The Correlation Structure for a Class of Scene-Based Video Models*

Marwan Krunz
Department of Electrical and Computer Engineering
University of Arizona
Tuscon, AZ 85721
Tel. (520) 621-8731
krunz@ece.arizona.edu

Abstract

We analyze the correlation structure for a general class of scene-based MPEG video models. These models explicitly capture the impact of scene dynamics on the variable bit rate (VBR). The autocorrelation function (ACF) for this class is derived at the GOP (course grain) and frame (fine grain) levels for an arbitrary scene-length distribution. Based on the GOP-level ACF, we establish the relationship between the scene-length statistics and the short-range/long-range dependence (SRD/LRD) of the underlying model. We formally show that when the intra-scene dynamics exhibit SRD (as often the case), the overall model exhibits LRD if and only if the second moment of the scene length is infinite. Our results provide the theoretical foundation for several empirically derived scene-based models. In particular, they prove the presence of LRD in a model with Pareto distributed scene lengths. Our GOP-level analysis is then used to derive the ACF for a generic frame-level MPEG model in which the three types of MPEG frames are interleaved according to a given GOP pattern. In this case, the repetitive application of the GOP pattern induces strong periodic components in the structure of the ACF. It is shown that this frame-level ACF *does not* converge to zero as the frame lag goes to infinity. This, somehow surprising, result can be extended to composite processes in which two drastically different sub-models are interleaved in a deterministic manner (e.g., composition of audio and video streams).

keywords: video modeling, MPEG, traffic correlations, VBR video.

*This work was supported by NSF through Career Award No. ANI-9733143.

1 Introduction

In this paper, we investigate the correlation structure for a general class of scene-based models that are used to characterize variable bit rate (VBR) MPEG-coded video streams. Scene-based models form an important family of video models in which scene dynamics are explicitly incorporated. These models are particularly capable of capturing the multiple-time-scale variations in a VBR video source and, consequently, on providing accurate estimates of the queueing performance at a buffering node [6].

Several scene-based video models have been proposed in the literature. Examples of these are given in [2, 3, 4, 6, 9, 10, 11] (also, see [5] for a survey of video models). In principle, a scene-based model could incorporate both inter- and intra-scene variations. However, intra-scene variations are often ignored to simplify the construction of the model [4, 2]. When both types of variations are present, they are often modeled independently using two individually autocorrelated stochastic processes. These processes, along with the scene-length statistics, distinguishes one scene-based model from another.

Most scene-based models have been developed for compression schemes in which one encoding technique is applied to all frames of a given video sequence. Such schemes result in *homogeneous* VBR sequences in which the fluctuations are primarily attributed to scene dynamics. In contrast, the MPEG algorithm applies different encoding techniques to produce three types of compressed frames (I, P, and B). The three types differ significantly in their bit rate characteristics, with I frames being larger, on average, than P frames, which in turn are larger than B frames. The complete VBR sequence is obtained by interleaving frames of different types according to a Groups-of-Pictures (GOP) pattern, which specifies the sequence of P and B frames between two successive I frames. The GOP pattern is applied repeatedly to a video sequence, resulting in *heterogeneous* frame sizes and significant periodicity in the traffic pattern of this sequence. To avoid the complexity of dealing with such periodicity, most scene-based models focus on characterizing MPEG-coded video at the GOP level (an exception is the model in [9]).

An important aspect of a video model is the form of its autocorrelation function (ACF). Because of their impact on the queueing performance at a buffering node, correlations in a real video sequence must be sufficiently captured. In typical scene-based models, the ACF is not given analytically but is obtained empirically using synthetic data. However, without an analytical expression for the ACF, it is not possible to make any formal statement about the SRD/LRD structure of the model. Such a statement provides first guidelines for efficient resource allocation and network dimensioning under QoS guarantees.

The contributions of this paper are as follows. First, we derive the GOP-level ACF for a general class of scene-based models with an arbitrary scene-length distribution and frame-size statistics. The only restrictions we impose on this class are that inter-scene and intra-scene variations are mutually independent and that scene lengths constitute an *i.i.d.* process. Such assumptions are satisfied by most existing scene-based models (for which the ACF has not been previously reported). From the derived ACF, we establish the relationship between the scene-length distribution of a model and its SRD/LRD structure. Our results indicate that when the intra-scene dynamics exhibit SRD (as often the case), the overall video model is LRD only if the second moment of the scene length is infinite (as in the case of a Pareto scene length with parameter $1 < \alpha < 2$).

Based on our generic GOP-level MPEG model, we introduce a frame-level counterpart that incorporates the three types of MPEG frames. We derive the ACF for this frame-level model. Our results indicate that due to the repetitive application of the GOP pattern, the ACF at the frame level exhibits (as expected) strong periodicity. More interestingly, because of this periodicity the ACF never drops off to zero (i.e., as the frame lag goes to infinity, the ACF converges to known positive and negative limits). The non-zero-convergence result can be extended to other types of media streams that are interleaved in a deterministic manner (e.g., the interleaving of audio and video packets in MPEG-2).

2 Autocorrelation Structure at the GOP Level

In this section, we investigate the ACF for a scene-based video model at the GOP level. While the GOP notion is specific to MPEG video, our "GOP-level" analysis applies, in general, to VBR sequences in which the frame sizes are homogeneous, i.e., produced by the same compression approach. For example, it applies to JPEG and H.261 video sequences, among others. Without loss of generality, we present our ideas in the context of MPEG video.

Consider an MPEG-coded video sequence. Let X_n be a random variable (rv) that models the size of the nth GOP in this sequence, $n = 1, 2, \ldots$. Let S_i be a discrete rv that models the length of the ith scene (measured in the number of GOPs). We assume that scenes are *i.i.d.* with common probability mass function f_s and cumulative distribution function F_s. Let S be a generic rv that describes the length of an arbitrary scene. Intuitively, GOPs that belong to the same scene are relatively close in size. In contrast, GOPs belonging to different scenes may have significantly different sizes. Accordingly, we model X_n by the sum of two random components:

$$X_n \stackrel{a.s.}{=} Y_n + Z_n \tag{1}$$

where Y_n is a rv that accounts for the impact of scene dynamics on the bit rate; $Y_i \stackrel{a.s.}{=} Y_j$ if GOPs i and j belong to the same scene. The rv Z_n represents the difference between the size of the nth GOP and the mean GOP size in the underlying scene. By construction, $E[Z_n] = 0$. We assume that Y_n and Z_n are mutually independent. The random processes $\{Y_n : n = 1, 2, \ldots\}$ and $\{Z_n : n = 1, 2, \ldots\}$ constitute two sequences of correlated and identically distributed rvs. We assume that both processes are second-order stationary, and we denote their corresponding ACFs at lag k by $\rho_Y(k)$ and $\rho_Z(k)$, respectively. Furthermore, we let σ_Y^2 and σ_Z^2 be the variances of Y_1 and Z_1, respectively. The above formulation encompasses many existing scene-based models.

Now consider the random process $\{X_n : n = 1, 2, \ldots\}$. The ACF for this process at lag k is given by

$$\rho_X(k) \triangleq \frac{E\left[(X_n - m)(X_{n+k} - m)\right]}{\sigma_X^2} = \frac{E\left[Y_n Y_{n+k}\right] + \sigma_Z^2 \rho_Z(k) - m^2}{\sigma_Y^2 + \sigma_Z^2} \tag{2}$$

where $m \triangleq E[X_1] = E[Y_1]$. Consider the term $E\left[Y_n Y_{n+k}\right]$ for $k = 1, 2, \ldots$, and an arbitrary n. The relationship between Y_n and Y_{n+k} depends on whether GOPs n and $n+k$ belong to the same scene. Let $\hat{S}$ be the forward recurrence time that is associated with the scene length S. The pmf of $\hat{S}$ is given by:

$$f_{\hat{s}}(i) \triangleq \Pr[\hat{S} = i] = \frac{\Pr[S \geq i]}{E[S]}, \quad i = 1, 2, \ldots \tag{3}$$

Since n is chosen arbitrarily, the two GOPs belong to the same scene if $\hat{S} > k$. Otherwise, they belong to different scenes (and are, thus, independent). Consequently,

$$\begin{aligned}
E\left[Y_n Y_{n+k}\right] &= \sum_{j=1}^{\infty} E\left[Y_n Y_{n+k} / \hat{S} = j\right] \cdot \Pr\left[\hat{S} = j\right] && (4)\\
&= \sum_{j=1}^{k} E\left[Y_n Y_{n+k} / \hat{S} = j\right] f_{\hat{s}}(j) + \sum_{j=k+1}^{\infty} E\left[Y_n Y_{n+k} / \hat{S} = j\right] f_{\hat{s}}(j) && (5)\\
&= m^2 \sum_{j=1}^{k} f_{\hat{s}}(j) + E\left[Y_1^2\right] \sum_{j=k+1}^{\infty} f_{\hat{s}}(j) = m^2 F_{\hat{s}}(k) + E\left[Y_1^2\right] (1 - F_{\hat{s}}(k)) && (6)
\end{aligned}$$

where $F_{\hat{s}}$ is the CDF of $\hat{S}$. Accordingly, (2) can be written as

$$\rho_X(k) = \frac{\sigma_Y^2 \left[1 - F_{\hat{s}}(k)\right] + \sigma_Z^2 \rho_Z(k)}{\sigma_Y^2 + \sigma_Z^2} \tag{7}$$

In the absence of the noise process $\{Z_n : n = 1, 2, \ldots\}$, (7) reduces to $\rho_X(k) = \Pr[\hat{S} > k]$, i.e., the ACF is simply given by the complementary distribution of $\hat{S}$.

Equation (7) can be used to construct a simple test for the LRD/SRD of a video model with a given scene-length distribution. Recall that a process exhibits LRD behavior if its ACF has an infinite sum. Taking the sum of $\rho_X(k)$ from $k = 0$ to ∞, we have

$$\sum_{k=0}^{\infty} \rho_X(k) = \frac{1}{(\sigma_Y^2 + \sigma_Z^2)} \left[\sigma_Y^2 \sum_{k=0}^{\infty} (1 - F_{\hat{s}}(k)) + \sigma_Z^2 \sum_{k=0}^{\infty} \rho_Z(k)\right] = \frac{\sigma_Y^2 E[\hat{S}] + \sigma_Z^2 \sum_{k=0}^{\infty} \rho_Z(k)}{(\sigma_Y^2 + \sigma_Z^2)} \tag{8}$$

It is easy to show that

$$E[\hat{S}] = \frac{1}{2} + \frac{E[S^2]}{2E[S]} \tag{9}$$

From (8) and (9), we arrive at the following result:

Proposition 1 *The video model* $\{X_n : n = 1, 2, \ldots\}$ *is long-range dependent if and only if at least one of the following conditions is satisfied:*

1. *The second moment of the scene length is infinite.*
2. *The noise process* $\{Z_n : n = 1, 2, \ldots\}$ *is LRD.*

Examples

I. Pareto Distribution

Some studies have reported the appropriateness of the Pareto distribution for modeling the scene duration [3, 6]. The complementary form of this distribution is given by

$$\Pr[S > k] = \left(\frac{w}{k}\right)^{\alpha}, \text{ for } k \geq w \tag{10}$$

where w and α are two positive parameters. Assuming that the noise process is SRD, then for $1 < \alpha < 2$, $E[S^2]$ is infinite and the model is LRD. A similar result has been provided for the superposition of ON/OFF sources in which the ON periods of one or more sources are Pareto distributed with $1 < \alpha < 2$ [1].

II. Frater's Scene-Length Distribution

In [2], Frater et al. introduced a model for JPEG video in which the scene duration has the following distribution:

$$f_s(k) \triangleq \Pr[S = k] = \frac{a}{k^n + b^2}, \quad k = 1, 2, \ldots \tag{11}$$

where a, b, and n are three positive constants. The noise process was ignored. To examine the SRD/LRD structure of this model, we consider $E[S^2]$:

$$E[S^2] = \sum_{k=1}^{\infty} k^2 f_s(k) = \sum_{k=1}^{\infty} \frac{ak^2}{k^n + b^2} \tag{12}$$

It is straightforward to show that $E[S^2]$ is finite for $n > 3$, and is infinite otherwise. In other words, Frater's video model is SRD if and only if $n > 3$. Two video sequences were examined in [2]: *Star Wars* and *Film*. Their corresponding n values were determined to be 2 and 3.8, respectively, implying that only the first sequence (*Star Wars*) exhibits LRD!

III. Gamma and Weibull Distributions

Gamma and Weibull distributions have also been proposed for the scene duration [3]. Both have finite second moments. Thus, under an SRD noise process, both distributions result in SRD models.

In typical scene-based modeling studies, the scene-length distribution is obtained by directly fitting the empirical scene lengths. The ACF in this case is only obtained empirically using synthetic sequences. Due to the small number of long scenes (typically, in the order of few tens), accurate fitting of the tail of the empirical scene-length distribution is not feasible, despite the importance of this tail in determining the SRD/LRD structure of the model. We remedy this issue by providing an alternative modeling approach for the scene-length distribution. Ignoring the noise process, we have $\rho_X(k) = \Pr[\hat{S} > k]$, from which we have

$$\begin{aligned}
\Pr[\hat{S} = j] &= \rho_X(j-1) - \rho_X(j) = \frac{\Pr[S \geq j]}{E[S]} && (13) \\
\Pr[S = j] &= \Pr[S \geq j] - \Pr[S \geq j+1] = E[S]\left(\Pr[\hat{S} = j] - \Pr[\hat{S} = j+1]\right) && (14) \\
&= E[S]\left[\rho_X(j-1) - 2\rho_X(j) + \rho_X(j+1)\right] && (15)
\end{aligned}$$

Evaluating (13) at $j = 1$, we obtain

$$E[S] = \frac{\Pr[S \geq 1]}{\rho_X(0) - \rho_X(1)} = \frac{1}{1 - \rho_X(1)} \tag{16}$$

From (16) and (15), we obtain the pmf for the scene length in terms of the ACF:

$$f_s(i) \triangleq \Pr[S = j] = \frac{\rho_X(j-1) - 2\rho_X(j) + \rho_X(j+1)}{1 - \rho_X(1)} \tag{17}$$

Accurate modeling of the scene length can proceed by first obtaining an adequate fit for the empirical ACF (which can be done relatively with high accuracy), and then using this fit to derive the corresponding scene-length distribution. We provide one example: In [8] an ACF of the form $\rho_X(k) = e^{-\beta\sqrt{k}}$ was found quite appropriate for modeling JPEG-coded video. Ignoring the noise process (i.e., intra-scene variations), the scene-length distribution in this case is simply given by:

$$f_s(k) = \frac{e^{-\beta\sqrt{k-1}} - 2e^{-\beta\sqrt{k}} + e^{-\beta\sqrt{k+1}}}{1 - e^{-\beta}}, \quad k = 1, 2, \ldots \tag{18}$$

3 Autocorrelation Structure at the Frame Level

GOP-level modeling is sufficient for evaluating the queueing performance at a buffering node (e.g., an ATM multiplexer) with a relatively large buffer size. In this case, synchronization between different MPEG sources has little impact on the performance. However, when the buffer size is small (e.g., it drains in less than a GOP time), then synchronization between MPEG sources will have a significant impact on the performance. Consider, for example, the multiplexing of two MPEG streams. If both streams are exactly aligned with respect to their GOPs (i.e., their I frames overlap), then the buffer overflow rate will be much higher than when one MPEG source lags the other by one frame period. Clearly, a frame-level model is needed to study the packet loss performance in such cases. In this section, we extend our generic GOP-level scene-based model to characterize the frame-level variations.

Consider the stationary process $\{X_n : n = 1, 2, \ldots\}$ that represents the GOP sequence of an MPEG stream. Let f_X be its marginal distribution. A GOP pattern is characterized by two parameters: the I-to-I frame distance (N) and the I-to-P frame distance (M). Not all MPEG sequences involve a repetitive GOP pattern. In fact, some encoders allow a new GOP to start before the completion of the previous one, typically in response to a large frame (i.e., the start of a high-action scene). However, for tractability purposes, we restrict our work to MPEG sequences that conform to repetitive GOPs, and many sequences do so in practice.

Denote the size of the kth frame in the MPEG sequence by U_k. Suppose that this frame belongs to the rth GOP. If the MPEG sequence starts with a complete GOP (i.e., the first frame is I), then $r = \lceil k/N \rceil$. However, this makes the process $\{U_n : n = 1, 2, \ldots\}$ nonstationary, precluding any analysis of the correlation structure. Instead, we will allow the first GOP of the video sequence to be incomplete by randomly selecting the first frame in the sequence from any location in the GOP pattern, and continuing thereafter according to the GOP pattern. This will have no effect on the long-term behavior of the model, but will ensure its stationarity. Accordingly, $\lceil k/N \rceil \leq r \leq \lceil k/N \rceil + 1$. We define U_k as follows:

$$U_k \triangleq \begin{cases} c_I X_r, & \text{if the } k\text{th frame is an } I \text{ frame} \\ c_P X_r, & \text{if the } k\text{th frame is a } P \text{ frame} \\ c_B X_r, & \text{if the } k\text{th frame is a } B \text{ frame} \end{cases} \tag{19}$$

for some constants c_I, c_P, and c_B. According to this model, B frames (also, P frames) that belong to the same GOP have the same size. An example of the resulting sample path based on this model is shown in Figure 1.

The constants c_I, c_P, and c_B are obtained as follows: $c_I = \frac{I_{avg}}{X_{avg}}$, $c_P = \frac{P_{avg}}{X_{avg}}$, and $c_B = \frac{B_{avg}}{X_{avg}}$, where I_{avg}, P_{avg}, and B_{avg} are the average (empirical) frame sizes for I, P, and B frames, respectively; and X_{avg} is the average size of a GOP. Note that $c_I + (N/M - 1)c_P + (N - N/M)c_B = 1$, ensuring that the sum of frame sizes in the rth GOP is equal to X_r. The marginal distributions for the sizes of the three frame types are given in terms of f_X as follows: $f_I(x) = f_X(x/c_I)$, $f_P(x) = f_X(x/c_P)$, and $f_B(x) = f_X(x/c_B)$. Let U_I, U_P, and U_B be three generic rvs that indicate the sizes of arbitrary I, P, and B frames, respectively. It readily follows that $E[U_I] = c_I m$, $E[U_P] = c_P m$, and $E[U_B] = c_B m$. Also, $\mathrm{var}(U_I) \triangleq \sigma_I^2 = c_I^2 \sigma_X^2$, $\mathrm{var}(U_P) \triangleq \sigma_P^2 = c_P^2 \sigma_X^2$, and $\mathrm{var}(U_B) \triangleq \sigma_B^2 = c_B^2 \sigma_X^2$. Let $\rho_U(k)$ be the ACF of $\{U_n : n = 1, 2, \ldots\}$ at lag k

$$\rho_U(k) \triangleq \frac{E\left[(U_1 - \tilde{m})(U_{1+k} - \tilde{m})\right]}{\sigma_U^2} = \frac{E\left[U_1 U_{1+k}\right] - \tilde{m}^2}{\sigma_U^2} \tag{20}$$

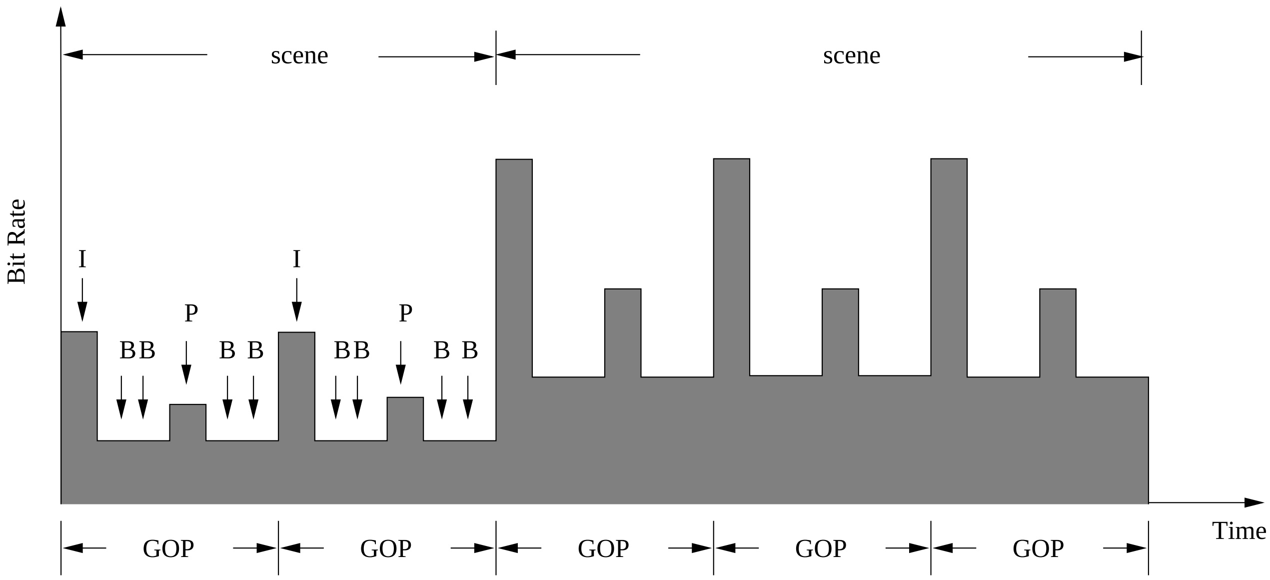

Figure 1: Bit-rate variations in frame-level model ($N = 6,\ M = 3$).

where $\tilde{m} \stackrel{\triangle}{=} E[U_1] = \frac{m}{N}$ and

$$\sigma_U^2 \stackrel{\triangle}{=} \mathrm{var}(U_1) = \frac{(\sigma_X^2 + m^2)}{N}\left[c_I^2 + (\frac{N}{M} - 1)c_P^2 + (N - \frac{N}{M})c_B^2\right] - \frac{m^2}{N^2} \tag{21}$$

Recall that according to our model, the first frame of an MPEG stream is selected randomly from the N frames of a GOP. Thereafter, the MPEG sequence proceeds according to the repetitive GOP pattern. This means that the first GOP may be incomplete. Consider $E[U_1 U_{1+k}]$ for $k > 1$:

$$E[U_1 U_{1+k}] = \sum_{i=1}^{N} E[U_1 U_{1+k}/T_1 = i]\,\Pr[T_1 = i] \tag{22}$$

where T_i is a discrete rv that reflects the location (and consequently, the type) of the ith frame in the GOP pattern. The sample space of T_i is $\Omega_T = \{1, 2, \ldots, N\}$. Thus, $T_j \stackrel{a.s.}{=} i$ means that the type of the jth frame is the same as the type of the frame in the ith location of the GOP pattern. Because of the repetitive application of the GOP pattern, the process $\{T_n : n = 1, 2, \ldots\}$ constitutes a deterministic Markov chain with transition probabilities

$$p_{ij} \stackrel{\triangle}{=} \Pr\left[T_n = j/T_{n-1} = i\right] = \begin{cases} 1 & \text{if } j = i + 1 \text{ and } i = 1, \ldots, N-1, \text{ or } i = N \text{ and } \ j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Our previous assumption related to the type of the first frame can now be stated formally by taking the initial distribution of the Markov chain to be its stationary distribution, i.e., $\pi_i \stackrel{\triangle}{=}$

$\Pr[T_1 = i] = 1/N$ for all $i \in \Omega_T$. Hence,

$$E[U_1 U_{1+k}] = \frac{1}{N} \sum_{i=1}^{N} E[U_1 U_{1+k}/T_1 = i] \tag{24}$$

Before proceeding with the computation of $E[U_1 U_{1+k}/T_1 = i]$, we need to define some related quantities. Let

$$g_N(i,k) \triangleq (i+k-1) \bmod N \tag{25}$$
$$g_M(i,k) \triangleq (i+k-1) \bmod M \tag{26}$$

where $i \in \Omega_T$ and k is a positive integer. Note that because N is a multiple of M, if $g_N(i,k) = 0$ then $g_M(i,k) = 0$ as well. Define the following two sets:

$$\Omega_P \triangleq \{1+M, 1+2M, 1+3M, \ldots, 1+(N/M-1)M\} \tag{27}$$
$$\Omega_B \triangleq \{2, 3, \ldots, M, M+2, M+3, \ldots, 2M, \ldots, (N/M-1)M, (N/M-1)M+2, \ldots, N\} \tag{28}$$

Note that $\Omega_T = \{1\} \bigcup \Omega_P \bigcup \Omega_B$ and the three sets $\{1\}$, Ω_P, and Ω_B are mutually exclusive. Next, we define the following function $\eta(i,k)$:

- **Case 1:** $i = 1$
  - If $g_N(1,k) = 0$, then $\eta(1,k) \triangleq c_I^2$.
  - If $g_N(1,k) \neq 0$ but $g_M(1,k) = 0$, then $\eta(1,k) \triangleq c_I c_P$.
  - If $g_M(1,k) \neq 0$, then $\eta(1,k) \triangleq c_I c_B$.
- **Case 2:** $i \in \Omega_P$
  - If $g_N(i,k) = 0$, then $\eta(i,k) \triangleq c_I c_P$.
  - If $g_N(i,k) \neq 0$ but $g_M(i,k) = 0$, then $\eta(i,k) \triangleq c_P^2$.
  - If $g_M(i,k) \neq 0$, then $\eta(i,k) \triangleq c_P c_B$.
- **Case 3:** $i \in \Omega_B$
  - If $g_N(i,k) = 0$, then $\eta(i,k) \triangleq c_I c_B$.
  - If $g_N(i,k) \neq 0$ but $g_M(i,k) = 0$, then $\eta(i,k) \triangleq c_P c_B$.
  - If $g_M(i,k) \neq 0$, then $\eta(i,k) \triangleq c_B^2$.

We now return to the problem of determining $E[U_1 U_{1+k}/T_1 = i]$. There are two cases to consider. First, when $i + k \leq N$ Frames 1 and $1+k$ must belong to the same (first) GOP, and

$$E[U_1 U_{1+k}/T_1 = i] = \eta(i,k) E[X_1^2] = \eta(i,k) \left[\sigma_X^2 + m^2\right] \tag{29}$$

When $i + k > N$, Frames 1 and $1 + k$ belong to different GOPs and possibly to different scenes. More specifically, the $(1 + k)$th frame belongs to the rth GOP, where $r \triangleq \lceil (i + k)/N \rceil > 1$. Thus,

$$E[U_1 U_{1+k}/T_1 = i] = \eta(i,k) E[X_1 X_{r-1}] = \eta(i,k) \left[\sigma_X^2 \rho_X(r-1) + m^2 \right] \tag{30}$$

Equations (29) and (30), combined with the expressions for $\eta(i,k)$, ρ_X, σ_X^2, and m can be used to compute $E[U_1 U_{1+k}]$ for various values of k, as shown next.

3.1 Computation of $E[U_1 U_{1+k}]$

Case I: $k = 1, 2, \ldots, N-1$

Based on the previous discussion, we have

$$\begin{aligned} E[U_1 U_{1+k}] &= \frac{1}{N} \left[\sum_{i=1}^{N-k} E[U_1 U_{1+k}/T_1 = i] + \sum_{i=N-k+1}^{N} E[U_1 U_{1+k}/T_1 = i] \right] & (31) \\ &= \frac{1}{N} \left[\sum_{i=1}^{N-k} \eta(i,k)(\sigma_X^2 + m^2) + \sum_{i=N-k+1}^{N} (\sigma_X^2 \rho_X(r-1) + m^2) \right] & (32) \end{aligned}$$

where, as before, $r \triangleq \lceil (i+k)/n \rceil$. For $k = 1, \ldots, N$ and $i = N-k+1, \ldots, N$, we have $r = 2$. Thus,

$$\begin{aligned} E[U_1 U_{1+k}] &= \frac{1}{N} \left[(\sigma_X^2 + m^2) \sum_{i=1}^{N-k} \eta(i,k) + \sum_{i=N-k+1}^{N} \eta(i,k)(\sigma_X^2 \rho_X(1) + m^2) \right] & (33) \\ &= \frac{1}{N} \left[m^2 \sum_{i=1}^{N} \eta(i,k) + \sigma_X^2 \left(\sum_{i=1}^{N-k} \eta(i,k) + \rho_X(1) \sum_{i=N-k+1}^{N} \eta(i,k) \right) \right] & (34) \end{aligned}$$

It can be shown that

$$\sum_{i=1}^{N} \eta(i,k) = \begin{cases} \eta_1^* \triangleq c_I^2 + c_P^2(N/M - 1) + c_B^2(N - N/M), & \text{if } k \bmod N = 0 \\ \eta_2^* \triangleq 2c_I c_P + (N - N/M)c_B^2 + (N/M - 2)c_P^2, & \text{if } k \bmod N \neq 0 \text{ but } k \bmod M = 0 \\ \eta_3^* \triangleq 2c_I c_B + 2(N/M - 1)c_P c_B + (N - 2N/M)c_B^2, & \text{if } k \bmod M \neq 0 \end{cases} \tag{35}$$

Case II: $k \geq N$

Starting with (30), this case is further divided into two subcases.

Case II-A: $k = pN$ **for** $p = 1, 2, \ldots$.

In this case, Frames 1 and $1 + k$ differ exactly by p GOPs, irrespective of the value of i. Thus, $r = \lceil i/N \rceil + p = 1 + p$. Accordingly,

$$E[U_1 U_{1+k}] = \frac{1}{N} \sum_{i=1}^{N} \eta(i,k) \left[\sigma_X^2 \rho_X(p) + m^2 \right] \tag{36}$$

$$= \frac{\sigma_X^2 \rho_X(p) + m^2}{N} \sum_{i=1}^{N} \eta(i,k) \tag{37}$$

Finally,

$$\rho_U(k) = \frac{\left[(\sigma_X^2 \rho_X(p) + m^2) \left(\sum_{i=1}^{N} \eta(i,k) \right) / N \right] - \widetilde{m}^2}{\sigma_U^2} \tag{38}$$

Case II-B: $k \neq pN$ **and** $k > N$

In this case, Frames 1 and $1 + k$ may differ by either p GOPs or by $p + 1$ GOPs, where $p \triangleq \lceil k/N \rceil - 1 = \lfloor k/N \rfloor$. More specifically, for $T_1 = 1, 2, \ldots, i^*$, where $i^* = N \lceil k/N \rceil - k$, the 1st and $(1 + k)$th frames differ by p GOPs. For $T_1 = i^* + 1, \ldots, N$, the two frames differ by $p + 1$ GOPs. Thus,

$$E[U_1 U_{1+k}] = \frac{1}{N} \sum_{i=1}^{i^*} \eta(i,k) \left[\sigma_X^2 \rho_X(p) + m^2 \right] + \frac{1}{N} \sum_{i=i^*+1}^{N} \eta(i,k) \left[\sigma_X^2 \rho_X(1+p) + m^2 \right] \tag{39}$$

which concludes the derivation of the ACF for the process $\{U_n : n = 1, 2, \ldots\}$.

3.2 Asymptotic Behavior of the Frame-Level ACF

From the analytical form of ρ_U, one can examine its asymptotic behavior, shedding light on the LRD/SRD structure. As $k \to \infty$, $p \to \infty$ and $\rho_X(p) \to 0$, so that

$$\lim_{k \to \infty} \rho_U(k) = \frac{\frac{m^2}{N} \lim_{k \to \infty} \sum_{i=1}^{N} \eta(i,k) - \widetilde{m}^2}{\sigma_U^2} \tag{40}$$

The limit of $\sum_i \eta(i,k)$ as $k \to \infty$ alternates between the three values given in (35). Substituting the values of $\widetilde{m}$ and σ_U^2 in (40), it is easy to see that $\lim_{k \to \infty} \rho_U(k)$ alternates between the following three values:

$$\frac{N\eta_j^* - 1}{N\eta_1^*(\sigma_X^2/m^2 + 1) - 1}, \quad j = 1, 2, 3 \tag{41}$$

Note that in general the above three values are non-zero, which justifies the persistent, periodic autocorrelations that are observed in empirical MPEG sequences. Note also that the above limit is valid for any scene-length distribution (including exponentially distributed scene lengths).

4 Numerical Examples

In this section, we demonstrate the validity of our analytical expressions using three numerical examples. We consider the ACF only in the presence of inter-scene variations. Additional examples related to the ACF in the presence of intra-scene variations can be found in [7].

Our validation approach is based on comparing the analytical ACF against the sample ACF of synthetically generated VBR sequences. In the first two examples, we investigate the ACF at the GOP level assuming a particular scene-length distribution and gamma distributed GOP sizes with scale and shape parameters of 0.05 and 25, respectively. These parameters correspond to GOP sizes having mean and standard variation of 500 and 100, respectively. Based on the scene length and GOP size distributions, we generate synthetic VBR sequences and compute their empirical ACFs.

In the first example, we use a shifted exponential scene-length distribution:

$$\Pr[S > k] = \Pr[\hat{S} > k] = e^{-\beta(k-1)}, \ \ k = 1, 2, \ldots \tag{42}$$

Note that in this case S and $\hat{S}$ have the same distribution. We set $\beta = 1/49$, so that $E[S] = 50$. Ten synthetic traces were generated, and their sample ACFs were computed and averaged. The average ACF for the synthetic traces is plotted in Figure 2 along with its theoretical counterpart. There is a clear match between the two plots.

Next, we consider a sub-geometric scene-length distribution of the form:

$$\Pr[S > k] = \alpha^{\sqrt{k}}, \ \ k = 1, 2, \ldots \tag{43}$$

for some $0 < \alpha < 1$. In this case, $\Pr[S > k] \neq \Pr[\hat{S} > k]$. To obtain the theoretical ACF, we proceed as follows:

$$\begin{aligned} \rho_X(k+1) &= 1 - \Pr[\hat{S} \leq k+1] = 1 - \Big(\Pr[\hat{S} \leq k] + \Pr[\hat{S} = k+1]\Big) && (44) \\ &= \rho_X(k) - \Pr[\hat{S} = k+1] = \rho_X(k) - \frac{\Pr[S \geq k+1]}{E[S]} && (45) \\ &= \rho_X(k) - \frac{\alpha^{\sqrt{k}}}{E[S]} && (46) \end{aligned}$$

If $E[S]$ is known, then the last expression can be used to compute $\rho_X(k)$ recursively. While an exact value for $E[S] = \sum_{k=0}^{\infty} \alpha^{\sqrt{k}}$ is not known, one can accurately approximate it: It is easy to see that

$$\int_0^\infty e^{\ln \alpha \sqrt{x}} dx \leq \sum_{k=0}^{\infty} e^{\ln \alpha \sqrt{k}} = \sum_{k=0}^{\infty} \alpha^{\sqrt{k}} \leq \int_0^\infty e^{\ln \alpha \sqrt{x}} dx + 1 \tag{47}$$

Since $\int_0^\infty e^{\ln\alpha\sqrt{x}}dx = \frac{2}{(\ln\alpha)^2}$, $E[S]$ is bounded by

$$\frac{2}{(\ln\alpha)^2} \leq E[S] \leq \frac{2}{(\ln\alpha)^2} + 1 \tag{48}$$

Setting $\alpha = 0.8$, we have $40.17 \leq E[S] \leq 41.17$. Thus, we take $E[S] \approx 40.67$. Figure 3 depicts the theoretical and empirical ACFs under a sub-geometric scene-length distribution. At small and large lags, the plots match very well. At intermediate lags, there is a slight difference that is attributed to the large variance of the empirical autocorrelations and to other approximations in the generation of sub-geometrically distributed random numbers.

Our last example is related to the frame-level ACF. Here, we use the same shifted exponential scene-length distribution as in the first example. We set $N = 12$, $M = 3$, $c_I = 5/22$, $c_P = 3/22$, and $c_B = 1/22$. The analytical and empirical ACFs are shown in Figure 4 for lags in the range 450 to 500. This range is chosen arbitrarily, and is representative of the behavior at large lags. The two ACFs almost match at all examined lags (similar trend is also observed at small lags). This match validates the correctness of our analysis. Note that although the scene-length distribution is exponential, the deterministic interleaving of three, drastically different processes (one for each frame type) induces strong correlations that determine the asymptotic shape of the ACF. These correlations do not die out to zero as the lag goes to infinity, but instead they converge to $\eta_1^* = 0.8912$, $\eta_2^* = 0.710$, and $\eta_3^* = -0.3776$. The impact of the scene-length distribution in this case is limited to the speed at which the ACF converges to these asymptotic quantities.

References

[1] O. J. Boxma and V. Dumas. Fluid queues with long-tailed activity distributions. technical report Report BS-R9705, Centrum voor Wiskunde en Informatica (CWI), 1997.

[2] M. R. Frater, J. F. Arnold, and P. Tan. A new statistical model for traffic generated by VBR coders for television on the Broadband ISDN. *IEEE Trans. on Circuits and Systems for Video Technology*, 4(6):521–526, Dec. 1994.

[3] D. P. Heyman and T. V. Lakshman. Source models for VBR broadcast-video traffic. *IEEE/ACM Transactions on Networking*, 4(1):40–48, Feb. 1996.

[4] D. P. Heyman, A. Tabatabai, and T. V. Lakshman. Statistical analysis and simulation study of video teleconferencing traffic in ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 2(1):49–59, Mar. 1992.

[5] M. Izquierdo and D. Reeves. A survey of statistical source models for variable bit-rate compressed video. Technical report TR 97-10, Center for Advanced Comput-

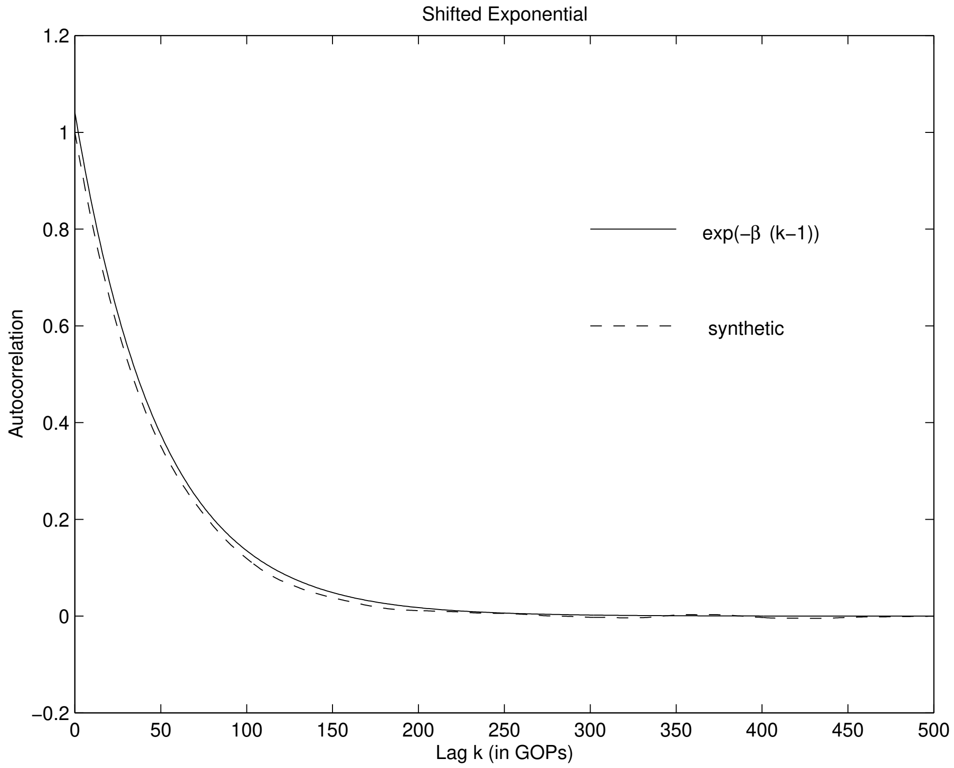

Figure 2: ACF for the GOP sequence with shifted exponential scene-length distribution ($\beta = 1/49$).

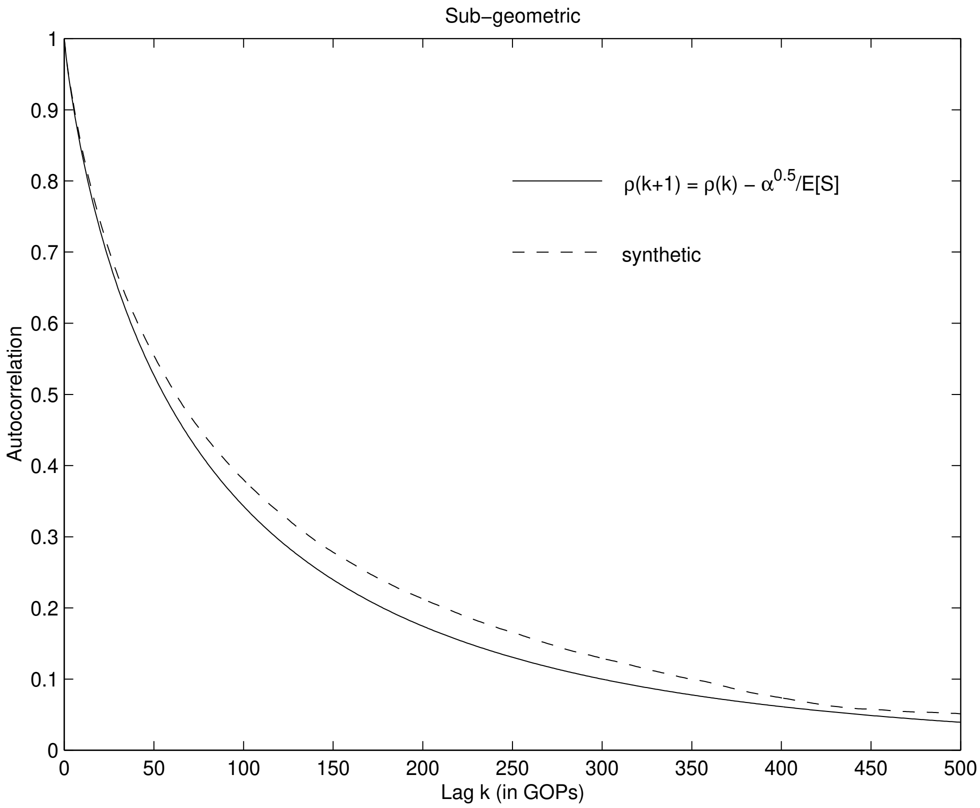

Figure 3: ACF for the GOP sequence with a sub-geometric scene-length distribution ($\alpha = 0.8$).

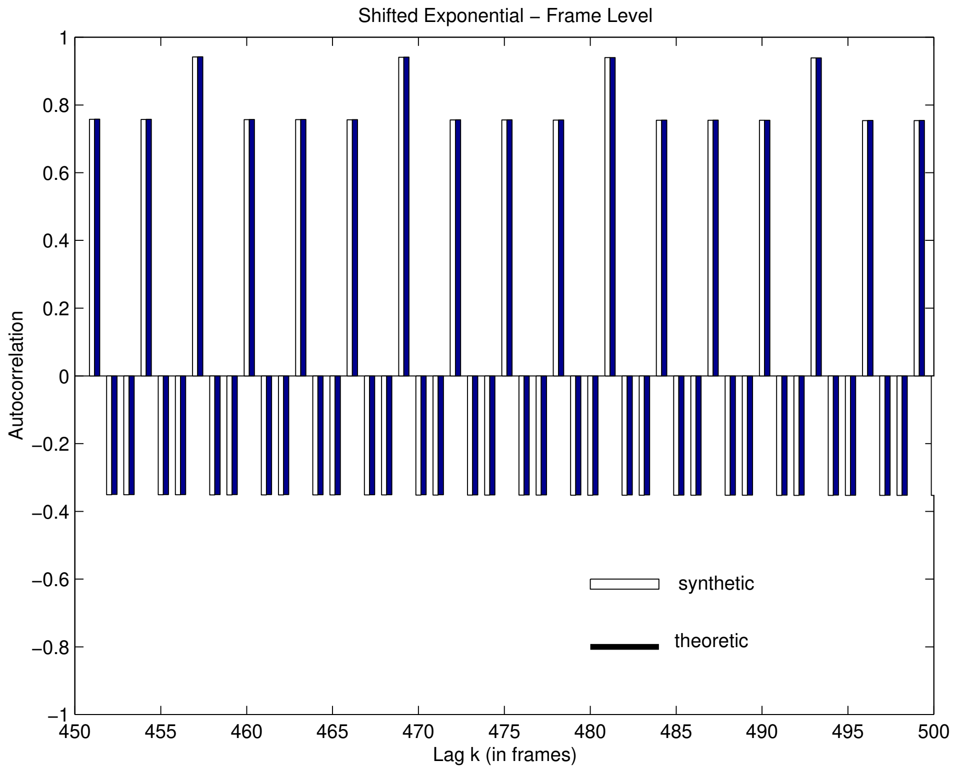

Figure 4: ACF at the frame level based on a shifted exponential scene-length distribution ($\beta = 1/49$, $N = 12,\ M = 3$).

ing and Communication, North Carolina State University, June 1997. (Available at ftp://ftp.csc.ncsu.edu/pub/rtcomm/video.html).

[6] P. R. Jelenkovic, A. A. Lazar, and N. Semret. The effect of multiple time scales and subexponentiality in MPEG video streams on queueing behavior. *IEEE Journal on Selected Areas in Communications*, 15(6):1052–1071, Aug. 1997.

[7] M. Krunz. The autocorrelation structure for a class of scene-based video models and its impact on the queueing performance. Technical report CENG-99-01, University of Arizona, Department of ECE, Jan. 1999.

[8] M. Krunz and A. Makowski. Modeling video traffic using $M/G/\infty$ input processes: A compromise between Markovian and LRD models. *IEEE Journal on Selected Areas in Communications*, 16(5):733–748, June 1998.

[9] M. Krunz and S. K. Tripathi. On the characterization of of VBR MPEG streams. In *Performance Evaluation Review (Proceedings of the ACM SIGMETRICS '97 Conference)*, volume 25, pages 192–202, June 1997.

[10] A. A. Lazar, G. Pacifici, and D. E. Pendarakis. Modeling video sources for real-time scheduling. *Multimedia Systems Journal*, 1(6):253–266, 1994.

[11] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski. TES-based video source modeling for performance evaluation of integrated networks. *IEEE Transaction on Communications*, 42(10):2773–2777, Oct. 1994.