A Stochastic Model for Video and its Information Rates

Arthur L. da Cunha   Minh Do
ECE Dept. and CSL, UIUC
{cunhada,minhdo}@uiuc.edu

Martin Vetterli
LCAV-EPF Lausanne
martin.vetterli@epfl.ch

Abstract

We propose a stochastic model for video and compute its information rates. The model has two sources of information representing ensembles of camera motion and visual scene data (i.e. "realities"). The sources of information are combined generating a vector process that we study in detail. Both lossless and lossy information rates are derived. The model is further extended to account for realities that change over time. We derive bounds on the lossless and lossy information rates for this dynamic reality model, stating conditions under which the bounds are tight. Experiments with synthetic sources suggest that in the presence of scene motion, simple hybrid coding using motion estimation with DPCM can be suboptimal relative to the true rate-distortion bound.

1 Introduction

Consider a moving camera that takes sample snapshots of an environment over time. The samples are to be coded for later transmission or storage. Because the movements of the camera are small relative to the scene, there are large correlations amongst multiple acquisitions.

Examples of such scenarios include video compression and the compression of light-fields. The practical aspects of these two examples have been studied extensively (see e.g. [1], [2], and references therein). But the theoretical aspects still lack a framework that enables the development of precise information rate results that are not tied to a particular coding scheme such as DPCM. We propose such a framework and derive information rate bounds. We develop a simple model that can be studied with the usual information theoretic tools, but that still bears the main elements of the practical case.

The general problem can be posed as shown in Fig. 1. There is a world or "reality" (e.g, panorama images, objects, moving objects), and a camera that generates a "view of reality" V. This "view of reality" (e.g. a video sequence) is coded with a source coder with memory M, giving an average rate of R bits. This bitstream is decoded with a decoder with memory M to reconstruct a view of reality $\hat{V}$ close to the original one in the MSE sense.

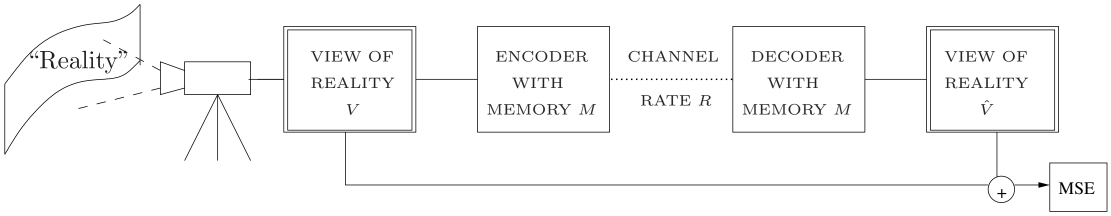
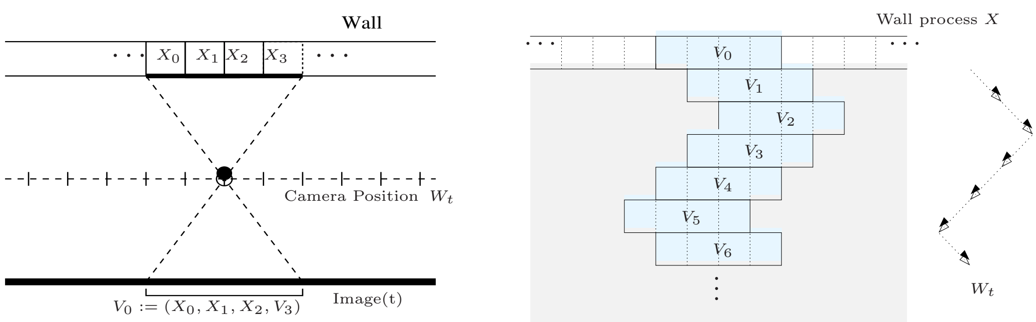


Figure 1: The problem under consideration. There is a world and a camera that produces a "view of reality" that needs to be coded with finite or infinite memory.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

2 Definitions and Problem Set Up

A simplified model for the problem described in Fig. 1 is as follows [3]. Consider a camera moving according to a 1-D discrete Bernoulli random walk. The random walk is the process $W = (W_t : t \in \mathbb{Z}^+)$ such that $\Pr\{W_0 = 0\} = 1$, and for $t \geq 1$, $W_t = \sum_{i=1}^{t} N_i$, where N_i is drawn i.i.d. from the set $\{-1, +1\}$ with $\Pr\{N_1 = +1\} = p_W$. We assume without loss of generality that $p_W \leq 0.5$. In front of the camera there is an infinite wall that represents a scene that is projected onto a screen in front of the camera path (i.e. we ignore occlusion). The wall is modelled as a 1-D strip "painted" with an i.i.d. process $X = (X_n : n \in \mathbb{Z})$ that is *independent* of the random walk W. The process X follows some probability law p_X drawing values from an alphabet $\mathcal{X}$. In the static case, the wall process X is drawn at $t = 0$. Fig. 2 (a) illustrates the proposed model.



(a) (b)

Figure 2: A stochastic model for video. (a) Simplified model. (b) The resulting vector process V. Each sample of the vector process is a block of L samples from the process $\mathcal{X}$ taken at the position indicated by the random walk W_t. In the figure $L = 4$.

At each random walk time step, the camera sees a block of L samples from the wall, where $L \geq 1$. This results in a vector process $V = (V_t : t \in \mathbb{Z}^+)$ indexed by the random walk positions.

Definition 1 *Let W be a random walk independent of X, and let L be an integer greater than one. The vector process $V = (V_t : t \in \mathbb{Z}^+)$ is defined as*

$$V_t := (X_{W_t}, X_{W_t+1}, \cdots, X_{W_t+L-1}). \tag{1}$$

The random walk is a simple stochastic model for an ensemble of camera movements and it includes *camera panning* as a special case. The discrete displacements of the random walk thus neglects other effects such as zooming, rotation, and change of angle. Consecutive samples of the vector process, which are vectors of length L, have at least $L - 1$ samples that are repeated. Fig. 2 (b) illustrates the vector process V.

The coding problem Given the vector process $V = (V_0, V_1, \cdots)$, the coding problem consists in finding an encoder/decoder pair that is able to describe and reproduce the process V at the decoder using no more than R bits per vector sample[1]. The

[1]The vector process can be shown to be stationary and mean-ergodic provided that the wall process X is stationary and mean-ergodic.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

IEEE COMPUTER SOCIETY

decoder reproduces the vector process $V = (\hat{V}_0, \hat{V}_1, \cdots)$ with some delay. The reproduction can be lossless or lossy with fidelity D. The encoder encodes each sample V_t based on the observation of M previous vector samples $V_{t-1}, \ldots, V_{t-M}$. Thus, M is the memory of the encoder/decoder. Since encoding is done jointly, there is a delay incurred. The information content of the process V provides the limiting rate needed to either perfectly reproduce the process V at the decoder, or to reproduce it within distortion D. The information content is usually only achievable at the expense of infinite memory and delay [4].

Properties of the random walk The following notions are needed in what follows.

Definition 2 *Let W be a random walk. We denote by $Pr\{\mathcal{R}^t\}$ the probability of recurrence at step t. That is, the probability of the event set*

$$\mathcal{R}^t = \{(W_0, W_1, \ldots, W_t) : W_t = W_s \ \text{ for some } 0 \leq s < t\}.$$

The probability of the complementary set $\Pr\left\{\overline{\mathcal{R}^t}\right\}$ is called the *first-passage* probability. When a site W_t has not occurred before, we refer to it as a *new site*. A related quantity is the probability of return.

Definition 3 *Let W be a random walk. The probability of return at step t after step $s < t$ is the probability of the event set*

$$\mathcal{T}_s^t = \{(W_0, W_1, \ldots, W_t) : W_t = W_s \text{ but } W_t \neq W_i, \text{ for } i \text{ such that } s < i < t\}.$$

When $s = 0$ we write $\mathcal{T}^t$ for $\mathcal{T}_0^t$. For the Bernoulli random walk one can check that

$$\mathcal{R}^t = \bigcup_{i=1}^{t} \mathcal{T}_{t-i}^t. \tag{2}$$

The sets $\mathcal{T}_s^t$ are shift invariant in the sense that $\mathcal{T}_s^t = \mathcal{T}^{t-s}$. Moreover, for the case of the Bernoulli random walk we have the following [5].

Lemma 1 *For the Bernoulli random walk with $p_W \leq 1/2$ the following holds:*

(i) $\lim_{t\to\infty} P(\overline{\mathcal{R}^t}) = 1 - 2p_W$.

(ii) *For $t > 0$, $Pr\{\mathcal{T}^{2t-1}\} = 0$, and $Pr\{\mathcal{T}^{2t}\} = 2C_{t-1}\left((1-p_W)p_W\right)^t$, where $C_t := \frac{1}{t+1}\binom{2t}{t}$.*

3 Information Rates for a Static Reality

Lossless Information Rates Denote $V^t = (V_1, \ldots, V_t)$. We assume without loss of generality that V_0 is known do the decoder. We seek to quantify the entropy rate of V [6]:

$$H(V) = \lim_{t\to\infty} \frac{1}{t} H\left(V^t\right) = \lim_{t\to\infty} H(V_t | V^{t-1}). \tag{3}$$

To characterize $H(V)$, we describe intuitively an upper and a lower bound (resp. sufficient and necessary rates) that will be formalized in Theorem 1 below. For a sufficient rate, note that V can be reproduced up to time t when both the trajectory $W^t = (W_1, \ldots, W_t)$ and the samples of the wall occurring at the new sites of W^t are

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

available. When t is large, this amounts to $H(W^t) = tH(p_W)$ bits for the trajectory, plus $t\Pr\left\{\overline{\mathcal{R}^t}\right\} H(X) \approx t(1-2p_W)H(X)$ for the new sites. So, a sufficient average rate is $H(p_W) + (1-2p_W)H(X)$. Moreover, the complexity of V is at least the complexity of the new sites, and so $(1-2p_W)H(X)$ is a necessary rate. This intuitive lower bound can be improved by examining the probability of correctly inferring the random path W^t from observing the vector process V^t. This probability is related to the probability of a repeated pattern represented by the following event:

$$A_L := \{\ (X_0, \ldots, X_L) = (x_0, x_1, x_0, x_1, \ldots),\ x_0, x_1\ \in \mathcal{X}\} \tag{4}$$

To see this, let $L = 4$ and consider inferring W_1 from (V_1, V_0). If $V_0 = (x_0, x_1, x_0, x_1)$ and $V_1 = (x_1, x_0, x_1, x_0)$, then it follows that W_1 cannot be unambiguously determined. Intuitively, if W^t can be determined from V^t, then the complexity of the trajectory is embedded in V^t and thus has to be fully described. If, however, there is ambiguity on W^t, then sets of W^t that are consistent with V^t can be indexed and coded with a lower rate. We are now ready to state and prove Theorem 1.

Theorem 1 *Consider the vector process V consisting of L-tuples generated by a Bernoulli random walk with transition probability $p_W \le 1/2$, and a wall process X, drawing values on a discrete alphabet, and that has entropy rate $H(X)$. The conditional entropy $H(V_t|V^{t-1})$ obeys*

$$Pr\left\{\overline{\mathcal{R}^t}\right\} H(X) + H(p_W) Pr\left\{\overline{A_L}\right\} \le H(V_t|V^{t-1}) \le \frac{1}{t}\sum_{i=1}^{t} Pr\left\{\overline{\mathcal{R}^i}\right\} H(X) + H(p_W), \tag{5}$$

where A_L is as in (4). In particular, the entropy rate $H(V)$ satisfies

$$(1-2p_W)H(X) + H(p_W) Pr\left\{\overline{A_L}\right\} \le H(V) \le (1-2p_W)H(X) + H(p_W). \tag{6}$$

Proof: For each t we have

$$\begin{aligned} H(V_t|V^{t-1}) &\overset{(a)}{\le} \frac{1}{t}\sum_{i=1}^{t} H(V_i|V^{i-1}) = \frac{H(V^t)}{t} \\ &\overset{(b)}{\le} \frac{H(W^t) + H(V^t|W^t)}{t} = \frac{H(W^t) + \sum_{i=1}^{t} H(V_i|V^{i-1}, W^t)}{t} \\ &\overset{(c)}{=} H(p_W) + \frac{1}{t}\sum_{i=1}^{t} H(V_i|V^{i-1}, W^i), \end{aligned} \tag{7}$$

where (a) follows because $H(V_t|V^{t-1})$ decreases with t, (b) holds because $H(W^t|V^t) \ge 0$, and (c) is true because $H(W^t) = tH(p_W)$ and $(W_{i+1}, \ldots, W_t)$ is independent of (V^i, W^i). Further, if $W^i = w^i$ recurs at i, then $H(V_i|w^i, V^{i-1}) = 0$. If otherwise w^i is such that w_i is new site, then $H(V_i|w^i, V^{i-1}) = H(X)$. Consequently,

$$H(V_i|V^{i-1}, W^i) = \sum_{w^i \in \overline{\mathcal{R}^i}} \Pr\left\{W^i = w^i\right\} H(V_i|V^{i-1}, W^i = w^i) = \Pr\left\{\ \overline{\mathcal{R}^i}\right\} H(X). \tag{8}$$

Combining (7) and (8) gives the upper bound in (5). We now turn to the lower bound. Using the chain rule for mutual information, and the information inequality

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

we have:

$$\begin{aligned} H(V_t|V^{t-1}) &= H(V_t|V^{t-1}, W^t) + I(W^t; V_t|V^{t-1}) \\ &= H(V_t|V^{t-1}, W^t) + I(W^{t-1}; V_t|V^{t-1}) + I(W_t; V_t|V^{t-1}, W^{t-1}) \\ &\geq H(V_t|V^{t-1}, W^t) + I(W_t; V_t|V^{t-1}, W^{t-1}). \end{aligned} \tag{9}$$

Moreover, because the random walk increment $W_t - W_{t-1}$ is independent of (V^{t-1}, W^{t-1}), it follows that

$$\begin{aligned} I(W_t; V_t|V^{t-1}, W^{t-1}) &= H(W_t|V^{t-1}, W^{t-1}) - H(W_t|V^t, W^{t-1}) \\ &= H(p_W) - H(W_t|V^t, W^{t-1}). \end{aligned} \tag{10}$$

We proceed by finding an upper bound for $H(W_t|V^t, W^{t-1})$. If $(\mathbf{v}^t, w^{t-1})$ is such that W_t is uniquely determined, then the conditional entropy is zero. Otherwise, if $(\mathbf{v}^t, w^{t-1})$ is such that W_t is ambiguous, then the conditional entropy is at most $H(p_W)$. So,

$$\begin{aligned} H(W_t|V^t, W^{t-1}) &= \sum_{(\mathbf{v}^t, w^{t-1})} \Pr\left\{w^{t-1}, \mathbf{v}^t\right\} H\left(W_t|V^t = \mathbf{v}^t, W^{t-1} = w^{t-1}\right) \\ &\leq H(p_W)\ \Pr\left\{\ (w^{t-1}, \mathbf{v}^t) \text{ such that } W_t \text{ is ambiguous}\right\} \end{aligned}$$

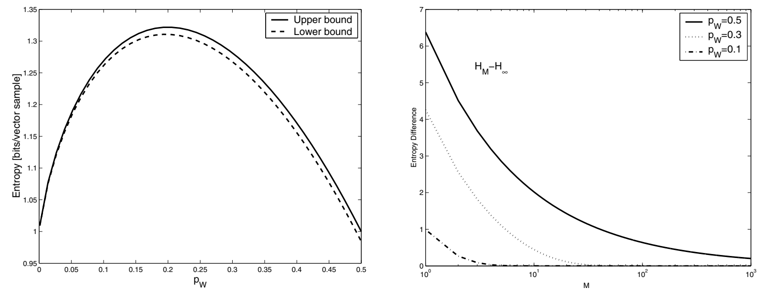
The event set on the right-hand side above is contained in the event set $\{V_{t-1} = (x_0, x_1, \ldots), V_t = (x_1, x_0, \ldots)\}$, which has probability $\Pr\{A_L\}$. So the right-hand side above is bounded by $H(p_W)\ \Pr\{A_L\}$. Combining this with (9 - 10) and (8) we assert the lower bound in (5). By letting $t \to \infty$ and using Lemma 1 (i) we obtain (6). ■

Example 1 Suppose that the X is uniformly distributed over $|\mathcal{X}|$ values. For simplicity let L be even. Then, it is easily seen that

$$\Pr\{A_L\} = |\mathcal{X}|^{-\lfloor \frac{L+1}{2} \rfloor}.$$

Consequently, the difference between upper and lower bounds decays exponentially fast when the block length $L \to \infty$. For fixed L, the difference also decays as $|\mathcal{X}|$ increases. At the limit, when the alphabet is continuous, then the lower and upper bounds coincide. Thus, for L and $|\mathcal{X}|$ sufficient large, we can estimate the entropy rate as $(1 - 2p_W)\log|\mathcal{X}| + H(p_W)$ bits per block. Fig. 3(a) illustrates the bounds when X is Bern(1/2), and $L = 8$.

Memory constrained coding The entropy rate $H(V)$ can be attained with an encoder-decoder pair with unbounded memory. In the finite memory case, the encoder has to code V_t based on the observation of $V_{t-1}, \ldots, V_{t-M}$, and the decoder proceeds accordingly. This situation is similar to one encountered in video compression, where a frame at time t is coded based on M previously coded frames. In this case, the average code-length is lower bounded by the conditional entropy $H(V_t|V_{t-1}, \ldots, V_{t-M}) = H(V_M|V_{M-1}, \ldots, V_0)$. The bound (5) in Theorem 1 describe the behavior of $H(V_M|V^{M-1})$. Intuitively, by looking at the stored samples from $t - M$ up to t, the encoder can separately code W_t and take advantage of recurrences existent from $t - M$ to $t - 1$. In effect, finite memory prevents the encoder to exploit long term recurrences that are not visible in the memory. Similar observations are



(a) (b)

Figure 3: Bounds of information rate. (a) Lower and upper bounds as a function of p_W for the binary wall with $p_X = 1/2$ and $L = 8$. (b) Difference $H(V) - H(V_M|V^M)$ as a function of M. When $p_W = 0.5$, the bit rate can be lowered significantly at the cost of large memory. A moderate bit rate reduction is obtained with small values of M when $p_W = 0.1$.

verified in practice for instance in [7] and [8]. Fig. 3(b) illustrates how memory influences coding when X is uniform over an alphabet of size $|\mathcal{X}| = 256$. The bounds of Theorem 1 are tight and hence are used to compute the entropy rates.

Lossy Information Rates Consider a t-tuple $(V_1, \ldots, V_t)$ and a reproduction t-tuple $(\hat{V}_1, \ldots, \hat{V}_t)$. The rate-distortion function for each t, and for the MSE distortion metric is given by:

$$R_{V^t}(D) = \inf \left\{ t^{-1} I(V^t; \hat{V}^t) : p(\hat{V}^t|V^t) \text{ such that } \mathbb{E}d(V^t, \hat{V}^t) \leq D \right\}, \qquad (11)$$

where $d(V^t, \hat{V}^t) = (tL)^{-1} \sum_{i=1}^{t} \|V_i - \hat{V}_i\|^2$. The rate-distortion function for the process $V = (V_1, V_2, \ldots)$ is given by [4]

$$R_V(D) = \lim_{t\to\infty} R_{V^t}(D). \qquad (12)$$

By coding the side information W^t separately, a constructive upper bound for $R_V(D)$ similar to Theorem 1 can be developed. This upper bound is shown to be tight in the limit of small distortions in the case where X is continuously distributed.

Theorem 2 *Let $R_X(D)$ be the rate-distortion function for the memoryless wall process X. The rate distortion function of the process V satisfies*

$$(1 - 2p_W)R_X(D) \leq R_V(D) \leq H(p_W) + (1 - 2p_W)R_X(D). \qquad (13)$$

Proof: The upper bound follows a constructive proof similar to the discussion prior to Theorem 1. The lower bound is a consequence of Theorem 3.1 in [9] and the fact that $R_{V|W}(D) = (1 - 2p_W)R_X(D)$, where $R_{V|W}(D) = \lim_{t\to\infty} R_{V^t|W^t}(D)$, and $R_{V^t|W^t}(D)$ is the *conditional rate distortion function* ([9], [4], Sec. 4.1). ■

4 Information Rates for The Dynamic Reality

In the static model, the wall process X drawn at $t = 0$ did not change with time. To develop a model for scenes that change over time, we model X as a 2-D random field indexed by $(n, t) \in \mathbb{Z} \times \mathbb{Z}^+$. A simple yet rich model for the field is that of a first order Markov model over time. The random field is defined as follows:

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

Definition 4 *The random field is the field $RF = \{X_n^{(t)} : (n,t) \in \mathbb{Z} \times \mathbb{Z}^+\}$, such that $(X_n^{(0)} : n \in \mathbb{Z})$ is i.i.d. and for each $n \in \mathbb{Z}$, the process $(X_n^{(t)} : t \in \mathbb{Z})$ is a first order time-homogeneous Markov process.*

The *dynamic vector process* V is defined similar to the static case, but now taking snapshots or vectors from the random field:

Definition 5 *Let $RF = \{X_n^{(t)} : (n,t) \in \mathbb{Z} \times \mathbb{Z}^+\}$ be a random field, and let W be a random walk. The dynamic vector process is the process $V = (V_t : t \in \mathbb{Z}^+)$ such that for each $t > 0$,*

$$V_t = (X_{W_t}^{(t)}, X_{W_t+1}^{(t)}, \ldots, X_{W_t+L-1}^{(t)}).$$

The random field and the corresponding vector process are illustrated in Fig. 4.

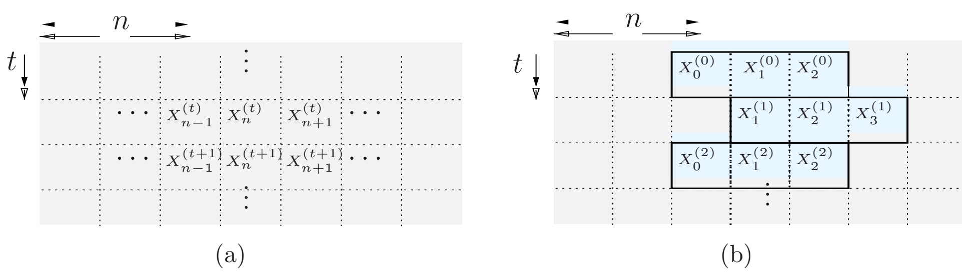


Figure 4: A model for the dynamic reality. (a) It entails a random field that is Markov in the time dimension t, and i.i.d. in the spatial dimension n. (b) Motion then occurs within this random field.

Lossless Information Rates To derive bounds for $H(V)$ in the dynamic reality case, we compute first the conditional entropy $H(V|W) := \lim_{t\to\infty} H(V_t|V^{t-1}, W^t)$. The key is to compute $H(V^t|W^t = w^t, V^{t-1})$ by splitting the set of paths into recurrent and non-recurrent, and further splitting the set of recurrent paths according to (2). Referring to Fig.4(b), let w^t be a given path. Each V_t has $L-1$ entries on the same location as $L-1$ entries from V_{t-1}. The remaining entry corresponds to either a non-recurrent or a recurrent location depending on w^t. If w^t is non-recurrent, then by the Markov property of the field, we have $H(V^t|W^t = w^t, V^{t-1}) = H(X_0^{(t)}) + (L-1)H(X_0^{(t)}|X_0^{(t-1)})$. If a path is recurrent at t, then there is $s < t$ such that $w_s = w_t$ but $w_{s+1} \neq w_t, \ldots, w_{t-1} \neq w_t$. Using the Markov property again it follows that $H(V^t|W^t = w^t, V^{t-1}) = H(X_0^{(t)}|X_0^{(s)}) + (L-1)H(X_0^{(t)}|X_0^{(t-1)})$. Averaging over all possible w^t, and by letting $t \to \infty$ using Lemma 1 (i) leads to:

$$H(V|W) = H(X_0^{(\infty)})(1-2p_W) + (L-1)H(X_0^{(1)}|X_0^{(0)}) + \sum_{i=1}^{\infty} H(X_0^{(i)}|X_0^{(0)}) \Pr\{\mathcal{T}^i\}, \tag{14}$$

where $\Pr\{\mathcal{T}^i\}$ is the probability of return given in Lemma 1 (ii).

With the conditional entropy we can derive lower and upper bounds on the entropy rate $H(V)$. The upper bound is obtained from the inequality $H(V) \leq H(p_W) + H(V|W)$. Because the process X changes at each step, W^t cannot be obtained from V^t with probability 1 even in the continuous alphabet case. As a result, we cannot use the event A_L to obtain a lower bound as in the static case. Because the development (9-11) is valid for the dynamic case, to obtain a lower

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

bound we upper bound $H(W_t|V^t, W^{t-1})$ using Fano's inequality. Denote by P_e the probability of error in estimating W_t based on observing $Y := (V_t, V_{t-1}, W_{t-1})$, i.e., $P_e = \Pr\left\{\hat{W}_t(Y) \neq W_t\right\}$. Since W^{t-1} is known, estimating W_t amounts to estimating the increment $N_t = W_t - W_{t-1}$. From Fano's inequality, we have that

$$H(W_t|V^t, W^{t-1}) \leq H(N_t|Y) \leq H(P_e) + P_e \log(1) = H(P_e). \tag{15}$$

Consequently, a lower bound is obtained [2] by combining (9-10) with (15) above and letting $t \to \infty$. We arrive at the following:

Theorem 3 *Consider the vector process V consisting of L-tuples generated by a Bernoulli random walk with transition probability p_W with $p_W \leq 1/2$, and the random field $RF = \{X_n^{(t)} : (n,t) \in \mathbb{Z} \times \mathbb{Z}^+\}$ that is i.i.d. on the n dimension and first-order Markov on the t dimension. The entropy rate of the process V obeys*

$$H(p_W) - H(P_e) + H(V|W) \leq H(V) \leq H(p_W) + H(V|W), \tag{16}$$

where P_e is the probability of error in estimating W_t based on the observation of (V_t, V_{t-1}, W_{t-1}), and $H(V|W)$ is as in (14).

Example 2 For each n, let $X_n^{(t)} = \rho X_n^{(t-1)} + \epsilon_t$ for $t \in \mathbb{Z}^+$, where $\epsilon_t \sim N(0, 1-\rho^2)$ i.i.d. and independent of X. Such a random field model is used for instance in [10] for bit allocation over multiple frames. Let $\phi(\sigma^2)$ denote the differential entropy of a Gaussian density with variance σ^2. It is then easy to check that $h(X_1^{(\infty)}) = \phi(1)$, and $h(X_1^{(i)}|X_1^{(0)}) = \phi(1-\rho^{2i})$, so that we obtain an upper bound on the entropy rate using Theorem 3. Notice that for L large and ρ close to 1, P_e and $H(P_e)$ are small so that the bound in Theorem 3 is tight.

Lossy Information Rates for the Gaussian AR(1) random field In the AR(1) case of the previous example, it is possible to extend the upper bound to the lossy information rate case. The key is to compute $R_{V|W}(D)$ and use the upper bound [9]:

$$R_V(D) \leq H(p_W) + R_{V|W}(D). \tag{17}$$

The conditional rate-distortion satisfies the *Shannon lower bound* (SLB) [4]:

$$R_{V|W}(D) \geq h(V|W) - L\phi(D). \tag{18}$$

The key observation is that for a given fixed trajectory w^t, the rate-distortion function of V^t is that of a gaussian vector consisting of the samples of the random field covered by W^t. For a Gaussian vector, the SLB is tight when the distortion is less than the minimum eigenvalue of the covariance matrix [4] pp. 111. The next proposition gives a condition under which (18) is tight, and thus when combined with (17) provides an upper bound on the rate-distortion function.

Proposition 1 *Consider the vector process V resulting from the Gaussian AR(1) random field with correlation coefficient $0 < \rho < 1$, and a Bernoulli random walk with probability $p_W \leq 1/2$. The Shannon Lower Bound for the conditional rate-distortion function is tight whenever the distortion satisfies*

$$0 < D < \frac{1-\rho}{1+\rho}. \tag{19}$$

[2] Sharper lower bounds can be obtained by estimating N_t using (V^t, W^{t-1}). However, the estimate using Y is easily computed and already leads to a sharp enough bound.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE

Proof: The SLB for each $t > 0$ is given by

$$R_{V^t|W^t}(D) \geq \frac{h(V^t|W^t)}{t} - L\phi(D). \tag{20}$$

Because $I(V^t;\hat{V}^t|W^t) = \sum_{w^t} \Pr\{w^t\} I(V^t;\hat{V}^t|w^t)$, it suffices to show that for each $t > 0$, and for $0 \leq D \leq \frac{1-\rho}{1+\rho}$, the bound

$$\frac{I(V^t;\hat{V}^t|w^t)}{t} \geq \frac{h(V^t|w^t)}{t} - L\phi(D), \text{ for } \mathbb{E}(d(V^t,\hat{V}^t)|w^t) \leq D$$

is achievable[3]. Given w^t, the above bound is attainable if D is smaller than the minimum eigenvalue of the covariance matrix of the random field samples covered by w^t. Denote this covariance by $\mathrm{C}_{w^t} := \mathrm{Cov}(V^t|w^t)$. Because the random field is independent in the spatial dimension n, the spectrum of the covariance matrix is the disjoint union of the spectra of the covariance matrices corresponding to the random field samples of V^t at similar location n. Each C_{w^t} is a submatrix of the $t \times t$ Toeplitz matrix $\mathrm{T}_t(\rho)$ with entries $[\mathrm{T}_t(\rho)]_{ij} = \rho^{|i-j|}$. Since $\lambda_{min}(\mathrm{T}_t(\rho))$ decreases to $(1-\rho)/(1+\rho)$ as $t \to \infty$ [11], by applying Theorem 4.3.15 in [12] pp. 189 we conclude that

$$\lambda_{min}(\mathrm{C}_{w^t}) \geq \lambda_{min}(\mathrm{T}_t(\rho)) \geq \frac{1-\rho}{1+\rho}. \tag{21}$$

Therefore, the bound (20) is achievable for each t and since the limit of $R_{V^t|W^t}(D)$ exists it follows that the bound is achievable for $t \to \infty$. ■

Example 3 We simulate the AR(1) dynamic reality model. To compress the process V^t, we estimate the trajectory and send it as side information. With the trajectory at hand, we encode the samples with DPCM, encoding the residual with entropy constrained scalar quantization (ECSQ). We build two encoders. In the first one, prediction is done utilizing only the previously encoded vector sample; in the second, all encoded samples up to time t are available to the encoder (and decoder). Fig. 5 illustrates the SNR as a function of rate when the block-length $L = 8$. In Fig. 5 (a) and (b) we have $\rho = 0.99$ and the upper bound is valid for SNR greater than $23dB$. Because the scene changes slowly and is highly recurrent, the infinite memory encoder ($M = \infty$) is about 3.5dB better than when $M = 1$. The same behavior is not observed when the scene is not recurrent (panning case, $p_W = 0.1$, Fig. 5 (b)), and when the background changes too rapidly ($\rho = 0.9$, Fig. 5 (c)).

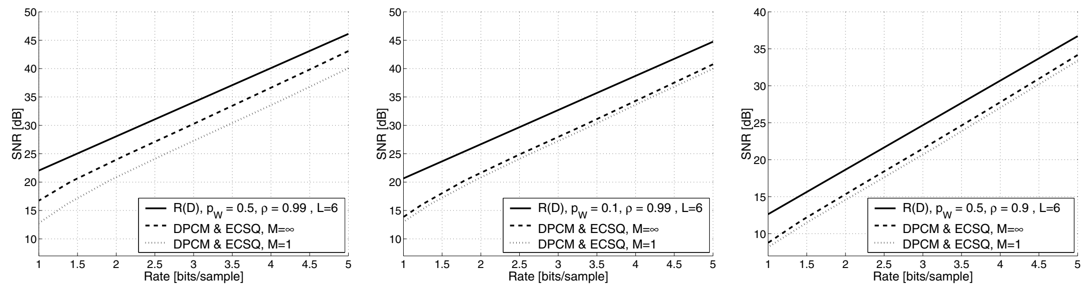
5 Conclusion

We have proposed a stochastic model for video that enables the precise computation of information rates. For the static case, we provided lossless and lossy information rate bounds that are tight in a number of interesting cases. The theoretical results support the ubiquitous hybrid coding paradigm of extracting motion and coding a motion compensated sequence. However, the use of DPCM to code such sequence can be suboptimal.

We extended the model to account for changes in the background scene, and computed bounds for the lossless and lossy information rates for the particular case of

[3]A simple argument can show that distortion allocation for each $V^t|w^t$ is not needed when $D < (1-\rho)/(1+\rho)$. We omit this argument for lack of space.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE



(a) (b) (c)

Figure 5: Performance of DPCM with motion. (a) Memory provide considerable gains, $p_W = 0.5, \rho = 0.99$. (b) Modest gains when $p_W = 0.1$. (c) Modest gains when $\rho = 0.9$, as background changes too rapidly.

AR(1) innovations. The bounds for this "dynamic reality" are tight in some scenarios, namely when the background scene changes slowly with time (i.e. ρ close to 1).

The model explains precisely how long-term motion prediction helps coding in both static and dynamic cases. In the dynamic model, this is related to the two parameters (p_W, ρ) that symbolizes the rate of recurrence in motion and the rate of changes in the scene. As $(p_W, \rho) \to (0.5, 1)$, long term memory predictions results in significant improvements (in excess of 3.5dB). By contrast if either ρ is away from 1, or p_W is away from 0.5, long term memory brings very little improvements.

Although we developed the results for the Bernoulli random walk, the model can be generalized to other random walks on $\mathbb{Z}$ and $\mathbb{Z}^2$. Our current work includes estimating ρ and p_W for real video signals and fitting the model to such signals.

References

[1] A. Telkap, *Digital Video Processing*, Prentice-Hall, Upper Saddle River, NJ, 1995.

[2] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Trans. on CSVT*, vol. 13, pp. 1038–1050, Nov. 2003.

[3] A. L. da Cunha, M. N. Do, and M. Vetterli, "On the information rates of the plenoptic function," in *ICIP*, Atlanta, GA, 2006.

[4] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*, Prentice-Hall, Englewood Cliffs, NJ, 1972.

[5] J. Rudnick and G. Gaspari, *Elements of the Random Walk*, Cambridge University Press, 2004.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, NY, 1991.

[7] H. Li and R. Forchheimer, "Extended signal-theoretic techniques for very low bit-rate video coding," in *Video Coding: The Second Generation Approach*. Kluwer, 1996.

[8] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. CSVT*, vol. 9, no. 1, pp. 70–84, February 1999.

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

[10] Y. Sermadevi, J. Chen, S. Hemami, and T. Berger, "When is bit allocation for predictive video coding easy?," in *Data Compression Conference (DCC)*, Snowbird, UT, 2005.

[11] R. M. Gray, "On the asymptotic eigen-value distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. 18, no. 6, pp. 725–730, 1972.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, 1999.

2007 Data Compression Conference (DCC'07)
0-7695-2791-4/07 $20.00 © 2007 IEEE