Modeling Bit Rate Variations in MPEG Sources*

Marwan Krunz† and Satish K. Tripathi‡

† University of Maryland Institute for Advanced Computer Studies

‡ Department of Computer Science

University of Maryland at College Park

College Park, MD 20742

{krunz,tripathi}@cs.umd.edu

Abstract

In this paper, we propose a traffic model for the characterization of VBR MPEG-coded video streams. This model provides the means to generate synthetic MPEG streams that can be used in performance studies of ATM networks. The model is appropriately fitted to three long empirical video sequences taken from different movies. We use multiple components to model bit rate variations in an MPEG stream. These components have different time scales. Long-term variations in the bit rate are captured at the scene level. Within a scene, the sizes of I frames tend to slightly fluctuate around some average. Hence, we measure the activity of the scene by the average size of I frames in that scene. This average varies from one scene to another, and its randomness is reasonably approximated by a lognormal distribution. For a given scene, the fluctuations of the sizes of I frames about their mean are modeled as an AR(2) time series. Finally, we show that the sizes of P and B frames can be appropriately fitted by lognormal distributions, with corresponding parameters. Using the compression pattern, the complete frame size sequence is formed by intermixing three subsequences, each of which describes the frame size sequence for a particular frame type. The validity of the model is demonstrated by the similarity between a synthetic stream and an actual trace, in terms of the correlation structure, the marginal distribution, the sample paths, and more importantly, the queueing performance.

keywords: compressed video, MPEG, variable bit rate, traffic models.

*This research was partially supported by the NSF grant # 9318933.

1 Introduction

A broadband network, such as B-ISDN/ATM, is being designed to transport diverse streams that have different traffic patterns and performance requirements. The content of these streams consists of combinations of text, voice, image, and video. A large proportion of the network capacity will be used to transport video traffic. This is partially due to the increasing number of network applications that involve digital video, and also the huge bandwidth requirement of full-motion digital video. For economical reasons, digital video is being compressed before entering the network. By means of compression, the number of simultaneously transported video streams over a single link can be increased by more than an order of magnitude.

We focus on the modeling and characterization of video streams that are compressed based on the Moving Picture Expert Group (MPEG) algorithm. MPEG is a compression standard that has recently gained a considerable attention. The first draft of the MPEG standard [6] defines the compressed bit stream for video and audio for transmission at an average rate of 1.5 Mbits/s. A more recent draft, known as MPEG-2 [7], is optimized for an average bit rate of 5 Mbits/s. MPEG-2 was accepted as the basis for the HDTV standard in the United States [1]. The standardization and widespread acceptance of MPEG are the main reasons for choosing it among other compression techniques to be the focus of our research. Additionally, the MPEG algorithm involves several modes of compression, some of which are the basis of other compression algorithms. For example, JPEG (the Joint Photographic Experts Group standard), which is a still-image compression algorithm, is already being incorporated in MPEG. Thus, a traffic model for an MPEG stream is applicable, as well, to video streams that are compressed using the JPEG algorithm.

Several studies investigated the characterization of compressed video traffic (see [3] for a survey). Only few of these studies dealt with characterization of MPEG traffic [13, 5, 8]. We believe that video models based on compression algorithms other than MPEG do not accurately characterize the variability of an MPEG stream. This is mainly due to the periodic generation of three types of frames in an MPEG stream. Pancha and El Zarki [13] described the statistical characteristics of a few minutes of full-motion video generated by an MPEG encoder (with no bidirectional frames). No traffic models were reported there. In [5] the authors suggest modeling MPEG video using self-similar models. Although these models partially capture the frame size distribution and the correlation structure of the empirical sequences, the generated sample are significantly different from the empirical ones. A deterministic approach to modeling an MPEG stream was suggested in [9] based on a bounding traffic envelop. Despite the simplicity of that model, it significantly overestimates the traffic behavior.

The MPEG compression algorithm relies on two basic techniques: block-based motion compensation for the reduction of the temporal redundancy and DCT-based compression for the reduction of the spatial redundancy [12]. Motion compensation techniques are applied with both causal (i.e., predictive) and noncausal (i.e., interpolative) coding. Three types of compressed frames are generated by an MPEG encoder: Intra-coded (I), Predictive (P), and Bidirectional (B) frames. An I frame is encoded independently of other frames based on DCT coding. A P frame is encoded with reference to a past I or P frame, and is used as a reference point for the next P frame. A B frame is an interpolated frame that requires both a past and a future reference frames (I or P), and it results in the highest amount of compression. B frames are never used as reference frames. From the point of view of the encoder, a stream of frames could look like the pattern shown in Figure 1-a. Since B frames require noncausal encoding, the transmission sequence has a different pattern than the encoding sequence, as shown in part (b) of the figure.

At the encoder side (part (a) of the figure), the sequence of frames are generated according to a predetermined pattern which we refer to as the *compression pattern* (or group of pictures (GOP)). This pattern is repeated periodically until the whole stream is compressed. The pattern defines the numbers and types of P and B frames between successive I frames. In Figure 1, the pattern is $IBBPBBPBB$ which continues to be repeated until the end of the video stream. The transmitted sequence (part (b) of Figure 1) is a permutation of the ordering of the original sequence. If we only consider the types of the frames (and not their encoding order), except for the first few frames, the encoded and transmitted sequences look alike.

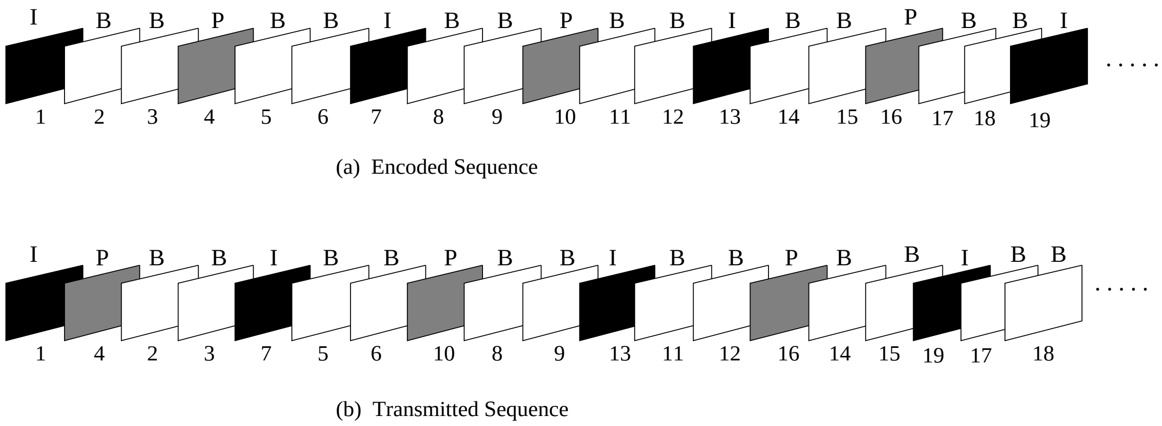

Figure 1: Encoding and transmission orders of an MPEG sequence.

In most cases, the compression pattern of an MPEG stream is "regular" in the following sense: the number of consecutive B frames in a given stream is fixed throughout that stream (e.g., this number is 2 for the pattern in Figure 1). This feature simplifies the parameterization of the compression pattern. A regular compression pattern can be characterized by two parameters:

L : the length of the compression pattern (the frame distance between two successive I frames

Q : the frame distance between an I frame and the following I or P frames (whichever comes first)

Note that the above parameterization is applicable even to a stream that consists of I frames only (a JPEG-like stream), in which case, $L = Q = 1$.

Our objective in this paper is to develop a *statistical* source model for an MPEG-coded video stream. Such a model can be used to generate synthetic MPEG streams that are needed in many performance studies. Examples include dimensioning the buffers at video servers and computing the effective bandwidth for multiplexed video streams. The ability to generate synthetic streams has two advantages. First, performance studies can be carried without requiring the actual traces (which may not be available in many cases). Second, a stochastic model encompasses many realizations (sample paths) which represent "structurally" similar, but not identical synthetic streams. Hence, these streams are ideal for statistical multiplexing studies (an single trace cannot be properly used to generate multiple streams that have the same stochastic structure, but are not identical).

We restrict our attention to video movies because of their rich VBR dynamics. The empirical data in our study are described in Section 2. In Section 3 we present our source model. Standard statistical techniques are used to derive and validate the various components of the model. We conclude the paper in Section 4.

2 Empirical Statistics of MPEG Video Sequences

The model presented in this paper was developed and tested based on three empirical video sequences. Each sequence consists of the sizes of MPEG-compressed frames taken from a video movie. The three traces (i.e., sequences) were taken from three different movies: *The Wizard of Oz*, *The Silence of the Lambs*, and *James Bond: Goldfinger*. For short, we refer to these movies as Oz, Lambs, and Bond. Oz trace was captured from a laser disk player at a rate of 30 frames/sec (details were provided in a previous paper). The other two traces were provided by Oliver Rose [14] [1]. Rose's movies were taken from VCR tapes, and were digitized at a rate of 25 frames/sec using a *SunVideo* card (Bond movie was recorded from cable TV, while Lambs movie was recorded from a commercial cassette [14]). The three movies were compressed using MPEG-I Berkeley's software encoder [4]. Such an encoder is ideal for our purposes since it does not limit the output peak rate (as is the case with hardware encoders). Consequently, the results presented in this paper are

[1] Lambs and Bond traces can be obtained from the ftp site ftp-info3.informatik.uni-wuerzburg.de in the directory /pub/MPEG/. Oz trace is available from the authors upon request.

related to constant-quality VBR MPEG video. Some of the parameters used in the compression process are shown in Table 1.

Parameter	Video Sequence	
	Oz	Lambs and Bond
Image Dimensions (in pixels)	640x480	384x288
Compression Pattern	*IBBPBBPBBPBBPBB*	*IBBPBBPBBPBB*
L	15	12
Q	3	3

Table 1: Encoding parameters for the three sequences.

Notice that the image dimensions in Oz are larger than Lambs and Bond. Hence, the frame sizes in Oz trace should be expected, in general, to be larger. The compression pattern in Oz is also different from the one used to obtain the other two traces.

The rationale for using three different traces is to ensure that the proposed model is applicable to various types of movies with different levels of dynamics (Oz has the least amount of scene activity followed by Lambs and finally Bond). Sample statistics for the three traces are shown in Table 2. A representative segment from these traces is plotted in Figure 2 (taken from the Oz trace). In this figure, the larger frame sizes (which are depicted by the lightly shaded areas) correspond to I frames. The darker parts correspond to P and B frames which are smaller in size than I frames.

Trace	Frame Type	Number of Frames	Frame Size Statistics (in kbits)			
			Mean	Std. Deviation	Maximum	Minimum
Oz	Total	41760	41.7	51.7	343.1	0.56
	I	2784	197.1	63.8	343.1	36.2
	P	11136	58.0	37.3	284.6	0.56
	B	27840	19.6	5.7	60.0	0.57
Lambs	Total	40000	7.312	11.20	134.22	0.288
	I	3334	38.02	12.813	134.22	15.120
	P	10000	7.436	8.052	88.60	0.528
	B	26666	3.425	3.184	54.976	0.288
Bond	Total	40000	24.308	25.695	244.592	1.912
	I	3334	83.297	26.072	244.592	22.226
	P	10000	41.428	20.167	208.328	5.304
	B	26666	10.513	4.194	55.368	1.912

Table 2: Sample statistics from the three empirical traces.

Histograms and autocorrelation functions (ACF) for the three traces are shown in Figures 3- 5 (in the ACF figures, the amount of autocorrelation at a given lag is indicated by the magnitude of the impulse at that lag). In obtaining these figures, frame sizes were first converted to ATM

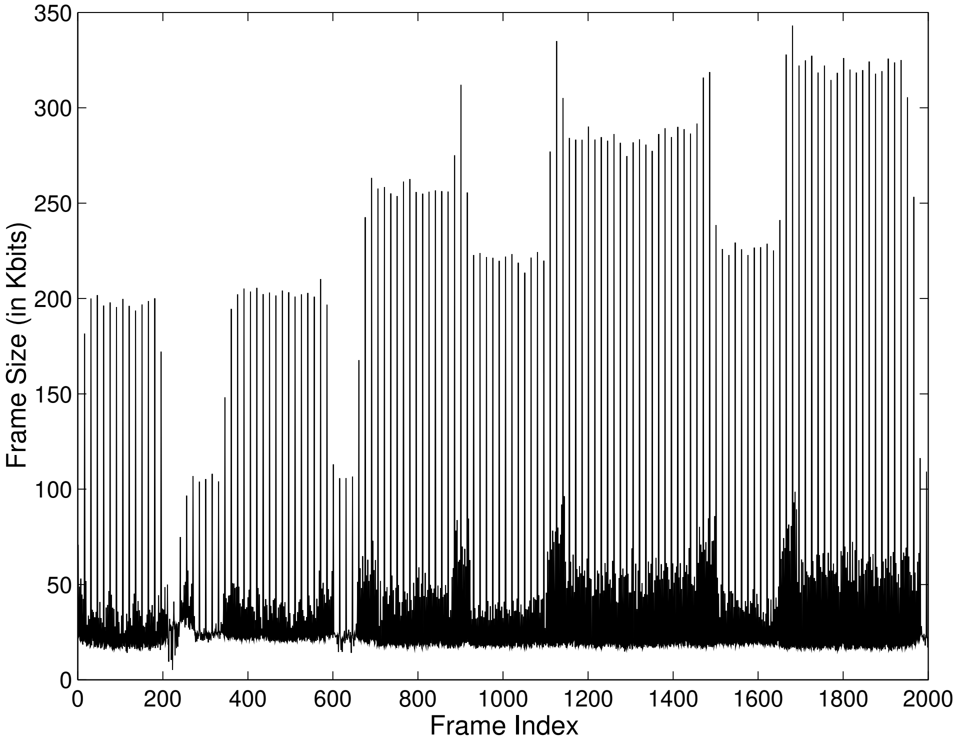

Figure 2: Segment of the frame size sequence for Oz trace.

cells (a frame of size X in bits is packetized into $\lceil X/(8*48) \rceil$ cells, where $\lceil . \rceil$ denotes the next largest integer operator). The empirical histogram for each sequence has the shape of a Pareto or a Pearson type V *pdf*. However, fitting the histogram by a theoretical *pdf* will hide the impact of the individual frame types on the overall frame size distribution. A better approach would be to provide theoretical fits to frame size histograms of same-type frames, and then superpose the individual fits to obtain the frame size distribution of an arbitrary frame. Such an approach provides more insight on the interaction among the different types of frames.

The ACF for each of the three sequences exhibits a complicated structure. This structure is strongly influenced by the form of the compression pattern used to encode each sequence. Two apparent periodic components are observed in each ACF: one at multiple lags of L (which is 15 frames for Oz and 12 frames for Lambs and Bond) and the other at multiple lags of $Q = 3$. The latter component is caused by correlations between P frames only. Both components decay at very slow rates. The first periodic component in the ACF is almost the same in the three traces. However, the relative magnitude of the second component (i.e., correlations due to P frames) is observed to vary significantly from one trace to another. The three traces also differ in the impact of B frames on the ACF (this impact is seen by the negative correlations).

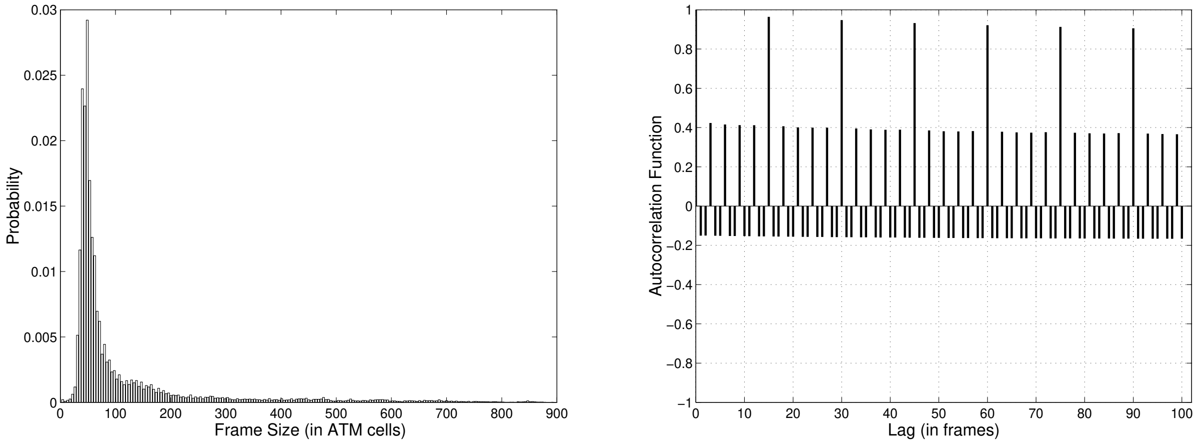

Figure 3: Frame size histogram and autocorrelation function for Oz trace.

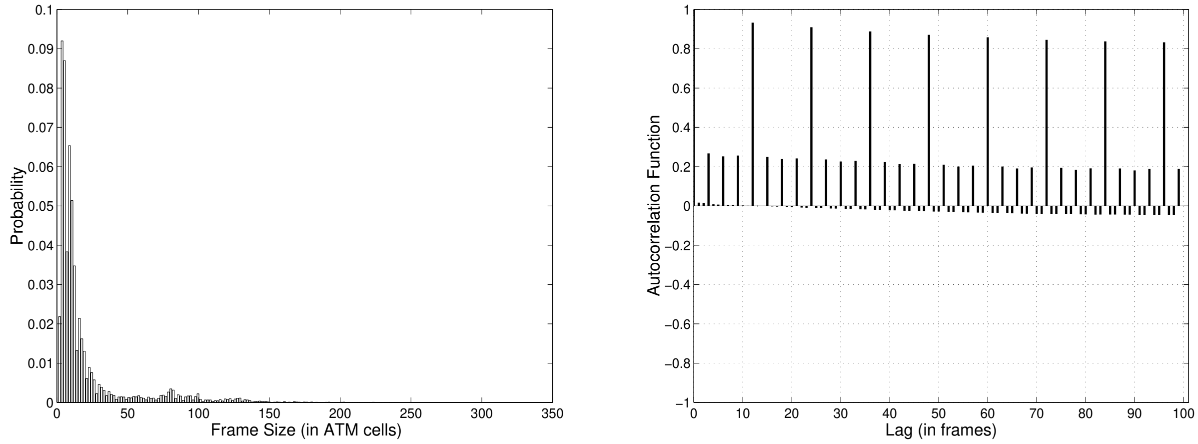

Figure 4: Frame size histogram and autocorrelation function for Lambs trace.

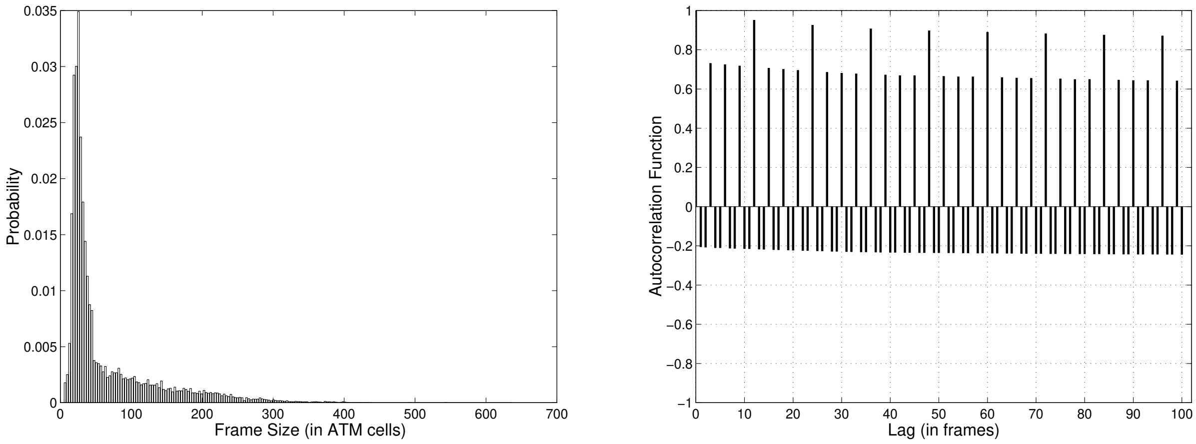

Figure 5: Frame size histogram and autocorrelation function for Bond trace.

3 Modeling a VBR MPEG Stream

In this section, we propose a traffic model that approximates the statistical behavior of an MPEG stream. Specifically, the model captures the general shapes of the frame size histogram and ACF of a given MPEG trace, in addition to producing sample paths similar to the actual traces. Our strategy is based on observing the VBR behavior of real MPEG streams, and then modeling this behavior by a stochastic process. Multiple components are used in our model to capture the complicated behavior of a VBR MPEG stream. These components take place at different time scales.

3.1 Scene Length Distribution

From Figure 2, it is observed that an MPEG trace consists of several segments such that the sizes of I frames in each segment are close in value (this behavior is also observed in the other traces which are not depicted here). By visual inspection of the examined movies, we verified that each of these segments corresponds to part of the movie with no *abrupt* view shifts. In [11], such segments were referred to as *scenes*. Note that only an abrupt view shift accounts for the beginning of a new scene. Camera panning and zooming are not considered as scene boundaries.

To model the length of a scene (in seconds or in successive frames), one must first obtain a set of empirical data that represent the lengths of actual scenes in a movie. This can be done by watching a movie and measuring the duration of each scene. A disadvantage of this approach is that it requires the availability of the actual movie which must be watched in slow motion to accurately measure scene durations. Moreover, it is sometimes difficult to visually determine whether a given view shift represents the start of a scene (e.g., when the camera is moving fast but in a continuous manner). We use a heuristic approach to compute scene durations, which only requires the availability of the VBR trace [2]. In this approach, we make use of the fact that a 'sufficient' difference between the sizes of two consecutive I frames is a strong indication of the start of a new scene. Thus, one can use the sizes of consecutive I frames in a trace to obtain the scene length data. Let $\{\widehat{X}_I(n) : n = 1, 2, ...\}$ be the I frames subsequence in an empirical MPEG stream (throughout the paper, we use the 'hat' symbol above a variable name to indicate an empirical quantity). This subsequence consists of the sizes of consecutive I frames in a given MPEG trace. Suppose that the current scene is the ith scene that started with the kth I frame. The $(n+k+1)$th

[2] A slightly different approach was used in [11] to determine scene boundaries in JPEG-based movies.

I frame of the sequence indicates the start of the $(i+1)$th scene if

$$\frac{\left|\widehat{X}_I(n+k+1) - \widehat{X}_I(n+k)\right|}{\left(\sum_{j=k}^{n+k} \widehat{X}_I(j)\right)/n} > T_1 \tag{1}$$

and

$$\frac{\left|\widehat{X}_I(n+k+2) - \widehat{X}_I(n+k)\right|}{\left(\sum_{j=k}^{n+k} \widehat{X}_I(j)\right)/n} > T_2 \tag{2}$$

where T_1 and T_2 are two thresholds. Clearly, the values of T_1 and T_2 determine the amount of correlations between successive scenes; the larger these values, the less correlated the scenes. However, increasing T_1 and T_2 will impact the shape of the scene length distribution (and the visual definition of a scene). Using $T_1 = 0.1$ and $T_2 = 0.15$, scene lengths were obtained for the three empirical traces, and were used to construct scene length histograms. Notice that the length of a scene is measured by the number of consecutive I frames in that scene. The values for T_1 and T_2 were chosen because they provided accurate mapping between the actual scenes in the movies and the computed ones. Additionally, they resulted in empirical histograms that have the shape of a geometric *pmf*. Figure 6 shows the scene length histogram (left) and the Q-Q plot for Oz trace. Similar plots were obtained for the other two traces. The simplicity and memoryless property of a geometric *pmf* make it possible to derive tractable analytical statistics of the model. Hence, we assume that the scene length in our model is geometrically distributed with parameter q ($q < 1$). Let N_j be the length of the jth scene ($j = 1, 2, \ldots$). Then,

$$\Pr\{N_j = n\} = q^{n-1}(1-q), \text{ for } n = 1, 2, \ldots \tag{3}$$

The mean scene length is $E[N_j] = 1/(1-q)$.

Sample statistics for the scene length in each trace are provided in Table 3. Observe that Oz has the smallest number of scenes, followed by Lambs, and finally Bond. The average scene length can be used as an indication of the relative activity of a movie (the higher the action of the movie, the smaller the average scene length). Hence, movies can be classified according to the average scene length which is obtainable from the empirical traces.

Trace	Number of Scenes	Scene Length Statistic (in consecutive I frames)			
		Mean	Std. Deviation	Max.	Min.
Oz	199	14.0	17.3	90	1
Lambs	327	10.2	18.1	209	1
Bond	441	7.6	9.4	66	1

Table 3: Sample statistics for the scene length in the three traces ($T_1 = 0.1$, $T_2 = 0.15$).

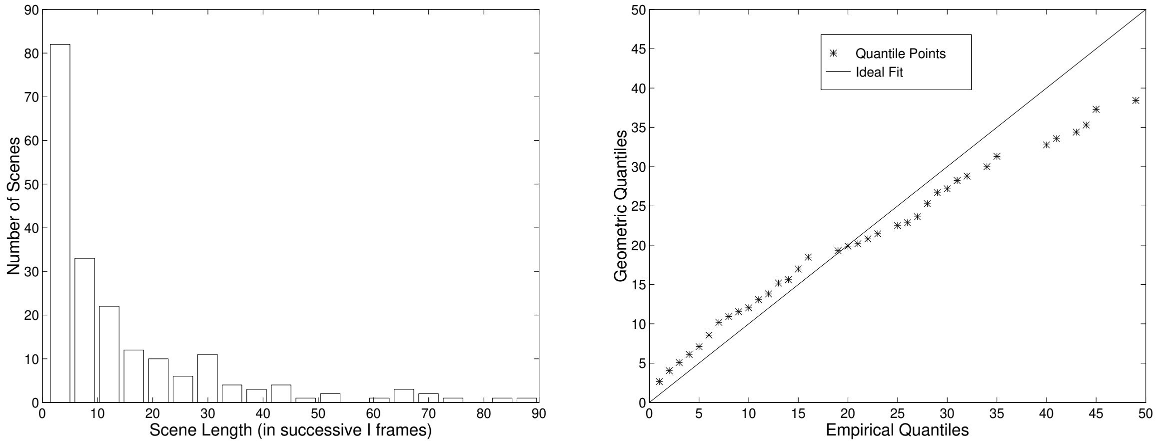

Figure 6: Scene length histogram and quantile plot for Oz trace.

We tested the hypothesis of independence between scene lengths based on the runs-up test [10]. At a 95% level of confidence, the hypothesis of independence cannot be rejected for either trace. Hence, we assume in our model that $\{N_j\}$ is a sequence of *i.i.d.* random variables with common geometric distribution, *GEO(q)*.

3.2 Frame Size Distributions

The two important factors that determine the size of an MPEG frame are the scene activity and the frame type. The effect of the scene activity occurs at a time scale of few seconds (the average scene length). Frame types are determined by the compression pattern, and hence, their impact on the bit rate occur at a smaller time scale (e.g., 1/30 of a second). In our model, the impact of the frame type is captured by modeling the frame size for each type of frames, and then using the compression pattern to produce the desired stream. As we demonstrate later, the deterministic and periodic nature of the compression pattern ensures that a certain degree of correlation exists between frame sizes of different types.

Although the effect of the scene activity exists in all types of frames, this effect is particularly obvious (and thus easier to capture) in the case of I frames. Because I frames are intra-coded, their sizes are independent of the sizes of P and B frames. On the contrary, the size of a P or a B frame is the result of a complex interaction between scene activity and the sizes of other reference (I or P) frames. In our model, we assume that scene activity affects the sizes of I frames only. Since I frames are the largest, incorporating the scene activity in their frame size variation amounts to the capturing most of the actual impact of scene activity. Next, we consider the different components of our model.

3.2.1 I Frames

By our definition of a scene (see (1) and (2)), the sizes of I frames *within* a scene are close. These sizes fluctuate about some average value that represents the level of activity of the scene. Let $X_I(n)$ be the size of the nth I frame. In our model, $X_I(n)$ is the sum of two independent random variables:

$$X_I(n) = X_I^*(n) + \Delta_I(n) \tag{4}$$

where $X_I^*(n)$ reflects the level of activity of the scene, while $\Delta_I(n)$ represents the fluctuation of the nth I frame about $X_I^*(n)$.

First, we consider $X_I^*(n)$. This quantity is constant for all I frames in the same scene, and it varies from one scene to another. Hence, for the jth scene that started with the ith I frame, we have

$$X_I^*(i) = X_I^*(i+1) = \cdots = X_I^*(i+N_j-1) \stackrel{\triangle}{=} \widetilde{X}_I(j) \tag{5}$$

It is assumed that $\{\widetilde{X}_I(j) \, : \, j = 1, 2, \ldots\}$ is a sequence of *i.i.d.* random variables with common *pdf* $f_{\widetilde{X}_I}$. This assumption was motivated by the lack (or weak presence) of correlations in the empirical sequence $\{\hat{\widetilde{X}}_I(j)\}$, where $\hat{\widetilde{X}}_I(j)$ is the sample average of the sizes of I frames in the jth scene in a given empirical trace. The histogram of $\{\hat{\widetilde{X}}_I(j)\}$ for Bond trace is shown in Figure 7. The shape of this histogram suggests the use of a lognormal fit. Similar conclusions were reached for the other two traces. Hence, we assume that $\widetilde{X}_I(j)$ is lognormally distributed with parameters $\widetilde{\mu}_I$ and $\widetilde{\sigma}_I$:

$$f_{\widetilde{X}_I}(u) = \begin{cases} \frac{1}{u\sqrt{2\pi\widetilde{\sigma}_I^2}} \exp\left[\frac{-(\ln u - \widetilde{\mu}_I)^2}{2\widetilde{\sigma}_I^2}\right], & u > 0 \\ 0, & otherwise \end{cases} \tag{6}$$

The mean ($m_{\widetilde{X}_I}$) and the variance ($v_{\widetilde{X}_I}$) of $\widetilde{X}_I$ are given by:

$$m_{\widetilde{X}_I} = \exp\left\{\widetilde{\mu}_I + \widetilde{\sigma}_I^2/2\right\} \tag{7}$$

$$v_{\widetilde{X}_I} = \exp\left\{2\widetilde{\mu}_I + \widetilde{\sigma}_I^2\right\}\left(\exp\left\{\widetilde{\sigma}_I^2\right\} - 1\right) \tag{8}$$

The maximum-likelihood estimates (MLE) are used to obtain the estimates of $\widetilde{\mu}_I$ and $\widetilde{\sigma}_I$. These estimates are given in Table 4 for the three traces.

The following proposition gives the probability distribution for $X_I^*(n)$.

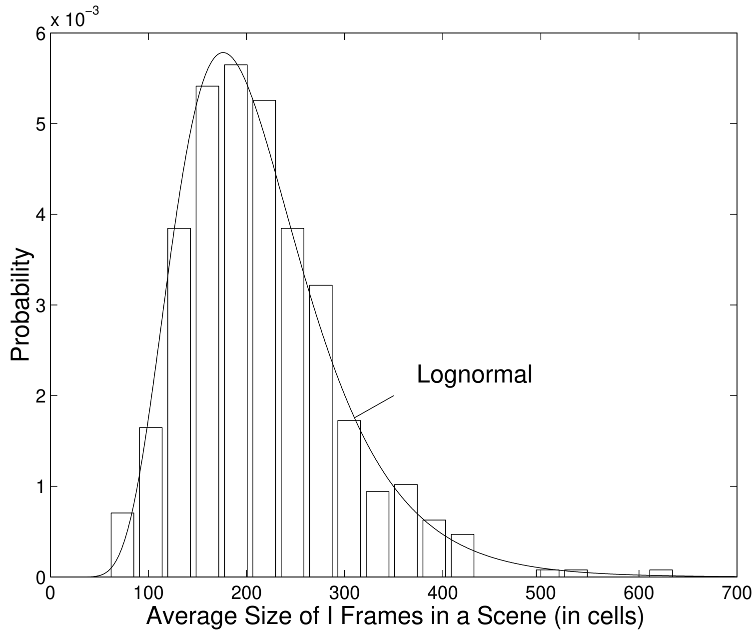

Figure 7: Histogram and lognormal fit for $\{\widehat{\widetilde{X}}_I(j)\}$ in Bond trace.

Proposition 1 *Let $f^{(n)}_{X^*_I}$ be the pdf for $X^*_I(n)$. Then*

$$f^{(n)}_{X^*_I} = f_{\widetilde{X}_I}, \text{ for all } n. \tag{9}$$

Proof: (by induction). It is clear that $f^{(1)}_{X^*_I} = f_{\widetilde{X}_I}$. Suppose that $f^{(n)}_{X^*_I} = f_{\widetilde{X}_I}$. The probability that the nth and $(n+1)$th I frames are in the same scene is q. In this case, $X^*_I(n) = X^*_I(n+1)$ w.p.1., which implies that the two random variables are also equal in distribution. With probability $1-q$, the $(n+1)$th I frame is the starting frame of a new scene, in which case $f^{(n+1)}_{X^*_I} = f_{\widetilde{X}_I} = f^{(n)}_{X^*_I}$. Thus, $f^{(n)}_{X^*_I} = f^{(n+1)}_{X^*_I}$. By induction on n, the assertion is true. □

Proposition 2 says that $\{X^*_I(n)\}$ is an identically distributed (but not independent) random sequence. The proof relies on the memoryless property of the geometric distribution.

Next, we consider the quantity $\Delta_I(n)$ which represents the variation of the size of the nth I frame about $X^*_I(n)$. For a given trace, we compute the empirical sequence $\{\widehat{\Delta}_I(n)\}$, where $\widehat{\Delta}_I(n) = \widehat{X}_I(n) - \widehat{\widetilde{X}}_I(j_n)$ (j_n is the scene index to which the kth I frame belongs). The sequence $\{\widehat{\Delta}_I(n)\}$ is then fitted by the stochastic process $\{\Delta_I(n)\}$. Notice that $\{\Delta_I(n)\}$ was assumed to be independent of $\{X^*_I(n)\}$, and thus invariant to scene changes.

The sample ACF for $\{\widehat{\Delta}_I(n)\}$ in each trace is shown in Figures 8-10 (the left-hand side). Its shape suggests considering an autoregressive process to characterize $\{\Delta_I(n)\}$. Based on the empirical partial autocorrelation function (PACF) for $\{\widehat{\Delta}_I(n)\}$ (right-hand side of Figures 8- 10), we assume that $\{\Delta_I(n)\}$ is an AR(2) process. In fitting a time series by an autoregressive model, the PACF of the time series provides a means to determine the number of autoregressive coefficients (i.e., the order of the autoregressive model) that are adequate for the fitting (see [2]). The sample PACF for the sequence $\{\widehat{\Delta}_I(n)\}$ in Oz and Lambs traces exhibits large first and second autoregressive coefficients, suggesting an AR(2) fit. In Bond trace, an AR(1) seems sufficient. For the purpose of generality, we assume that the sequence $\{\Delta_I(n)\}$ is characterized by an AR(2) model:

$$\Delta_I(n) = a_1\Delta_I(n-1) + a_2\Delta_I(n-2) + \varepsilon(n) \tag{10}$$

where $\{\varepsilon(n)\}$ is a sequence of *i.i.d.* random variables with mean zero and variance σ_ε^2. To determine an appropriate distribution for the residuals, let f_ε and f_Δ be the *pdf* for $\varepsilon(n)$ and $\Delta_I(n)$, respectively (note that $\{\Delta_I(n)\}$ is a stationary process). The shape of f_Δ can be approximated from the histogram of $\{\widehat{\Delta}_I(n)\}$, which is shown in Figure 11 for Oz trace (similar shapes are observed in Lambs and Bond traces). A normal fit seems quite appropriate. Hence, we let $f_\Delta = N(0, \sigma_\Delta^2)$. A normal distribution for $\{\Delta_I(n)\}$ is naturally obtained if the residuals are also normally distributed. Hence, we let $f_\varepsilon = N(0, \sigma_\varepsilon^2)$.

Estimates of the parameters a_1 and a_2 in (10) can be obtained using Yule-Walker equations [2]:

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{11}$$

where r_k is the empirical autocorrelation for $\{\widetilde{\Delta}_I(n)\}$ at lag k. These estimates are given in Table 4 for the three traces. Using these estimates, we fitted the sequence $\{\widetilde{\Delta}_I(n)\}$ in each trace to an AR(2) model and computed the experimental residuals (see [2], Chapter 7). The ACF for these residuals in the Oz trace is shown in Figure 12. The absence of correlations among the residuals is a strong evidence of the appropriateness of our model (similar behavior was also observed in Lambs and Bond traces). An estimate of σ_ε^2 is given by the the sum-of-squares of the experimental residuals divided by the number of samples (I frames). These estimates are provided in Table 4.

The variance of $\Delta_I(n)$ is given by [2]:

$$\sigma_\Delta^2 = \frac{(1-a_2)\sigma_\varepsilon^2}{(1+a_2)\left[(1-a_2)^2 - a_1^2\right]} \tag{12}$$

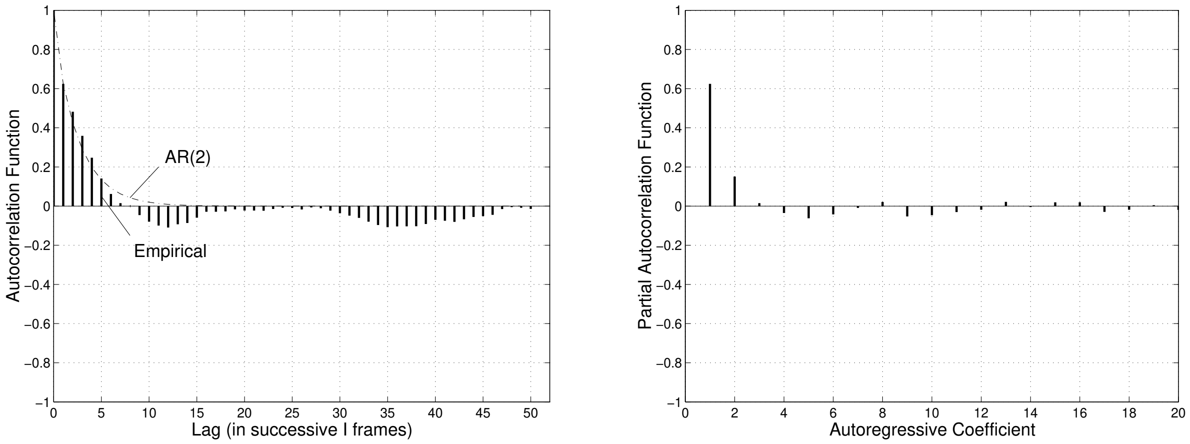

Figure 8: Autocorrelation and partial autocorrelation functions for $\{\widehat{\Delta}_I(n)\}$ in Oz trace.

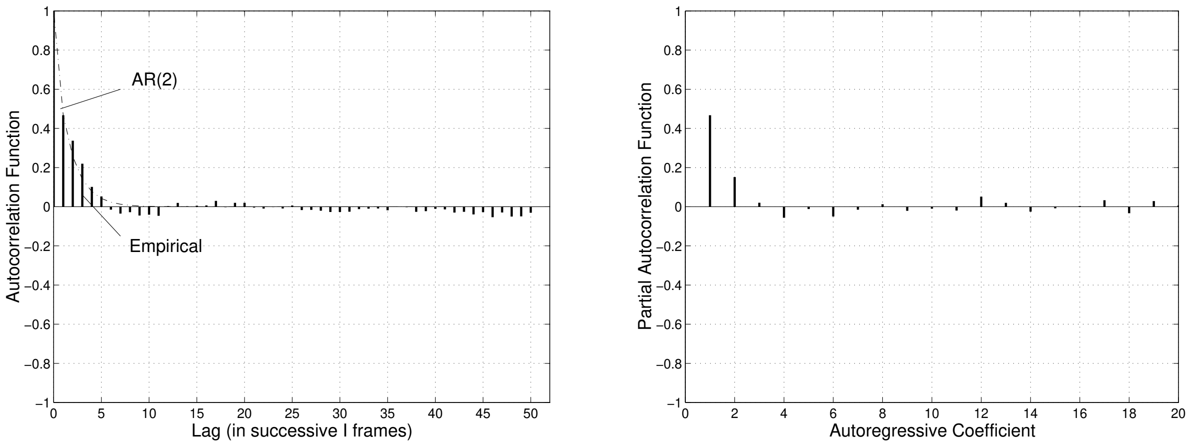

Figure 9: Autocorrelation and partial autocorrelation functions for $\{\widehat{\Delta}_I(n)\}$ in Lambs trace.

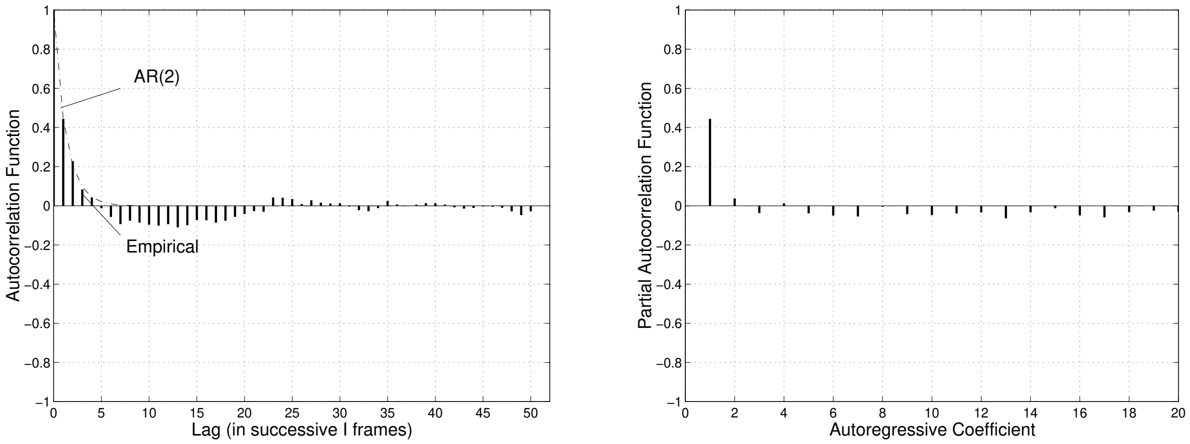

Figure 10: Autocorrelation and partial autocorrelation functions for $\{\widehat{\Delta}_I(n)\}$ in Bond trace.

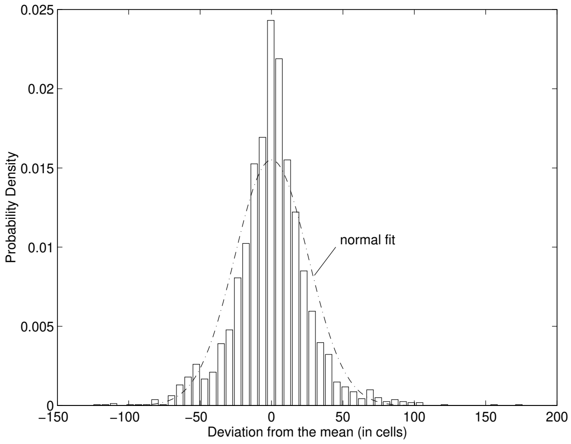

Figure 11: Histogram for the fitted sequence $\{\widehat{\Delta}_I(n)\}$ in Oz trace.

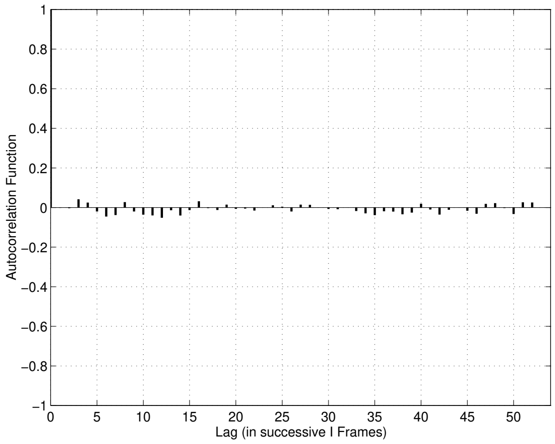

Figure 12: ACF for the empirical residuals of the AR(2) fit (Oz trace).

Let κ_m be the autocorrelation in $\{\Delta_I(n)\}$ at lag m ($m > 0$). Then (see [2]),

$$\kappa_m = \frac{A(1-B^2)A^m - B(1-A^2)B^m}{(A-B)(1+AB)} \tag{13}$$

where A and B are the roots of the characteristic equation $x^2 - a_1 x - a_2 = 0$.

Since $\{X_I^*(n)\}$ and $\{\Delta_I(n)\}$ are two independent and stationary random processes, then $\{X_I(n)\}$ is also a stationary process. Moreover, the random variables $X_I(1), X_I(2), \ldots$, are identically distributed with common *pdf* f_{X_I} that is given by the convolution of $f_{\widetilde{X}_I}$ and f_Δ. The following proposition provides the ACF for $\{X_I(n)\}$.

Proposition 2 *Let $\rho_I(m)$ denote the autocorrelation of $\{X_I(n)\}$ at lag m for $m \geq 0$. Then,*

$$\rho_I(m) = \frac{q^m v_{\widetilde{X}_I} + \kappa_m \sigma_\Delta^2}{v_{\widetilde{X}_I} + \sigma_\Delta^2}. \tag{14}$$

Proof: see appendix. □

3.2.2 P and B Frames

As mentioned earlier, our methodology approach is based on characterizing the frame size subsequence for each frame type, and then use the compression pattern to compose the complete frame size sequence. It is shown in the next section that constructing the frame size sequence in this manner captures the average amount of correlations between different types of frames. Let $\{X_P(n)\}$

and $\{X_B(n)\}$ be the frame size subsequences for P and B frames. We assume that $\{X_P(n)\}$ (also $\{X_B(n)\}$) is a sequence of *i.i.d.* random variables with common *pdf* f_{X_P} (f_{X_B} for B frames). Although in an actual MPEG stream positive correlations between the sizes of P frames (similarly, between B frames), these correlations become negligible when compared to the correlations between different types of frames (which are imposed by the structure of the compression pattern). In a previous study, we found that the functions f_{X_P} and f_{X_B} can be appropriately characterized by lognormal distributions with parameters (μ_P,σ_P) and (μ_B,σ_B), respectively. In fact, a gamma or Weibull *pdf* can also be used, with slightly less accurate fitting. Histograms for $\{\widetilde{X}_P(n)\}$ and $\{\widetilde{X}_B(n)\}$, along with their fitting distributions, are shown in Figures 13 and 14 for Oz trace. The fits in these figures are based on the MLE of the parameters of these distributions (see Table 4 for the values of these estimates). It is clear that the lognormal fit is more accurate than the other two fits (for both P and B types). Similar observations were made using Lambs and Bond traces.

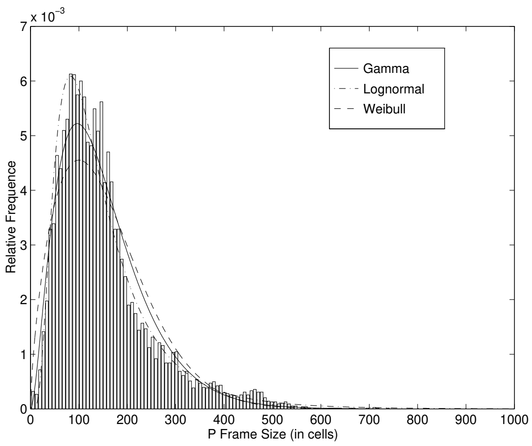

Figure 13: Histogram and fitting *pdf*s for $\{\widetilde{X}_P(n)\}$in Oz trace.

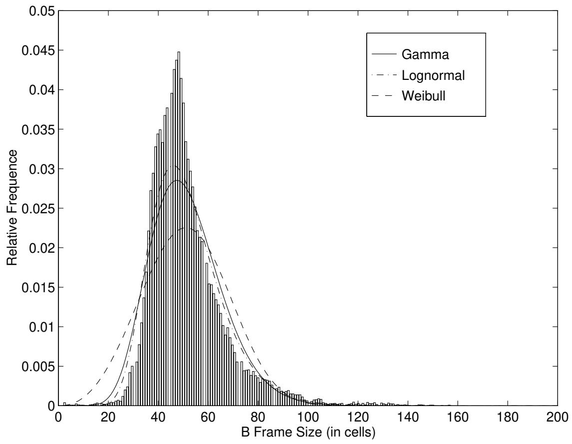

Figure 14: Histogram and fitting *pdf*s for $\{\widetilde{X}_B(n)\}$in Oz trace.

Trace	$\widetilde{\mu}_I$	$\widetilde{\sigma}_I$	a_1	a_2	σ_ε	μ_P	σ_P	μ_B	σ_B
Oz	5.9651	0.4832	0.5300	0.1512	19.8	4.8294	0.6443	3.9039	0.2746
Lambs	4.5813	0.3852	0.3965	0.1516	4.36	2.5884	0.8721	1.9861	0.6937
Bond	5.3050	0.3663	0.4271	0.0380	10.3	4.5656	0.5111	3.2572	0.3761

Table 4: Estimates of the various parameters used in the model. The MLE are used for $\widetilde{\mu}_I$, $\widetilde{\sigma}_I$, μ_P, σ_P, μ_B, and σ_B. Yule-Walker estimates are used for a_1 and a_2. The estimate of σ_ε^2 is the sum-of-square of the experimental residuals divided by the sample size.

3.3 The Autocorrelation Structure

In this section, we derive the autocorrelation function for the sequence $\{X_n\}$ where X_n is the size of the nth frame based on the model. The sequence $\{X_n\}$ is the composition of the three subsequences that were described in the previous sections. Before proceeding with our derivation, we must first check if the random process $\{X_n\}$ is stationary.

It is clear that if the sequence $\{X_n\}$ is generated such that the first frame is an I frame, followed by two B, etc., then $\{X_n\}$ is nonstationary. However, if we let $\{X_n\}$ start in any location in the compression pattern (with equal probability), and from that point continue to generate frames according to the compression pattern, then we can show that $\{X_n\}$ is stationary. This scenario is quite realistic considering, for example, that MPEG streams at at ATM multiplexer are admitted at different times. Thus, when the first frame (which is an I frame) of a newly admitted stream arrives at the multiplexer, the other ongoing streams could be in any location with respect to their compression pattern. From the perspective of the new connection (and possibly the traffic management control that is activated when new connections arrive), the other streams look as though they are initialized at different locations of the compression pattern.

Let S_n be a discrete random variable which describes the location of the nth frame within the compression pattern. The state space for S_n is $\Omega_S = \{1, 2, \ldots, L\}$. The value of S_n determines the frame type: if $S_n = 1$, then the type is I; if $S_n = 1 + Q,\ 1 + 2Q, \ldots,\ 1 + (L - Q)$, then the type is P; otherwise, it is B. Given $S_n = i$, the distribution of X_n is stationary, and only depends on i. For $i = 1, \ldots, L$, let $F_i(u) \stackrel{\triangle}{=} \Pr[X_n \le u / S_n = i]$, and let $f_i(u)$ be the corresponding *pdf*. It is easy to see that $\{S_n\}$ is an irreducible and homogeneous Markov chain, with transition probabilities:

$$p_{ij} \stackrel{\triangle}{=} \Pr[S_n = j / S_{n-1} = i] = \begin{cases} 1 & \text{if } j = i + 1,\ i = 1, \ldots, L - 1, \text{ and } i = L,\ j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

All states of the chain are positive recurrent with mean recurrence time equals to L. Thus, there is a unique stationary distribution Π such that $\Pi\mathcal{P} = \Pi$. Let $\Pi = [\pi_1\, \pi_2\, \ldots\, \pi_L]$. Then, $\pi_i = 1/\text{mean recurrence time} = 1/L$ for all i. The mean, variance, and distribution of X_n are given by the following proposition.

Proposition 3 *If $S_1 \sim \Pi$, then the random process $\{X_n\}$ is first-order stationary. In this case, the pdf for X_n is given by:*

$$f_X(u) = \frac{1}{L} \sum_{i=1}^{L} f_i(u) = \frac{1}{L} \left(f_{X_I}(u) + (L/Q - 1) f_{X_P}(u) + (L - Q) f_{X_B}(u) \right). \tag{16}$$

The mean and variance of X_n are given by:

$$m_X \triangleq E[X_n] = \frac{1}{L}\left(m_{X_I} + (L/Q-1)m_{X_P} + (L-Q)m_{X_B}\right) \tag{17}$$

$$v_X \triangleq E\left[(X_n - m_X)^2\right] = \frac{1}{L}\left(v_{X_I} + m_{X_I}^2 + (L/Q-1)(v_{X_P} + m_{X_P}^2) + (L-Q)(v_{X_B} + m_{X_B}^2)\right) - m_X^2.$$

Proof: If $S_1 \sim \Pi$, then the Markov chain $\{S_n\}$ is stationary, i.e., $\Pr[S_k = i] = \pi_i = 1/L$ for all i and k.

$$\Pr[X_n \le u] = \sum_{i=1}^{L} \Pr[X_n \le u/S_n = i] \Pr[S_n = i] \tag{18}$$

$$= \frac{1}{L}\sum_{i=1}^{L} F_i(u). \tag{19}$$

Taking the derivative of (19), we get (16), which shows that $\{X_n\}$ is 1-order stationary. The expressions for the mean and variance follow in a straightforward manner. □

In fact, it is possible, but algebraically tedious, to show that $\{X_n\}$ is also second-order stationary. We skip the proof and proceed to provide the ACF of $\{X_n\}$.

Proposition 4 *Let ρ_k denote the ACF of $\{X_n\}$ at lag k, and let $h(i,k) \triangleq (i+k-1) \bmod L + 1$. Then,*

$$\rho_k = \frac{(1/L)\sum_{i=1}^{L} E[X_1/S_1 = i]\, E[X_1/S_1 = h(i,k)] - m_X^2}{v_X}, \text{ for all } k,\ k \ne L,\ 2L, \ldots \tag{20}$$

$$= \frac{(1/L)\sum_{i=2}^{L} E^2[X_1/S_1 = i] + (1/L)E[X_1 X_{1+k}/S_1 = 1] - m_X^2}{v_X} \text{ for } k = L,\ 2L, \ldots \tag{21}$$

Proof: see appendix. □

All the quantities in the RHS of (20) and (21) are known. The middle term in the numerator of (21) reflects the contribution of the correlations between I frames only, and is given by:

$$E[X_1 X_{1+k}/S_1 = 1] = q^m E\left[(X_I^*)^2\right] + (1-q^m)E^2[X_I^*] + \kappa_m \sigma_\Delta^2 \tag{22}$$

where $m = k/L$ (see the appendix for more details).

Initialization: set *Scene_max* := desired number of scenes
set values for compression pattern: L, Q
set scene length parameter: q
set the values for $\widetilde{\mu}_I$, $\widetilde{\sigma}_I$, μ_P, σ_P, μ_B, σ_B, a_1, a_2, and σ_ε^2
$X \triangleq$ frame size (in cells)

Loop:

for $j = 1$ to *Scene_max* **do**
    generate the length of scene j: $N_j \sim$ Geometric(p)
    generate $\widetilde{X}_I(j) \sim \text{lognormal}(\widetilde{\mu}_I, \widetilde{\sigma}_I)$
    for $i = 1$ to N_j **do**
        for $k = 1$ to L **do**
            if $k = 1$, **then** /* an I frame */
                $X = \widetilde{X}_I(j) + \Delta_I(i)$, where
                $\Delta_I(i) = \hat{a}_1 \Delta_I(i-1) + \hat{a}_2 \Delta_I(i-2) + \varepsilon(i)$
            else
                if remainder$(k/Q) = 1$, **then** /* a P frame */
                    $X \sim \text{lognormal}(\hat{\mu}_P, \hat{\sigma}_P)$
                else /* a B frame */
                    $X \sim \text{lognormal}(\hat{\mu}_B, \hat{\sigma}_B)$
                end if
            end if
        end for
    end for
end for

Figure 15: Generating a synthetic stream based on the model.

4 Generating Synthetic MPEG Streams

The algorithm in Figure 15 can be used to generate synthetic MPEG streams based on our model. If the maximum sizes of I, P and B frames (or possibly some upper bounds of these values) are known, then the computer-generated frame sizes should be limited to these maximum values.

To illustrate the usefulness of the model, we show an example of a synthetic stream being generated with parameter values taken from the fitting of Oz trace. A sample path from this stream in shown in Figure 16. The empirical frame size distribution is shown in Figure 17 (notice that no closed-form solution is available for this distribution). The analytical ACF and the sample ACF for this synthetic trace are shown in Figures 18 and 19. Comparing these figures to their counterparts in the original trace, it can be observed that the model is capable of capturing the general shapes of the frame size distribution and ACF of the actual trace. Additionally, the sample path generated by the model has a similar structure to the empirical trace; a property which could not be obtained without incorporating the effect of scene changes.

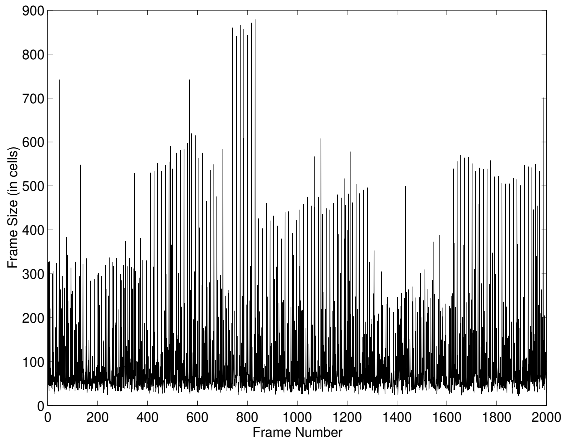

Figure 16: A segment of the frame size sequence in a synthetic MPEG stream (using Oz parameters).

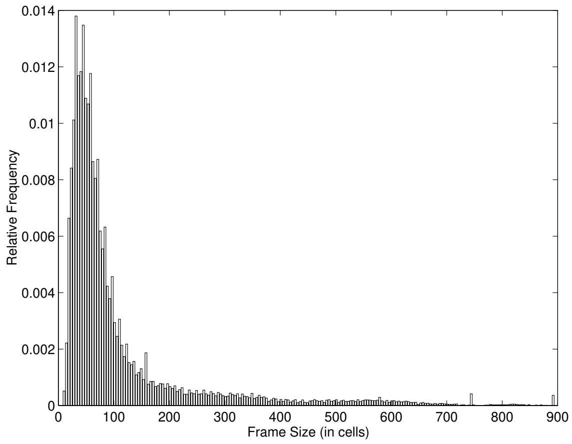

Figure 17: Frame size histogram in a synthetic stream.

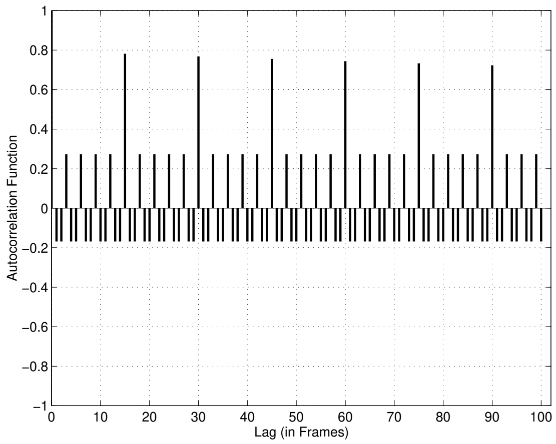

Figure 18: The analytical ACF of the frame size sequence.

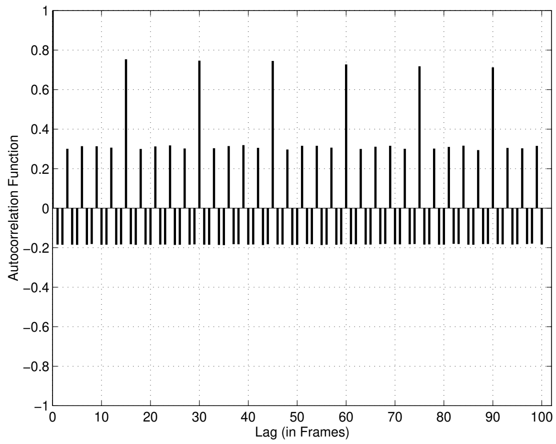

Figure 19: Sample ACF of the frame size sequence in a synthetic stream.

In many performance studies, traffic models are mainly used to evaluate the queueing performance in network buffers. Hence, the most important criterion for testing the appropriateness of a given traffic model should be the degree of similarity between the queueing performance under the examined model and that obtained under the actual traffic. Other criteria such as the similarity in the ACF and the marginal distribution are important but may not be sufficient to ensure similar queueing behavior. To verify the appropriateness of our model, we conducted simulation experiments of a finite-capacity FIFO queueing system. The input to the queue consists of a single MPEG stream which can be either the actual empirical trace, or a synthetic stream generated according to our model. Figures 20 and 21 depict the cell loss rate and the average queue length, respectively, based on Oz trace at two levels of utilization: $U = 40\%$ and $U = 60\%$. The performance results for the model-based stream were obtained using five simulation replications, each having a different random number seed. The 95% confidence intervals for the model-based results are shown by the short vertical lines. For small buffers, the agreement between the two curves is remarkable. As buffer size increases, the cell loss rate under the model becomes more pessimistic than the trace-driven loss rate. Nevertheless, the difference between the two is considered small (less than an order of magnitude). Moreover, the confidence interval for the model-based results becomes larger when buffer size increases; thus, making the trace-driven results closer to the lower limit of the confidence interval. In general, simulation results become less reliable at small loss rates. Notice that our simulations were conducted at relatively light loads. For the loss rate, the agreement between the model and trace results is more obvious at higher levels of utilization. Observe that the results based on the model always upper-bound the trace-driven results. Hence, the model can be quite useful in providing performance guarantees (i.e., QoS) in ATM networks.

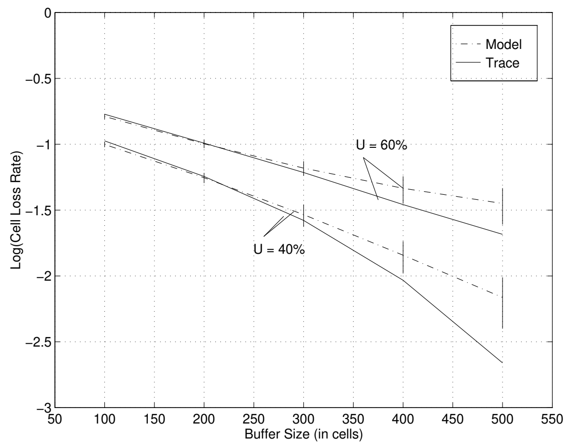

Figure 20: Cell loss rate in a FIFO buffer versus the buffer size.

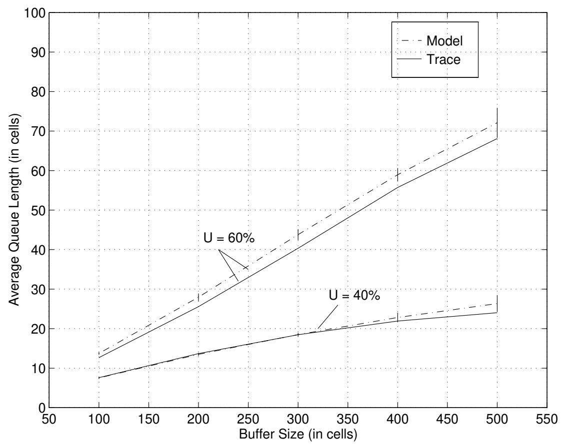

Figure 21: Average queue length versus buffer size.

5 Summary

A traffic model for MPEG-coded video streams was presented in this paper. The model was fitted to three empirical video traces, taken from movies with different levels of action. In each case, the fit was found adequate. The model captures bit rate variations in an MPEG stream at different time scales. Long-term variations are mainly caused by scene changes, and are incorporated in our model. First, we presented a heuristic to measure scenes using the trace data only. The heuristic is based on the observation that the sizes of I frames fluctuate in small amounts *within* a scene. A considerable change in the size of consecutive I frames is used to indicate the start of a new scene. Scenes were modeled by a geometric distribution. The average size of I frames within a scene was appropriately fitted by a lognormal distribution. Fluctuations about this average were modeled using an AR(2) time series. Based on a previous study, we modeled the sizes of P and B frames by lognormal distributions (with different parameters). We demonstrated the validity of our model by showing the similarity between a synthetic stream and an actual trace, in terms of the general shapes of the autocorrelation function, the marginal distribution, and the sample paths. More importantly, we showed that both the synthetic stream and the actual trace exhibit sufficiently similar queueing behavior.

Appendix

A Proof of Proposition 2

For $m > 0$,

$$\rho_I(m) \triangleq \frac{E\left[(X_I(1) - m_{X_I})(X_I(1+m) - m_{X_I})\right]}{v_{X_I}} \tag{23}$$

$$= \frac{E\left[X_I^*(1)X_I^*(1+m)\right] + E\left[\Delta_I(1)\Delta_I(1+m)\right] - m_{X_I}^2}{v_{X_I}} \tag{24}$$

(since $\{X_I^*(n)\}$ and $\{\Delta_I(n)\}$ are independent). The variance of $X_I(1)$ is given by $v_{X_I} = v_{\widetilde{X}_I} + \sigma_\Delta^2$. Consider the numerator of the RHS of (24). The probability that the first I frame and the $(1+m)$th I frame belong to the same scene is q^m. Thus, $X_I^*(1) = X_I^*(1+m)$ with probability q^m. Otherwise, $X_I^*(1)$ and $X_I^*(1+m)$ are independent with identical distribution $f_{\widetilde{X}_I}$. Therefore,

$$E\left[X_I^*(1)X_I^*(1+m)\right] = \begin{cases} E\left[X_I^*(1)^2\right], & w.p. q^m \\ E\left[X_I^*(1)\right]E\left[X_I^*(m+1)\right], & w.p. (1-q^m) \end{cases} \tag{25}$$

Also, $E\left[\Delta_I(1)\Delta_I(1+m)\right] = \kappa_m \sigma_\Delta^2$. Thus, (24) becomes

$$\rho_I(m) = \frac{q^m E\left[X_I^*(1)^2\right] + (1-q^m)E^2\left[X_I^*(1)\right] + \kappa_m\sigma_\Delta^2 - m_{X_I}^2}{v_{\widetilde{X}_I} + \sigma_\Delta^2} \tag{26}$$

which reduces to the form of Proposition 2 after some manipulations. □

B Proof of Proposition 4

To show that the ACF for $\{X_n\}$ is given by (20) and (21), we consider two cases:

Case 1: $k \neq L,\ 2L, \ldots$

$$\rho_k \triangleq \frac{E\left[(X_1 - m_X)(X_{1+k} - m_X)\right]}{v_X} \tag{27}$$
$$= \frac{E\left[X_1 X_{1+k}\right] - m_X^2}{v_X} \tag{28}$$

Consider the term $E\left[X_1 X_{1+k}\right]$.

$$\begin{aligned} E\left[X_1 X_{1+k}\right] &= \int_0^\infty \Pr\left[X_1 X_{1+k} > u\right] du \\ &= \int_0^\infty \sum_{i=1}^{L} \Pr\left[X_1 X_{1+k} > u / S_1 = i\right] \cdot \Pr\left[S_1 = i\right] du \\ &= \sum_{i=1}^{L} \pi_i E\left[X_1 X_{1+k} / S_1 = i\right] \\ &= \frac{1}{L} \sum_{i=1}^{L} E\left[X_1 X_{1+k} / S_1 = i\right] \end{aligned} \tag{29}$$

The two conditional events $[X_1 / S_1 = i]$ and $[X_{1+k} / S_1 = i]$ are independent for $k \neq$ a multiple of L. Hence,

$$E\left[X_1 X_{1+k}\right] = \frac{1}{L} \sum_{i=1}^{L} E\left[X_1 / S_1 = i\right] E\left[X_{1+k} / S_1 = i\right] \tag{30}$$
$$= \frac{1}{L} \sum_{i=1}^{L} E\left[X_1 / S_1 = i\right] E\left[X_{1+k} / S_{1+k} = h(i,k)\right] \tag{31}$$
$$= \frac{1}{L} \sum_{i=1}^{L} E\left[X_1 / S_1 = i\right] E\left[X_1 / S_1 = h(i,k)\right] \tag{32}$$

which proves the assertion.

Case 2: $k = mL$ **for** $m = 1, 2, \ldots$

In this case, $E\left[X_1 X_{1+k}\right]$ can be written as

$$E\left[X_1 X_{1+k}\right] = \frac{1}{L} \sum_{i=2}^{L} E\left[X_1 X_{1+k} / S_1 = i\right] + \frac{1}{L} E\left[X_1 X_{1+k} / S_1 = 1\right] \tag{33}$$

For $i = 2, 3, \ldots, L$,

$$E\left[X_1 X_{1+k}/S_1 = i\right] = E\left[X_1/S_1 = i\right] E\left[X_{1+k}/S_1 = i\right] = E^2\left[X_1/S_1 = i\right] \tag{34}$$

(the first equality holds because each of the subsequences $\{X_P(n)\}$ and $\{X_B(n)\}$ are *i.i.d.*. The second equality holds because, in this case, $S(1+k) = S(1+mL) = S(1)$). The term $E\left[X_1 X_{1+k}/S_1 = 1\right] = E\left[X_I(1) X_I(1+m)\right]$ which is given in the proof of Proposition 2. $\square$

References

[1] T. G. Alliance. The U.S. HDTV standard: the Grand Alliance (special report on digital TV). *IEEE Spectrum*, pages 36–45, Apr. 1995.

[2] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forcasting and Control.* Holden-Day, Inc., 1976. (Revised Edition).

[3] V. S. Frost and B. Melamed. Traffic modeling for telecommunications networks. *IEEE Communications Magazine*, 32(3):70–81, Mar. 1994.

[4] K. L. Gong. Berkeley MPEG-1 video encoder, user's guide. Technical report, Computer Science Division-EECS, 1994.

[5] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. Kaye. Modeling and simulation of self-similar variable bit rate compressed video: A unified approach. In *Proc. of SIGCOMM '95*, 1995.

[6] ISO/MPEG. ISO CD11172-2: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 mbits/s, Nov. 1991.

[7] ISO/MPEG II. ISO CD11172-2: Coding of moving pictures and associated audio, Dec. 1992.

[8] B. Jabbari, F. Yegengolu, Y. Kuo, S. Zafar, and Y.-Q. Zhang. Statistical characterization and block-based modeling of motion-adaptive coded video. *IEEE Trans. on Circuits and Systems for Video Technology*, 3(3):199–207, June 1993.

[9] E. W. Knightly and H. Zhang. Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model. In *Proc. of IEEE INFOCOM '95*, pages 1137–1145, 1995.

[10] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, Inc., second edition, 1991.

[11] A. A. Lazar, G. Pacifici, and D. E. Pendarakis. Modeling video sources for real-time scheduling. Technical Report 324–93–03, Columbia University, Department of Electrical Engineering and Center for Telecommunications Research, Apr. 1993.

[12] D. Le Gall. MPEG: A video compression standard for multimedia applications. *Communications of the ACM*, 34(4):47–58, Apr. 1991.

[13] P. Pancha and M. El Zarki. MPEG coding for variable bit rate video transmission. *IEEE Communications Magazine*, 32(5):54–66, May 1994.

[14] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. In *Proceedings of the 20th Annual Conference on Local Computer Networks*, Minneapolis, MN, 1995.