A CONVEX-OPTIMIZATION FRAMEWORK FOR FRAME-LEVEL OPTIMAL RATE ALLOCATION IN PREDICTIVE VIDEO CODING

Aniello Fiengo[1], *Giovanni Chierchia*[1], *Marco Cagnazzo*[1], *and Béatrice Pesquet-Popescu*[1]

[1] Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI, 75014 Paris, France

ABSTRACT

Optimal rate allocation is among the most challenging tasks to perform in the context of predictive video coding, because of the dependencies between frames induced by motion compensation. In this paper, we derive an analytical rate-distortion model that explicitly takes into account the dependencies between frames. The proposed approach allows us to formulate the frame-level optimal rate allocation as a convex optimization problem. Within this framework, we are able to achieve the exact solution in limited time (even for large-size problems), thanks to the flexibility offered by recent convex optimization techniques. Experiments on standard sequences demonstrate the interest of considering the proposed rate-distortion model and confirm that the optimal rate allocation ensures a better distribution of the total bit budget, with superior results (in the rate-distortion sense) with respect to the standard H.264/AVC rate control.

Index Terms— Video coding, rate distortion, convex optimization, resource allocation.

1. INTRODUCTION

Predictive coding is one of the best tools at hand to exploit temporal redundancies induced by motion. As a matter of fact, motion-compensated prediction plays a key role to reduce significantly the bit rate in state-of-the-art video codecs such as H.264/AVC [1] and H.265/HEVC [2]. Several techniques have been proposed in the literature to select the coding parameters that achieve an optimal trade-off between rate and distortion. Many conventional coding schemes tend to make these choices frame by frame [3, 4, 5, 6, 7, 8]. However, it is widely recognized that, from a rate-distortion standpoint, the optimal choice for a single frame may be potentially suboptimal for encoding the remaining frames, because of the chain of predictions created by motion compensation. Consequently, an optimization that jointly takes into consideration the dependencies between frames may yield a significant bit reduction [9, 10, 11, 12, 13].

In the last decades, a substantial research effort has been made to enlarge the optimization scope from single frames to groups of frames. A first class of techniques amounts at either considering long-term reference frames [14, 15] or segmenting video frames into objects to be coded separately [16, 17]. However, these approaches exploit only indirectly the temporal dependencies between frames. A more theoretical approach consists in embedding temporal dependencies in a rate allocation problem. In this context, one of the most relevant work dates back to Uz, Shapiro and Czigler [18], who provided an optimal strategy for performing frame-level bit allocation in presence of quantizer feedback. The point is that motion-compensated frames carry the quantization errors affecting the previous frames and, consequently, these errors propagate towards the prediction residuals of successive frames, resulting in multiple quantization errors when such residuals are coded. The basic result shown in [18] is that both easily-predicted frames and good predictors need to be allocated a higher bit rate, in order to limit the propagation of quantization errors. Similar insights were provided in [9] within an operational rate-distortion (R-D) framework, but this method requires to evaluate a set of operational R-D points for each frame, which makes the computational burden prohibitive in many applications, because of the involved multi-pass coding. Recent work [12] extended at pixel-level the rate allocation model in [18] and used it to drive an heuristic for selecting the optimal quantization parameters. Interestingly, the authors observed that a motion estimation approach which exploits temporal dependencies may lead to observable gains in the rate-distortion sense.

Related work. The optimal rate allocation is, in general, a non trivial problem. In the seminal work [18], to take into account the temporal dependency between a frame I_n and its predictor I_{n-1}, the rate-distortion function was decoupled into two terms, the prediction error and the distortion of I_{n-1}, leading to a recursive model that can be managed analytically. The same decoupling idea was recently rediscovered in [13], but instead of considering a recursive model, the allocation problem was formulated in terms of rate as a function of distortion and solved by resorting to a series of convex relaxations based on a first-order Taylor approximation. A similar error-splitting model was also employed in [19] within the distributed video coding (DVC) framework, in order to analyse the R-D performance of decoding strategies and propose new schemes for multi-view DVC.

Contributions. In this paper, we propose an efficient solution to exactly solve the frame-level bit allocation problem.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

We use the same theoretical foundations as in [18], but

1. we extend the R-D model by letting the exponential decay vary at each frame, allowing us to better represent the intrinsic non-stationarities in a group of frames (in [18], it is explicitly required that the exponential decay needs to be the same for all the frames);
2. we provide an efficient solution based on proximal tools [20] in order to find the *exact* solution of the R-D problem (within the limit of the accuracy of the model), in limited time even for hundreds of frames.

The paper is organized as follows. Sec. 2 illustrates the proposed R-D model and discusses the differences with [13]. Sec. 3 describes the bit allocation problem and the proposed algorithm to solve it. Sec. 4 provides model validation and illustrates the performance of the proposed rate allocation method, and the conclusions are given in Sec. 5.

2. RATE-DISTORTION MODEL

Accurate Rate-Distortion (R-D) modelling plays a fundamental role in optimal bit allocation. Due to the different characteristics of frames, as well as the sophisticated compression techniques employed in coding algorithms, analytic R-D modelling is still an open problem. According to the classical R-D theory [21], it is possible to express the relationship between R and D for a frame compressed at high bit rate as

$$D = \alpha\sigma^2 2^{-2R} \tag{1}$$

where α and σ^2 are, respectively, the p.d.f. shape factor and the variance of the residual DCT coefficients, and D is measured as the Mean Square Error between the original frame and the reconstructed frame. As shown in [22], such type of analytic formula is quite inaccurate for hybrid video coders. We therefore present an alternative R-D model based on the work [18]. To do so, we consider a group of frames $I = (I_0, \dots, I_{N-1})$ of size $N > 0$ and we denote by $r = (r_0, \dots, r_{N-1})$ the rates used to encode these frames. We further assume that frame I_0 is intra encoded (I-frame) and, for $n \geq 1$, frame I_n is predicted from frame I_{n-1} and the prediction residual is spatially encoded (P-frames).

I-frame. For frame I_0, we follow the model proposed in [18] and employ the following R-D function:

$$D_0 = \alpha_0 \, M_0 \, 2^{-\beta_0 \, r_0}, \tag{2}$$

where $M_0 = \sigma_0^2$ is the variance of I_0 and (α_0, β_0) are model parameters estimated as explained in Sec. 4. It is worth noting the difference with the R-D function proposed in [13], which reads (after a simple inversion):

$$\widehat{D}_0 = \frac{a_0 \cdot G}{r_0 - c_0 \cdot G} - b_0. \tag{3}$$

where (a_0, b_0, c_0) are model parameters and G is the average gradient of a frame.

P-frame. For frames I_n, with $n \geq 1$, we choose a model that is very close to [18, 19, 13], but with a different parametrization, which consists in expressing the R-D function as:

$$D_n = \alpha_n \, (M_n + D_{n-1}) \, 2^{-\beta_n \, r_n}, \tag{4}$$

where M_n is the *innovation* of frame I_n, i.e. the residual that would result if I_n was predicted (through a prediction function f_n) from a non-quantized reference frame I_{n-1}

$$M_n = \mathbb{E}\left\{ \left[I_n - f_n(I_{n-1}) \right]^2 \right\}, \tag{5}$$

and (a_n, β_n) are model parameters estimated in Sec. 4. The distortion of I_n is hence decoupled in two terms: the error due to the motion-compensated prediction of I_n and the distortion due to the quantization of the reference frame I_{n-1}.

3. RATE ALLOCATION ALGORITHM

Optimal rate allocation consists in finding the vector of rates that minimizes the global distortion while keeping the total rate under a given budget $\eta > 0$,

$$\underset{r \in [0, \infty[^N}{\text{minimize}} \ \sum_{n=0}^{N-1} D_n(r_n, \dots, r_0) \quad \text{s.t.} \quad \sum_{n=0}^{N-1} r_n \leq \eta. \tag{6}$$

The R-D function of frame I_0 is given in Eq. (2), while the one for the n-th P-frame is given in Eq. (4). Therefore, we can demonstrate by the induction principle that the operational R-D function D_n actually depends on all the frames involved in the chain of predictions leading to I_n, yielding

$$D_n(r_n, \dots, r_0) = \sum_{\ell=0}^{n} \alpha^{(n,\ell)} \, M_\ell \, 2^{-\sum_{j=\ell}^{n} \beta_j r_j}, \tag{7}$$

where $\alpha^{(n,\ell)} = \prod_{j=\ell}^{n} \alpha_j$. Note that Eq. (7) reduces to Eq. (2) when $n = 0$.

To gain some insight into the solution of Problem (6), we introduce a vector $u = (u_{n\ell})_{0 \leq n \leq N-1, 0 \leq \ell \leq n}$ defined as

$$u_{n\ell} = \sum_{j=\ell}^{n} \beta_j r_j, \tag{8}$$

which allows us to express the global distortion as a separable sum of exponentials

$$F(u) = \sum_{n=0}^{N-1} \sum_{\ell=0}^{n} \alpha^{(n,\ell)} \, M_\ell \, 2^{-u_{n\ell}}. \tag{9}$$

Therefore, Problem (6) can be reformulated as follows

$$\underset{r \in \mathbb{R}^N}{\text{minimize}} \ F(Lr) \quad \text{s.t.} \quad r \in C, \tag{10}$$

where $L\colon \mathbb{R}^N \mapsto \mathbb{R}^{\frac{N(N+1)}{2}}$ is the linear operator that maps the vector $r \in \mathbb{R}^N$ into the vector $u \in \mathbb{R}^{\frac{N(N+1)}{2}}$ defined in (8), and C is the nonempty closed convex set defined as

$$C = \{r \in [0, +\infty[^N \quad | \quad \sum_{n=0}^{N-1} r_n \leq \eta\}. \qquad (11)$$

Among the many approaches proposed in the literature to solve this class of problems, we do not transform the constrained problem in Eq. (6) to a Lagrangian formulation, but rather we manage the bit budget as a hard constraint, in order to bypass the need for determining the corresponding Lagrangian multiplier. Consequently, we resort to proximal algorithms [20, 23], which can handle a wide class of convex optimization problems involving smooth and non-smooth penalizations, as well as hard constraints. In particular, we employ the M+LFBF algorithm proposed in [24], which guarantees the convergence in a reasonable time even for large-scale problems, offers robustness to numerical errors and its structure makes it suitable for parallel implementations.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm, we selected eight video sequences composed of 84 frames at resolution 352×288. We encoded all the sequences with a GOP structure IPP...P of size $N = 12$, a frame rate of 30 fps and CABAC as entropy coder. For the motion estimation, we set the maximum search range to ± 16. Moreover, we estimated the model parameters $(\alpha_n, \beta_n, M_n)_{0 \leq n \leq N-1}$ by encoding the sequences in H.264/AVC with the rate control disabled, in order to manually fix seven different quantization parameters: $10, 12, 14, \ldots, 22$. For each frame, we recorded the values of D and R produced at the encoder output and, after a logarithmic transformation of D, we estimated the model parameters by resorting to a linear regression. We then encoded all the sequences with the rate control of H.264/AVC enabled, setting seven different target bit-rates ranging from 2.2 to 4.0 Mb/s and reported the results in the following.

Model validation. We start our experimental analysis by comparing the proposed R-D model with the one in [13]. The validation was performed on the basis of the R^2 metric[1] [25], which was designed to quantitatively measure the degree of deviation from a given model (the closer the value of R^2 is to 1, the more accurate the model). Table 1 shows the R^2 values associated to the R-D function of I-frames given in Eq. (1)-(3). For all the sequences, the proposed model (2) shows superior fitting performance, giving R^2 values very close to 1 and higher than (1) and (3). We skip the comparison between the R-D functions associated to P-frames, as we got R^2 values very close to each other (the R-D models are very similar).

Table 1: R^2 values of the R-D functions for Intra-frames

Sequence	R^2 with (2)	R^2 with (3)	R^2 with (1)
hall	0.997	0.455	0.975
foreman	0.995	0.912	0.908
coastguard	0.984	0.596	0.395
akiyo	0.9799	0.7447	0.9738

Table 2: PSNR increase for several bitrates

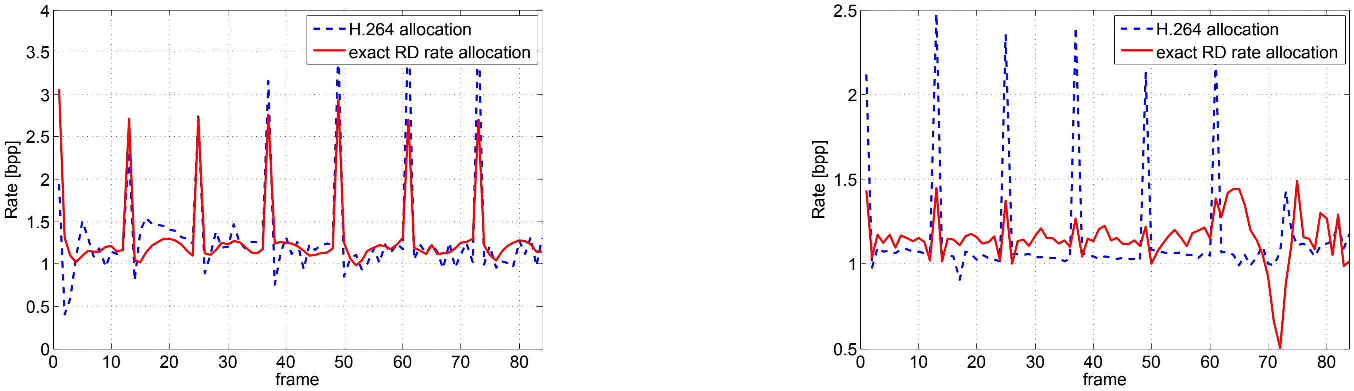
Sequence	1.31 bpp	1.15 bpp	0.76 bpp
akiyo	3.38	6.78	3.23
football	2.28	0.95	2.26
coastguard	0.99	0.55	1.11
eric	0.10	0.12	0.19

Comparison with H.264/AVC. To assess the validity of our rate allocation method, we compared it with the standard rate control algorithm of H.264/AVC [26]. For the sequences “akiyo” and “coastguard”, Fig. 1 reports the rates, as a function of frame numbers, which are allocated by the proposed method (solid line) and by H.264/AVC (dashed line), while Fig. 2 displays the corresponding distortions. As we can see in Fig. 1, the allocation by the proposed method is quite different from the reference, especially for I-frames. Also, our distribution of rates is more uniform within the GOP. This is an important result, because often the H.264/AVC encoder shows a sub-optimal greedy behaviour: it allocates the largest part of the bit budget to the first frames of the GOP and hence it rapidly runs out of bits for the remaining GOP frames, causing an increase of the global distortion. This greedy behaviour is especially noticeable for static sequences such as *akiyo*. Moreover, we collected in Fig. 3 the average distortion as function of the average rate, and in Tab. 2 the corresponding PSNR increments. These results show that the distortion achieved with the proposed method is always lower than the one obtained by the standard rate-control algorithm.

5. CONCLUSIONS

We have proposed a new algorithm to exactly solve the frame-level rate allocation problem arising in predictive video coding. The obtained results demonstrate that the analytical R-D model presented in Sec. 2 allows us to accurately describe the temporal dependencies in a group of frames. Furthermore, our experiments indicate that the optimal rate allocation, when supported by an accurate R-D model, attains better results (in the R-D sense) than the standard rate control in H.264/AVC. The higher performance of our approach is related to its ability to *see beyond* the first frames of the GOP and to keep the rate budget for the successive frames when necessary. This is in contrast with the greedy behaviour of H.264/AVC rate controller, which tends to allocate the largest part of the bit budget to the first frames of the GOP.

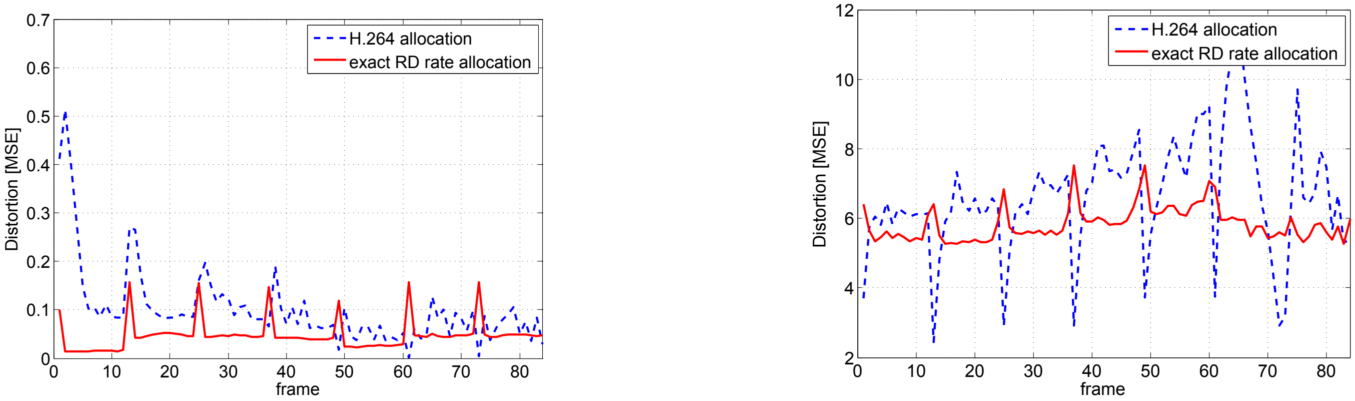
[1] $R^2 = 1 - \frac{\sum_i (X_i - \hat{X}_i)^2}{\sum_i (X_i - \bar{X})^2}$ where X_i and $\hat{X}_i$ are the real and the estimated values of one data point, and $\bar{X}$ is the mean of all data points.



(a) Sequence *akiyo*.

(b) Sequence *coastguard*.

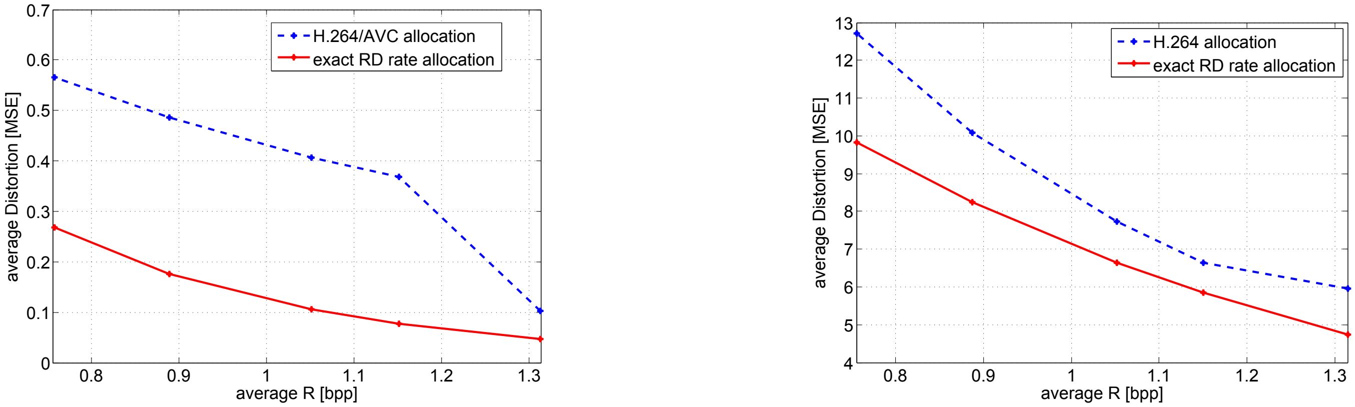
Fig. 1: Rates vs frame numbers.



(a) Sequence *akiyo*.

(b) Sequence *coastguard*.

Fig. 2: Distortion vs frame numbers.



(a) Sequence *akiyo*.

(b) Sequence *coastguard*.

Fig. 3: Comparison between actual and estimated distortion.

6. REFERENCES

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

[2] G. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[3] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, 1998.

[4] A. Ortega and K. Ramchandran, "Rate-distortion techniques in image and video compression," *Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, 1998.

[5] Z. He and S. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–849, 2002.

[6] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst.*, vol. 13, no. 7, pp. 688–703, July 2003.

[7] Siwei Ma, Wen Gao, and Yan Lu, "Rate-Distortion Analysis for H . 264 / AVC Video Coding and its Application to Rate Control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1533–1544, 2005.

[8] Zhan Ma, Meng Xu, YF Ou, and Yao Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 671–682, 2012.

[9] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 533–545, 1994.

[10] B. Schumitsch, H. Schwarz, and T. Wiegand, "Optimization of transform coefficient selection and motion vector estimation considering interpicture dependencies in hybrid video coding," in *Proc. SPIE*, 2005, pp. 327–334.

[11] V. Chellappa, P. Cosman, and G. Voelker, "Dual frame motion compensation with uneven quality assignment," *IEEE Trans. Circuits Syst.*, vol. 18, no. 2, pp. 23–50, 2008.

[12] G. Valenzise and A. Ortega, "Improved video coding efficiency exploiting tree-based pixelwise coding dependencies," in *SPIE Visual Information Processing and Communication*, San Jose, USA, Jan. 2010.

[13] C. Pang, O. Au, F. Zou, J. Dai, X. Zhang, and W. Dai, "An analytic framework for frame-level dependent bit allocation in hybrid video coding," *IEEE Trans. Circuits Syst.*, vol. 23, no. 6, pp. 990–1002, June 2013.

[14] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. Circuits Syst.*, vol. 9, no. 1, pp. 70–84, 1999.

[15] M. Gothe and J. Vaisey, "Improving motion compensation using multiple temporal frames," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, May 1993, vol. 1, pp. 157–160.

[16] A. Vetro, H. Sun, and Y. Wang, "Mpeg-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst.*, vol. 9, no. 1, pp. 186–199, 1999.

[17] A. Vetro, T. Haga, K. Sumi, and H. Sun, "Object-based coding for long-term archive of surveillance video," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2003, vol. 2, pp. 417–420.

[18] K. Uz, J. Shapiro, and M. Czigler, "Optimal bit allocation in the presence of quantizer feedback," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 1993.

[19] T. Maugey and B. Pesquet-Popescu, "Side information estimation and new schemes for multiview distributed video coding," *J. Visual Communication and Image Representation*, vol. 19, no. 8, pp. 589–599, Dec. 2008.

[20] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer-Verlag, New York, 2011.

[21] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley-Interscience, New York, USA, 1991.

[22] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Trans. Circuits Syst.*, vol. 12, no. 11, pp. 970–982, 2002.

[23] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools," 2013, Submitted, http://arxiv.org/pdf/1210.5844.pdf.

[24] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, 2011.

[25] J. L. Devore and N. R. Farnum, *Applied Statistic for Engineers and Scientists*, Duxbury, 1999.

[26] G. Sullivan, T. Wiegand, and K.P. Lim, "Joint model reference encoding methods and decoding concealment methods; section 2.6: Rate control," *JVT-I049*, Sept. 2003.