ITERATIVE R-D OPTIMIZATION OF H.264

Cheolhong An and Truong Q. Nguyen

UCSD, ECE Dept, La Jolla, California 92093-0407
chan@ucsd.edu, nguyent@ece.ucsd.edu
http://videoprocessing.ucsd.edu/

ABSTRACT

In this paper, we apply the primal-dual decomposition and subgradient projection methods to solve the rate-distortion optimization problem with the constant bit rate constraint. The primal decomposition method enables spatial or temporal prediction dependency within a Group Of Picture (GOP) to be processed in the master primal problem. As a result, we can apply the dual decomposition to minimize independently the Lagrangian cost of all the MBs using the reference software model of H.264. Furthermore, the optimal Lagrange multiplier λ^* is iteratively derived from the solution of the dual problem. As an example, we derive the optimal bit allocation condition with the consideration of temporal prediction dependency among the pictures. Experimental results show that the proposed method achieves better performance than the reference software model of H.264 with rate control for given bit constraint.

Index Terms— H.264, R-D optimization, optimal bit allocation, primal-dual decomposition, rate control.

1. INTRODUCTION

After Rate-Distortion (R-D) optimization is introduced for video compression using the Lagrange multiplier [1][2], there are many methods to reduce the complexity in deciding Macro Block (MB) modes, Motion Vectors (MVs) for a given Lagrange multiplier λ. Even though R-D optimization method is not mandatory for standard video compression such as H.264 [3], it is the main part of video coding to improve the coding efficiency [2][4]. Therefore, we review the relation between R-D optimization and previous works. R-D optimization with inequality constraint in a frame is mathematically formulated as follows:

$$\min_{\mathbf{m}} \quad \sum_{n=1}^{N} d_n(\mathbf{m}_n) \quad s.t. \qquad \sum_{n=1}^{N} x_n(\mathbf{m}_n) \leq X_F \tag{1}$$

where $\mathbf{m}_n = (M_n, \mathbf{MV}_n, QP_n, \mathbf{Ref}_n)$ is a vector of MB mode, MVs, Quantization Parameter (QP) and reference frames for inter prediction. N is the number of MBs in a frame and X_F is the bit constraint of a frame. d_n and x_n are distortion and coded bits of the n-th MB, respectively. The optimization problem (1) can be solved by the Lagrangian duality in order to obtain the optimal solution if the problem is a convex optimization problem and satisfies the Slater's condition [5]. After the Lagrange duality is applied, the dual function of the primal problem (1) is

$$q(\lambda) = \min_{\mathbf{m}} \quad \sum_{n=1}^{N} \Big(d_n(\mathbf{m}_n) + \lambda x_n(\mathbf{m}_n) \Big) - \lambda X_F \tag{2}$$

and its dual problem is $\max_{\lambda \geq 0} q(\lambda)$. If we know the optimal solution of the dual problem, we can obtain the solution of the primal problem (1) after solving (2). However, in order to simplify the above optimization problems, the relation between λ and QP was derived in [2][6][7][8] and the estimation of X_F from QP was studied in [9]. The reference software model of H.264 (simply denoted as JM model) [10] has the following relation:

$$\lambda = \kappa 2^{\frac{(QP-12)}{3}}, \tag{3}$$

$$X_F = aQP^{-1} + bQP^{-2} \tag{4}$$

where κ is a function of picture types (I, P, B), the number of referenced frames and QP, and a and b are estimated using the linear regression based on Mean Absolute Difference (MAD) and target bits. Equations (3) and (4) give estimated solution for λ of the dual problem, that is, QP[1] is estimated from (4) for a given constraint X_F and then λ is induced from (3). Thus, JM model does not directly solve the dual problem. For a given λ, JM model minimizes the Lagrangian function, that is, solves the problem (2). However, if there is no bit constraint, we can just choose any QP to derive λ from (3). Consequently, JM model has two coding modes: one is a coding mode without a rate constraint and the other is with a rate constraint.

Without a rate constraint, users just specify any QP and Group Of Picture (GOP) structure, and then JM model solves the problem (2). As a result, users do not know how many bits are generated after encoding. In this case, reference frames, QP and λ are given, the optimization variables are MB modes and MVs for all MBs of a frame. This problem can be simplified by the independent assumption among the MBs. Consequently, the problem (2) is

$$q(\lambda) = \sum_{n=1}^{N} \min_{\mathbf{m}_n} \Big(d_n(\mathbf{m}_n) + \lambda x_n(\mathbf{m}_n) \Big) - \lambda X_F \tag{5}$$

where the optimization variables are MB mode and MVs for each MB. This optimization problem is solved by the following method: First, fix a MB mode of all the inter MB modes and then find optimal MVs with or without considering both residual bits and MV bits for the MB mode. Next, given the optimal MVs for inter MB modes, find the optimal MB mode which minimizes the Lagrangian cost $l_n(\mathbf{m}_n)$, that is, $d_n(\mathbf{m}_n) + \lambda x_n(\mathbf{m}_n)$ among the inter and the other MB modes such as intra MB and direct MB modes. In order to reduce the loss of coding efficiency of independent assumption, references [1][11][12] solve the dependent optimization problem (2) using dynamic programming without considering frame-level dependency or λ. Reference [13] considers the frame-level dependent

This work is supported by CWC and matching fund from UC Discovery program

[1]For simplicity, we directly denote QP instead of Qstep in (4), and QP is derived from the mapping between QP and Qstep.

coding problem using the Viterbi Algorithm (VA), but it considers that distortion and coded bits are only function of QP.

With a rate constraint, users specify the coded bit rate and GOP structure, and then JM model solves the problem (5) with independent assumption. Although λ and QP are obtained from equations (3) and (4), bit constraint X_F in (2) should be derived from user bit rate constraint because user bit rate constraint is average bits per second, but not MB-level or frame-level bit constraint. Therefore, MB-level or frame-level which are generalized as a basic unit (a group of MBs) [14] and GOP-level bit constraints need to be derived from a given user bit rate constraint. JM model uses the basic unit for a bit constraint. Without loss of generality, the basic unit is considered as a MB or a frame in this paper. If we find target bits for a basic unit, the other parameters can be obtained from equations (3) and (4). References [14][15][16] show how to estimate target bits of a basic unit from user bit rate constraint, video frame rate, buffer fullness, picture type and some other information.

Thus, JM model of H.264 mainly focuses on real-time or low complexity rate control scheme with the constant or variable bit rate constraint. Therefore, the rate control method induces loss of coding efficiency, and it cannot tightly satisfy the bit constraint which are shown in this paper. In case of non-real time applications with the constant bit rate constraint, the loss can be reduced. In this paper, we apply the primal-dual decomposition and subgradient projection methods to solve directly the problem (1) with the constant GOP bit constraint. Although this method can be used for the optimal bit allocation of any basic unit with consideration of spatial and temporal prediction dependency, we show the frame-level bit allocation within a GOP with considering temporal prediction dependency as an example. Thus, we ignore spatial prediction dependency, that is, MBs which have intra and spatial direct modes are independent. In addition, the optimal bit allocation condition can be applied for video coding without rate control.

The rest of this paper is organized as follows. In section 2, we explain primal-dual decomposition and subgradient projection, and we apply these methods for MB-level bit allocation of an intra-sliced picture with independent assumption. In section 3, frame-level bit allocation is considered with temporal dependency. Experimental results are shown in section 4.

2. PRIMAL-DUAL DECOMPOSITION

In this section, we introduce the general framework to solve optimization problems using the primal-dual decomposition [17][18]. The primal decomposition corresponds to deciding the optimal bit constraint of a basic unit and dual decomposition and the Lagrangian duality are equivalent to obtaining the optimal primal and dual solution for the given optimal bit constraint as a result of the primal decomposition. For simplicity, we simplify the notation of problem (1) as follows:

$$\min_{\mathbf{x}} \sum_n d(x_n), \quad s.t. \sum_n x_n \le X \tag{6}$$

$$\min_{\mathbf{y}} \min_{\mathbf{x}} \sum_n d(x_n) \ \ s.t.\ x_n \le y_n, \sum_n y_n \le X, \ \forall n \tag{7}$$

$$\min_{\mathbf{x}} \sum_n d(x_n), \quad s.t.\ x_n \le y_n, \ \forall n \tag{8}$$

$$\min_{\mathbf{y}} \ q^*(\mathbf{y}), \quad s.t. \sum_n y_n \le X \tag{9}$$

where $q^*(\mathbf{y}) = \min_{\mathbf{x}} \ \sum_n d(x_n) + \lambda_n^*(x_n - y_n)$ which is the optimal value of the problem (8). The original problem (6) can be reformulated into the problem (7) by introducing auxiliary variables $\mathbf{y}$. Then the problem (7) can be decomposed into two optimization problems (8) and (9) with respect to (w.r.t.) optimization variables $\mathbf{x}$ and $\mathbf{y}$, respectively. The decomposition from problem (6) to problem (9) is called as a master primal decomposition, and the decomposition from problem (8) to problem (11) is the dual decomposition. Problem (8) is solved by the Lagrangian duality as follows:

$$q(\mathbf{y}, \lambda) = \min_{\mathbf{x}} \sum_n d(x_n) + \lambda_n(x_n - y_n) \tag{10}$$

$$= \sum_n \min_{x_n} d(x_n) + \lambda_n(x_n - y_n) \tag{11}$$

$$q^*(\mathbf{y}) = \max_{\lambda \succeq 0} q(\mathbf{y}, \lambda) = \max_{\lambda \succeq 0} \sum_n q_n(y_n, \lambda_n) \tag{12}$$

$$= \sum_n \max_{\lambda_n \ge 0} q_n(y_n, \lambda_n) = \sum_n q_n^*(y_n) \tag{13}$$

where $q_n(y_n, \lambda_n) = \min_{x_n} d(x_n) + \lambda_n(x_n - y_n)$. Equations (11) and (13) are derived from independent assumption. The problem (11) is solved by the R-D optimization which is implemented in JM model [10] and the dual problem (13) can be solved by the subgradient projection method [17] as follows:

$$\lambda_n^{k+1} = \left[\lambda_n^k + \eta^k \mathrm{g}_n^k\right]^+ = \max(\lambda_n^k + \eta^k \mathrm{g}_n^k, 0) \tag{14}$$

where η^k is a positive step size at iteration k and $[\cdot]^+$ denotes the projection onto the nonnegative orthant. The projection operation guarantees that the Lagrange multipliers λ_n satisfy their nonnegative conditions. The subgradient g_n^k of $q_n(\lambda_n^k, y_n)$ is $x_n^k - y_n$ which is derived in [17]. Thus, the subgradient of $q_n(\lambda_n^k, y_n)$ is just the difference between coded bits and the constraint bits at iteration k. If coded bits x_n^k are smaller than the constraint bits y_n, the current Lagrangian multiplier λ_n^k decreases, otherwise, λ_n^k increases. From the R-D optimization, smaller λ increases the coded bits. Therefore, the coded bits are getting close to the constraint bits after several iterations.

The master primal problem (9) is also solved by the subgradient projection method. However, the constraint is not as simple as in (14). The solution of the problem (9) is obtained from two procedures. First, the optimization variables y_n are updated by the subgradient as follows:

$$\tilde{y}_n^{k+1} = y_n^k - \eta^k g_n^k \tag{15}$$

and then $\tilde{\mathbf{y}}$ is projected onto the feasible constraint set as:

$$\min_{\mathbf{y}} \quad \| \tilde{\mathbf{y}} - \mathbf{y} \|^2, \ \ s.t. \ \sum_n y_n \le X \tag{16}$$

which is formulated from the fact that the projected point $\mathbf{y}$ from $\tilde{\mathbf{y}}$ minimizes the distance between two points. This problem can be solved using a very efficient algorithm discussed in [19]. Reference [17] shows that the subgradient g_n^k of $q_n^*(y_n)$ at y_n^k is $-\lambda_n^k$ where λ_n^k is the optimal dual variable of the sub-problem in (8). Consequently, more bits are allocated to MBs or pictures (basic units) which have larger λ since larger λ implies that distortion of a MB or a picture decreases further according to the unit bit increment of the constraint. As a result, if we reallocate bits, sum of distortion can be decreased. Furthermore, all the subgradients of MBs or pictures should be equal for the optimal bit allocation. Because the sensitivity of the optimal values of all the MBs or pictures are equal, there is no way to reallocate bits to decrease sum of distortion. This can be clearly observed from (15) since $\tilde{y}_n^{k+1}$ increase equally if their subgradient g_n^k and step size η^k are equal, and then $\tilde{y}_n^{k+1}$ are projected onto the feasible set shown in (16). The projected y_n^{k+1} are the same as y_n^k which results from [19].

3. OPTIMIZATION WITH TEMPORAL DEPENDENCY

In the previous section, we assume that all the distortion and coded bit function are independent. In this section, independent assumption among the basic units is removed. Even though we only consider a frame-level optimization problem with temporal dependency within a GOP, there is no restriction in applying MB-level optimization with spatial dependency. However, we still assume that all the MBs which have spatial prediction dependency are independent for simplicity, but temporal coding dependency is considered among the basic units (frames). With this assumption, a single λ for all the MBs in a frame is optimal which is explained in the previous section.

Equation (6) is reformulated for the frame-level optimization with a GOP bit constraint as follows:

$$\min_{\mathbf{s}} \sum_{f=1}^{F} D_f(\mathbf{s_f}, \mathbf{x^f_{ref}}) \;\; s.t. \sum_{f=1}^{F} X_f(\mathbf{s_f}, \mathbf{x^f_{ref}}) \le X_{gop} \qquad (17)$$

where $D_f(\mathbf{s_f}, \mathbf{x^f_{ref}}) = \sum_{n=1}^{N} d_n(\mathbf{m}_n)$, $X_f(\mathbf{s_f}, \mathbf{x^f_{ref}}) = \sum_{n=1}^{N} x_n(\mathbf{m}_n)$ and $\mathbf{s_f} = (\mathbf{M_f}, \mathbf{MV_f}, \mathbf{QP_f})$. $\mathbf{M_f}, \mathbf{MV_f}, \mathbf{QP_f}$ and $\mathbf{x^f_{ref}}$ are MB modes, MVs, QPs and bits of reference frames for all the MBs in a frame f, X_{gop} is a GOP bit constraint and F is the number of frames within a GOP. Distortion D_f and bits X_f of a frame f depend on all the MB modes, MVs and QPs as well as bits of reference frames. Therefore, every frame can not be optimized independently in the problem (17) due to the dependency of bits of reference frames $\mathbf{x^f_{ref}}$.

As a specific example, we only consider the 1^{st} GOP structure which starts with the Instantaneous Decoder Refresh (IDR) frame and the close GOP. However, the open GOP and any number and prediction dependency of B and P frames within a GOP are not limited to the frame-level optimization with dependency. Here, all the B frames within a GOP are predicted from the same reference frames I and P, and the P frame is predicted from the I frame. When we perform the master primal decomposition, dependency among the frames are considered. As in section 2, auxiliary variables y_f are induced for each frame bits. Consequently, the problem in (17) is decomposed into one master primal problem (19) and F sub-problems (18) which are solved by the Lagrangian duality:

$$\min_{\mathbf{s}} \; D_f(\mathbf{s_f}, \mathbf{y^f_{ref}}) \;\; s.t. \;\; X_f(\mathbf{s_f}, \mathbf{y^f_{ref}}) \le y_f \qquad (18)$$

$$\min_{\mathbf{y}} \; \sum_f Q^*_f(y_k, \mathbf{y^f_{ref}}), \;\; s.t. \;\; \sum_f y_f \le X_{gop} \qquad (19)$$

where $Q^*_f(y_f, \mathbf{y^f_{ref}})$ are the optimal values of sub-problems (18) for a given $\mathbf{y}$, and the reference frame bits are $\mathbf{y^w_{ref}} = (y_1, y_F)$ where $w \in \{2, .., F-1\}$ for B frames, $\mathbf{y^F_{ref}} = y_1$ for the P frame and $\mathbf{y^1_{ref}} = \varnothing$ since the I frame has no reference frames. Comparing (17) to (18), we can recognize the main benefit from the primal decomposition. In the formulation (17), the reference frame bits $\mathbf{x^f_{ref}}$ prevents independent optimization, but in the formulation (18), $\mathbf{x^f_{ref}} = \mathbf{y^f_{ref}}$ because given y_f, $X_f(\mathbf{s_f}^*)$ are y_f due to the complementary slackness condition [5]. Therefore, we reuse the same reference software model of H.264 to minimize the Lagrangian cost in problem (18) with the consideration of dependency. However, the dependency among the frames is processed in the master primal problem as shown in (19) which is a much simpler optimization problem.

In order to solve the problem in (18), we use the Lagrangian duality and subgradient projection which are explained in the section 2. Therefore, we only discuss the master primal problem (19) in this section. Due to the limitation of space, we only show the subgradients of $\sum_{f=1}^{F} Q^*_f(y_f, y^f_{ref})$ w.r.t. B, P and I pictures y_f at $\hat{y}_f$ without proof as follows:

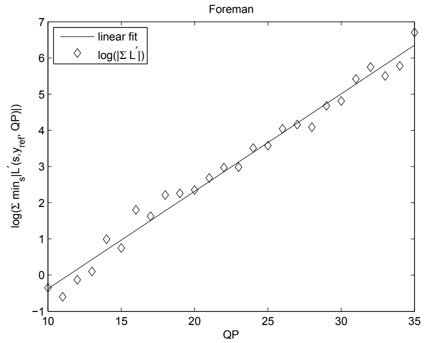

Fig. 1. $\log \left| \sum_f \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, \hat{y}^f_{ref}) \right|$ and its linear fit at $\hat{\lambda}(QP)$.

$$-\hat{\lambda}_k, \;\; k \in \{2, .., F-1\}, -\Big(\hat{\lambda}_F - \sum_{f=2}^{F-1} \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, \hat{y}_F)\Big)$$

$$\text{and} \;\; -\Big(\hat{\lambda}_1 - \sum_{f=2}^{F} \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, \hat{y}_1)\Big)$$

where $L'_f(\mathbf{s_f}, \hat{y}^f_{ref_i}) = \frac{\partial L_f(\mathbf{s_f}, y^f_{ref_i})}{\partial y^f_{ref_i}} \Big|_{y^f_{ref_i} = \hat{y}^f_{ref_i}}$, $\hat{y}^f_{ref} \in \{\hat{y}_F, \hat{y}_1\}$ which shows the variation of Lagrangian cost of a frame f w.r.t. the bit variation of its reference frame. Thus, $L'_f(\mathbf{s_f}, \hat{y}^f_{ref})$ is generally negative because increasing of reference frame bits induces decreasing of the Lagrangian cost of the frame f. As a result, the subgradients of referenced frames which are used as reference frames for prediction are smaller than independent frames for given equal λs. It means that more bits are allocated to referenced frames from equation (15). This result matches with intuition, that is, referenced frames are more important than non-referenced frames because they are used for prediction. As explained in section 2, the subgradients of all the frames are equal for the optimal bit allocation. Therefore, the relation among the λ of pictures is derived as follows:

$$\lambda_I - \sum_{f=2}^{F} \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, y_I) = \lambda_P - \sum_{f=2}^{F-1} \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, y_P) = \lambda_B$$

where $\lambda_I = \lambda_1$, $\lambda_P = \lambda_F$ and $\lambda_B = \lambda_k$, $k \in \{2, .., F-1\}$. Consequently, $\lambda_I \le \lambda_P \le \lambda_B$. This result explains the reason why JM model uses different κ of λ in (3) for different picture types as well as the number of prediction dependency. Furthermore, if B frames are used for prediction, the referenced B frames have different λs from non-referenced B frames. However, current JM model [10] uses the same κ for I and P pictures. $\min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, \hat{y}^f_{ref})$ is experimentally estimated. Figures 1 represents sum of the variation of the Lagrangian cost is closer to exponential decrement. Exponential decrement explains that bits of reference frames (quality of reference frames) are more important at low bit rate, that is, at large λ. As a result, $\sum_f \min_{\mathbf{s_f}} L'_f(\mathbf{s_f}, \hat{y}^f_{ref})$ is modeled as $-\exp^{(\alpha QP + \beta)}$ where α and β are constants.

4. EXPERIMENTAL RESULTS

In this section, we compare the performance of two coding modes (without rate control and with rate control) of JM model [10] with the proposed method. We set QP=35 in JM model without Rate Control (RC) and then the coded bits after encoding are used for the constraint bits. Initial QP is set to JM model with RC and proposed encoder. First, we assume all the frames are independent. Therefore, a global λ for all the frames within a GOP induces an optimal

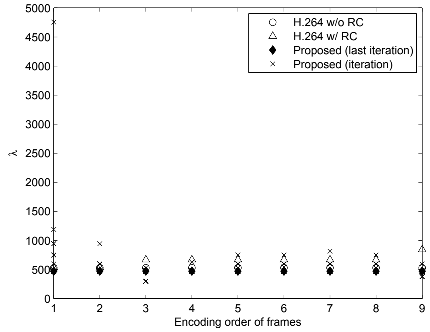

Fig. 2. λ of frames with independent assumption.

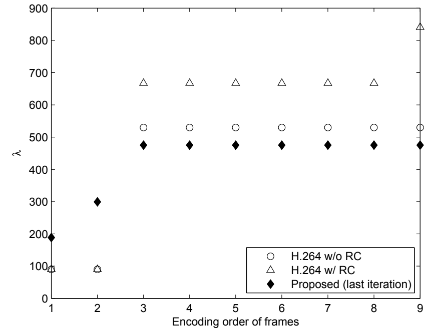

Fig. 3. λ of frames with temporal dependency.

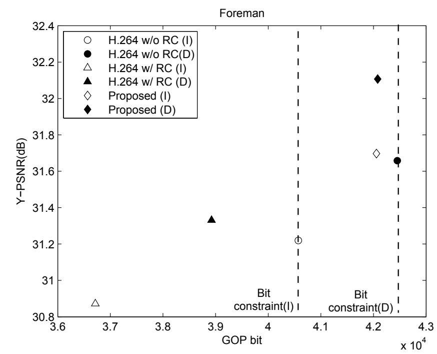

Fig. 4. PSNR vs. bit with dependent and independent cases.

bit allocation since all the frames have equal λs. In order to compare the performance, we set equal κ of (3) of JM model for all the picture types according to independent assumption. A GOP consists of one I and P pictures and seven B pictures. Figure 2 illustrates that in JM model without RC, every frame has equal λ and the proposed method also has equal λ after iterations, but JM model with RC has different λ, especially at the last B frame. However, JM model with RC predicts encoded bits and QP to satisfy bit constraint and its dual variable λ is derived from (3). Consequently, it has some performance degradation.

In the next experiment, we consider temporal dependency among the frames within a GOP. Therefore, κ of (3) of JM model is set to have different values of the reference software model [10]. Proposed method performs QP optimization within ± 1 at center QP which is derived from λ to give more achievable bit region. Here, we only show the last iteration. Figure 3 shows the different λs among I, P and B pictures. All the B pictures have the same λ except JM model with RC. Thus, bit allocation of JM model with RC for B frames is not optimal. Proposed method use different λs for I and P pictures as a result of section 3, but JM model without RC uses the same λ for I and P pictures. Figure 4 illustrates the overall encoded bits and Y-PSNR(dB) with independent assumption (I) and temporal dependency (D). The bits of JM model without RC are constraints to JM model with RC and the proposed method. The constraint bits of temporal dependency (coded bits of JM model without RC) increase since λs of I and P pictures become smaller. Due to R-D optimization, smaller λ increases coded bits. In the independent experiment, proposed method allows constraint violation within 4% of allocated bits, but in the dependent case, only 0.1% constraint violation is allowed. Therefore, the dependent experiment meets more tightly bit constraint. Small bit constraint violation is allowed to consider the convex-hull point around the constraint. JM model with RC does not satisfy the constraint well in addition to having lower PSNR.

5. REFERENCES

[1] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression: An overview," *IEEE Signal Processing Mag.*, vol. 15, pp. 23–50, Nov. 1998.

[2] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74–99, Nov. 1998.

[3] "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, 2003.

[4] A. Puria, X. Chenb, and A. Luthrac, "Video coding using the h.264/mpeg-4 avc compression standard," *Signal Processing: Image Communication*, vol. 19, pp. 793–849, Oct. 2004.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[6] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE ICIP*, 2001.

[7] K. Takagi, "Lagrange multiplier and rd-characteristics," JVT-C084, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, May 2002.

[8] T. Weigand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688–703, July 2003.

[9] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 246–250, Feb. 1997.

[10] *H.264/AVC reference software (JM11.0)*, HHI. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/

[11] T. Wiegand, M. Lightstone, T. G. Campbell, and S. k. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging h.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 182–190, Apr. 1996.

[12] Y. Yang and S. S. Hemami, "Generalized rate-distortion optimization for motion-compensatedvideo coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 942–955, Sept. 2000.

[13] K. Ramchandran, A. Ortega, and M. vetterli, "Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders," *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sept. 1994.

[14] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for jvt," JVT-G012-r1, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Mar. 2003.

[15] H. Lee, T. Chiang, and Y. Zhang, "Scalalble rate control for mpeg-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 878–894, Sept. 2000.

[16] Z. Li, C. Zhu, N. Ling, X. Yang, G. Feng, S. Wu, and F. Pan, "A unified architecture for real-time video-coding systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 472– 487, June 2003.

[17] D.P.Bertsekar, *Nonlinear Programming*, 2nd ed. Athena Scientific, 2003.

[18] B. Johansson and M. Johansson, "Primal and dual approaches to distributed cross-layer optimization," in *Proc. 16th IFAC World Congress*, Prague, Czech republic 2005.

[19] D. Palomar, "Convex primal decomposition for multicarrier linear mimo transceivers," *IEEE Trans. Signal Processing*, vol. 53, no. 12, pp. 4661–4674, Dec. 2005.