RATE-DISTORTION BOUNDS FOR MOTION COMPENSATED RATE SCALABLE VIDEO CODERS

Gregory W. Cook, J. Prades-Nebot†, and Edward J. Delp**

*Video and Image Processing Laboratory (VIPER)
Purdue University
West Lafayette, IN 47907-1285, USA
ace@ecn.purdue.edu

†Departamento de Comunicaciones
Universidad Politécnica de Valencia
Valencia 46071, SPAIN
jprades@dcom.upv.es

ABSTRACT

In this paper, we derive and evaluate theoretical rate-distortion performance bounds for scalable video compression algorithms which use a single motion-compensated prediction (MCP) loop. These bounds are derived using rate-distortion theory based on an optimum mean-square error (MSE) quantizer. By specifying translatory motion and using an approximation of the predicted error frame power spectral density, it is possible to derive parametric versions of the rate-distortion functions which are based solely on the input power spectral density and the accuracy of the motion-compensated prediction. The theory is applicable to systems which allow prediction drift, such as the SNR-scalability in MPEG-2, as well as those with zero prediction drift such as the MPEG-4 fine grained scalable standard.

1. INTRODUCTION

Scalable video coders allow us to decode compressed video at two or more rates in an interval ($R_{\min}$, $R_{\max}$) to achieve a desired quality. These qualities are generally grouped into three categories: signal-to-noise ratio (SNR), spatial resolution, and temporal resolution. In the following we consider only SNR-scalable video coders. Scalable video coders are also distinguished by how the different rates are achieved. In Layered Scalable (LS) codecs, the bit stream is divided into a *base layer*, that provides a minimum level of quality, and one or more *enhancement layers* that improve the quality provided by the base layer. The number of layers in LS codecs, and so, the number of decoding rates, is usually small. By using embedded coding, Fine-Grained Scalable (FGS) codecs (*e.g.*, MPEG-4 FGS [1]) allow decoding of the bit stream for a very large set of different rates.

Motion-compensated prediction (MCP) is used in video compression to reduce redundant temporal information [2]. In MCP-based coders, the MCP loop works with a version of the input signal decoded at the MCP rate (R_{MCP}). The value of this rate determines the main features of MCP-based scalable codecs. For instance, in some scalable strategies, as the SNR-scalable part of the MPEG-2 standard [3], R_{MCP} is set to $R_{\max}$, which provides a prediction with the highest possible quality but introduces prediction drift when the decoding rate R is below $R_{\max}$. In other coders, such as the FGS part of MPEG-4 [1], R_{MCP} is set to $R_{\min}$ which guarantees the absence of prediction drift but also decreases the quality of the prediction. In this paper we obtain the rate-distortion (RD) bounds for single-loop MCP-based SNR-scalable video coders, for both decoding below and above R_{MCP}.

This work was supported in part by an Indiana Twenty-First Century Research and Technology Fund grant and by a grant from the Secretaría de Estado de Educación y Universidades of the Spanish Government.

2. BACKGROUND

In this section we provide some preliminary results useful in obtaining the RD functions of SNR-scalable MCP-based video coders. For the notation that follows lower case letters denote the signals and upper case indicate the Fourier transform. Signals which are functions of spatial variables x and y, and temporal variable t, are written as $s = s(\lambda, t)$, where $\lambda = (x, y)$. The resulting Fourier transform is denoted $S = S(\Lambda, \omega_t)$, where $\Lambda = (\omega_x, \omega_y)$, and ω_x, ω_y are the spatial frequency variables and ω_t denotes the temporal frequency variable.

2.1. Optimum Intraframe Encoding

Given a two dimensional, stationary, jointly Gaussian, input random process $s = s(\lambda)$, its associated power spectral density (PSD) $S_{ss}(\Lambda)$, and the output of the codec $s' = s'(\lambda)$, for a mean-squared error (MSE) criterion, the RD function [4] can be expressed in parametric form:

$$D_{\text{O}}^{\theta} = E\left\{(s - s')^2\right\} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\left[\theta, S_{ss}(\Lambda)\right] d\Lambda. \quad (1)$$

$$R_{\text{O}}^{\theta} = \frac{1}{8\pi^2} \iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ss}(\Lambda)}{\theta}\right] d\Lambda, \quad (2)$$

where $0 < \theta < \infty$ and the rate is measured in bits/(unit length)^2. The optimum coding is equivalent to the "optimum forward channel" of Fig. 1 [4], where the frequency response of the filter is

$$G(\Lambda) = \max\left[0, 1 - \frac{\theta}{S_{ss}(\Lambda)}\right] \quad (3)$$

and $n(\lambda)$ is an independent, zero mean, Gaussian random process with a PSD given by

$$S_{nn}(\Lambda) = \max\left[0, \theta\left(1 - \frac{\theta}{S_{ss}(\Lambda)}\right)\right]. \quad (4)$$

For an optimum MSE codec with differential output (Fig. 2):

$$S_{\tilde{s}\tilde{s}}(\Lambda) = |1 - G(\Lambda)|^2 S_{ss}(\Lambda) + S_{nn}(\Lambda) \quad (5)$$

$$= \min\left[\theta, S_{ss}(\Lambda)\right]. \quad (6)$$

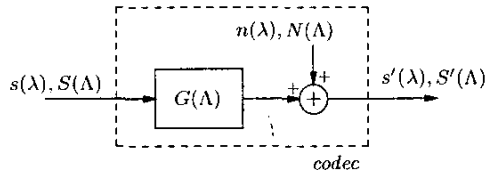

Fig. 1. Block diagram of an optimum MSE codec.

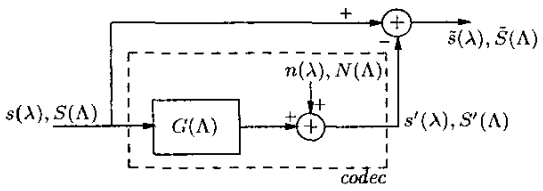

Fig. 2. Block diagram of differential optimum MSE codec.

2.2. Alternate Optimum MSE Encoding Models

In this section we explore two alternate encoding models which use *two* of the optimum forward channels shown in Fig. 1.

2.2.1. Optimum Layered Encoding

Fig. 3 shows the block diagram for a layered codec using two optimum MSE codecs. The distortion ($D_{\rm I}$) in this scheme is

$$D_{\rm I} \triangleq E\left[(s-s'')^2\right] = E\left[(\tilde{s}-\tilde{s}')^2\right], \tag{7}$$

which, assuming $\tilde{\theta} \le \theta$, and considering (6), provides

$$D_{\rm I}^{\theta,\tilde{\theta}} = \frac{1}{4\pi^2}\iint_{\Lambda} \min\left[\tilde{\theta}, S_{ss}(\Lambda)\right] d\Lambda \tag{8}$$

where variables θ and $\tilde{\theta}$ have been added to show the dependence of $D_{\rm I}$ on these variables. We note if $\tilde{\theta} > \theta$, the system is no longer operating in a layered fashion and (8) no longer holds.

The rate of the layered codec ($R_{\rm I}$), is the sum of the rate of the codec associated with θ, and the rate of the codec associated with $\tilde{\theta}$. Then, if $\tilde{\theta} \le \theta$ as in (8), we can obtain

$$R_{\rm I}^{\theta,\tilde{\theta}} = \frac{1}{8\pi^2}\iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ss}(\Lambda)}{\tilde{\theta}}\right] d\Lambda. \tag{9}$$

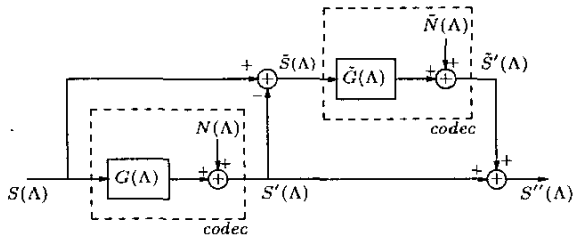

Fig. 3. Block diagram an of optimum MSE layered codec.

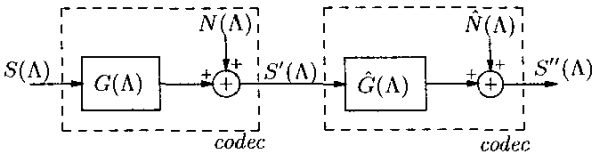

Fig. 4. Block diagram of an optimum MSE cascaded codec.

As for $\tilde{\theta} \le \theta$, $D_{\rm I}^{\theta,\tilde{\theta}} = D_{\rm O}^{\tilde{\theta}}$ and $R_{\rm I}^{\theta,\tilde{\theta}} = R_{\rm O}^{\tilde{\theta}}$, the layered codec has a RD function which is equivalent to a single optimum MSE codec.

2.2.2. Optimum Cascaded Encoding

In the cascaded system shown in Fig. 4 the distortion ($D_{\rm II}$), is

$$D_{\rm II} \triangleq E\left\{\left(s-s''\right)^2\right\} = E\left\{(s-s')^2\right\} + E\left\{(s'-s'')^2\right\} \tag{10}$$

where (10) is only true if $\tilde{s} = s - s'$ and $\hat{s} = s' - s''$ are uncorrelated. While this is in general not true for cascaded systems, in [5] this is shown to be true when using optimum MSE codecs. Then from (10), we can derive

$$D_{\rm II}^{\theta,\hat{\theta}} = \frac{1}{4\pi^2}\iint_{\Lambda} \min\left[\theta+\hat{\theta}, S_{ss}(\Lambda)\right] d\Lambda. \tag{11}$$

In (2) it is assumed the maximum value of the PSD to be transmitted is exactly the maximum value in the input PSD, and bits are predicted relative to this value. In [5], we demonstrate that this same principle holds for the the cascaded encoder, and find

$$R_{\rm II}^{\theta,\hat{\theta}} = \frac{1}{8\pi^2}\iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ss}(\Lambda)}{\hat{\theta}+\theta}\right] d\Lambda. \tag{12}$$

As $D_{\rm II}^{\theta,\hat{\theta}} = D_{\rm O}^{\theta+\hat{\theta}}$ and $R_{\rm II}^{\theta,\hat{\theta}} = R_{\rm O}^{\theta+\hat{\theta}}$, the cascaded system has a RD function which is equivalent to a single optimum MSE codec.

2.3. Interframe Encoding Using MCP

This section is a summary of [2] which describes the properties of an MCP non-scalable video system using an optimum MSE codec and displacement estimates. The variables are now extended to include time, *e.g.*, $s = s(\lambda, t)$, and the corresponding Fourier transform is designated by $S = S(\Lambda, \omega_t) = S(\Omega)$. Fig. 5 shows the block diagram of the system, where the codec is the optimum MSE codec of Section 2.1 and the properties of the MCP loop are captured by the stochastic filter $H(\Omega)$. Since $s - s' = e - e'$, by substituting $S_{ee}(\Lambda)$ for $S_{ss}(\Lambda)$ in (1) and (2) [2], we obtain:

$$D_{\rm O}^{\theta} = E\left\{(e-e')^2\right\} = \frac{1}{4\pi^2}\iint_{\Lambda} \min\left[\theta, S_{ee}^{\theta}(\Lambda)\right] d\Lambda \tag{13}$$

$$R_{\rm O}^{\theta} = \frac{1}{8\pi^2}\iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ee}^{\theta}(\Lambda)}{\theta}\right] d\Lambda, \tag{14}$$

where the dependence of $S_{ee}(\Lambda)$ on θ is explicitly denoted.

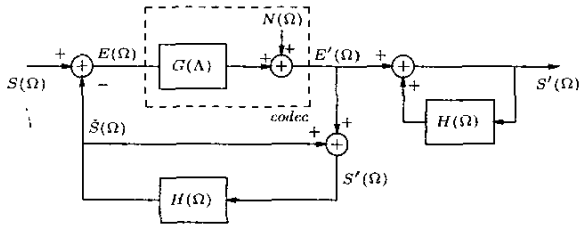

Fig. 5. Block diagram of an MCP optimum MSE codec.

A reasonable approximation to $S_{ee}^{\theta}(\Lambda)$ is [2]

$$S_{ee}^{\theta}(\Lambda) \approx S_{ee}^{\text{appr},\theta}(\Lambda) \triangleq \begin{cases} \max\left[\theta, S_{ee}^{\text{I},\theta}(\Lambda)\right] & \{\Lambda : S_{ss}(\Lambda) > \theta\} \\ S_{ss}(\Lambda) & \{\Lambda : S_{ss}(\Lambda) \le \theta\} \end{cases}. \quad (15)$$

where $S_{ee}^{\text{I},\theta}(\Lambda)$ is found below. The frequency response response of the stochastic filter is

$$H(\Omega) = H(\Lambda, \omega_t) = F(\Lambda) \exp\left(-j\Lambda \cdot \hat{d} - j\omega_t \Delta_t\right), \quad (16)$$

where $\hat{d}$ is the 2D estimated displacement vector, Δ_t is the time interval between consecutive frames, and $F(\Lambda)$ is the frequency response of the spatial filter. For constant, translatory displacement with the optimum $F(\Lambda)$ given by

$$F(\Lambda) = P^*(\Lambda) \frac{S_{ss}(\Lambda)}{S_{ss}(\Lambda) + \theta}, \quad (17)$$

$S_{ee}^{\text{I},\theta}(\Lambda)$ is found to be [2]

$$S_{ee}^{\text{I},\theta}(\Lambda) = S_{ss}(\Lambda)\left[1 - \frac{|P(\Lambda)|^2 S_{ss}(\Lambda)}{S_{ss}(\Lambda) + \theta}\right]. \quad (18)$$

where $P(\Lambda)$ is the 2-D Fourier transform of the probability density function (pdf) $p_{\Delta d}(\Delta d)$ with $\Delta d = d - \hat{d}$, and d is the true displacement. In this analysis the data rate needed to represent the motion vectors is ignored.

3. RD FUNCTION FOR MCP SCALABLE VIDEO

Based on Sections 2.2.1 and 2.2.2, here we extend the theory of Section 2.3 for MCP SNR-scalable video compression.

3.1. Case I: Scalable Video Operating above the MCP Rate

When decoding scalable video above the MCP rate, there are in essence two data sources: a MCP base layer, and an enhancement layer which is an encoding of the difference between the original signal and the base layer signal without MCP, *e.g.*, MPEG-4 FGS [1]. Then, we can model this system as shown in Fig. 6 and the RD function is obtained by substituting $S_{ee}^{\theta}(\Lambda)$ for $S_{ss}(\Lambda)$ in (7) and (8). Consequently, for $\tilde{\theta} \le \theta$

$$\begin{aligned} D_{\text{I}}^{\theta,\tilde{\theta}} &= \frac{1}{4\pi^2} \iint_{\Lambda} \min\left[\tilde{\theta}, S_{ee}^{\theta}(\Lambda)\right] d\Lambda \\ R_{\text{I}}^{\theta,\tilde{\theta}} &= \frac{1}{8\pi^2} \iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ee}^{\theta}(\Lambda)}{\tilde{\theta}}\right] d\Lambda. \end{aligned} \quad (19)$$

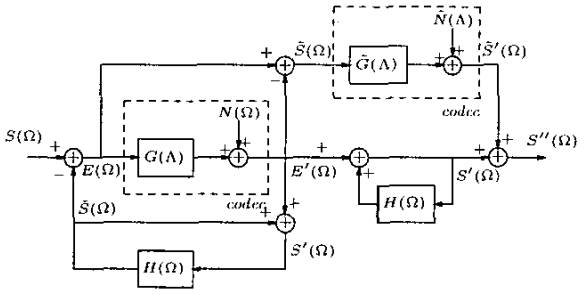

Fig. 6. Block diagram of an MCP scalable codec with $R < R_{\text{MCP}}$

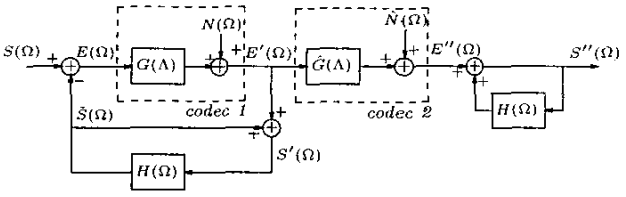

Fig. 7. Block diagram of a MCP scalable codec with $R < R_{\text{MCP}}$.

3.2. Case II: Scalable Video Operating below the MCP Rate

When scalable video is used below the MCP rate, the enhancement layer is completely eliminated and only part of the base layer information is transmitted. Here there is only one data source, but two sources of distortion: one from the usual source of the encoder in the MCP loop (codec 1 in Fig. 7), and another because the entire bit stream is not being sent (codec 2 in Fig. 7).

To determine the distortion (D_{II}), note (10) is still true, if $\tilde{s}$ and $\hat{s}$ (defined as $\tilde{s} = s - s'$ and $\hat{s} = s' - s''$) are uncorrelated. This has been shown to be true in [5]. If an optimum spatial filter as given in (17) is used, then $|F(\Lambda)| < 1$ which guarantees that the decoder is stable irrespective of $S_{ss}(\Lambda)$ and $P(\Lambda)$ [5]. By taking into account these considerations, we can arrive at:

$$\begin{aligned} D_{\text{II}}^{\theta,\tilde{\theta}-\theta} &= \frac{1}{4\pi^2} \iint_{\Lambda} \min\left[\theta, S_{ee}^{\theta}(\Lambda)\right] \quad (20) \\ &\quad + \frac{1}{1 - |F(\Lambda)|^2} \min\left[\tilde{\theta} - \theta, \max\left[0, S_{ee}^{\theta}(\Lambda) - \theta\right]\right] d\Lambda \\ R_{\text{II}}^{\theta,\tilde{\theta}-\theta} &= \frac{1}{8\pi^2} \iint_{\Lambda} \max\left[0, \log_2 \frac{S_{ee}^{\theta}(\Lambda)}{\tilde{\theta}}\right] d\Lambda. \quad (21) \end{aligned}$$

4. RD FUNCTIONS USING APPROXIMATIONS TO S_{EE}^{θ}

By approximating $S_{ee}^{\theta}(\Lambda)$ with (15) we can obtain RD functions, for both above and below the MCP rate cases, based entirely on $S_{ss}(\Lambda)$, $F(\Lambda)$ and the motion-compensation method.

For Case I, that is $\tilde{\theta} \leq \theta$ and $R > R_{\text{MCP}}$, we obtain

$$D_{\text{I}}^{\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\left[\tilde{\theta}, S_{ss}(\Lambda)\right] d\Lambda$$

$$R_{\text{I}}^{\theta,\tilde{\theta}} = \frac{1}{8\pi^2} \iint_{\{\Lambda: S_{ss}(\Lambda) > \theta\}} \log_2 \frac{\max\left[\theta, S_{ee}^{\text{I},\theta}(\Lambda)\right]}{\tilde{\theta}} d\Lambda + \frac{1}{8\pi^2} \iint_{\{\Lambda: S_{ss}(\Lambda) \leq \theta\}} \max\left[0, \log_2 \frac{S_{ss}(\Lambda)}{\tilde{\theta}}\right] d\Lambda, \tag{22}$$

and for Case II, that is, $\tilde{\theta} > \theta$ and $R < R_{\text{MCP}}$ we have:

$$D_{\text{II}}^{\theta,\tilde{\theta}-\theta} = \frac{1}{4\pi^2} \iint_{\{\Lambda: S_{ss}(\Lambda) > \theta\}} \theta + \frac{1}{1 - |F(\Lambda)|^2} \times \min\left[\tilde{\theta} - \theta, \max\left[0, S_{ee}^{I,\theta}(\Lambda) - \theta\right]\right] d\Lambda + \frac{1}{4\pi^2} \iint_{\{\Lambda: S_{ss}(\Lambda) \leq \theta\}} S_{ss}(\Lambda)\, d\Lambda$$

$$R_{\text{II}}^{\theta,\tilde{\theta}-\theta} = \frac{1}{8\pi^2} \iint_{\{\Lambda: S_{ss}(\Lambda) > \theta\}} \max\left[0, \log_2 \frac{\max[\theta, S_{ee}^{\text{I},\theta}(\Lambda)]}{\tilde{\theta}}\right] d\Lambda, \tag{23}$$

where $S_{ee}^{\text{I},\theta}(\Lambda)$ is given by (18).

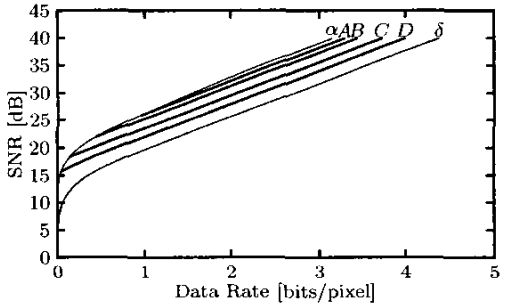

Fig. 8. RD functions $D_{\text{I}}^{\theta,\tilde{\theta}}$ and $R_{\text{I}}^{\theta,\tilde{\theta}}$ for $\sigma_{\Delta d}^2 = 0.04/f_{sx}^2$ with various MCP rates. Curves α and δ are plots of the RD functions for an optimum MCP NS video codec. In Curve α $\sigma_{\Delta d}^2 = 0.04/f_{sx}^2$, and Curve δ has no MCP. The MCP rates in bits/pixel are: $R_{\text{MCP}}^A = 0.96$, $R_{\text{MCP}}^B = 0.45$, $R_{\text{MCP}}^C = 0.15$, and $R_{\text{MCP}}^D = 0.04$.

5. EVALUATION OF SCALABLE VIDEO RD FUNCTIONS

In this section, the results in Section 4 are solved numerically by assuming a model for the video signal and the MCP accuracy [2]. Every frame is modeled as a continuous random field with an isotropic autocorrelation. The random field is then limited in band and sampled at the horizontal and vertical Nyquist frequencies f_{sx} and f_{sx} respectively. Parameters are set to obtain a good model for video conference signals at rates less than 2 Mb/s [2]. We assume that between consecutive frames there is a translatory displacement d and that Δd has a zero mean, Gaussian isotropic pdf with variance $\sigma_{\Delta d}^2$. To give these results some practical grounding, sequences with low motion are the equivalent of having an accurate displacement estimate; conversely, sequences with high motion tend to not have good motion estimates.

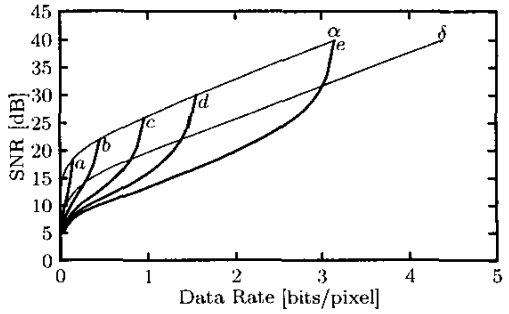

Fig. 9. RD functions $D_{\text{II}}^{\theta,\tilde{\theta}-\theta}$ and $R_{\text{II}}^{\theta,\tilde{\theta}-\theta}$ for $\sigma_{\Delta d}^2 = 0.04/f_{sx}^2$ for various MCP rates. Curves α and δ are repeated from Fig. 8. The respective MCP rates in bits/pixel are: $R_{\text{MCP}}^a = 0.15$, $R_{\text{MCP}}^b = 0.45$, $R_{\text{MCP}}^c = 0.96$, $R_{\text{MCP}}^d = 1.55$, and $R_{\text{MCP}}^e = 3.15$. The location of each letter marking the curve indicates the MCP rate.

Fig. 8 and 9 shows the effectiveness of encoding above the MCP rate when the motion estimation is accurate ($\sigma_{\Delta d}^2 = 0.04/f_{sx}^2$), or equivalently, when video sequences have low motion. Notice that while in the $R > R_{\text{MCP}}$ case, the loss with respect to the NS coder is low except when decoding below the "knee" of the NS function (Curve α), the contrary effect happens when $R < R_{\text{MCP}}$: the lower the R_{MCP}, the lower the loss. Also notice that when decoding below R_{MCP}, the loss can be significantly greater than if simple intraframe coding is employed. In both cases, decoding above and below R_{MCP}, similar graphs are obtained when the accuracy of the motion estimation is decreased (or alternatively, high motion sequences are encoded) but the loss with respect to the NS coder is reduced.

6. CONCLUSIONS

Presented here was a closed-form expression of the rate-distortion function which serves as a lower bound for all MCP SNR or rate scalable video compression systems. Further insight is gained through deriving these results for fixed translatory motion with uncertainty in the displacement prediction.

7. REFERENCES

[1] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 video standard," *IEEE Trans. CSVT*, vol. 11, no. 3, pp. 301–317, March 2001.

[2] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE Journal SAC*, vol. SAC-5, no. 7, pp. 1140–1154, August 1987.

[3] B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman and Hall, 1997.

[4] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1971.

[5] G. W. Cook, *A Study of Scalability in Video Compression: Rate-Distortion Analysis and Parallel Implementation*, Ph.D. thesis, Purdue University, December 2002.