Youjip Won · Soohan Ahn

GOP ARIMA: Modeling the nonstationarity of VBR processes

Abstract In this work, we develop a stochastic model, GOP ARIMA (autoregressive integrated moving average for a group of pictures) for VBR processes with a regular GOP pattern. It explicitly incorporates the deterministic time-dependent behavior of frame-level VBR traffic. The GOP ARIMA model elaborately represents the inter- and intra-GOP sample autocorrelation structures and provides a physical explanation of observed stochastic characteristics of the empirical VBR process. We explain stochastic characteristics of the empirical VBR process, e.g., slowly decaying sample autocorrelations and strong correlations at the lags, based on the aspect of nonstationarity of the underlying process. The GOP ARIMA model generates synthetic traffic, which has the same multiplicative periodic sample autocorrelation structure as well as slowly decaying autocorrelations of the empirical VBR process. The simulation results show that the GOP ARIMA process very well captures the behavior of the empirical process in various respects: packet loss, packet delay, and frame corruption. Our work makes a contribution not only toward providing a theoretical explanation of the observed characteristics of the empirical VBR process but also toward the development of an efficient method for generating a more realistic synthetic sequence for various engineering purposes and for predicting future bandwidth requirements.

Keywords Periodicity · VBR · GOP ARIMA · Multimedia · MPEG

1 Introduction

1.1 Motivation

Recent advancements in microprocessor speed, communication media, and storage technologies enable computers to store and transport huge amounts of information. Furthermore, advances in 3G wireless communication service, home networking, and wireless personal area network (WPAN) technologies, accompanied by a rapid decrease in hardware form factor, have opened up a new era of ubiquitous dissemination and consumption of multimedia contents. Variable bit rate (VBR) video traffic is expected to be one of the major workloads of a substantial fraction of network traffic in wired and wireless environments [28]. Despite all these technical achievements and the initial success in prototype service, realization of streaming service is still being challenged by a number of technical issues. One of the major factors that practically prohibits the cost-effective deployment of streaming services is the fact that real-time video streaming still consumes too many resources to compete with legacy removable-storage-based entertainment media, e.g., video cassettes and DVDs [29]. Comprehensive understanding of VBR traffic can be used to effectively exploit the resources involved in transporting a real-time video stream, e.g., buffer allocation in the server, I/O scheduling, network bandwidth management, etc. Henceforth, obtaining the invariant characteristics of VBR traffic can no longer be emphasized.

One of the most widely used coding schemes, MPEG [41], exploits the spatial and temporal information redundancy in successive frames. There are three types of frames in the MPEG coding scheme: I, P, and B. Frame type I is self-contained, frame type P carries the information difference from the preceding I type frame, and frame type B contains the interpolated information between consecutive I or P frame type pairs. One of the important features in the MPEG encoding process is that the GOP structure has been built as a way to preserve a degree of random access and to achieve a high level of compression. The GOP structure
specifies the number and temporal order of P and B frames between two successive I frames. The GOP structure is represented by \( \text{GOP}(N,M) \), where \( N \) is the distance between successive I frames and \( M \) denotes the distance between successive (P or I) and (P or I) frames. For example, \( \text{GOP}(15,3) \) denotes the frame sequence “IBBPBBPBBPBBPBB.” It is worth noting that the MPEG coding standard does not require regularity of the GOP structure, or even a repetitive one. However, it is common practice in industry to use the regular GOP pattern for source coding\(^1\) [18, 35]. A fixed GOP pattern is used primarily to meet the random access granularity requirement (e.g., less than 0.5 s in the ATSC standard, www.atsc.org) with minimum decoder complexity and maximum compression ratio. The MPEG-2 coding scheme with GOP structure, \( \text{GOP}(15,3) \), is chosen as the source coding standard for digital broadcasting services. The MPEG-2 transport stream, which is an interleaved collection of MPEG-2 elementary streams, can be envisioned as one of the dominant forms of multimedia traffic in broadband networks. Several important characteristics have been commonly observed in most empirical VBR processes. The first characteristic is the slowly decaying sample autocorrelations. The second characteristic is the periodicity in frame size sequence. The cause of this phenomenon is clear: a regular GOP structure, e.g., \( \text{GOP}(15,3) \).

Our work is motivated by the need for a good stochastic model for MPEG VBR traffic with a regular GOP structure. We develop a stochastic model, GOP ARIMA (autoregressive integrated moving average for a group of pictures), for VBR processes with a regular GOP pattern. “MPEG-2 video” is a typical example of this type of traffic. The GOP ARIMA model explicitly incorporates deterministic time-dependent behavior of regular GOP patterns and successfully represents the sample correlation structures of the underlying process. Our model can be used to generate synthetic MPEG streaming traffic for network performance studies and for resource allocation or buffer dimensioning in the server or network router or at the client. Using a synthetic traffic generator in a performance study has many advantages. Actual traces are not required for the study. Collecting a trace requires a significant amount of time and effort, and it may not even be available. Also, a synthetic model facilitates the generation of different kinds of traffic with similar stochastic structures.

The novelty of the proposed GOP ARIMA model lies in the fact that it elaborately incorporates the inter- and intra-GOP correlation structures of the empirical VBR process and provides a rigorous explanation of the observed stochastic characteristics of the empirical process: slowly decaying sample autocorrelations and cyclicity due to frame type. Since our model explicitly models the correlations at the seasonal lags, it can be effectively used to predict future bandwidth usage when combined with an adaptive optimization algorithm, e.g., Kalman filter, exponential smoothing, double exponential smoothing, linear prediction method, etc. This property is particularly valuable for online traffic smoothing or online resource allocation for live video feed.

1.2 Related work

Stochastic characterization of the workload plays an important role in designing the hardware and software components of a system [19]. A large number of studies have used the notion of long-range dependence (LRD) and successfully explained the observed characteristics of aggregated packet traffic [6, 24, 31, 42]. The results of source-level analysis [43] are also well aligned with the observed characteristics. However, when the time scale of interest is relatively small, it has been reported that the existence of the LRD property may be insignificant [12, 34]. Feldmann et al. [9] provide a more sophisticated model for network traffic that incorporates the multifractal behavior of the network traffic on a different time scale.

A fair number of studies have been devoted to developing a good model for compressed video traffic, and several stationary time series models, e.g., AR(1) [8], DAR(1) [15], ARMA [13], Markov renewal process [26], have been proposed. These models take into account the marginal distributions and the first few autocorrelations of video frame size. These models leave much to be desired with respect to properly representing the slowly decaying ACF (autocorrelation function) and cyclicity in the empirical process. The notion of self-similarity was then introduced to explain the slowly decaying sample autocorrelations in MPEG video dynamics [3, 11]. The notion of self-similarity requires that the underlying process be covariance stationary. Therefore, it cannot model the regular GOP structure of the MPEG VBR process. AR(1)-, DAR(10)-, and ARMA-based modeling approaches suffer from the same limitation. These time series models can only model a stationary process.

Layered modeling approaches were proposed to resolve the limitation of the above-mentioned time series model. Doulamis et al. [7] used the AR(1) process to represent the relationship between each type of frame in consecutive GOPs and added another layer to represent inter-GOP behavior. Turaga et al. [37] enhanced this model by incorporating doubly a Markov process to model irregular GOP patterns. The problem with these approaches is that the AR(1) process still has a geometrically decaying autocorrelation function and thus cannot properly reflect the slowly decaying autocorrelation structure. Krunz et al. [20] examined the size histograms and correlation structures of each frame type sequence. Rose [33] suggests separating scene layer, GOP layer, and frame sequence layer. He used a Markov model in generating scene length sequence and the GOP size sequence within a scene. Within GOP, constant scale factors are applied to obtain the size of individual frames. Lombardo et al. [25] modeled the frame size correlations within GOP and proposed generating MPEG traces using an algorithm that has long-range dependent properties.
at the GOP level. The drawback of the layering approach is that a certain amount of correlation information may get lost due to layering. For example, separating the GOP level (aggregation at GOP scale) modeling and intra-GOP frame level modeling may not represent the sample correlations at lags of 15 (GOP size). Improper characterization of the correlation structure can result in biased performance estimates.

Most of the above-mentioned modeling studies assume that the bandwidth process of VBR compressed video is covariance stationary, i.e., for time series $X = (X_t : t = 0, 1, 2, \ldots)$ with $\mu = E[X_t]$, the autocorrelation function

$$
\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{E[(X_t - \mu)^2]}
$$

depends only on $k$. However, this assumption does not hold in the frame level process with regular GOP patterns. Recent efforts on VBR traffic modeling have adopted the regularity in the GOP structure. Frey and Nguyen-Quang [10] proposed a nonstationary process called the GOP GEAR model. It is based on Heyman’s gamma–beta autoregression (GBAR) model [16]. The GOP GEAR model formulates the VBR traffic as the sum of a varying number of stationary GEAR models with respect to B, P, and I frame types. This model captures well the cyclicity of the VBR sequence. However, sample autocorrelations decay very fast in this model because the GEAR model has a geometrically decaying autocorrelation function. In addition, the GOP GEAR model requires that the variance of B frame sizes be less than the variance of P frame sizes and that the variance of P frame sizes is less than the size of I frame sizes. We find that this assumption does not hold in our sports video clip (Fig. 1). Including our earlier work, a number of studies have proposed to apply nonstationary models to represent or to predict the deterministic time-dependent behavior of underlying processes: VBR process [44], user behavior [14], and I/O workload [36].

An important application of traffic modeling is to predict the bandwidth behavior of the process and to use this knowledge in resource allocation [2], traffic smoothing, etc. Manzoni et al. [27] proposed to allocate the resources based upon the workload characteristics of I frame sequences. They used the aggregate characteristics, e.g., first- and second-order statistics of I frame size sequence. However, they did not consider the autocorrelation structure of the underlying sequence. A number of studies have proposed to smooth the bandwidth process by adjusting the source coding and channel coding scheme [22, 38] or by adjusting the packet transmission schedule [5, 23, 32]. Traffic smoothing has been shown to significantly enhance the quality of service (QoS) of a stream via empirical study [45]. In determining the source/channel coding scheme and packet transmission schedule, knowledge of future bandwidth requirements can greatly enhance the QoS of the resulting VBR process.

By explicitly incorporating the deterministic time-dependent behavior, we are able to successfully provide a robust and consistent explanation of the observed stochastic characteristics: slow decay and cyclicity in sample autocorrelations. The GOP ARIMA model can be used to generate various kinds of traffic with different bandwidth characteristics, i.e., different playback bandwidth and different degrees of interframe dependency, but with structural similarities.

This paper is organized as follows. In Sect. 2, we present a primitive statistical analysis of sample MPEG video traffic data and provide estimated Hurst parameters for later numerical comparisons. In Sect. 3, we present the GOP ARIMA model. In Sect. 4, we fit the GOP ARIMA model against three frame size sequences. In Sect. 5, we present the results of performance analysis. We compare the performance behavior of three different process models: empirical VBR process, synthetic process with GOP ARIMA model, and self-similar process generated by the fractional Gaussian model (FGN). Under a given single server queuing model with FIFO discipline, we examine the queuing behavior of each sequence using various performance measures: packet loss, frame corruption, and queue length. Section 6 concludes the paper.

2 Statistical analysis

2.1 Empirical MPEG traffic

The MPEG coding scheme exploits the temporal and spatial differences between successive frames. Initially, we examined a total of 11 video traces from 2 different encoders: 3 traces from encoder A and 8 traces from encoder B. The original video clips were carefully chosen to represent different degrees of interframe dependency. We primarily deal with the traces from encoder A and selectively introduce the analysis results of traces from encoder B. Table 1 summarizes the parameters for VBR video traces from encoder A. These video traces are publicly available at http://www.dmclab.hanyang.ac.kr/data/data.htm. The mean and variance of each empirical VBR process are summarized in Table 2.

![Fig. 1 Mean and standard deviation of each frame type, B, P, and I, for sample MPEG video traces](image-url)
Table 1 Parameters for generating VBR video traces: encoder A

<table>
<thead>
<tr>
<th>Stream type</th>
<th>MPEG2 elementary</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of frames</td>
<td>Drama VBR.mp4</td>
</tr>
<tr>
<td>Frame rate (length)</td>
<td>8869</td>
</tr>
<tr>
<td>Avg/min/max bandwidth</td>
<td>332 MB</td>
</tr>
<tr>
<td>GOP pattern</td>
<td>Cropped full D1, 704 x 480</td>
</tr>
</tbody>
</table>

Table 2 Mean and variance of frame-level empirical VBR processes

<table>
<thead>
<tr>
<th>Video</th>
<th>Mean (byte)</th>
<th>Variance (byte^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drama VBR</td>
<td>37537.35</td>
<td>182026701.19</td>
</tr>
<tr>
<td>News VBR</td>
<td>37603.86</td>
<td>212952461.24</td>
</tr>
<tr>
<td>Sports VBR</td>
<td>37508.96</td>
<td>146507073.78</td>
</tr>
</tbody>
</table>

Table 2 shows the basic statistics of the frame size sequences. We plot complete traces of three VBR compressed video streams in Fig. 3 for illustrative purposes. The average frame size is approximately the same for all three video clips, which corresponds to a 9-Mbits/s encoding rate. The news video clip has the largest sample variance and the sports video clip the smallest one. We suspect that this is because the news video clip has a higher degree of interframe dependency and thus the encoder can achieve a higher compression ratio for interframe coded frames (P or B). The nature of our football video clip lies at the other end of the spectrum. It has the smallest frame size variance. We cautiously conjecture that this is because football content is dynamic in nature, and therefore the temporal and spatial redundancy between consecutive frames is relatively low. In eight traces from encoder B (two news clips, one sports game clip, one infomercial clip, two drama clips, one lecture, and one comedy clip), we observed the same phenomenon: a sports (basketball) clip again yields the smallest variance in frame size sequence. Detailed statistics on these eight traces can be found in Appendix 2.

2.2 Nonstationarity of empirical VBR process

Figure 4a–c illustrates the frame size sequence on a sub-second time scale. Given that the video is encoded at 30 frames/s, the frames in these figures correspond to 3.5-s playback. A simple visual inspection shows that the frame sequence yields a strictly regular pattern: a big spike every 15 frames (I frame) and a medium-size spike every 3 frames (P frame). Individual spike sizes vary slightly. Nonetheless, this cyclicity of the spikes persists through the entire sequence across the three video clips. The notion of stationarity is pretty well defined in the context of a stochastic process. However, whether the given empirical process is stationary or not is a subject of debate. A certain empirical process can be stated as nonstationary only when it exhibits strict time-dependent behavior and only when there is a clear cause for it. In our test sequences, we observed deterministic and strict time-dependent behavior as in Fig. 4. Further, we can easily conjecture that the cause of this time-dependent behavior is a fixed GOP pattern.

Figure 5 illustrates a sample autocorrelation of the test sequences. The sample correlation structures in Fig. 5 exhibit multiplicative periodic characteristics. Our video sequences are compressed using the GOP(15,3) pattern. From the interframe distance between different types of frames we can find an explanation for the multiplicative periodic autocorrelations in Fig. 5. Similar characteristics were observed in [21]. There are large positive correlations in the lags of multiples of 15. This is due to the correlations between I frames. The medium-scale correlations that occur at the lags of multiples of 3 are due to the correlations between P frames or between P and I frames. As shown in Fig. 5, the sample autocorrelations decrease slowly with the lag. Many previous studies attempted to explain this phenomenon in the context of long-range dependence. A process $X_t$ is called a long-range dependent (or long-memory) process if a constant $C_\rho > 0$ and $0.5 < H < 1$ exist such that $\rho_k = \frac{E[(X_t-\mu)(X_{t+k-\mu})]}{E[(X_t-\mu)^2]} \sim C_\rho k^{2H-2}$, as $k \to \infty$. The parameter $H$ is called the Hurst parameter and is used to indicate the degree of long-range dependency. Informally, a given time series is said to be long-range dependent if the autocorrelations decay hyperbolically rather than exponentially fast and are therefore nonsummable, i.e., $\sum_{k=0}^\infty \rho_k = \infty$. The notion of long-range dependence requires that the underlying process be covariance stationary. However, as we observed, the VBR process with a regular GOP pattern is not stationary. It exhibits strict deterministic time-dependent behavior that stems from interframe dependency coding. In fact, slowly decaying sample autocorrelations can easily
Fig. 3 Empirical VBR process. a Drama. b News. c Sports. d Drama I, B, P. e News I, B, P. f Sports I, B, P

Fig. 4 Evidence of nonstationarity: sequence of frame size on smaller time scale. a News. b Drama. c Sports
be found in some nonstationary processes. Random walk \( W_t, W_t - W_{t-1} = \varepsilon_t \), where \( \varepsilon_t \) is an i.i.d. random variable with mean 0 and variance \( \sigma^2 \), is a typical example [4]. A periodic structure in the sample autocorrelations is widely observed in seasonal data, e.g., the sequence of monthly average temperatures.\(^2\) We cautiously believe that the nonstationarity (or, more precisely, cyclicity) of the underlying VBR process carries significant implications for its observed stochastic behavior, e.g., slowly decaying sample autocorrelations.

2.3 LRD property of empirical VBR process

A number of methods exist for estimating the Hurst parameter of a traffic sequence, e.g., \( R/S \) plot (rescaled adjusted range plot), variance time plot, periodogram-based MLE estimate. These are well summarized in [24]. Recently, a wavelet-transform-based method was proposed for estimating the Hurst parameter [1]. Hurst [17] showed that \( T^{-H} R_T \) converges to \( C \) in probability as \( T \to \infty \) for some constant \( C \), where \( R_T = \max_{1 \leq k \leq T} \left\{ \sum_{i=1}^{k} X_i - k \bar{X}_n \right\} - \min_{1 \leq k \leq T} \left\{ \sum_{i=1}^{k} X_i - k \bar{X}_n \right\} \) and \( S_T = \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X}_n)^2 \right\}^{1/2} \). We estimate the \( H \) parameter using the pox diagram of an \( R/S \) plot and a simple least-squares fit [43]. That is, for \( n = 10, 11, \ldots, T \), where \( T \) is the sample size, we subdivide all samples into nonoverlapping blocks with a sample size \( n \) and compute \( R_n \) and \( S_n \) for each block. The pox diagram of \( R/S \) is the plot of \( \log_{10} \frac{R_n}{S_n} \) versus \( \log_{10} n \).

Figure 6 shows the pox diagrams of the \( R/S \) plot of our VBR processes. The estimate \( \hat{H} \) of the Hurst parameter \( H \) is obtained by street’s asymptotic slope, which is typically

---

\(^2\) Seasonal processes are a special part of nonstationary processes.
obtained by a simple least-squares fit to \( (\log_{10}(n), \log \frac{R_n}{S_n}) \). In Fig. 6, \( H \)-parameter estimates of the three test sequences are all greater than 0.5. Judging solely from these \( H \)-parameter estimates, one may conclude that the empirical VBR traffic is self-similar. However, we observed that the underlying sample process exhibits strong cyclic behavior and therefore does not satisfy the condition of covariance stationarity. For this reason, an LRD-based modeling approach may not be feasible for providing a satisfiable explanation on the observed stochastic characteristics of the underlying VBR process. A good VBR traffic model mandates proper incorporation of inter- and intra-GOP frame correlations.

When we aggregate a sample sequence at the GOP level, the resulting sequence will most likely yield a stationary process. Short-term correlations in the frame sequence will be lost. We find that after being aggregated at the GOP level, the sample sequence still exhibits an \( H \)-parameter greater than 0.5. This is convincing evidence that the underlying sequence has an LRD property. Figure 7 illustrates the result of GOP-level aggregation. As has been shown, our VBR process has a very interesting nature. It exhibits both long-range dependency and nonstationarity. Each of these manifests itself on different time scales. Both the stationary and the nonstationary aspect of the underlying process should contribute to the observed slowly decaying sample autocorrelations. The question of which aspect of the two is more important is quite a subjective one and should be dealt with in a separate study. However, it is clear that the model that more closely simulates the behavior of the original process should be the model of choice.

### 3 GOP ARIMA: ARIMA model for groups of pictures

#### 3.1 Nonstationarity and the ARIMA model

A number of time series models have been used to represent the marginal distribution of the VBR frame size sequence: AR(1) [8], DAR(1) [15], and even ARMA [13]. The autoregressive moving average (ARMA) model is a generalization of the autoregressive (AR) and moving average (MA) models. While the ARMA model is a more powerful one than the MA or AR models, it still leaves much to be desired as a model for effectively modeling MPEG VBR traffic. Representing VBR traffic data using the ARMA process is suitable if data (i) exhibit no apparent deviations from stationarity and (ii) have a rapidly decreasing autocorrelation function. Unfortunately, there is strong evidence that our VBR process is not stationary. Also, many recent studies have shown that (i) does not hold, i.e., VBR traffic data have slowly decaying sample autocorrelations. Sample autocorrelations of our VBR sequences not only decay very slowly but also exhibit cyclicity with period I-to-I frame distance and P-to-P frame distance. These characteristics cannot be modeled properly with the ARMA model since it has a geometrically decaying autocorrelation structure and is stationary.

The autoregressive integrated moving average (ARIMA) is a generalization of the ARMA class and incorporates a wide range of nonstationary series. The ARIMA model is widely used to represent empirical data with slowly decaying sample autocorrelations [40] or data with seasonality, e.g., time series data of monthly fuel consumption. A process is an AREVIA process when it leads to an ARMA
Fig. 7 GOP level aggregation: $R/S$ plot for estimating the Hurst parameter of the empirical VBR process

process by differencing transformation. For example, random walk, $W_t = W_{t-1} + \epsilon_t$, where $\epsilon_t$ is a sequence of i.i.d. random variables with a mean of zero and a variance of $\sigma^2$, is an AREVIA$(0,1,0)$ process. This process is not covariance stationary and has slowly decaying sample autocorrelations [4]. Appendix 1 presents the basics of time series models: AR, MA, ARMA, and AREVIA models. Interested readers are referred to [4].

3.2 Decomposition and removal of cyclicity

In classical decomposition of the time series, sample $X_t$ consists of a seasonal component $s_t$, a trend component $m_t$, and a random noise component $\epsilon_t$, i.e., $X_t = s_t + m_t + \epsilon_t$. In our model, we exclude the trend component $m_t$. In VBR traces with a regular GOP structure, the seasonal component $s_t$ actually consists of two components: (i) a cyclic component related to P frame size and (ii) a cyclic component related to I frame size. In our empirical data, we subpartition the seasonal component $s_t$ into $s_{3,t}$ and $s_{15,p}$. We can decompose the sample process as in Eq. 1:

$$X_t = s_{3,t} + s_{15,t} + \epsilon_t,$$  \hspace{2cm} (1)

where $s_{3,t}$ and $s_{15,t}$ have nonzero values when $t$ is a multiple of 3 and 15, respectively.

In the first step, we need to remove the cyclic component in $X_t$ to transform the original process into a more analyzable one. In the ARIMA$(p, d, q)$ model, we take the difference of consecutive samples repeatedly until the differenced process looks stationary and until it can be formulated with the ARMA$(p, q)$ process. $d$ here corresponds to the difference order. Nonstationary components $s_{3,t}$ and $s_{15,t}$ appear in every 3 samples and in every 15 samples, respectively. To remove the cyclic components from the samples, we take the difference at a lag of 3 and 15. We introduce backward operator $B$, which is widely used to make the time series expression more concise. $B^kX_t$ denotes $X_t - X_{t-k}$. $(1 - B)X_t$ thus denotes the differenced time series $Y_t = X_t - X_{t-1}$. The differenced process at lags of 3 and 15 can be obtained by applying the $(1 - B^3)^d(1 - B^{15})^D$ operator, and $d$ and $D$ correspond to the difference order for their respective lags. The backward shift operator $B$ satisfies the distributive law. For example, differencing the original process twice is equivalent to applying the $(1 - B)^2$ operator to the original process. The important issue here is to find the appropriate values for $d$ and $D$. In applications, $d$ and $D$ are rarely greater than 1. A number of methods exist for finding the difference orders, $d$ and $D$, and this will be discussed in depth in Sect. 4.1. We can formulate the differenced process $Y_t$ as in Eq. 2:

$$Y_t = (1 - B^3)^d(1 - B^{15})^D X_t.$$  \hspace{2cm} (2)

Figure 8 illustrates how we remove the cyclicity from the original process and transform the original process into a more analyzable one. In Fig. 8, we take the difference at a lag of 15 and then take the difference at a lag of 3. We can take the difference multiple times for each lag. In practice, however, the difference order rarely goes beyond 1. These values are called the difference order and are denoted as $D$ and $d$ for a lag of 15 and a lag of 3, respectively.
3.3 Modeling inter-GOP correlations

To effectively model the apparent remaining correlations at the seasonal lags, i.e., multiples of 3 and 15, we analyze $Y_t$ from the point of view of inter- and intra-GOP correlations. We can organize $Y_t$ using the two-dimensional matrix in Table 3. Each column of Table 3 may itself be viewed as a realization of a time series. Our objective is to find a model that effectively represents the correlation structure of the samples within $S_i$ as well as between $S_i$s. The correlation structures of the samples within $S_i$ as well as between $S_i$s correspond to the inter-GOP and intra-GOP correlation structures, respectively. Since the samples in each $S_i$ are frames of the same type, we can safely assume that each $S_i$ is a stationary process. Samples in $S_i$ correspond to $Y_{i+15j}$, $i = 0, \ldots, 14$, $j = 0, \pm 1, \pm 2, \ldots$. We can model $S_i$s as in Eq. 3, which is based on the definition of the ARMA process of Eq. 16:

$$
Y_{i+15t} - \Phi_1 Y_{i+15(t-1)} - \cdots - \Phi_P Y_{i+15(t-P)} = \Upsilon_{i+15(t-1)} + \Theta_1 \Upsilon_{i+15(t-1)} + \cdots + \Theta_Q \Upsilon_{i+15(t-Q)}.
$$

(3)

where $i = 0, 1, \ldots, 14$, $t = 0, \pm 1, \pm 2, \ldots$ and $\{\Upsilon_{i+15t}\}$ is a white noise process with mean 0 and variance $\sigma^2_t$. Note that Eq. 3 is used to capture the correlation structure within $S_i$s, and $\Upsilon_i$s are uncorrelated to each other. Since the same $ARMA(P,Q)$ model is assumed to apply to each series, Eq. 3 can be rewritten for all $t$ as

$$
Y_t - \Phi_1 Y_{t-15} - \cdots - \Phi_P Y_{t-15P} = \Upsilon_t + \Theta_1 \Upsilon_{t-15} + \cdots + \Theta_Q \Upsilon_{t-15Q}.
$$

(4)

Equation 4 can be written in compact form as $\Phi(B^{15}) Y_t = \Theta(B^{15}) \Upsilon_t$, where $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P Z^P$ and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q Z^Q$.

3.4 Modeling intra-GOP correlations

It is unlikely that the consecutive $U_i$s, i.e., $U_i$ and $U_{i+1}$, would also be uncorrelated to each other. Possible nonzero correlations between consecutive $U_i$s imply nonzero correlations between the size of consecutive frames within GOP. In practice, samples in each row of Table 3 exhibit strong correlations at a lag of 3. This is caused by P-to-P or P-to-I frame distances. As a final step in the effort to model the nonstationarity of the VBR process, we need to establish the appropriate relationship between the consecutive $U_i$s in Eq. 4. This is equivalent to modeling the correlation structure within GOP where strong correlations exist at a lag of 3. We use the ARMA process with a lag of 3 to model $U_i$s.

---

**Fig. 8** Removing the seasonality from the original process
Equation 5 illustrates the ARMA process at a lag of 3:
\[
U_t - \phi_1 U_{t-3} - \cdots - \phi_p U_{t-3p} = \varepsilon_0 + \theta_1 \varepsilon_{t-3} \\
+ \cdots + \theta_q \varepsilon_{t-3q} + \varepsilon_t, \{\varepsilon_t\} \sim WN(0, \sigma^2).
\] (5)

Equation 5 can be written in a more compact form as
\[\phi(B^3)U_t = \theta(B^3)\varepsilon_t, \text{ where } \phi(z) = 1 - \phi_1 z^{-1} - \cdots - \phi_p z^{-p}, \quad \text{and } \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^{-q}.\]

3.5 GOP ARIMA model

Equations 4 and 5 represent an inter-GOP correlation structure and an intra-GOP correlation structure, respectively. Combining Eqs. 4 and 5, we can finally obtain the following equation for an empirical VBR process \(X_t\) with a regular GOP pattern:
\[\phi(B^3)\phi(B^{15})X_t = \theta(B^3)\theta(B^{15})Z_t, \{Z_t\} \sim WN(\sigma^2).\] (6)

We call the model in Eq. 6 the GOP ARIMA (ARIMA) for a group of pictures model. The GOP ARIMA model effectively incorporates the slowly decaying autocorrelation structure as well as the inter- and intra-GOP correlation structure of the empirical VBR process. We now introduce the general definition of GOP ARIMA.

**Definition 1** GOP ARIMA \((p, d, q)_n \times (P, D, Q)_S\) Process. Let \(S\) and \(s\) be P-to-P and 1-to-I frame distances, respectively. If \(d\) and \(D\) are nonnegative integers, then \(X_t\) is said to be a GOP-ARIMA \((p, d, q)_n \times (P, D, Q)_S\) process with period \(s\) and \(S\) if the differenced process \(Y_t(1 - B^d)(1 - B^S)X_t\) is a causal ARIMA process,
\[\phi(B^3)\Phi(B^S)Y_t = \theta(B^3)\Theta(B^S)\varepsilon_t,\] (7)
where \(\phi(z) = 1 - \phi_1 z^{-1} - \cdots - \phi_p z^{-p}, \quad \Phi(z) = 1 - \Phi_1 z^{-S} - \cdots - \Phi_P z^{-PS}, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^{-q}, \quad \Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^{-QS}\), \(\varepsilon_t\) is an independent and identically normal distributed random sequence with mean 0 and variance \(\sigma^2\).

4 Modeling the empirical process

4.1 Determining orders and parameters

The accuracy of the GOP ARIMA model depends on finding the right values for (i) difference orders \((D)\) and \(d\), (ii) autoregressive orders \((P)\) and \(p\), and (iii) moving average orders \((Q)\) and \(q\). We now introduce general guidelines for determining the orders and model parameters from the sample correlation function of the data. First, we find \(d\) and \(D\) so as to make the differenced observations \(Y_t = (1 - B^d)(1 - B^S)X_t\) stationary in appearance. Next, we examine the sample autocorrelation function at the lags of multiples of 15 in order to identify the orders \(P\) and \(Q\) in Eq. 4. If \(\hat{\rho}(k), k = 0, 1, 2, \ldots\) is the sample autocorrelation function of \(Y_t\), then the \(P\) and \(Q\) should be chosen such that \(\hat{\rho}(15i), i = 1, 2, \ldots\) is compatible with the autocorrelation function of an ARMA \((P, Q)\) process. The orders \(p\) and \(q\) are then selected by attempting to match \(\hat{\rho}(3i), i = 1, 2, \ldots\), with the autocorrelation function of an ARMA \((p, q)\) process. Then, we can estimate model parameters \(\Phi, \Theta, \phi,\) and \(\theta\) using the least-squares or maximum-likelihood method. Ultimately, the best model is identified by statistical goodness-of-fit tests. For details, refer to [4].

Although it is easy to understand conceptually, this method requires that a data analyst visually inspect the time series to determine the orders and parameters. In practice, the Schwartz Bayesian Criterion (SBC) is widely used to quantify the accuracy of the stochastic model. The model is said to be more accurate the smaller its SBC value. SBC is defined as in Eq. 8. Using SBC, the procedure for determining the orders and the parameters can be performed without human interaction:

\[\text{SBC} = \log(\hat{\sigma}^2) + (\gamma + 1) \log(n)/n.\] (8)

In Eq. 8, \(\gamma\) is the number of autoregressive and moving average parameters used in a fitted model and thus corresponds to \(P + Q + p + q\) in the GOP ARIMA model, \(n\) is the number of samples, and \(\hat{\sigma}^2\) is the maximum-likelihood estimator of \(\sigma^2\) (variance of \(\varepsilon_t\) in Eq. 6). \(\hat{\sigma}^2\) in Eq. 8 is the amount of error in a fitted model. The second term on the right-hand side of Eq. 8 is the penalty for using more parameters in the fitted model. With more parameters, we can obtain a more accurate model. However, using more parameters increases the computational complexity of the model and negatively affects the model efficiency. Thus, the number of parameters is included in the SBC as a penalty term.

In applying the SBC method, we first establish a set of GOP ARIMA models, \(\mathcal{L}\), as possible choices for the given empirical VBR process. \(\mathcal{L}\) is usually obtained by limiting the possible values of orders for \(d, D, p, q, P,\) and \(Q\). Actual values for \(\phi, \theta, \Phi, \Theta,\) and \(\sigma\) are estimated for each combination of the orders using a maximum-likelihood estimator (MLE). Once we obtain the values for \(\phi, \theta, \Phi, \Theta,\) and \(\sigma\), we can compute the SBC value for the respective model. From \(\mathcal{L}\) the combination of orders that yields the minimum SBC value is selected for the given empirical process. We establish \(\mathcal{L}\) as in Eq. 9:

\[\mathcal{L} = \{	ext{GOP ARIMA} (p, d, q)_n \times (P, 1, Q)_S : p, q, P, Q = 0, 1, 2, 3, d = 0, 1\},\] (9)

where 15 and 3 denote the I-to-I frame distance and I-to-P (or P-to-P) frame distance, respectively.

4.2 GOP ARIMA model fitting

We obtain a GOP ARIMA model for three frame size sequences: news, drama, and football game, which are from encoder A. Each video sequence contains 8000 samples,
which reflects 5 min playback. We use the SBC method introduced in Sect. 4.1 to determine the orders and parameters of the GOP ARIMA model. The fitted GOP ARIMA models for drama, news, and sports VBR processes are obtained as in Eqs. 10, 11, and 12, respectively:

[Drama VBR] \((1 - B^{15})(X_t - 10.511) = (1 - 0.644B^{15})\epsilon_t, \hat{\sigma}^2 = (5835.5)^2,\) \((10)\)

[News VBR] \((1 - B^{15})(X_t - 2.757) = (1 - 0.730B^{15})\epsilon_t, \hat{\sigma}^2 = (6680.3)^2,\) \((11)\)

[Sports VBR] \((1 + 0.036B^{15} - 0.033B^{30})(1 - B^3)\times (1 - B^{15})(X_t - 2.187) = (1 - 0.863B^{15})\epsilon_t, \hat{\sigma}^2 = (4853.0)^2.\) \((12)\)

Figure 9a–c illustrates empirical VBR processes and the synthetic sequences generated by fitted GOP ARIMA models. Synthetic sequences are generated using Eqs. 10–12. Visually inspecting the graphs in Fig. 9, we find that a synthetic process generated by the GOP ARIMA model well captures the deterministic time-dependent behavior of the empirical VBR process. Slowly decaying sample autocorrelations and strong sample correlations at the seasonal lags are all well reflected in our model via Eq. 6. Figure 10 illustrates the R/S plot of the GOP ARIMA generated samples. In theory, the Hurst parameter cannot be defined for the GOP ARIMA model since it is not a covariance-stationary process. Whether Hurst parameters are meaningless or not, we can still compute them. With the R/S plot, Hurst parameters of GOP ARIMA-generated processes are all greater than 0.5. Ignoring the fact that the underlying process is nonstationary, it may be categorized as an LRD process. The GOP ARIMA-generated process successfully captures the slowly decaying sample autocorrelation of the original process.

Figure 11 illustrates the sample autocorrelations of the original MPEG trace, the fitted GOP GEAR data [10], and the fitted GOP ARIMA data. The figures in the second row illustrate the sample ACF when the samples are aggregated at the GOP scale. In Sect. 2, we observe that when the original process is aggregated at the GOP level, the resulting process has an \(H\)-parameter value greater than 0.5. Without aggregation, all three sequences exhibit a similar correlation structure. However, when the samples are aggregated in GOP scale, autocorrelations of a GOP GBAR fitted process decay very fast as we pointed out in Sect. 1.2. Meanwhile, an ACF of GOP ARIMA fitted processes (aggregated at GOP level) decay as slowly as the ACF of the original samples.

4.3 The invariants

Traces from encoders A and B have identical GOP structures. From eight traces from encoder B we select five traces whose scene nature is similar to the traces from encoder A and compare the various characteristics. Even though the same compression standard is used, the stochastic characteristics of frame size sequence is highly subject to the given encoder and the nature of the scene. Thus, a GOP ARIMA
Fig. 10 Hurst parameter of GOP ARIMA fitted processes. a Drama. b News. c Sports

Fig. 11 Left to right: News MPEG video sample data, fitted GOP GEAR data, fitted GOP ARIMA data. a Autocorrelation of original data. b Autocorrelation of aggregated data with respect to GOP cycle
model fitted from one VBR trace most likely does not necessarily yield the best performance in other traces. To verify this property, we examine the mean squared error of the fitted GOP ARIMA model against two different video scenes. The GOP ARIMA model for Drama1 and Drama2 (Fig. 20a,b) are \((1 - B^{15})Y_t = (1 - 0.644B^{15})\epsilon_t, \sigma^2(\epsilon_t) = 2500^2\) and \((1 - B^{15})Y_t = (1 - 0.865B^{15})\epsilon_t, \sigma^2(\epsilon_t) = 2500^2\), respectively. We compute the MSE of the GOP ARIMA model for Drama1 and and Drama2, respectively, and the values are illustrated in Figure 12. The sum of MSEs for Drama1 are 60% larger. Appendix 2.2 presents the GOP ARIMA model for each frame sequence from encoder B.

While the orders and the parameters of individual GOP ARIMA models are all different, there exists a commonly observed phenomenon: there exists a strong relationship between the magnitudes of white noise variance in fitted GOP ARIMA models and the motion dynamics of the underlying video clip, \(\sigma^2\) in the GOP ARIMA model inversely proportional to the motion dynamics, i.e., the interframe dependency of the scene. Unlike drama and news video clips, sports video requires differencing at a lag of 3, i.e., \(d = 1\), to make the sequence stationary. Proper exploitation of these properties can enable us to generate different VBR traffic patterns with similar stochastic structures.

4.4 Modeling the different GOP structures

GOP ARIMA aims at modeling the intra- and intercorrelation structures, and thus it can model any regular GOP structure, e.g., GOP(6,3), GOP(9,3). The empirical data with GOP(6,3) and GOP(9,3) structure can be decomposed as \(X_t = s_{3,t} + s_{6,t} + \epsilon_t\) and \(X_t = s_{3,t} + s_{9,t} + \epsilon_t\), respectively. We compress a video scene in Fig. 13 using two different GOP structures, GOP(6,3) and GOP(9,3) (Fig. 14a,b). GOP ARIMA models for these frame sequences are \((1 - B^3)(1 - B^6)Y_t = (1 - 0.773B^3)(1 - 0.227B^6)\epsilon_t, \sigma^2 = (828.7)^2\) and \((1 - B^3)(1 - B^6)Y_t = (1 - 0.445B^3)(1 - 0.604B^6)\epsilon_t, \sigma^2 = (978.7)^2\).

5 Performance analysis

A good VBR source model should closely capture the network behavior of the empirical process. This is particularly important in predicting network performance, e.g., packet loss, delay, etc. We perform queuing simulation to examine how well the GOP ARIMA process closely models the original VBR process. We perform comprehensive queuing analysis with three input processes: (i) empirical VBR process, (ii) GOP ARIMA fitted process, and (iii) self-similar process that has the same \(H\) value, sample mean, and sample variance as the empirical VBR process from encoder A. A simple queuing system with a single server and finite queue is used in this experiment. This experiment enables us to evaluate the accuracy of the proposed model more quantitatively. Synthetic processes are generated using a GOP ARIMA model based upon Eqs. 10–12. We use fractional Gaussian noise (FGN) [30] in generating the self-similar process. The \(H\)-values of VBR processes for the drama, news, and sports video clips are 0.86, 0.74, and 0.97, respectively (Fig. 6). The unit of simulation time is frame time. The frame rate of our VBR process is 30 frames/s, and thus in our experiment, one simulation time unit corresponds to 30 s.

5.1 Packet loss

We first examine the packet loss behavior under different service times. Queue length is 100 (packets). Figure 15 illustrates the packet loss behavior. The X-axis denotes service time in the unit of frame time and the Y-axis denotes the packet loss probability. The GOP ARIMA process exhibits a very similar packet loss behavior to the empirical VBR process. Meanwhile, the self-similar process exhibits larger packet loss over all service time scales. In the drama and news clips, GOP ARIMA processes generate slightly fewer packet losses. In the sports clip, the GOP ARIMA process exhibits the same packet loss behavior as the empirical process. We include the log scaled plot for closer examination in Fig. 15.
5.2 Frame corruption behavior

A frame is said to be corrupt if one or more packets of the frame are missing. The VBR trace used in this work was obtained from DVD quality MPEG-2 video clips with a 9-Mbits/s playback rate. In the original trace, a single frame can be as large as 120 KB (I frame in most cases). This means that a single frame consists of more than 100 packets. In the course of transporting the entire frame from source to destination, it is inevitable that packets are exposed to delay, and loss and thus some of the packets may get lost. Advances in modern error concealment techniques and error-resilient coding algorithms [39] enable the streaming client to reconstruct the entire frame, even with an absence of some packets. However, it is not possible to avoid a certain degradation in the visual quality of the frame in the reconstruction process.
Fig. 16 Frame corruption behavior, queue capacity = 100 packets. a DramaVBR. b NewsVBR. c SportsVBR

Fig. 17 Normalized corruption factor (NCF), queue capacity = 100 packets. a DramaVBR. b NewsVBR. c SportsVBR

Fig. 18 Queue length distribution
process. In this regard, we examine the frame corruption probability. Figure 16 illustrates the frame corruption behavior of drama, news, and sports clips. The $X$-axis denotes the service time in the unit of frame time and the $Y$-axis denotes the $\text{P}(\text{corrupt frame})$. We observe in frame level analysis that the GOP ARIMA model closely captures the frame corruption behavior of the original empirical process. On the other hand, a larger number of frames get corrupt under a self-similar-process-driven queuing system. The level of quality degradation depends largely on the amount of lost packets in a frame. Even though the aggregate packet losses in two different streams are similar, it is possible that the burstiness of packet loss may exhibit quite different characteristics and eventually affect the perceivable quality of the stream in a different way. A uniform distribution of packet loss implies that a relatively larger number of frames are corrupt. On the other hand, if packet loss occurs in a burstier fashion with aggregate packet loss being the same, then the number of corrupt frames will be relatively smaller. We believe that there is a perceivable difference in the QoS depending on the burstiness of packet loss. Figure 17 reveals another level of detail of packet loss behavior. We introduce a new metric, $\text{NCF}$, normalized corruption factor, which is defined as $\text{P}(\text{corrupt frame}) / \text{P}(\text{packet loss})$. We normalize the $\text{P}(\text{corrupt frame})$ by packet loss probability. NCF can be thought of as the second-order information of intervals between consecutive packet losses. $\text{P}(\text{corrupt frame})$ being the same, the process with larger packet loss will have a smaller NCF $\text{NCF}$ and thus is inversely proportional to burstiness of packet loss. Figure 17 illustrates the NCF for three different types of video clips. Each figure plots the NCF of the empirical process, the GOP ARIMA process, and the self-similar process. Note that NCF values for the self-similar process is always smaller than NCF values for the GOP ARIMA and original processes. The GOP ARIMA process exhibits a burstiness of packet loss that is very similar to the empirical VBR process in all three of the test clips. This means that packet loss occurs in a burstier fashion in the self-similar-process-driven queuing system. Examining the FGN-based synthetic process, we found that it occasionally generates very bursty packet traffic and consequently is exposed to burstier packet losses. This phenomenon (occasional big frames) stems from the fundamental nature of the self-similar process, which is a stationary process with slowly decaying autocorrelations.

5.3 Queue length distribution

Loss and delay are two widely used metrics for multimedia QoS. To examine how the proposed model closely captures the delay characteristics of the empirical process, we examine the queue length distribution of the system. The queue length distribution is the probability that queue length is less than $x$, i.e., $\text{P}(Q \leq x)$, where $Q$ is the number of packets in the queue. $\text{P}(Q \leq x)$ denotes the fraction of simulation period during which the queue length is less than $x$. Figure 18 illustrates the result of our experiments. Service time and queue length are set to 0.02 and 100, respectively. We can see that the queue length distribution, when the system is fed by the self-similar process, is statistically larger than the others, i.e., $\text{P}(Q_{\text{SS}} > x) \geq \text{P}(Q_{\text{GOP ARIMA}} > x)$.
and \( P(Q_{SS} > x) \geq P(Q_{VBR} > x) \) for all \( x = 0, \ldots, 100 \). Therefore, an average queue length of the system with a self-similar input process is longer than the others. This experimental result confirms our observation in Sect. 5.2 that the self-similar process is actually burstier than the empirical VBR process.

5.4 Generation of the synthetic VBR process

GOP ARIMA explicitly models correlation structures of the underlying sequence. This property enables us to generate various VBR frame size sequences with structural similarities. We modify the parameters of the GOP ARIMA model in Eqs. 10–12 and build new models: Eqs. 13–15. Figure 19 illustrates synthetic traces generated by new models.

\[
\text{Synthetic Drama: } (1 - B^{15})Y_t = (1 - 0.644B^{15})\varepsilon_t, \sigma^2 = (2500)^2; \quad (13)
\]

\[
\text{Synthetic News: } (1 - B^{15})Y_t = (1 - 0.730B^{15})\varepsilon_t, \sigma^2 = (2500)^2; \quad (14)
\]

\[
\text{Synthetic Sports: } (1 + 0.036B^{15} - 0.033B^{30})(1 - B^3) \times (1 - B^{15})Y_t = (1 - 0.863B^{15})\varepsilon_t, \sigma^2 = (4800)^2. \quad (15)
\]

6 Conclusion

In this work, we focus our efforts on developing a stochastic model for VBR frame size sequence. The commonly observed characteristics of the VBR frame size process are slowly decaying sample autocorrelations and strong correlations at the seasonal lags. We exploit the fact that slowly decaying autocorrelations can be found in nonstationary (or, more precisely, cyclicity) process. We cautiously conjecture that the observed sample autocorrelation property of the VBR process can be due to the cyclic behavior of the underlying process. The cyclicity in the VBR process is due to the regular GOP pattern of the MPEG coding scheme.

The GOP ARIMA model elaborately harbors the inter- and intraframe sample correlation structures. The GOP ARIMA model effectively represents the multiplicative seasonal nature of the VBR process, i.e., cyclicity appears at a lag of 3 as well as 15. The GOP ARIMA model provides the physical explanation of the slowly decaying and periodic sample autocorrelations of the empirical VBR process. The synthetic sequence generated by the proposed GOP ARIMA model well captures the deterministic and time-dependent behavior of the empirical VBR process. It exhibits the same slowly decaying and multiplicative periodic autocorrelation structure that is seen in the empirical VBR process. The GOP ARIMA process closely models the original sample sequence in GOP scale aggregation. The queuing simulation results show that the GOP ARIMA process accurately captures the behavior of empirical processes in a number of respects: packet loss, frame corruption, and queue length behavior. From the simulation-based experiment, we found that packet loss occurs in a burstier fashion in the self-similar process even though it has the same mean and variance of the original VBR process. This suggests that the traffic generated by the self-similar process is much burstier than the empirical VBR processes, even though they have the same H values.

Stochastic characteristics of the VBR process can be analyzed from several standpoints: frame level autocorrelations, bit rate distribution, frame size distribution for each frame type, etc. These are different manifestations of the same property. A good stochastic model should be able not only to represent the characteristics of the VBR process but also to provide a rigorous explanation of their mutual relationships. The novelty of the GOP ARIMA model is that it can effectively capture the inter- and intraframe correlation structure and provide a robust explanation for the observed stochastic characteristics of empirical VBR traffic. Our work contributes not only to the theoretical explanation of the observed characteristics of the empirical VBR process with a regular GOP pattern but also to the development of an efficient method for generating a synthetic sequence for various engineering purposes, e.g., generating the various kinds of video traffic with structural similarities and predicting the bandwidth requirements of future VBR traffic.

Appendix 1: Synopsis: ARMA and ARIMA models

We revisit the basic materials of the time series that are required to understand the stochastic model used in this work. Let \( \varepsilon_t, t = \pm 1, \pm 2, \ldots \) be white noise with mean 0 and variance \( \sigma^2 \). White noise \( \varepsilon_t \) is a simple stochastic process whose mean is 0 and whose sample \( \varepsilon_t \) is independently and identically distributed. Since the samples in \( \varepsilon_t \) are independently and identically distributed, its autocovariance function \( r(h) \) is \( \sigma^2 \) if \( h \) is zero and is 0 otherwise. The autoregressive process of order \( p \), AR(p), is a stochastic process where each sample, \( X_t \), is an linear aggregation of past \( p \) samples plus white noise. Sample \( X_t \) in an AR(p) process can formally be written as \( X_t = \varepsilon_t + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \). In other words, in an AR(p) process, \( X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} \) is white noise \( \varepsilon_t \). Another basic time series is a moving average process. A moving average process with order \( q \), MA(q), is the stochastic process where \( X_t \) is the linear combination of \( q \) consecutive white noise samples, i.e., \( X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \). We can easily see the application of moving average processes in our daily lives, e.g., stock market trends, temperature trends, etc. The autoregressive moving average (ARMA) model introduced by Box and Jenkins can be regarded as a generalization of moving average processes and autoregressive processes and is represented as a linear aggregation of random shocks [4]. The process \( X_t, t = 0, \pm 1, \pm 2, \ldots \) is called an ARMA process with autoregressive order \( p \) and moving average
order $q$, denoted by $ARMA(p, q)$ if $X_t$ is stationary and satisfies Eq. 16 for every $t$:

$$
X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.
$$

(16)

We introduce the backward operator $B$, which is widely used to make time series expressions more concise. $B^k X_t$ denotes $X_{t-k}$. Using $B$, $\phi(B)X_t$ corresponds to $1 - \phi_1 B X_t - \cdots - \phi_p B^p X_t$ and subsequently is equivalent to $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$. Thus, the definition of ARMA process Eq. 16 can be written in a more compact form as

$$
\theta(B) X_t = \theta(B) \epsilon_t,
$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$. $\phi(z) \neq 0$ for $|z| \leq 1$. When $p = 0$, the $ARMA(p, q)$ process becomes a moving average process of order $q$. Similarly, when $q = 0$ in the $ARMA(p, q)$ process, it becomes an autoregressive process with order $p$. The ARMA process is often referred to as a short memory process since its autocorrelation decreases exponentially with lag. In fact, the autocorrelation function, $\rho(k)$, is geometrically bound in the ARMA process, i.e.,

$$
\rho(k) \leq C r^k,
$$

where $C > 0$ and $0 < r < 1$. Meanwhile, in the LRD process, $\rho(k)$ decreases polynomially, as mentioned in Sect. 2.2. It is worth noting that the LRD process, which has very slowly decaying autocorrelations, can always be approximated by the ARMA process. However, the orders $p$ and $q$ can be so large that achieving a reasonably good approximation can be practically infeasible.

The autoregressive integrated moving average (ARIMA) process is a generalization of the ARMA class that incorporates a wide range of nonstationary series. A process is an ARMA process when it leads to an ARMA process by differencing transformation. When the series is stationary only after differencing a number of times, the series is called an $I(d)$ series and $d$ is called the differencing order or integration order. The character $I$ in ARIMA denotes the integration. Backward operator $B$ can be used to represent the differenced process in concise form. Let $Y_t$ be the second-order differenced process of $X_t$, i.e., $Y_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$. Then, $Y_t = (1 - B)^2 X_t$.

The definition of ARMA processes requires three parameters: autoregression order $p$, differencing order $d$, and
moving average order \( q \). \( X_t \) is said to be an ARIMA\((p, d, q)\) process if its differenced process \( Y_t \) with difference order \( d \) is an ARMA\((p, q)\) process. This definition means that \( X_t \) satisfies a difference equation of the form 

\[
\phi(B)(1 - B)^d X_t = \theta(B) \varepsilon_t,
\]

where \( \phi(z) \) and \( \theta(z) \) are polynomials of degree \( p \) and \( q \), respectively. An ARMA model manifests itself in characterizing samples with a slowly decaying positive sample autocorrelation function.

### Appendix 2: Comprehensive frame sequence statistics

#### 2.1 Basic statistics

In this work, we examine an extensive set of frame sequence data: 11 sequences. In the main body of this paper, we present three representative results. This section presents the comprehensive results. Table 4 illustrates the basic statistics of the frame size sequences of various video clips, and a sample scene of each video clip is illustrated in Fig. 20. Each clip is approximately 6 min long and consists of about 11,000 samples. The home-shopping clip and sports clip have the largest and the smallest frame size variance, respectively. The infomercial video clip from the Home-Shopping Channel has a very static scene that consists of an almost still image on the sales item and commentary from the announcer. On the other hand, the sports video clip of a basketball game has very dynamic movement.

#### 2.2 GOP ARIMA fitting

Following are the GOP ARIMA fitted models for each video clip in Table 4:

\[
\text{News 1: } (1 - B^{15})X_t = (1 - 0.70297B^{15})\varepsilon_t, \quad \hat{\sigma}^2 = (4104.3)^2,
\]

\[
\text{News 2: } (1 - B^{15})X_t = (1 - 0.62471^{15})\varepsilon_t, \quad \hat{\sigma}^2 = (4193.3)^2,
\]

\[
\text{Drama 1: } (1 - B^{15})X_t = (1 - 0.88968^{15})\varepsilon_t, \quad \hat{\sigma}^2 = (3731.8)^2,
\]

\[
\text{Sports 1: } (1 - 0.242B^{15} - 0.0368B^{30})(1 - B^3)(1 - B^{15})
\]

\[\times X_t = (1 - 0.981B^{15})\varepsilon_t, \quad \hat{\sigma}^2 = (3388.3)^2. \quad (21)
\]

### Acknowledgements

Supported by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University. The work of Soohan Ahn was supported by the University of Seoul in 2004.

### References

35. Sony, Philips, Matsushita, and JVC. While Book VCD spec. (1993)