590.86	31.95	2581.04	31.98		
City	119723.08	48.73	119955.45	49.29	0.10	-2.17
	77718.95	45.49	78034.36	45.84		
	10845.05	37.66	10879.08	37.68		
	1851.61	31.09	1849.68	31.09		
Shuttle Start	56096.32	49.89	56158.14	50.12	0.09	-1.72
	27255.46	47.00	27095.45	47.08		
	1497.99	38.90	1481.28	38.93		
	159.33	27.88	160.76	28.15		
Big Ships	104309.03	48.72	104252.83	49.01	0.07	-1.60
	26150.79	41.97	25651.33	42.02		
	2863.85	35.37	2843.29	35.39		
	247.77	26.33	255.36	26.39		

VI. Conclusions

This paper proposes a statistical reconstruction solution, similar to the one used in DVC, to replace the ISQ in the context of the H.264/AVC standard. Experimental results show that the proposed solution allows achieving BD-Rate savings up to 2.4% regarding H.264/AVC High profile. As future work, it is planned to extend the proposed statistical reconstruction solution to the recently emerged HEVC (High Efficiency Video Coding) standard aiming to improve the RD performance.

Acknowledgment

This work was supported by FCT project MUVIS with reference PTDC/EEA-TEL1119804/2010.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] X. Artigas et al., “The DISCOVER codec: architecture, techniques and evaluation,” in *Proc. PCS’07*, Lisbon, Portugal, Nov. 2007.

[3] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, “Low-complexity transform and quantization in H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 598–602, Jul. 2003.

[4] F. Bellifemine et al., “Statistical analysis of the 2D-DCT coefficients of the differential signal for images,” *Signal Process.: Image Commun.*, vol. 4, no. 6, pp. 477–488, Nov. 1992.

[5] E. Kreyszig, *Advanced Engineering Mathematics*, 7th ed., New York: Wiley, 1993.

[6] G. Bjontegaard, “Calculation of average PSNR differences between RD curves,” Doc. VCEG-M33, Austin, TX, USA, Apr. 2001.