HAL
open science

Convex optimization for frame-level rate allocation in MV-HEVC

Aniello Fiengo, Giovanni Chierchia, Marco Cagnazzo, Béatrice Pesquet-Popescu

To cite this version:

Aniello Fiengo, Giovanni Chierchia, Marco Cagnazzo, Béatrice Pesquet-Popescu. Convex optimization for frame-level rate allocation in MV-HEVC. 2016 IEEE International Conference on Image Processing (ICIP), Sep 2016, Phoenix, United States. 10.1109/ICIP.2016.7532740 . hal-01718476

HAL Id: hal-01718476

https://hal.science/hal-01718476v1

Submitted on 27 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONVEX OPTIMIZATION FOR FRAME-LEVEL RATE ALLOCATION IN MV-HEVC

Aniello Fiengo[1], *Giovanni Chierchia*[2], *Marco Cagnazzo*[1], *and Béatrice Pesquet-Popescu*[1]

[1] Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI, 75014 Paris, France
[2] Université Paris-Est, LIGM, Équipe A3SI, ESIEE Paris, France

ABSTRACT

Optimal rate allocation is among the most challenging tasks to perform in the context of multi-view video coding, because of the dependency between frames induced by motion compensation and depth image-based rendering. In this paper, using a recursive rate-distortion model that explicitly takes into account these dependencies, we approach the frame-level rate allocation as a convex optimization problem. Within this framework, we provide an efficient algorithm for exactly solving the above problem with recent convex optimization tools. Experiments on standard sequences demonstrate the interest of considering the proposed rate allocation method and confirm that our approach ensures a better performance (in rate-distortion sense) than the standard MV-HEVC rate control.

***Index Terms*—** MV video coding, rate distortion, convex optimization, resource allocation.

1. INTRODUCTION

Digital 3D video technology has recently experienced a proliferation of new applications, such as auto-stereoscopic 3D TV and free-viewpoint video. In order to respond to the increasing requirement for an efficient 3D video coding, the new standard High Efficiency Video Coding (HEVC) [1] has been recently extended with the Multiview Video-plus-Depth (MV) format [2]. A typical MV involves multiple video sequences of the same scene taken at different angles, along with the disparity between adjacent views.

Just like in regular videos, a key role in improving the visual quality of MV coding is played by the rate control, which is responsible to efficiently distribute the bit budget among frames. In particular, the rate control involves two steps: *bit allocation*, where the total bit budget is allocated to frames, and *quantization selection*, where the quantization parameter (QP) is adjusted in function of the allocated bits.

The goal of rate control is to achieve an optimal trade-off between rate and distortion. To this end, a large panel of techniques have been proposed in the literature to select the coding parameters. Many conventional approaches tend to make these choices frame by frame [3, 4, 5, 6, 7]. However, it is widely recognized that, from a rate-distortion (R-D) standpoint, the optimal choice for a single frame may be potentially suboptimal for encoding the remaining frames. This is due to the motion compensation and depth image-based rendering, which carry the quantization error of a coded frame into the prediction of adjacent ones (either in the same view or different ones), resulting in multiple quantization errors when the latter are coded. Consequently, an allocation scheme that takes into account the dependency between frames yields a significant bit reduction [8, 9, 10, 11, 12].

The focus of this paper is on optimal rate allocation in MV coding, for which it is of paramount importance to exploit the dependency between frames induced by motion compensation and depth image-based rendering. To this end, we propose a convex optimization approach for *exactly* and *efficiently* solving the frame-level rate allocation problem (within the limit of the accuracy of our model).

1.1. Related work

In the last decades, a substantial research effort has been made to enlarge the optimization scope from single frames to groups of frames (GOP). One of the first attempts in this sense was made by Ramchandran *et al.* [8], who used an operational R-D framework to search the optimal combination of quantization parameters for the whole GOP. Although it was conceived for the allocation at frame level, this method was also extended to different coding levels, such as blocks within the same frame [13, 14]. However, the parameter search requires to evaluate a set of R-D points for each frame, making the computational burden very high due to the fact that a multi-pass coding is required.

The obstacle encountered by search-based techniques is that the R-D behavior of a residual frame is unknown before encoding the frames from which it was predicted. Such a difficulty can be circumvented by using a parametric R-D model to describe the dependency between frames [15, 16, 17, 18]. This approach was recently followed by Pang *et al.* [11], who formulated the rate allocation as a non-convex optimization problem, and solved it through a sequence of approximations. The same R-D model was also used in distributed video coding (DCV) to propose new coding schemes for MV-DVC [19].

Some other works focus on minimizing the maximum distortion of encoded frames (MINMAX), instead of the usual average of distortions (MINAVE), as the former cri-

terion might lead to some benefits to the visual perception. The MINMAX criterion can be optimized through dynamic programming [20], multi-pass approaches [21], or ad-hoc iterative methods with lower encoding complexity [22]. However, it is not clear whether the MINMAX is always a better criterion than the MINAVE.

1.2. Contributions

The methods proposed in the literature to deal with the frame-level rate allocation are either theoretical in nature [15], computationally demanding [8], approximated [11], or based on the MINMAX criterion [22]. In this paper, we propose an efficient solution to exactly solve the frame-level rate allocation problem formulated with the MINAVE criterion. Our approach is based on a recursive R-D model in which the error variance of a residual frame is decoupled in two terms: the distortion of the frame used to build the prediction, and the inaccuracy of the prediction itself.

This paper extends our previous work [23] with two significant additions. Firstly, we generalize our approach to the scenario in which the dependency between frames is represented by a bi-dimensional graph (instead of a linear chain), in order to accommodate the more intricate prediction schemes used in MV coding. Secondly, we integrate the proposed algorithm in the MV-HEVC encoder, so as to compare the R-D performance of our approach with the standard MV-HEVC rate control.

The paper is organized as follows. Sec. 2 illustrates the proposed R-D model. Sec. 3 describes the rate allocation problem and the proposed algorithm to solve it. Sec. 4 illustrates the performance of our approach. Finally, the conclusion is drawn in Sec. 5.

2. RATE-DISTORTION MODEL

In order to effectively reduce the spatial and temporal redundancy of a video sequence, it is crucial to optimally distribute the bit budget within the sequence, because this directly affects the visual distortion of encoded frames. In this regard, the rate control of HEVC allocates the bit budget at three different levels: GOP level, frame level, and coding-unit level.

2.1. Background (2D videos)

The frame-level rate control allocates the bit budget within a GOP by looking at the R-D behavior of the different frames. To do so, a possible approach consists of resorting to a parametric R-D model. Let us consider a GOP $I = (I_0, \ldots, I_{N-1})$, where the frame I_0 is intra encoded, while for $n \geq 1$ the frame I_n is predicted from the previously encoded frame $\widetilde{I}_{n-1}$, and the prediction residual E_n is spatially encoded (IP...P structure). According to the R-D model proposed in [23], the distortion of the frame $\widetilde{I}_n$ encoded at a high rate r_n reads

$$D_n = \mathbb{E}\{(I_n - \widetilde{I}_n)^2\} = \alpha_n \, \sigma_n^2 \, \exp\{-\beta_n \, r_n\}, \qquad (1)$$

where (α_n, β_n) are free parameters, and σ_n^2 is the variance of the frame I_0 when $n = 0$ or the residual E_n for $n \geq 1$.

However, the distortion D_n also depends on the (encoded) frame $\widetilde{I}_{n-1}$ used to build the motion-compensated prediction. This fact can be accounted for by decomposing the variance of the residual frame as $\sigma_n^2 = M_n + D_{n-1}$ [15], where M_n is the prediction square mean error. Consequently, the R-D behavior of a frame I_n can be modeled as [23]

$$D_n = \alpha_n \, (M_n + D_{n-1}) \, \exp\{-\beta_n \, r_n\}, \qquad (2)$$

where, for $n = 0$, M_0 denotes the variance of I_0 and $D_{-1} = 0$. Hence, for $n \geq 1$, the distortion D_n is controlled by two terms: the rate r_n used to encode $\widetilde{I}_n$, and the distortion D_{n-1} of its encoded reference $\widetilde{I}_{n-1}$, leading to a recursive model.

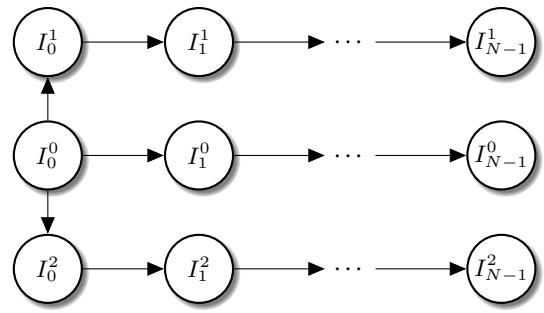


Fig. 1: Dependency graph used in the paper.

2.2. Proposed approach (MV videos)

In this paper, we consider the scenario where the dependency between frames can be represented as the graph depicted in Fig. 1, where I_n^k denotes the n-th frame of the k-th view, with $k = 0$ being the central view and $k \geq 1$ being the side views. Using the R-D model in (2), we can demonstrate by the induction principle that the distortion D_n^k actually depends on all the frames involved in the chain of predictions leading to I_n^k. In the central view, the distortion D_n^0 is thus a function of the rates $r_0^0, \ldots, r_n^0$ allocated in the central view, yielding

$$D_n^0 = \sum_{\ell=0}^{n} \alpha_{\ell,n}^0 \, M_\ell^0 \, \exp\Big\{ -\sum_{j=\ell}^{n} \beta_j^0 \, r_j^0 \Big\}, \qquad (3)$$

where $\alpha_{\ell,n}^0 = \prod_{j=\ell}^{n} \alpha_j^0$. In the k-th side view (with $k \geq 1$), the distortion D_n^k depends not only from the rate $r_0^k, \ldots, r_n^k$ allocated in the same view, but also from the rate r_0^0 allocated to the first frame I_0^0 of the central view, leading to

$$D_n^k = \sum_{\ell=0}^{n} \alpha_{\ell,n}^k \, M_\ell^k \, \exp\Big\{ -\beta_0^0 \, r_0^0 - \sum_{j=\ell}^{n} \beta_j^k \, r_j^k \Big\}. \qquad (4)$$

where $\alpha_{\ell,n}^k = \prod_{j=\ell}^{n} \alpha_0^0 \, \alpha_j^k$.

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 $\eta > 0$, leading to

$$\underset{r\in[0,+\infty[^{KN}}{\text{minimize}} \quad \sum_{k=0}^{K-1}\sum_{n=0}^{N-1} D_n^k(r) \quad \text{s.t.} \quad \sum_{k=0}^{K-1}\sum_{n=0}^{N-1} r_n^k \leq \eta, \tag{5}$$

where r is the vector of rates defined as

$$r = \left[r_0^0, \ldots, r_{N-1}^0, \ldots, r_{N-1}^{K-1}\right]^\top, \tag{6}$$

and D_n^k is defined in (3)-(4). To gain some insight into the solution of Problem (5), we introduce a vector $u = (u_{n,\ell}^k)_{0\leq k\leq K-1,\, 0\leq n\leq N-1,\, 0\leq \ell\leq n}$ defined as

$$u_{n,\ell}^k = \begin{cases} \displaystyle\sum_{j=\ell}^{n} \beta_j^k\, r_j^k, & \text{if } k = 0, \\ \displaystyle\beta_0^0\, r_0^0 + \sum_{j=\ell}^{n} \beta_j^k\, r_j^k, & \text{if } k \geq 1, \end{cases} \tag{7}$$

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

$$F(u) = \sum_{k=0}^{K-1}\sum_{n=0}^{N-1}\sum_{\ell=0}^{n} \alpha_{n,\ell}^k\, M_\ell^k\, \exp\{-u_{n,\ell}^k\}. \tag{8}$$

Therefore, Problem (5) can be reformulated as follows

$$\underset{r\in\mathbb{R}^{KN}}{\text{minimize}} \;\; F(Lr) \quad \text{s.t.} \quad r \in C, \tag{9}$$

where $L\colon \mathbb{R}^N \mapsto \mathbb{R}^{\frac{KN(N+1)}{2}}$ is the linear operator that maps the vector $r \in \mathbb{R}^N$ into the vector $u \in \mathbb{R}^{\frac{KN(N+1)}{2}}$ defined in (7), and C is the nonempty closed convex set defined as

$$C = \left\{ r \in [0,+\infty[^{KN} \;\mid\; \sum_{k=0}^{K-1}\sum_{n=0}^{N-1} r_n \leq \eta \right\}. \tag{10}$$

Among the approaches proposed in the literature to solve this class of problems, we do not transform the constrained problem in (5) 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. To do so, we resort to proximal algorithms [24, 25, 26], 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 [27], 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.

Table 1: Sequence characteristic and used target bit rates

Sequence	Resolution	Frame Rate (fps)	Target Bit Rate (Mbps)
Balloons	1024x768	30	1.5 2.7 3.6 4.5
Kendo			2.1 2.4 3.0 4.5
Newspaper_CC			9 12 15 21
Undo_Dancer	1920x1088	25	10.2 25.2 40.2 60
GT_Fly			4.5 10.5 15 21
Poznan_Hall2			0.6 1.8 2.4 4.2
Poznan_Street			10.2 21 30 51

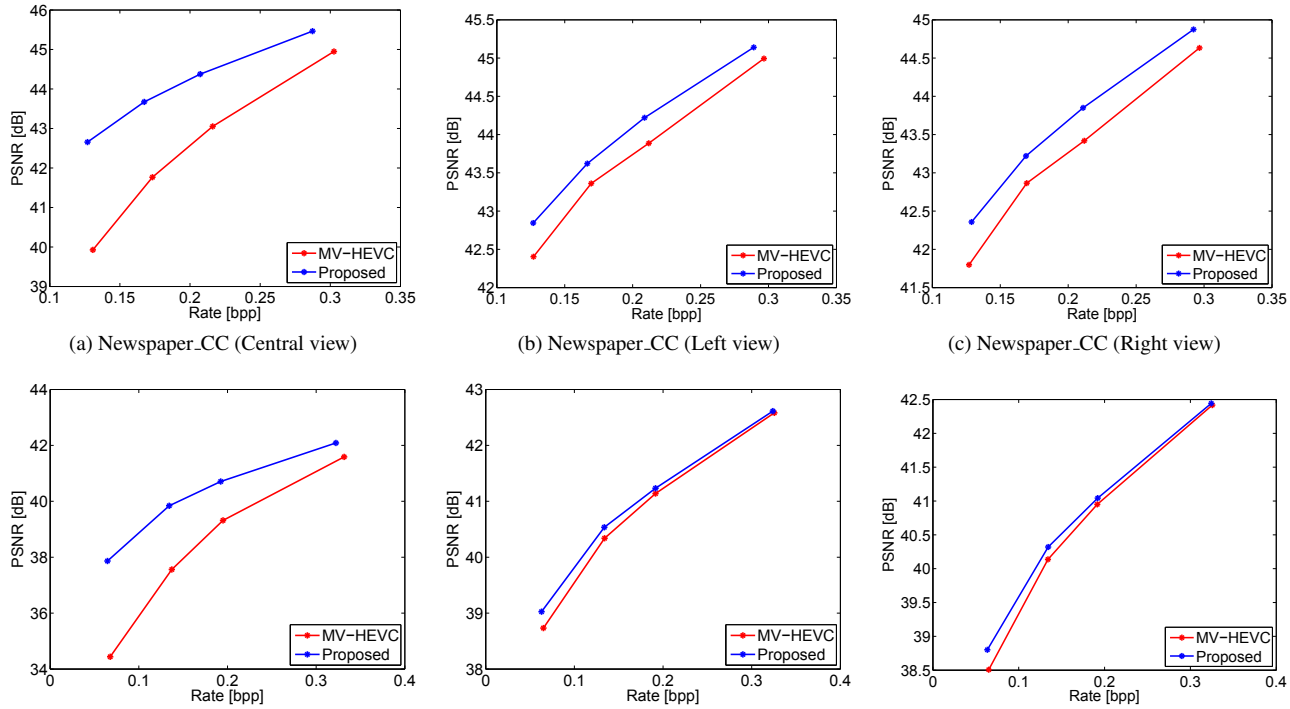
4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm, simulations were conducted using the reference software and experimental evaluation methodology has been developed. The experimental framework is composed of three steps, that are parameter estimation, bit allocation, and encoding, as explained in the following.

Parameter estimation. The model parameter estimation is performed by multiple encodings of the sequences at different quantization parameter (QP) values: $10, 12, 14, \ldots, 40$. A standard version of the reference software HM 16.0 was used in this phase. 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. The parameter $(M_n)_{0\leq n\leq N-1}$, which represents the inter-view residual/motion prediction, is obtained by encoding the sequences with a modified version of HM 16.0 that extrapolates M_n for each frame, using a $QP = 1$.

Bit allocation. The algorithm takes the estimated parameters of the sequence as input, providing as output the number of bits for each frame that minimizes the total distortion over each GOP, under the constrain of a defined target bit rate. Table 1 reports the target bit rates used for each sequence.

Encoding. The encoder used in the final step of our experiments is a modified version of HM 16.0 where the rate control module is able to accept an external file containing the result of the previous step and set the number of bits allocated for each GOP and for each frame on these values. The base view is coded as the center view and the two dependent views are encoded as left and right view, with a configuration that recalls the dependency graph of Fig. 1. The average luma PSNR values and bit rates for all three views are reported in Table 2 as total results.



(a) Newspaper_CC (Central view) (b) Newspaper_CC (Left view) (c) Newspaper_CC (Right view)

(d) Poznan_Street (Central view) (e) Poznan_Street (Left view) (f) Poznan_Street (Right view)

Fig. 2: The rate distortion curves of the proposed method as well as the rate control in the reference software.

4.1. Rate-Distortion performance comparison

To assess the validity of our bit rate allocation method, we compared it with the standard rate control algorithm of MV-HEVC. The R-D curves of the proposed method and MV-HEVC are shown in Fig. 2 for two sequences "*Poznan_Hall2*" and "*Newspaper_CC*" for the three view used. From that, we can see that our method can achieve a larger PSNR value at each target bit rate.

Table 2 shows the rate-distortion performance of our method compared to straightforward MV-HEVC evaluated with the Bjontegaard metrics [28]. For each sequence an average value over the three views of the gain in terms of Δ PSNR and bit rate reduction is presented. The results indicate that our method provides an average bit rate savings of 20% relative to standard MV-HEVC with an average gain of 0.70 dB in terms of Δ PSNR. In terms of allocation accuracy, the difference between the target bit rate and the achieved bit rate is for each sequence lower than 10%, with an average value among all sequence of 7.5%.

Sequence	Δ PSNR (dB)	Bit Rate (%)
Balloons	1.20	-21.52
Kendo	0.35	-7.86
Newspaper_CC	0.65	-23.35
Undo_Dancer	0.49	-12.88
GT_Fly	0.83	-20.34
Poznan_Hall2	0.59	-34.72
Poznan_Street	0.82	-23.61
Average	0.70	-20.61

Table 2: Average PSNR gain and bit rate savings compared to straightforward HEVC multi-view extension (MV-HEVC).

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 we developed in Sec. 2 allows us to accurately describe the temporal and inter-view 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 MV-HEVC. The higher performance of our approach is related to its ability to take into account the characteristics of the video sequence and the temporal and inter-view correlation between the frames. The results of our experiments, conducted on different standard test sequences, demonstrate the effectiveness of our proposition.

6. REFERENCES

[1] 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.

[2] K.-R. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, 2013.

[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] Z. Ma, M. Xu, YF Ou, and Y. 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] S. Tan, J. Si, S. Ma, S. Wang, and W. Gao, "Adaptive frame level rate control in 3D-HEVC," in *Visual Communications and Image Processing*, Dec 2014, pp. 382–385.

[13] Y. Yang and S. S. Hemami, "Generalized rate-distortion optimization for motion-compensated video coders," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 10, no. 6, pp. 942–955, 2000.

[14] C. Wang, X. Mou, and L. Zhang, "Optimization of the block-level bit allocation in perceptual video coding based on MINMAX," http://arxiv.org/abs/1511.04691, 2015.

[15] 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.

[16] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 8, no. 4, pp. 446–459, Aug. 1998.

[17] J. Liu, Y. Cho, Z. Guo, and C. C. Jay Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 20, no. 7, pp. 967–981, July 2010.

[18] S. Hu, S. Kwong, T. Zhao, and C. C. Jay Kuo, "Rate control optimization for temporal-layer scalable video coding," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 21, no. 8, pp. 11521162, Aug. 2011.

[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] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," vol. 1, no. 1, pp. 3–17, 1999.

[21] B. Xie and W. Zeng, "A sequence-based rate control framework for consistent quality real-time video," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 16, no. 1, pp. 56–71, Jan. 2006.

[22] N. Cherniavsky, G. Shavit, M. F. Ringenburg, R. E. Ladner, and E. A. Riskin, "Multistage: A MINMAX bit allocation algorithm for video coders," *IEEE Trans. Circuits Syst. for Video Technology*, vol. 17, no. 1, pp. 59–67, 2007.

[23] A. Fiengo, G. Chierchia, M. Cagnazzo, and B. Pesquet-Popescu, "A convex-optimization framework for frame-level optimal rate allocation in predictive video coding," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 2014, pp. 7328–7332.

[24] 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.

[25] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical projection and proximal tools for solving constrained convex optimization problems," *Signal Image Video P.*, vol. 9, no. 9, pp. 1737–1749, Nov. 2015.

[26] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, 2015, To appear.

[27] 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.

[28] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *VCEG Meeting*, 2001.