H.264/AVC Inter-Frame Rate-Distorsion dependency analysis based on independent regime switching AR models

Nesrine Changuel, Bessem Sayadi, Michel Kieffer

H.264/AVC INTER-FRAME RATE-DISTORTION DEPENDENCY ANALYSIS BASED ON INDEPENDENT REGIME-SWITCHING AR MODELS

N. Changuel, B. Sayadi

Alcatel-Lucent - Bell-Labs France
Route de Villejust, 91620 Nozay, France

M. Kieffer

L2S, CNRS - SUPELEC - Univ Paris-Sud
91192 Gif-sur-Yvette, France

ABSTRACT

The control of the trade-off between encoding rate and quality of compressed video is a challenging task. This control requires efficient rate and distortion (R-D) models, able to describe accurately the behavior of the compressed video. R-D models should account for the dependencies induced by the choice of frame-varying quantization parameters (QP). This paper proposes a dependent R-D model involving two independent regime-switching autoregressive (IRS-AR) models. Experimental results show that the proposed model is able to represent accurately the distortion dependency between frames. For the part of the rate due to the texture, the model fits experimental curves reasonably well. The model has to be completed to account also for the part of the rate due to motion vectors and signalization.

***Index Terms*—** Rate distortion theory, Autoregressive processes, Video codecs

1. INTRODUCTION

Among the various problems raised by the efficient delivery of real-time video contents via a wireless channel, the control of the encoding rate is a challenging task. This control has to cope (*i*) with a time-varying quality of the channel which results in variations in the channel capacity and (*ii*) with changing complexity of the scenes to be encoded. Buffers are usually present at several places of the communication chain, *e.g.*, at the output of the source coder, at intermediate nodes of the network, and at the receiver. They smooth variations of capacity and complexity, but their size has to remain small when a strict delivery delay has to be satisfied, for example in the case of visiophony. In such situations, the rate has to be controlled with a time scale of very few frames. Quality constraints on the compressed video (limited distortion, smooth variations of quality) increase the difficulty of the rate control [1].

Obtaining models for the rate and distortion (R-D) characteristics of the video sequence to be encoded is thus instrumental to build an efficient control of the source coder. Among the wide variety of available models, one may identify the *independent* and *dependent* R-D models. In the first family of models, the R-D characteristics of each frame, or group of frames (GOF), are assumed to be independent of those of the other frames, or GOFs. One gets then simple parametric model with few tuning parameters, in which the rate and distortion are the logarithmic [2], the power [3], or the exponential [4] of some input parameter. Such models are quite efficient to represent the R-D characteristics of INTRA-encoded frames, *i.e.*, frames encoded without reference to any other frame, or of a GOF which frames are encoded independently of the frames not belonging to the GOF. Nevertheless, it is well known that the quality at which a first frame is encoded impacts significantly that of the next frames when they are encoded with the first one as reference, as is done in most video coding standards based on the predictive coding principle [5], such as H.26X or MPEG X video coders [6].

In [2] a Cauchy density based R-D models for quantized DCT coefficients are proposed and analyzed in a frame bit allocation application for H.264 video coding, while in [7] a Gaussian distribution for the DCT coefficients is considered with an adaptive model-driven bit allocation method based on a parametric R-D model for MPEG2 video coding. In [8] a rate model is proposed for intra frames by considering Laplacian distribution for the DCT coefficients and controlled by the quantization parameter value for MPEG2 video coding.

Dependent R-D models take the dependency between frames into account, but involve usually much more tuning parameters. In [9] a dependent R-D model is proposed handling texture and motion information. This model involves the previous and current quantization parameters but requires a large number of R-D measurements to fit its parameters. In [10], different R-D models for Intra and Inter-coded frames are proposed. R-D models for the Inter-coded frames captures the quantization error propagation effect caused by motion compensation in the reference frame. Nevertheless, these models do not involve the coder quantization parameters, which are the actual control input. Moreover, the dependent R-D models proposed in [9] and [10] are based on experimental analysis. In [11] a theoretic analysis of optimal bit allocation in prediction-based video coders is presented. A Laplacian model is used to describe the distribution of DCT coefficients of residual frames and to analyze the dependency between reference frames and predicted frames at different bit rates.

This paper focuses on the construction of an efficient dependent R-D model for video coders involving motion compensation and texture encoding such as H.264/AVC. The dependency between a transform coefficient of the frame to be encoded and that of its prediction obtained from past (uncoded) frames is described by an independent regime-switching autoregressive (IRS-AR) model [12]. This model, introduced in Section 2 allows to represent coefficients linked to parts of the frame where motion compensation performs well (via the AR part), and coefficients related to parts not well predicted (with the regime switching part). The R-D characteristics of such IRS-AR model are then studied in Section 2. To evaluate the R-D characteristics of a video sequence encoded with a video coder such as H.264/AVC, many IRS-AR models would be necessary. A compound model consisting of two IRS-AR models (one for the all DC coefficients, and one for all AC coefficients) is introduced in Section 3. Experimental results, detailed in Section 4, show that this simple compound model is able to accurately represent the distortion dependency between frames. For the rate, additional adjustments have to be performed to account for the cost of motion vectors and

signalization.

2. IRS-AR MODEL

In a video coder involving block-based motion compensation such as H.264/AVC, to encode the j-th frame F_j of a video sequence, previously encoded frames $\widetilde{F}_k$, $k \neq j$, belonging to the same GOF are used to build a prediction $\widehat{F}_j$ of F_j via motion estimation and compensation [5, 6]. The prediction residual $F_j - \widehat{F}_j$ is then transformed, quantized, and entropy-coded. The reconstructed frame is then $\widetilde{F}_j$.

Neglecting the quantization noise, when the motion estimation and compensation is efficient, many collocated pixels in F_j and in $\widehat{F}_j$ are very similar. For some parts of the picture, however, the motion estimation may be less efficient, due, *e.g.*, to scene change, to appearing objects, or to motion of the camera. Collocated pixels in such areas may be quite differing. Similar observations may be done when considering transformed coefficients (TC). Collocated TC (DC and AC) may be similar or different depending on the quality of the motion estimation/compensation.

The effect of quantization noise is usually to decrease the efficiency of motion compensation. A prediction $\widehat{F}_j$ based on $\widetilde{F}_k$, $k \neq j$ is usually much less efficient than a prediction based on F_k, $k \neq j$. One aim of this paper is precisely to study the impact of the quantization noise of one frame on the next ones.

2.1. Definition

The dependency between collocated TC (or collocated pixels) is modeled here using an IRS-AR

$$Y_j = a_{X_j} Y_{j-1} + b_{X_j} U_j, \tag{1}$$

where Y_j and Y_{j-1} represent some collocated TC of a given block in frames j and $j-1$, X_j is a sequence of independent and identically distributed (iid) binary-valued random variables with $\Pr(X_j = 0) = 1 - \rho$ and $\Pr(X_j = 1) = \rho$, and U_j is a sequence of iid zero-mean and unit variance Gaussian variables. The sequences X_j and U_j are assumed to be independent. This model allows to take into account the fact that in most cases (when $X_j = 0$), collocated TC are quite similar, in which case a_0 is close to one, and that sometimes (when $X_j = 1$), there is fewer correlation, in which case a_1 is closer to zero ($a_1 = 0$ in what follows).

Y_j is assumed wide-sense stationary. This imposes some constraints on the parameters a_0, b_0, and b_1. With $X_j = 0$, one gets

$$\sigma_y^2 \quad = \quad a_0^2 \sigma_y^2 + b_0^2 \sigma_u^2 = b_0^2 \sigma_u^2 / (1 - a_0^2) \tag{2}$$

and if $X_j = 1$, one obtains

$$\sigma_y^2 = b_1^2 \sigma_u^2. \tag{3}$$

Combining (2) and (3), one gets

$$b_1^2 = b_0^2 / (1 - a_0^2). \tag{4}$$

The R-D characteristics of the proposed model when scalar quantized with two different steps for Y_{j-1} and Y_j is studied in the following section.

2.2. Rate and distortion characteristics

Assume that Y_{j-1} is quantized with a scalar uniform midtread quantizer q_{j-1} with step size Δ_{j-1} (to mimic intra coding). For Y_j a predictive coding is performed with

$$\widehat{Y}_j = a_{X_j} q_{j-1}(Y_{j-1}) \tag{5}$$

as prediction for Y_j (the value of X_j is assumed to be known). The prediction residual

$$E_j^{X_j} = Y_j - \widehat{Y}_j = a_{X_j} \left(Y_{j-1} - q_{j-1} \left(Y_{j-1} \right) \right) + b_{X_j} U_j \tag{6}$$

is then quantized with a stepsize Δ_j. Depending on X_j, (6) may become

$$E_j^0 = a_0 \left(Y_{j-1} - q_{j-1} \left(Y_{j-1} \right) \right) + b_0 U_j$$

when $X_j = 0$ corresponding to inter-coding and

$$E_j^1 = b_1 U_j$$

when $X_j = 1$ corresponding to intra-coding. The aim of the remainder of this section is to provide R-D curves for the model (1) as a function of Δ_{j-1} and Δ_j. In the H.264/AVC standard [13], characteristics of the quantizers depend on a quantization parameter Q. The relation between the quantization stepsize Δ and Q may be approximated as

$$\Delta(Q) = 2^{\frac{Q-4}{6}} / P_{\mathrm{F}}, \tag{7}$$

where P_{F} is a constant which value depends of the subband, see [14]. The distortion for Y_j may be written as

$$D_j^{X_j}(\Delta_{j-1}, \Delta_j) = \int_{-\infty}^{+\infty} \left(x - q_j(x) \right)^2 f_{E_j^{X_j}}(x)\, dx, \tag{8}$$

where $f_{E_j^{X_j}}(x)$ is the probability density function of $E_j^{X_j}$. The rate required to represent the quantized Y_j is evaluated as the entropy of the output of the quantizer fed with $E_j^{X_j}$

$$R_j^{X_j}(\Delta_{j-1}, \Delta_j) = -\sum_{k=-\infty}^{+\infty} P_k(\Delta_{j-1}, \Delta_j) \log(P_k(\Delta_{j-1}, \Delta_j)) \tag{9}$$

where

$$P_k(\Delta_{j-1}, \Delta_j) = \int_{(k-\frac{1}{2})\Delta_j}^{(k+\frac{1}{2})\Delta_j} f_{E_j^{X_j}}(x)\, dx. \tag{10}$$

When $X_j = 1$, E_j^1 is zero-mean Gaussian with variance $\sigma_y^2 = b_1^2 \sigma_u^2$ and does not depend on Δ_{j-1},

$$f_{E_j^1}(x) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{x^2}{2\sigma_y^2} \right). \tag{11}$$

High-rate approximations for (8) and (9) are easily obtained. At medium to low rates (large values of Δ_j compared to σ_u), such high-rate approximation becomes coarse, but (8) and (9) contain only few significant terms.

When $X_j = 0$, $f_{E_j^0}(x)$ is the convolution of the pdfs of $a_0 \left(Y_{j-1} - q_{j-1} \left(Y_{j-1} \right) \right)$ and of U_j. One may show that

$$f_{E_j^0}(x) = \frac{1}{\sqrt{8\pi\sigma_y^2}} \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{\left(x + k\Delta_{j-1} a_0 \right)^2}{2\sigma_y^2} \right) G(x, \Delta_{j-1}, k), \tag{12}$$

where

$$G(x,\Delta_{j-1},k) = \text{erf}\left(\frac{2a_0x+\Delta_{j-1}\left(1-2k\left(1-a_0^2\right)\right)}{2\sqrt{2}\sigma_y\sqrt{1-a_0^2}}\right) - \text{erf}\left(\frac{2a_0x-\Delta_{j-1}\left(1+2k\left(1-a_0^2\right)\right)}{2\sqrt{2}\sigma_y\sqrt{1-a_0^2}}\right). \quad (13)$$

Since $f_{E_j^0}(x)$ and $f_{E_j^1}(x)$ are known, one may evaluate numerically $D_j^{X_j}(\Delta_{j-1},\Delta_j)$ and $R_j^{X_j}(\Delta_{j-1},\Delta_j)$ using (8) and (9). The expectation of the rate and the distortion for a TC with respect to X_j is then

$$D_j(\Delta_{j-1},\Delta_j) = (1-\rho)\, D_j^0(\Delta_{j-1},\Delta_j) + \rho D_j^1(\Delta_j) \quad (14)$$

$$R_j(\Delta_{j-1},\Delta_j) = (1-\rho)\, R_j^0(\Delta_{j-1},\Delta_j) + \rho R_j^1(\Delta_j), \quad (15)$$

since D_j^1 and R_j^1 do not depend on Δ_{j-1}.

3. COMPOUND MODEL

Assuming that the size of the transform is the same for the whole frame, N_{DC} blocks of TC have to be considered, each of which containing a single DC coefficient and N_{AC} AC coefficients. Assume that each of these coefficients is represented by IRS-AR models, and that the quantization steps do not change within a frame, one gets a total distortion and rate for the j-th frame expressed as follows

$$D(\Delta_{j-1},\Delta_j) = \frac{1}{N_{\text{DC}}(1+N_{\text{AC}})}\sum_{n=1}^{N_{\text{DC}}}\left(D_j^{\text{DC},n}+\sum_{\ell=1}^{N_{\text{AC}}}D_j^{\text{AC},n,\ell}\right) \quad (16)$$

$$R(\Delta_{j-1},\Delta_j) = \sum_{n=1}^{N_{\text{DC}}}\left(R_j^{\text{DC},n}+\sum_{\ell=1}^{N_{\text{AC}}}R_j^{\text{AC},n,\ell}\right), \quad (17)$$

The number of parameters of the resulting compound model would be prohibitively large. Several simplifications have thus to be considered. First, it is assumed that when for a given TC block, the model for the DC and AC coefficients are switching simultaneously. This is reasonable, since when for a block, a motion compensation is not efficient, there is energy at all frequencies. Second, all IRS-AR models for the DC coefficients are aggregated within a single *average* IRS-AR model. Similarly, N_{AC} *averaged* IRS-AR models for the AC coefficients may be considered (one per frequency). This would lead to $1+N_{\text{AC}}$ models, which still leads to a large number of parameters. Usually, most of the energy is gathered in the low-frequency AC coefficients. One thus represents the R-D behavior of the AC coefficients by a single IRS-AR model which is quite coarse approximation.

As a result, one obtains a compound model consisting of two IRS-AR models (one for the DC coefficients and one for all AC coefficients)

$$Y_j^{\text{DC}} = a_{X_j}^{\text{DC}}Y_{j-1}^{\text{DC}} + b_{X_j}^{\text{DC}}U_j, \quad (18)$$

and

$$Y_j^{\text{AC}} = a_{X_j}^{\text{AC}}Y_{j-1}^{\text{AC}} + b_{X_j}^{\text{AC}}U_j. \quad (19)$$

The expression of the rate and distortion becomes then

$$R_j = \kappa(\delta R_j^{\text{DC}} + (1-\delta)R_j^{\text{AC}}) \quad (20)$$

end

$$D_j = \delta D_j^{\text{DC}} + (1-\delta)D_j^{\text{AC}}. \quad (21)$$

A multiplicative constant δ is introduced to weight the contributions of DC and AC coefficients to the total rate and distortion. An additional parameter κ is introduced in the rate to take into account the number of blocks of the frame. One may take $\kappa = N_{\text{DC}}$, but introducing an additional degree of freedom in the model of the rate helps to mitigate all approximations which were previously considered.

The parameter vector $\mathbf{p} = (a^{DC}, a^{AC}, b_1^{DC}, b_1^{AC}, \delta, \rho)$ is estimated as follows

$$\widehat{\mathbf{p}} = \arg\min \sum_{k=1}^{N}(D_j(\Delta_{j-1}^k,\Delta_j^k) - D_j^{\text{exp}}(\Delta_{j-1}^k,\Delta_j^k))^2 \quad (22)$$

here D_j is the distortion calculated using (8), (14), and (21) at frame j, whereas D_j^{exp} is the experimental distortion using H.264/AVC at the same frame j with N values of the pair of quantization steps $(\Delta_{j-1}^k,\Delta_j^k)$ or equivalently of the quantization parameters (Q_{j-1}^k,Q_j^k), $k=1,\ldots,N$.

The parameter κ is then adjusted using the same experimental points on the rate curve, but with $\mathbf{p}$ fixed at $\widehat{\mathbf{p}}$.

4. EXPERIMENTAL PART

The rate and distortion characteristics for several consecutive frames of various video sequences (Soccer, Foreman, News, Container) have been considered. Here the results for two first frames of Soccer and Container are reported. Similar behaviors are observed for the other sequences.

Experimental characteristics are obtained with the H.264/AVC encoder. A subset of $N=20$ measurements are considered to fit the parameter of the proposed compound model and its rate and distortion characteristics are compared to the experimental ones. A fit with $N=9$ is also considered and presented a comparable performance compared to those with 20 measurements. Figures 2 and 1 represent the evolution of the rate and distortion for the second frame (Inter-coded) of the Soccer and Container sequences in CIF format while varying the QP of the first reference (Intra-coded) frame.

Figure 1 shows the compound model is able to describe quite accurately the evolution of the distortion (for the luminance) and its various regimes, see [15]. Figure 2 illustrates the evolution of the rate. For the experimental part (obtained with H264/AVC), the total rate (solid) and the rate due to the texture (dotted) are both represented. Dots on the solid curves correspond to the experimental values used for the parameter estimation as defined in equation (22) Here the fit with the texture part of the rate is satisfying, except for large values of the quantization parameter. This is mainly due to the fact that the dependency between DC coefficients has not been taken into account in the proposed compound model. When large values of QP are considered, DC values get closer, and are mode efficiently entropy-coded. This is why the rate is overestimated.

5. CONCLUSION

In this paper, we proposed an analytical model that describes the inter-frame rate-distortion dependency between two successively encoded frames. For that purpose, a compound IRS-AR model is constructed and analyzed. By comparing the R-D performance between the analytical model and the real H.264/AVC curves, we have shown the accuracy of the proposed model to describe the H264 encoder behavior. The resulting distortion curves using IRS-AR fits quite well the distortion behavior of the video encoder. The rate is less accurately described. One of the reasons for this is that the parameters of the IRS-AR model focuses mainly on the distortion. In addition, the rate due to signaling and transmission of motion vectors is not taken into account. Additional adjustments have to be performed on the rate model to improve its R-D modeling efficiency, allowing an improved rate and quality control.

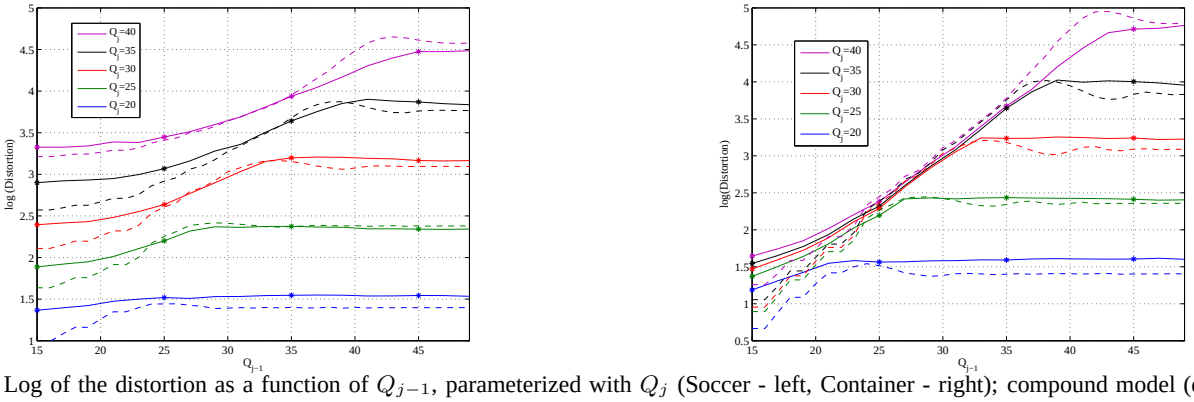

Fig. 1. Log of the distortion as a function of Q_{j-1}, parameterized with Q_j (Soccer - left, Container - right); compound model (dashed), H.264/AVC (solid)

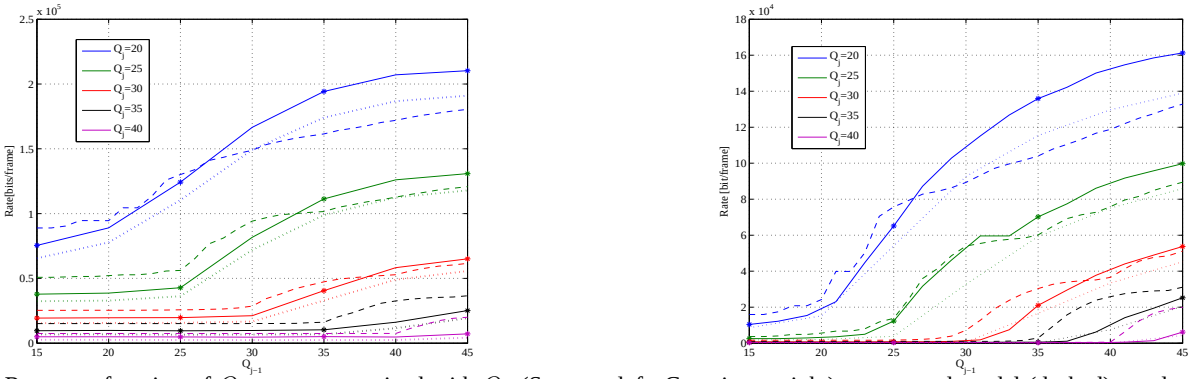

Fig. 2. Rate as a function of Q_{j-1}, parameterized with Q_j (Soccer - left, Container - right); compound model (dashed), total rate using H.264/AVC (solid), texture rate using H.264/AVC (dotted)

6. REFERENCES

[1] L. Ying, L. Zhu, C. Mung, and A.R. Calderbank, "Content-aware distortion-fair video streaming in networks," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-Dec. 4 2008, pp. 1–6.

[2] N. Kamaci, Y. Altunbasak, and R.M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models," *IEEE Trans on Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[3] H. Shen, X. Sun, F. Wu, and S. Li, "Rate-distortion optimization for fast hierarchical b-picture transcoding," *Proc. IEEE International Symposium on Circuits and Systems*, pp. 5279–5282, 2006.

[4] N. Changuel, B. Sayadi, and M. Kieffer, "Predictive control for efficient statistical multiplexing of digital video programs," in *Packet Video*, May 2009, pp. 1–9.

[5] K. Sayood, *Introduction to Data Compression, Third Edition*, Morgan Kaufmann, San Francisco, 2005.

[6] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space-Time Coding for Transmission over Fading Channels*, Wiley, Chichester, 2002.

[7] Bo Tao, B.W. Dickinson, and H.A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding," *IEEE Trans on Circuits and Systems for Video Technology*, vol. 10, no. 1, pp. 147–157, Feb 2000.

[8] J. Bai, Q. Liao, and X. Lin, "Hybrid models of the rate-distortion characteristics for MPEG video coding," in *Communication Technology Proc*, 2000, vol. 1, pp. 363–366 vol.1.

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

[10] X. Minghui, A. Vetro, S. Huifang, and L. Bede, "Rate-distortion optimized bit allocation for error resilient video transcoding," in *ISCAS*, May 2004, vol. 3, pp. III–945–8 Vol.3.

[11] Y. Wang, J. Sun, S. Ma, and W. Gao, "Theoretic analysis of inter frame dependency in video coding," in *PCM '08*, Berlin, Heidelberg, 2008, pp. 935–939, Springer-Verlag.

[12] L. Baum and T. Pétrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statist.*, vol. 37, pp. 15541563, 1966.

[13] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," Tech. Rep., ITU-T Rec. H.264, and ISO/IEC 14496-10 AVC, nov. 2003.

[14] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for h.264/avc video coding and its application to rate control," *IEEE Trans on Circuits and Systems for Video Technology,*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[15] N. Changuel, B. Sayadi, and M. Kieffer, "Statistical multiplexing of video programs," *IEEE Vehicular Technology Magazine*, vol. 4, no. 3, pp. 62–68, Sept. 2009.