Rate distortion bounds for blocking and intra-frame prediction in videos

Jing Hu
Digital Signal Processing Group
Cisco Systems
jinghu@cisco.com

Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara
gibson@ece.ucsb.edu

***Abstract*—Recently we proposed a block-based conditional correlation coefficient model for natural videos in the spatial-temporal domain. The conditioning is on local texture and the optimal parameters can be calculated for a specific video with a mean absolute error (MAE) usually smaller than 5%. We used this conditional correlation model and the classic results on conditional rate distortion functions to calculate new theoretical rate distortion bounds for videos which appear to be the only valid theoretical rate distortion bounds with regard to the current cutting-edge video compression technologies such as those standardized in AVC/H.264. In this paper, we focus on utilizing the new block-based local-texture-dependent correlation model to derive rate distortion bounds for blocking and optimal prediction across neighboring blocks. We study the penalty paid in average rate when the correlation among the neighboring blocks is discarded completely or is incorporated partially through predictive coding. We calculate the thresholds in average rate and distortion when incorporating the correlation among the neighboring blocks through optimal predictive coding becomes worse than completely discarding this correlation. We also discuss the role of local texture in inter-frame prediction.**

I. Introduction

Parsimonious statistical models of natural images and videos can be used to calculate the rate distortion functions of these sources as well as to optimize particular image and video compression methods. Although they were studied extensively in the 1970s and 1980s, the statistical models and their corresponding rate distortion theories have fallen behind the fast advancing image and video compression schemes of the past two decades. In this period, the emphasis of rate distortion analysis for images and videos has been on setting up operational models for practical image and video compression systems to realize rate control [1]–[7] and to implement quality optimization algorithms [8]–[12]. In the meantime studying the theoretical rate distortion bounds for images and videos is often considered infeasible [9].

Recently we revisited the classic problem of developing a correlation model for natural videos by proposing a block-based local-texture-dependent correlation coefficient model for natural videos in the spatial-temporal domain. We define the correlation coefficient of two pixels in two nearby video frames as the product of the spatial correlation coefficient of these two pixels, as if they were in the same frame, and a variable to quantify the temporal correlation between these two video frames. The spatial correlation model for pixels within one video frame is a conditional correlation model. The conditioning is on local texture and the optimal parameters can be calculated for a specific video with a mean absolute error (MAE) usually smaller than 5%. We use this conditional correlation model to calculate the conditional rate distortion function when universal side information on local texture is available at both the encoder and the decoder. We demonstrate that this side information, when available, can save as much as 1 bit per pixel for a single video frame and 0.5 bits per pixel for multiple video frames. This rate distortion bound with local texture information taken into account while making no assumptions on coding, is shown indeed to be a valid lower bound with respect to the operational rate distortion curves of both intra-frame and inter-frame coding in AVC/H.264. The results also suggest a potential coding efficiency improvement in video codec design by involving texture information even for inter-coded frames.

In this paper, we focus on utilizing the new block-based local-texture-dependent correlation model to derive rate distortion bounds for blocking and optimal prediction across neighboring blocks. The "blocking" scheme, referring to breaking an image frame into 16×16 pixel MBs and processing one MB at a time, has been employed in the most popular image coding standards such as JPEG and almost all video coding standards such as MPEG-2/4 and the H.26x series [13]–[16]. In AVC/H.264 a new coding technique called intra-frame prediction is integrated to reduce the spatial redundancy in the intra-coded frames. Blocking and intra-frame prediction have opposite effects on compression efficiency since blocking completely disregards the correlation among the neighboring blocks while intra-frame prediction restores, partially, this correlation. With the new block-based local-texture-dependent correlation model, an explicit study of the rate distortion behavior of these two key coding schemes is feasible. In this paper we study the penalty paid in average rate when the correlation among the neighboring MBs or blocks is disregarded completely by blocking or is incorporated partially through the predictive coding.

The remainder of this paper is organized as follows. In Section II we review the novel new correlation model based

This work was supported by the California Micro Program, Applied Signal Technology, Cisco, Inc., Dolby Labs, Inc., Marvell, Inc. and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc..

on local texture and the theoretical rate distortion bound with the local texture as the side information. This section ends on a discussion of the role of local texture in inter-frame prediction. In Section III we derive the rate distortion bounds for the blocking scheme alone and in Section IV we derive the rate distortion bounds for blocking and prediction across the blocks. These various rate distortion bounds are compared to the operational rate distortion curves of intra-frame and inter-frame coding in AVC/H.264 throughout Sections II-IV. We conclude this paper and provide insights into future research in Section V.

II. A TEXTURE DEPENDENT CORRELATION MODEL AND THEORETICAL RATE DISTORTION BOUNDS FOR VIDEOS

We assume that all pixel values within one natural video form a three dimensional Gaussian random vector with memory, and each pixel value is of zero mean and the same variance σ^2. To quantify the effect of the surrounding pixels on the correlation between pixels of interest, we utilize the concept of local texture, which is simplified as local orientation, i.e., the axis along which the luminance values of all pixels in a local neighborhood have the minimum variance. The local texture is similar to the intra-prediction modes in AVC/H.264, but with a generalized block size and an arbitrary number of total textures. The block can have any rectangular shape as long as its size is small compared to the size of the image. To calculate the local texture of a block, we employ the pixels on the top and to the left of this block as surrounding pixels. Once the block size and the available local textures are fixed, the local texture of the current block is chosen as the one that minimizes the mean absolute error (MAE) between the original block and the prediction block constructed based on the surrounding pixels and the available local textures. It is important to point out that even through we choose a very simple and computationally inexpensive way to calculate the local texture, there are other, more sophisticated schemes of doing so, as summarized in [17], which should produce even better results in correlation modeling.

The following is the formal definition of the new correlation coefficient model that is dependent on the local texture.

Definition 2.1: The correlation coefficient of two pixel values within a digitized natural video, with spatial offsets Δi and Δj, and temporal offset Δk, is defined as

$$\rho(\Delta i, \Delta j, \Delta k|y_1, y_2) = \rho_s(\Delta i, \Delta j|y_1, y_2)\rho_t(\Delta_k). \quad \text{(II.1)}$$

$\rho_s(\Delta i, \Delta j|y_1, y_2)$ is the spatial correlation coefficient and

$$\rho_s(\Delta i, \Delta j|y_1, y_2) = \frac{\rho_s(\Delta i, \Delta j|y_1) + \rho_s(\Delta i, \Delta j|y_2)}{2}, \quad \text{(II.2)}$$

where

$$\rho_s(\Delta i, \Delta j|y) = a(y) + b(y)e^{-|\alpha(y)\Delta i + \beta(y)\Delta j|^{\gamma(y)}}. \quad \text{(II.3)}$$

y_1 and y_2 are the local textures of the blocks the two pixels are located in, and the parameters a, b, α, β and γ are functions of the local texture y. We restrict $b(y) \geq 0$ and $a(y) + b(y) \leq 1$. $\rho_t(\Delta_k)$ quantifies the temporal correlation and can be calculated by averaging the approximate temporal correlation coefficients $\hat{\rho}_t(\Delta_k|y)$, over all local texture y's.

For each local texture, we choose the combination of the five parameters a, b, α, β and γ that jointly minimizes the MAE between the approximate correlation coefficients, averaged among all the blocks in a video frame that have the same local texture, denoted by $\hat{\rho}_s(\Delta i, \Delta j|y)$, and the correlation coefficients calculated using the new model, $\rho_s(\Delta i, \Delta j|y)$. These optimal parameters for one frame in Paris.cif and their corresponding MAEs are presented in Table I. (The local textures are calculated for each one of the 4 by 4 blocks; the available local textures are chosen to be those implemented in AVC/H.264; Δi and Δj range from -7 to 7.) We can see from this table that the parameters associated with the new model are quite distinct for different local textures while the MAE is always less than 0.05.

TABLE I
THE OPTIMAL PARAMETERS FOR ONE FRAME IN PARIS.CIF AND THEIR CORRESPONDING MEAN ABSOLUTE ERRORS (MAES)

Paris.cif						
	a	b	γ	α	β	MAE
texture #0	0.3	0.6	0.7	0.0	0.6	0.022
texture #1	0.3	0.6	0.9	-0.2	0.0	0.024
texture #2	0.6	0.3	0.9	0.0	-0.1	0.035
texture #3	0.6	0.3	0.9	-0.2	-0.1	0.043
texture #4	0.6	0.3	0.7	0.1	-0.2	0.034
texture #5	0.6	0.3	0.7	0.2	-0.6	0.028
texture #6	0.6	0.4	0.5	-1.3	0.4	0.026
texture #7	0.6	0.4	0.5	0.4	1.1	0.030
texture #8	0.6	0.4	0.6	0.4	0.1	0.046

In Fig. 1 we plot $\hat{\rho}_s(\Delta i, \Delta j|y)$ (shown in the plots as the loose surfaces, i.e., the mesh surfaces that look lighter with fewer data points) and $\rho_s(\Delta i, \Delta j|y)$ (shown in the plots as the dense surfaces, i.e., the mesh surfaces that look darker with more data points) of all the local textures for the same image from paris.cif using the optimal parameters. We can see that the new spatial correlation model does capture the dependence of the correlation between two pixels on the local texture and fits the average approximate correlation coefficients $\hat{\rho}_s(\Delta i, \Delta j|y)$ very well. In [18] we further compare the optimal values of the parameters a, b, α, β and γ and their respective MAEs for different videos, different frames throughout the same video, and for different block sizes and different spatial offsets Δi's and Δj's.

Having established the correlation model, we construct the video source in a frame k by two parts: $\underline{X}_k$ as an M by N block (row scanned to form an MN by 1 vector) and $\underline{S}_k$ as the surrounding $2M+N+1$ pixels ($2M$ on the top, N to the left and the one on the left top corner, forming a $2M+N+1$ by 1 vector). If we investigate the rate distortion bounds of a few frames k_1, k_2, ..., k_l, the video source across all these

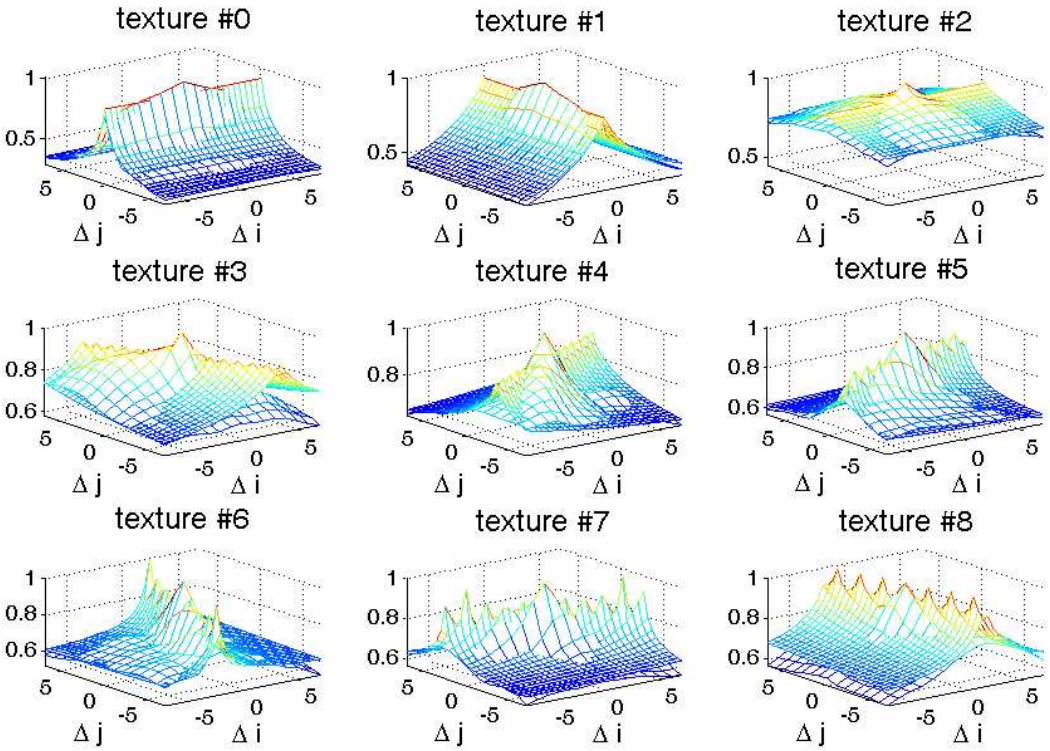


Fig. 1. The loose surfaces (the mesh surfaces that look lighter with less data points) are $\hat{\rho}_s(\Delta i, \Delta j|y)$, the approximate correlation coefficients of two pixel values in the first frame from paris.cif, averaged among the blocks that have the same local texture; the dense surfaces (the mesh surfaces that look darker with more data points) are $\rho_s(\Delta i, \Delta j|y)$, the correlation coefficients calculated using the proposed conditional spatial correlation model, along with the optimal set of parameters

frames is defined as a long vector $\underline{V}$, where

$$\underline{V} = [\underline{X}_{k_1}^T, \underline{S}_{k_1}^T, \underline{X}_{k_2}^T, \underline{S}_{k_2}^T, \ldots, \underline{X}_{k_l}^T, \underline{S}_{k_l}^T]^T. \quad \text{(II.4)}$$

For the local textures we use a variable Y to denote the information of local textures formulated from a collection of natural videos and Y is considered as universal side information available to both the encoder and the decoder. We only employ the first order statistics of Y, $P[Y = y]$, i.e., the frequency of occurrence of each local texture in the natural videos. In simulations, when available, $P[Y = y]$ is calculated as the average over a number of natural video sequences commonly used as examples in video coding studies.

The rate distortion bound of the video source $\underline{V}$ without taking into account the texture Y as side information, depicted by $R_{\textit{no texture}}(D)$, is a straightforward rate distortion problem of a source with memory which has been studied extensively. The rate distortion bound with the local texture as side information is a conditional rate distortion problem of a source with memory. It is defined as [19, Sec. 6.1]

$$R_{\underline{V}|Y}(D) = \min_{p(\underline{\hat{v}}|\underline{v},y):d(\underline{V},\underline{\hat{V}}|Y)\leq D} I(\underline{V};\underline{\hat{V}}|Y). \quad \text{(II.5)}$$

It can be proved [20] that the conditional rate distortion function in Eq. (II.5) can also be expressed as

$$R_{\underline{V}|Y}(D) = \min_{D_y' s:\sum_y D_y p(y)\leq D} \sum_y R_{\underline{V}|y}(D_y)p(y), \quad \text{(II.6)}$$

and the minimum is achieved by adding up $R_{\underline{V}|y}(D_y)$, the individual, also called marginal, rate-distortion functions, at points of equal slopes of the marginal rate distortion functions, i.e., when $\frac{\partial R_{\underline{V}|y}(D_y)}{\partial D_y}$ are equal for all y and $\sum_y D_y P[Y = y] = D$. These marginal rate distortion bounds can also be calculated using the classic results on the rate distortion bound of a Gaussian vector source with memory and a mean square error criterion, where the correlation matrix of the source is dependent on local texture y.

In Fig. 2 we plot these marginal rate distortion bounds for the first frame of paris.cif. This plot shows that the rate distortion curves of the blocks with different local textures are very different. Without the conditional correlation coefficient model proposed in this paper, this difference could not be calculated explicitly.

In Fig. 7 we plot the two rate distortion bounds $R_{\underline{V}|Y}(D)$ and $R_{\textit{no texture}}(D)$ as dashed and solid lines, respectively, as well as the operational rate distortion functions of *intra-frame* coding in AVC/H.264, for the first frame of paris.cif. In AVC/H.264, for both intra-frame and inter-frame coding, we choose the main profile with context-adaptive binary

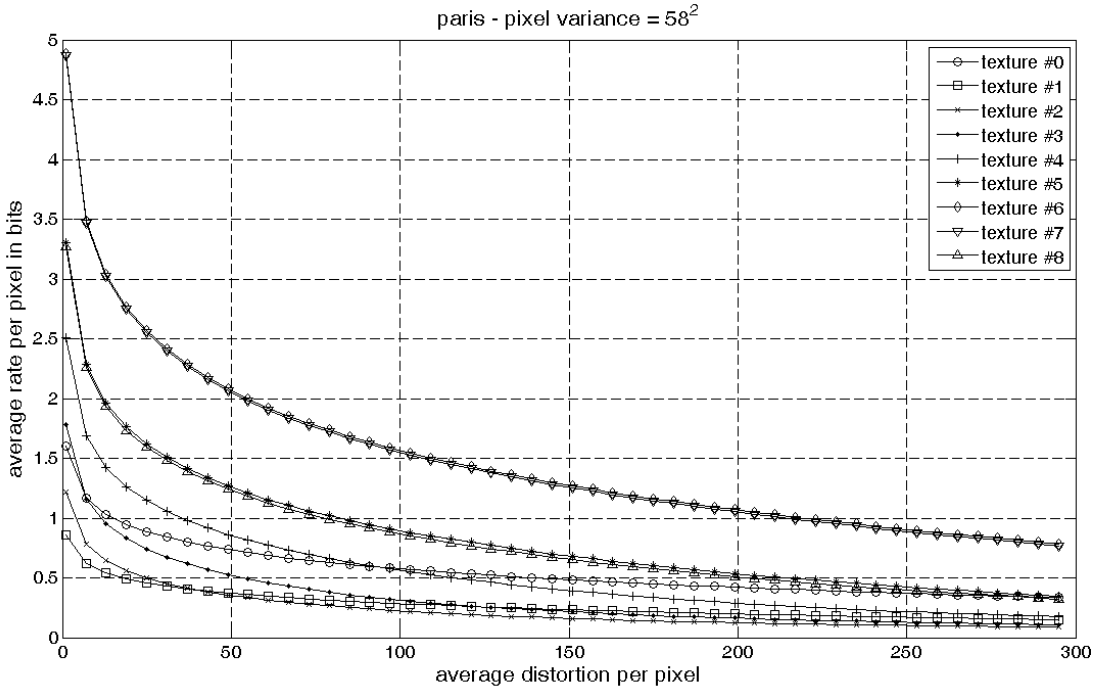


Fig. 2. Marginal rate distortion functions for different local textures, $R_{\underline{V}|Y=y}(D_y)$, for a frame in paris.cif

arithmetic coding (CABAC), which is designed to generate the lowest bit rate among all profiles. Rate distortion optimized mode decision and a full hierarchy of flexible block sizes from MBs to 4x4 blocks are used to maximize the compression gain. For the rate distortion bounds, we choose the block size 16x16 and the spatial offsets as from -16 to 16.

Comparing the two rate distortion bounds $R_{\underline{V}|Y}(D)$ and $R_{\textit{no texture}}(D)$ as dashed and solid lines, respectively, for paris.cif in Fig. 7 shows that engaging the first-order statistics of the universal side information Y saves at least 1 bit per pixel at low distortion levels (distortion less than 25, PSNR higher than 35 dB), which corresponds to a reduction of about 100 Kbits per frame for the CIF videos and 1.5 Mbps if the videos only have intra-coded frames and are played at a medium frame rate of 15 frames per second. This difference decreases as the average distortion increases but remains between a quarter of a bit and half a bit per pixel at high distortion level (distortion at 150, PSNR at about 26 dB), corresponding to about 375 Kbps to 700 Kbps in bit rate difference. Furthermore, the rate distortion bound without local texture information, $R_{\textit{no texture}}(D)$, plotted as a solid line, is higher than the actual operational rate distortion curve of *intra-frame* coding in AVC/H.264 at all distortion levels. The rate distortion bound with local texture information taken into account while making no assumptions in coding, i.e., $R_{\underline{V}|Y}(D)$, as in Eq. (II.5), plotted as a dashed line, is indeed a lower bound with respect to the operational rate distortion curves of *intra-frame* coding in AVC/H.264.

In Fig. 4 we plot the two rate distortion bounds $R_{\underline{V}|Y}(D)$ and $R_{\textit{no texture}}(D)$ as dashed and solid lines, respectively, as

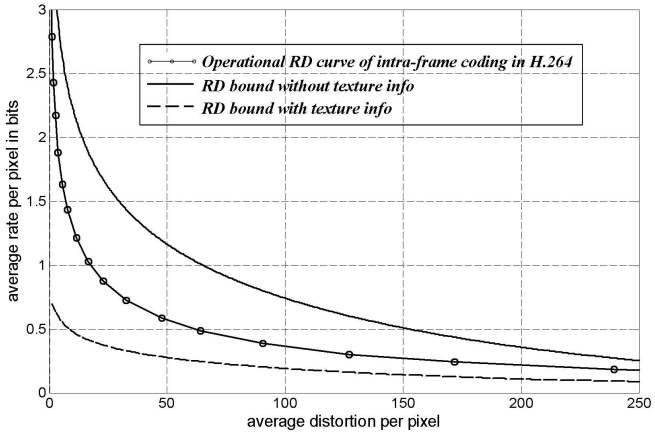
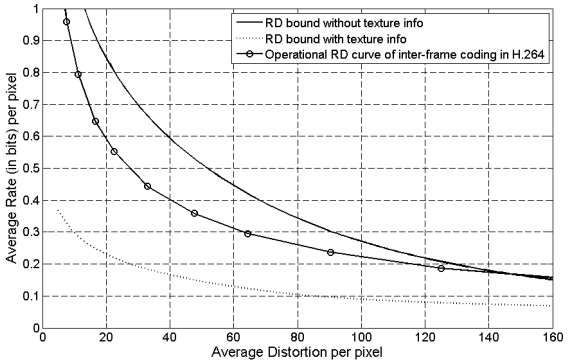
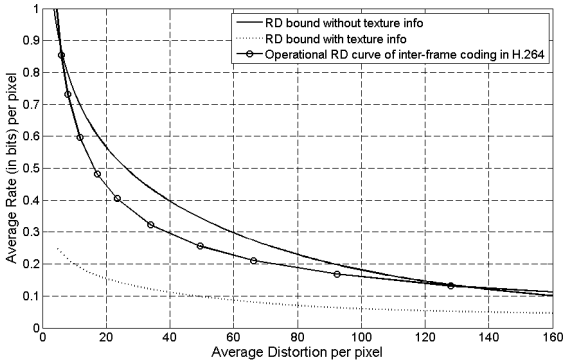


Fig. 3. Comparison of the rate distortion bounds and the operational rate distortion curves of paris.cif intra-coded in AVC/H.264

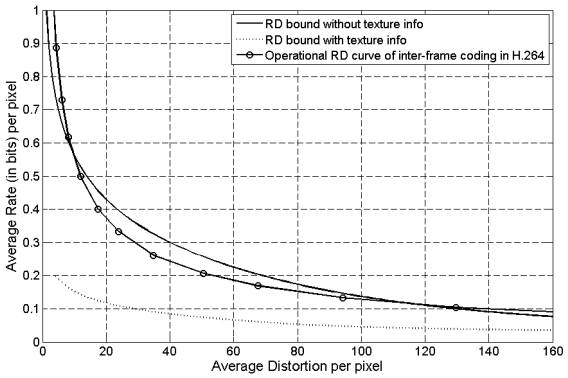
well as the operational rate distortion functions of *inter-frame* coding in AVC/H.264, for the first few frames in paris.cif. As shown in Fig. 4, the rate distortion bound without local texture information, plotted as solid lines, are higher than, or intersect with, the actual operational rate distortion curve of AVC/H.264. The rate distortion bounds with local texture information taken into account while making no assumptions in coding, plotted as dotted lines, are indeed lower bounds with respect to the operational rate distortion curves of AVC/H.264. comparing the two rate distortion bounds $R_{\underline{V}|Y}(D)$ and $R_{\textit{no texture}}(D)$ in Fig. 4(a) shows that by engaging the first-order statistics of the universal side information Y saves 0.5 bit per pixel at low distortion levels (distortion less than 25, PSNR higher than 35 dB), which corresponds to a reduction



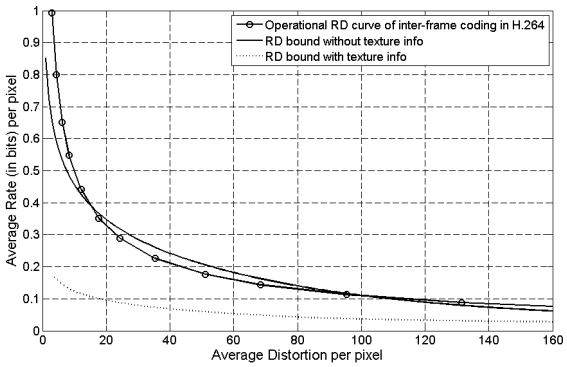
(a) Frames 1 and 2



(b) Frames 1, 2 and 3



(c) Frames 1, 2, 3 and 4



(d) Frames 1, 2, 3, 4 and 5

Fig. 4. Theoretical rate distortion bounds and the rate distortion curves of inter-frame coding in AVC/H.264

of about 50 Kbits per frame for the CIF videos and 750 Kbps if the videos have a group of picture size equal to 2 and are played at a medium frame rate of 15 frames per second. This difference decreases as the average distortion increases but remains 0.1 bit per pixel at high distortion level (distortion at 150, PSNR at about 26 dB), corresponding to about 150 Kbps in bit rate difference.

Another interesting observation of Fig. 4 is that as more video frames are coded, the actual operational rate distortion curves of inter-frame coding in AVC/H.264 become closer and closer to the theoretical rate distortion bound when no texture information is considered. This is because in AVC/H.264, only the intra-coded frames (i.e., only the 1^{st} frame in our simulation) take advantage of the local texture information through intra-frame prediction, while the inter-coded frames are blind to the local texture information. Therefore, when more frames are inter-coded, the bit rate saving achieved by intra-frame prediction in the 1^{st} frame is averaged over a larger number of coded frames. This suggests a possible coding efficiency improvement in video codec design by involving texture information even for inter-coded frames.

III. Rate distortion bounds for blocking only

In this section we are interested in the penalty paid in average rate when the correlation among the neighboring MBs or blocks are discarded completely. The basic set up for this section and the next section can be summarized in the block diagram in Fig. 5. In this block diagram $\underline{X}$ denotes the M by N block currently being processed in a video frame. The surrounding $2M+N+1$ pixels ($2M$ on the top, N to the left and the one on the left top corner), denoted by $\underline{S}$, are used to form a prediction block for each one of the available local textures, as

$$\underline{Z} = \underline{X} - P_d^{(A)} \underline{S}, \tag{III.7}$$

where $P_d^{(a)}$ is a $M \times N$ by $2M + N + 1$ matrix, different for each local texture. A is the local texture chosen for the current block which yields the smallest prediction error. $\underline{Z}$ and A are further coded and transmitted to the decoder, where the predicted value is added in to obtain

$$\underline{\hat{X}} = \underline{\hat{Z}} + P_d^{(\hat{A})} \underline{\hat{S}}. \tag{III.8}$$

In this Section we use the separate distortion measure on $\underline{X}$ and $\underline{S}$ since in video coding each MB is processed sequentially and only local distortion is considered. The rate distortion bounds calculated using a separate distortion measure should be slightly higher than those when a joint distortion measure on $\underline{X}$ and $\underline{S}$ is used.

The total rate to code $\underline{X}$ and $\underline{S}$ separately with a separate

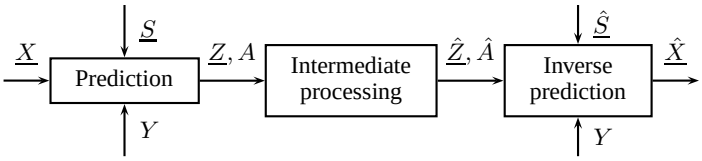


Fig. 5. Coding of one M by N block $\underline{X}$ and the surrounding $2M+N+1$ pixels $\underline{S}$

distortion constraint can be calculated as

$$R_{blocking}(D) = \frac{R_{\underline{X}}(D)|\underline{X}|+R_{\underline{S}}(D)|\underline{S}|}{|\underline{S}|+|\underline{X}|}, \tag{III.9}$$

which is simply the average of the rate distortion functions of $\underline{X}$ and $\underline{S}$. We plot $R_{blocking}(D)$ as dotted lines in Fig. 6 for two videos paris.cif and football.cif. Not surprisingly for both videos, coding $\underline{S}$ and $\underline{X}$ separately costs more bits than coding them jointly. We also find out that the difference in bit rate decreases as the block size increases, since for smaller block sizes information on stronger correlation across the blocks is disregarded. With the new correlation coefficient model and the corresponding rate distortion curves, we can calculate explicitly the bit rate increase caused by blocking. For example, this penalty is one sixth bit per pixel in this plot at all distortion levels in Fig. 6(a), which is quite significant.

IV. Rate distortion bound for blocking and optimal prediction across neighboring blocks

In this section we focus on the scenario when the video frames are processed block by block sequentially but the correlation among the blocks is utilized through predictive coding. We shall restrict ourselves to the separate distortion measure and therefore $\underline{S}$ is first coded with no consideration of $\underline{X}$. After that $\underline{Z}$ and A are calculated through intra-prediction in Eq. (III.7). Therefore the rate distortion function for this scenario is

$$R_{\underline{S},\underline{Z},A\ separately-without Y}(D) = \left(\min_{p(\hat{\underline{s}}|\underline{s}):\frac{E[||\underline{S}-\hat{\underline{S}}||^2]}{|\underline{S}|}\le D} I(\underline{S};\hat{\underline{S}}) + \min_{p(\hat{\underline{z}},\hat{a}|\underline{z},a,\underline{s},\hat{\underline{s}}):\frac{E[||\underline{X}-\hat{\underline{X}}||^2]}{|\underline{X}|}\le D} I(\underline{Z},A;\hat{\underline{Z}},\hat{A})\right)/(|\underline{S}|+|\underline{X}|) \tag{IV.10}$$

If we restrict that $A=\hat{A}$, i.e., we code the local texture A losslessly, the second part in Eq. (IV.10) becomes

$$\min_{p(\hat{\underline{z}},\hat{a}|\underline{z},a,\underline{s},\hat{\underline{s}}):\frac{1}{|\underline{X}|}E[||\underline{X}-\hat{\underline{X}}||^2]\le D} I(\underline{Z},A;\hat{\underline{Z}},\hat{A}) = \min_{p(\hat{\underline{z}}|\underline{z},a,\underline{s},\hat{\underline{s}}):\frac{1}{|\underline{X}|}E[||\underline{X}-\hat{\underline{X}}||^2]\le D} I(\underline{Z};\hat{\underline{Z}}|A)+H(A), \tag{IV.11}$$

which forms an upper bound for all the scenarios when A is coded either losslessly or subject to a fidelity criterion. Also when $A=\hat{A}$, we have

$$E[||\underline{X}-\hat{\underline{X}}||^2] = \sum_a Pr(a)E[||(\underline{Z}+P_d^{(a)}\underline{S})-(\hat{\underline{Z}}+P_d^{(a)}\hat{\underline{S}})||^2|a] = \sum_a Pr(a)\int_{\underline{s}}\int_{\hat{\underline{s}}}\int_{\underline{z}}\int_{\hat{\underline{z}}} p(\underline{z},\hat{\underline{z}},\underline{s},\hat{\underline{s}}|a)(\hat{\underline{z}}-\underline{z})^T(\hat{\underline{z}}-\underline{z}) + (\hat{\underline{s}}-\underline{s})^T P_d^{(a)T}P_d^{(a)}(\hat{\underline{s}}-\underline{s}) + 2(\hat{\underline{s}}-\underline{s})^T P_d^{(a)T}(\hat{\underline{z}}-\underline{z})d\underline{s}d\hat{\underline{s}}d\underline{z}d\hat{\underline{z}}. \tag{IV.12}$$

In order to investigate the lowest rate when predictive coding is employed, we use the optimal linear predictor $P_{opt}^{(a)} = E[\underline{X}\underline{S}^T|a](E[\underline{S}\underline{S}^T|a])^{-1}$ assuming that $E(\underline{S}\underline{S}^T|a)$ is non-singular. Since the source is assumed to be zero-mean Gaussian, the optimal linear predictor is also the optimal conditional mean predictor. The optimality is in the sense of minimizing MSE of $\underline{X}$. When the optimal linear predictor $P_{opt}^{(A)}$ is used, the cross-product term in Eq. (IV.12) disappears. Let

$$D'_{\underline{S}} = \sum_a Pr(a)\int_{\underline{s}}\int_{\hat{\underline{s}}} p(\underline{s},\hat{\underline{s}}|a)(\hat{\underline{s}}-\underline{s})^T P_{opt}^{(a)T}P_{opt}^{(a)}(\hat{\underline{s}}-\underline{s})d\underline{s}d\hat{\underline{s}}. \tag{IV.13}$$

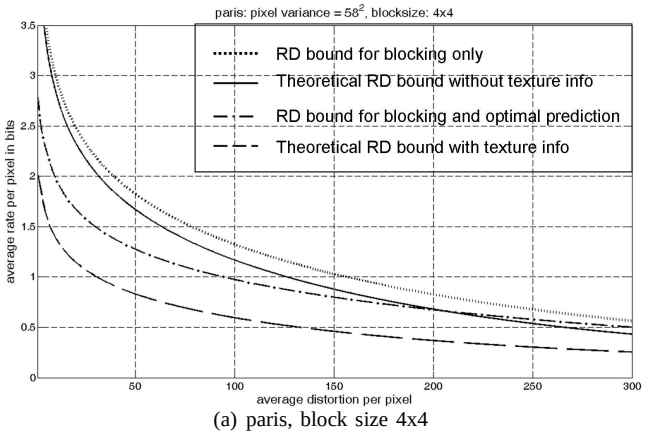
Eq. (IV.12) becomes

$$E[||\underline{X}-\hat{\underline{X}}||^2] = |\underline{Z}|D_{\underline{Z}} + D'_{\underline{S}}. \tag{IV.14}$$

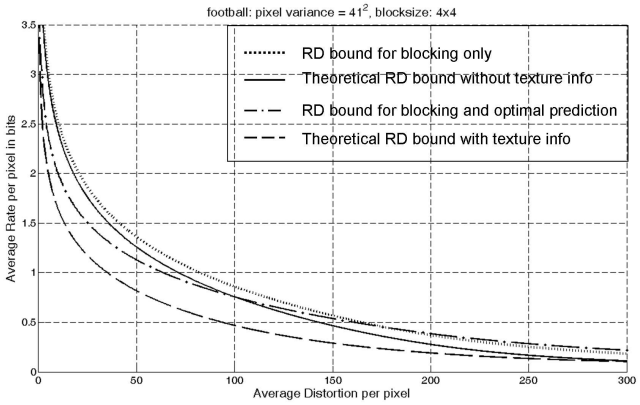
Since $\underline{S}$ is optimally coded without consideration of $\underline{X}$ as in the first part of Eq. (IV.10), $D'_{\underline{S}}$ is fixed as well. The constraint on the distortion of $\underline{Z}$ becomes

$$D_{\underline{Z}} \le (|\underline{X}|D - D'_{\underline{S}})/|\underline{Z}|. \tag{IV.15}$$

An upper bound for Eq. (IV.10), depicted by



(a) paris, block size 4x4



(b) football, block size 4x4

Fig. 6. Comparison of rate distortion bounds for two videos paris.cif and football.cif

$R_{\textit{opt-pred-upperbound}}$, is thus

$$R_{\textit{opt-pred-upperbound}}(D) = \frac{1}{|\underline{S}|+|\underline{X}|} \left(|\underline{S}|R_{\underline{S}}(D) + |\underline{Z}|R_{\underline{Z}|A}(\frac{|\underline{X}|D-D'_{\underline{S}}}{|\underline{Z}|}) + H(A)\right) \quad \text{(IV.16)}$$

The conditional rate distortion function $R_{\underline{Z}|A}(D_{\underline{Z}})$ in Eq. (IV.16) is again calculated based on the "equal slope" theorem of the marginal rate distortion functions $R_{\underline{Z}|A=a}(D_a)$ [20]. In this case since the actual local texture A is coded without any loss, the exact statistics of A are available at both the encoder and the decoder; therefore, whether the universal side information Y is available or not becomes insignificant. The only complexity in computation is caused because $E(\underline{S}\underline{S}^T|a)$ is usually singular when the direction of the local texture is DC, horizontal, vertical, or too close to horizontal/vertical. In these cases we use the pseudo-inverse matrix of $E(\underline{S}\underline{S}^T|a)$ in the calculation.

$R_{\textit{opt-pred-upperbound}}(D)$ is also plotted in Fig. 6 for the two videos paris.cif and football.cif. As seen from this figure, the bit rate decrease from the dotted lines (coding $\underline{S}$ and $\underline{X}$ separately, Eq. (III.9)) to the dash-dotted lines (the upper bound of coding $\underline{S}$, $\underline{Z}$ and A separately with optimal prediction, $R_{\textit{opt-pred-upperbound}}(D)$) is truly phenomenal in both plots at low distortion levels. The bit rate difference is about 1 bit per pixel for paris and between half a bit to 1 bit per pixel for for football at distortion 25 (corresponding to PSNR 35 dB). This bit rate saving decreases as the distortion increases, and interestingly, it vanishes for football at certain distortions. This is because spending bits coding the local texture A losslessly becomes unjustifiable at high distortion levels. This is especially true when the bit rate is low and the processing block size is small. We can see that in Fig. 6(b) the dash-dotted line and the dotted line intersect at a distortion of about 180, corresponding to an average rate of 0.4 bits per pixel. The average bit rate spent on coding the local texture A losslessly is simply the entropy of A, divided by the number of pixels per block, which is 16 in Fig. 6(b) since 4×4 blocks are investigated. This average rate is about 0.2 bits per pixel, or 50% of the total average rate. This is to say that for this particular video football.cif, processed in 4×4 blocks, 0.4 bits per pixel is the threshold in average rate that depicts when incorporating the correlation among the neighboring blocks through optimal predictive coding and coding the local texture A losslessly, becomes worse than discarding the correlation among the neighboring blocks. This crossover average rate is different for different videos and different processing blocksizes. It can be calculated along with the rate distortion bounds we derive in this paper and be utilized in real video codecs.

In Fig. 7 we plot the three rate distortion bounds derived in this paper for paris.cif and the operational rate distortion functions for paris.cif intra-coded in AVC/H.264. As shown in Fig. 7, the rate distortion bound calculated based on the new texture dependent correlation model for the scenario where optimal predictive coding is engaged to code $\underline{S}$, $\underline{Z}$ and A separately with separate distortion constraint, i.e., $R_{\textit{opt-pred-upperbound}}(D)$ as in Eq. (IV.16), plotted as a dash dotted line, is a reasonably tight lower bound, especially at medium to high distortion levels. In Fig. 8(a) we plot this lower bound $R_{\textit{opt-pred-upperbound}}(D)$ (Eq. (IV.16)) and the operational rate distortion function using AVC/H.264 for two other videos. We can see that although the lower bounds are calculated based on only five parameters generated from each video, they do agree with the operational rate distortion curves of the corresponding video reasonably well. If we further plot these lower bounds as average rate per pixel versus PSNR of a video frame as in Fig. 8(b), the lower bounds appear to be nearly linear which shows promises in codec design.

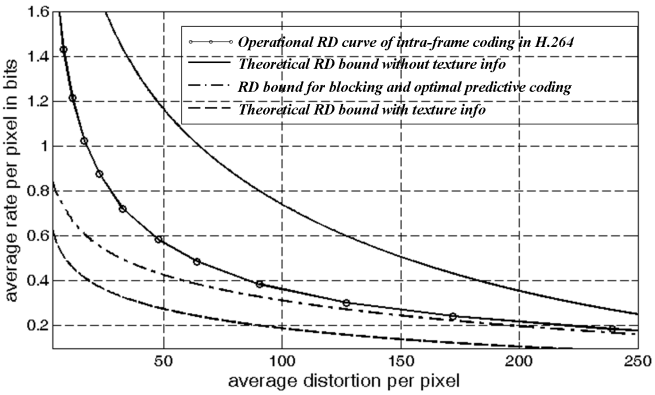


Fig. 7. Comparison of the rate distortion bounds and the operational rate distortion curves of paris.cif intra-coded in AVC/H.264

V. Conclusions

We utilize a recently proposed block-based local-texture-dependent correlation model to derive rate distortion bounds for blocking and optimal prediction across neighboring blocks. We study the penalty paid in average rate when the correlation among the neighboring blocks is discarded completely or is incorporated partially through predictive coding. We calculate the thresholds in average rate and distortion when incorporating the correlation among the neighboring blocks through optimal predictive coding becomes worse than completely discarding this correlation. We also discuss the role of local texture in inter-frame prediction. All of these results are derived from a correlation model in the spatial-temporal domain of videos that is independent of any other specific video coding scheme and therefore are very different from the operational rate distortion analysis of videos.

References

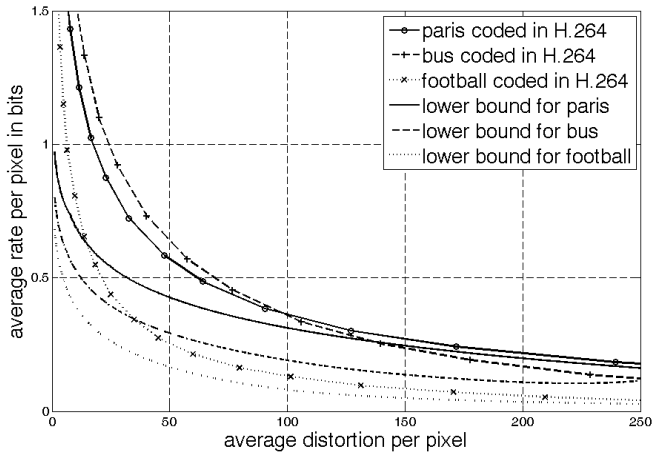
[1] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–251, Feb. 1997.

[2] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

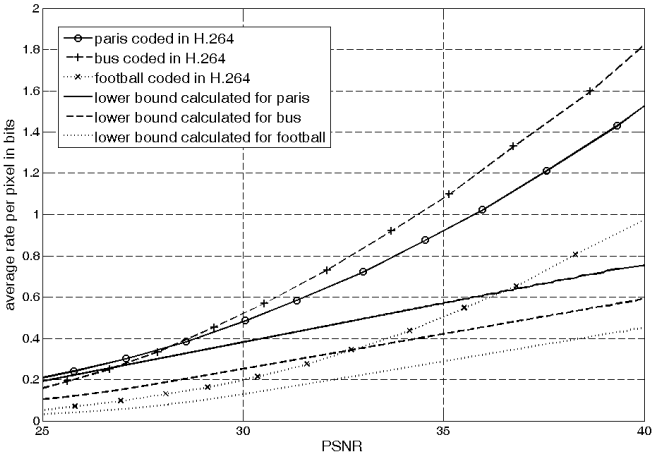
[3] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[4] S. Ma, W. Gao, and Y. Lu, "Rate control on JVT standard," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-D030*, Jul. 2002.

[5] Z. G. Li, F. Pan K. P. Lim, X. Lin and S. Rahardj, "Adaptive rate control for h.264," *IEEE International Conference on Image Processing*, pp. 745–748, Oct. 2004.



(a) average rate vs. average distortion



(b) average rate vs. PSNR

Fig. 8. The lower bounds calculated based on the new correlation coefficient model and its corresponding optimal parameters for three videos, compared to the operational rate distortion curves of these videos coded in AVC/H.264

[6] Y. Wu et al., "Optimum bit allocation and rate control for H.264/AVC," *Joint Video Team of ISO/IEC MPEG & ITU-T VCEG Document*, vol. JVT-O016, Apr. 2005.

[7] D.-K. Kwon, M.-Y. Shen and C.-C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 517–529, May 2007.

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

[9] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, p. 2350, Nov. 1998.

[10] Z. He and S. K. Mitra, "From rate-distortion analysis to resource-distortion analysis," *IEEE Circuits and Systems Magazine*, vol. 5, no. 3, pp. 6–18, Third quarter 2005.

[11] Y. K. Tu, J.-F. Yang and M.-T. Sun, "Rate-distortion modeling for efficient H.264/AVC encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 530–543, May 2007.

[12] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, 2000.

[13] ISO/IEC 13818-1:2000, "Information technology – generic coding of moving pictures and associated audio information: Systems," 2000.

[14] ISO/IEC 14496-1:2001, "Information technology – coding of audio-visual objects – part 1: Systems," 2001.

[15] ITU Recommendations, "Video coding for low bit rate communication," *ITU-T rec. H.263*, Jan. 2005.

[16] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audio-visual services," 2003.

[17] T. Aach, C. Mota, I. Stuke, M. Mhlich, and E. Barth, "Analysis of superimposed oriented patterns," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3690–3700, Dec. 2006.

[18] J. Hu and J. D. Gibson, "New rate distortion bounds for natural videos based on a texture dependent correlation model," *to appear, IEEE Transactions on Circuits and Systems for Video Technology.*

[19] T. Berger, *Rate distortion theory*. New York: Wiley, 1971.

[20] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Tran. Inform. Theory*, vol. IT-19, no. 4, pp. 480–489, Jul. 1973.