International Conference on
Information, Communications and Signal Processing
ICICS '97
Singapore, 9-12 September 1997

Characterization and Source Modeling of MPEG-2 VBR Video Source

Wenjing, Lou   Liang Tien, Chia   Bu Sung, Lee

Division of Computing Systems, School of Applied Science
Nanyang Technological University
Nanyang Avenue, Singapore 639798
Republic of Singapore

Abstract

This paper presents the results of our study on the statistical characterization of MPEG-2 VBR video stream and the modeling of such source. The frame trace, marginal distribution of the frame size and the high autocorrelation existing in the continuous frames are analyzed. An ARMA process is used to model MPEG-2 video source. The different frame types of MPEG-2 video are considered together by performing a normalization transformation during modeling. The performance of the model is examined and the results show that the model is good in fitting the histogram and preserving the property of autocorrelation. We also indicate that the proper order of the ARMA model depends on the frame pattern of the MPEG source.

1 Introduction

Traffic characterization and source modeling of Variable Bit Rate (VBR) coded video are active research areas because VBR video traffic would be a major media in future B-ISDN. Proposed statistical models for compressed video source have fallen into two main categories: Auto Regressive (AR) and Markov. AR model has been shown to produce good results in capturing the bitrate statistics at the picture layer for video conferencing type video with little motion. And, in general, it appears to capture the autocorrelation behavior of compressed video source well which is an important prerequisite for any model of compressed video source [4][5]. Markov chains provide a compact way of generating the probability distribution function which fits the video data well. The process of calculating the state transition probabilities is straight-forward. It does appear that one can model video at lower layer than the picture layer using a Markov chain. Markov chain also tends to capture the correlated behavior of the data well [6][7].

It has also been recognized that the MPEG based video compression algorithms will play a very important role in future B-ISDN [1][2][3]. However, little work has been focused on modeling MPEG-2 video source, especially source with all the three types of pictures. Therefore, in this paper, we present our study on the statistical characteristics of MPEG-2 video stream and the modeling of such video source.

2 Statistical Characteristics of MPEG-2 VBR Video Stream

2.1 Description of the Video Sequences

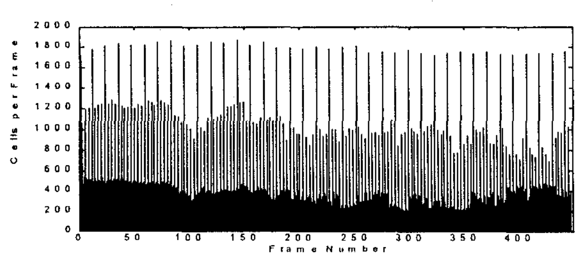
The test sequences used in this research are two MPEG-2 VBR video sequences which were encoded by a modified software MPEG-2 encoder using open loop coding mode. The two sequences are Mobl and Susi. Mobl scene shows a toy train slowly moving across the picture from the right to the left and knocking into a rolling ball. The background of the scene is complex and changing smoothly with the motion of the train. Susi sequence shows a young lady making a call. It consists of head and shoulder scenes with a uniform background and much less activities between frames.

Each sequence contains a total of 450 frames and has a frame rate of 30 fps (frame per second). The Mobl sequence was encoded with the quantization scale, q=12 for I and P frames and 16 for B frames. The Susi sequence was encoded with the quantization scale, q=4 for I and P frames and 5 for B frames. The selection of the above quantization scale is to maintain the average bit rate of about 8 Mbps. And the reason why the Q factor for B frames is not as fine as that of I or P frames is because B frames are never used as reference frames. A reasonable proportion for $Q_I:Q_P:Q_B$ is 1:1:1.4. Finally, the GOP pattern is IBBPBBPBBPBB (12 frames per GOP) for both sequences.

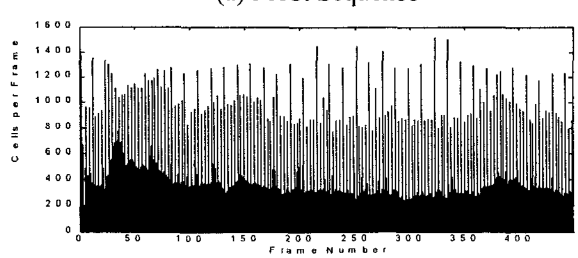
A program was written to parse the MPEG-2 video streams according to the MPEG syntax. The program scans the file for the Start Code Prefix consisting of a three byte 001 hex value followed by a one byte Start Code. Once the Start Code is found, the Start Code header is decoded and pertinent information within the header is written to an output file. Information extracted from the sequence header includes the image resolution, frame rate and bit rate, and so on. The number of bytes in each frame is also calculated and converted to number of cells, and this number is stored in the output file.

2.2 Frame Traces

Figure 1 presents the frame size traces of the Mobl and Susi sequences. The lines with different heights correspond to the different types of frames. The highest lines correspond to the biggest I (Intra-) frames. The midsize group of lines corresponds to the P (Predictive) frames. And the shortest ones correspond to B (Bidirectionally-predicted) frames. It is noticed that in Mobl sequence, the trace is rather smooth and similar at different time interval. While in the Susi sequence, there are irregular short traces among the whole plot. It is assumed that the similarity of the traces is due to the stationarity of the sequence and the stationarity may be assumed for smooth pan and zoom scenes such as Mobl. On the other hand, scenes containing sudden scene changes or fast camera panning or fast movement of the objects will cause a short irregular trace with bigger B or P frames. In Susi sequence, the bigger B frames between frame 40 and 80 are due to the fast head and eye movement because additional data has to be encoded to correct for the prediction errors during interframe coding.



(a) Mobl Sequence

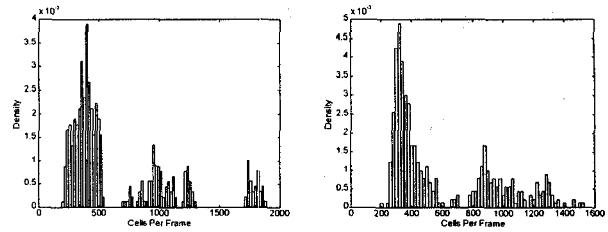


(b) Susi Sequence

Figure 1 Cells Per Frame Time Series

2.3 Distribution

Figure 2 presents the probability distribution function (pdf) of the frame size for the Mobl sequence and the Susi sequence. The distribution of frames among I, P, and B types and the relative average size of each frame type is shown.



(a) Mobl Sequence (b) Susi Sequence

Figure 2 Distribution of Cells Per Frame

2.4 Correlation

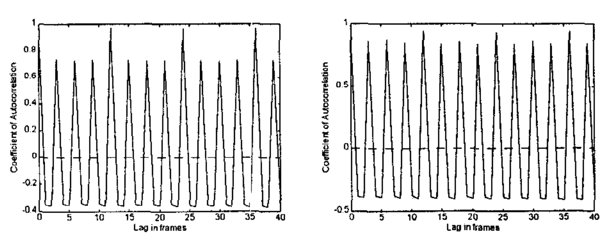
The number of bits generated in a MPEG frame is correlated to its previous frames due to the structure of the frame sequence and the continuity of the scene. The autocorrelation is defined as

$$\rho_k = E\left[\frac{Z_t - \mu}{\delta} \cdot \frac{Z_{t-k} - \mu}{\delta}\right],$$

where E[] is the expectation, μ is the mean of series Z, and δ^2 is the variance of series Z.

Autocorrelation function is an important time-dependent statistics in the case of video traffic, because correlation of the data streams can be utilized to improve performance of an ATM network.

The autocorrelation function of the frame size is presented in figure 3. The frame-by-frame correlation depends on the pattern of the GOP, and, in principle, always looks like figure 3. The larger positive peaks stem from the I frames, the smaller positive ones from the P frames, and the negative ones from the B frames. This shape reflects the relationship of the mean frame sizes of the frame types. A large I frame is followed by two small B frames. Then a midsize P frame is produced by the encoder, which is followed by two small B frames again. The pattern between two I frame peaks is repeated with slowly decaying amplitude of the peaks.



(a) Mobl Sequence (b) Susi Sequence

Figure 3 Autocorrelation Function of the Frame Size

3 Modeling MPEG-2 Video Source

3.1 Normalization of the Frame Size

The frames in an MPEG video source can be grouped into three types, I frames, P frames, and B frames. These three kinds of frames have obviously different size. A MPEG-2 video stream usually contains the mixture of the three types with a periodical frame sequence structure. With the distribution of frame size in the sequence shown in figure 2, it is impossible to find a distribution which fits the shape well. Therefore, in order to compare the sizes of different frame types on a fair level, a normalization procedure is applied to all I frames and P frames so that all the frames have equal mean size.

The normalization procedure is a linear transformation and can be described as follows:

Let

$\overline{Z}_I$ denote the mean size of all I frames in the sequence

$\overline{Z}_P$ denote the mean size of all P frames

$\overline{Z}_B$ denote the mean size of all B frames

$\hat{Z}(i)$ be the original size of the ith frame in the sequence

$Z(i)$ be the size of the ith frame after the normalization.

Define

$$S_{IB} = \frac{\overline{Z}_I}{\overline{Z}_B}, \qquad S_{PB} = \frac{\overline{Z}_P}{\overline{Z}_B}.$$

The following linear transformation is applied to each frame in the sequence.

$$Z(i) = \frac{\hat{Z}(i)}{S_{IB}} \qquad \text{if the } i\text{th frame is a I frame}$$

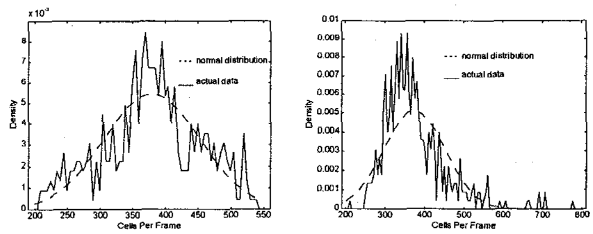
$$Z(i) = \frac{\hat{Z}(i)}{S_{PB}} \qquad \text{if the } i\text{th frame is a P frame}$$

$$Z(i) = \hat{Z}(i) \qquad \text{if the } i\text{th frame is a B frame}$$

After the normalization, the number of cells per frame in the whole sequence is in a comparable level (In order to simplify the description, the size after normalization is still called as the frame size or the number of cells per frame in this section).

3.2 Distribution

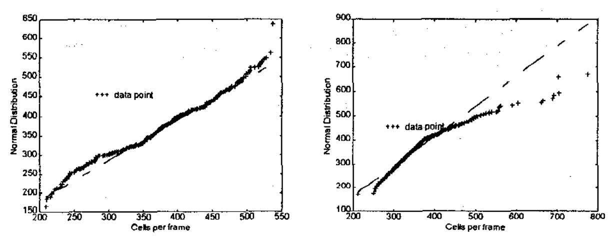
It is assumed that the data series of the frame size after normalization can be described by a normal distribution. To test whether the assumption is true, the histogram of the above series of frame size is examined. The histograms of the number of cells per frame for Mobl sequence and Susi sequence are shown in figure 4. An approximate normal distribution is also presented in the figure for comparison. It is observed that the distribution of the frame size fits the normal distribution fairly well.



(a) Mobl Sequence (b) Susi Sequence

Figure 4 Histogram of Cells Per Frame vs. Normal Distribution

To test whether the marginal distribution of the frame size is indeed a normal distribution, a Q-Q plot which plots the quantiles of the data vs. the quantiles of the fitted distribution is also used. The Q-Q plot is a powerful goodness-of-fit test. Figure 5 shows the Q-Q plot of number of cells per frame for the test sequence and their approximate normal distribution. The fit is fairly good except for a few points. Thus, conclusion drawn is that frame size in the test sequences, Mobl and Susi, can be described by a normal distribution after the normalization process.



(a) Mobl Sequence (b) Susi Sequence

Figure 5 Q-Q Plot of Cells Per Frame vs. Normal Distribution

Here we should point out that a linear transformation does not change the distribution type of a series if it follows a normal distribution. The linear transformation on the size of I and P frames does not change its distribution type and the conclusion on the distribution of frames after normalization is reasonable according to this mathematical theorem.

3.3 ARMA Model

Auto Regressive Moving Average (ARMA) process is widely used to model video source because it can effectively characterize the autocorrelation property of the video scenes. Basically, an ARMA model of orders p and q, denoted as ARMA(p,q), is defined as

$$Z(k) = \phi_0 + \phi_1 Z(k-1) + \phi_2 Z(k-2) + \cdots + \phi_p Z(k-p)$$
$$+ e(k) - \theta_1 e(k-1) - \theta_2 e(k-2) - \cdots - \theta_q e(k-q)$$

where $Z(k)$ is the size generated in the kth frame and $\phi_0, \phi_1, \phi_2, \cdots, \phi_p, \theta_1, \theta_2, \cdots, \theta_q$ are the coefficients of the ARMA process. It is a recursive procedure which generate a series of values. The current value of the series is a linear combination of the p most recent past values of itself plus a linear combination of the q most recent values of series e, which is a Gaussian random process. $\phi_0, \phi_1, \phi_2, \cdots, \phi_p, \theta_1, \theta_2, \cdots, \theta_q$ are constant coefficients and are derived empirically.

The autocorrelation property is dominated by the first few order ARMA. ARMA(1,1), ARMA(2,2), ARMA(3,3), and ARMA(4,4) are studied, and the results show that ARMA(3,3) reflects the autocorrelation property of the frames much more approximately than ARMA(1,1) and ARMA(2,2) while there is no significant improvement between ARMA(3,3) and ARMA(4,4). This may be because the distance between any two I or P frames is 3. Higher order ARMA model will introduce complexity in deriving the parameters but it does not improve by much the autocorrelation property. Thus, in this study, ARMA(3,3) is selected to model the MPEG2 video source.

Based on the frame size of the sequences, the coefficients are estimated and the ARMA equations are derived as

$$Z(t) = 13.1368 + 0.0486Z(t-1) + 0.0327Z(t-2)$$
$$+ 0.8841Z(t-3) + e(t) + 0.1772e(t-1)$$
$$+ 0.0438e(t-2) - 0.3203e(t-3)$$

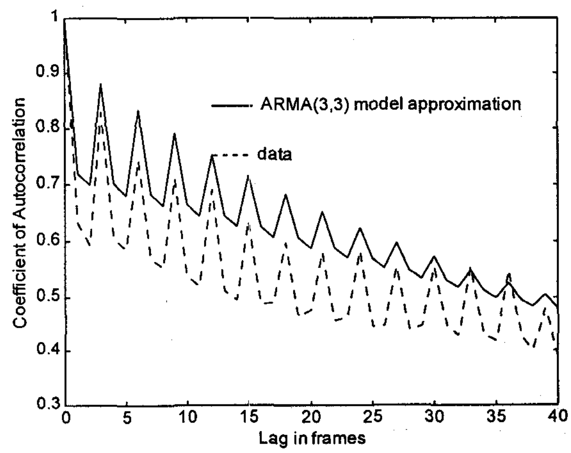
for Mobl sequence, and

$$Z(t) = 27.9223 + 0.0871Z(t-1) + 0.0277Z(t-2)$$
$$+ 0.8106Z(t-3) + e(t) - 0.0827e(t-1)$$
$$- 0.1484e(t-2) + 0.3482e(t-3)$$

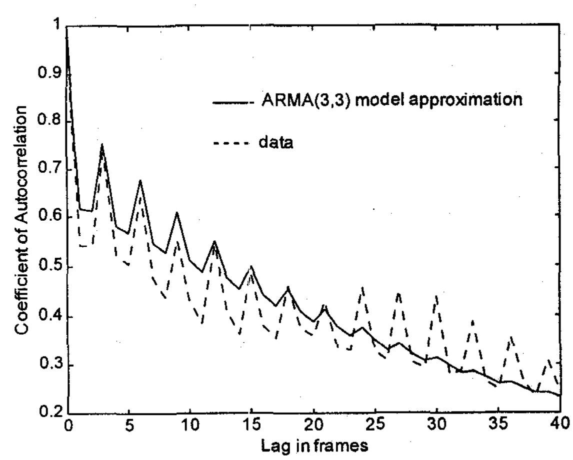
for Susi sequence.

3.4 Verification of the Model

To verify the validity and accuracy of the model, two important statistical characteristics, distribution and autocorrelation, of the model are compared with that of actual data.. The theoretical distribution of the ARMA model is a normal distribution whose parameters, mean and variance, approximate the parameters of the actual data distribution very well. The autocorrelation characteristics of the test sequences and the theoretical autocorrelation curve of the model are depicted in Figure 6. Ripples in the graph are due to the coexistence of different frame type. It can be observed that the model preserves the autocorrelation of the original data quite well, especially when the autocorrelation with a lag of less than the GOP size (12 or 15). The autocorrelation with a lag of greater than the GOP size is not very important because a new I frame is generated.



(a) Mobl Sequence



(b) Susi Sequence

Figure 6 Autocorrelation of Frame Size and ARMA(3,3) Model

4 Conclusion

The characteristics of VBR video traffic is very complex and it is not easy to derive a suitable analytic model. In this paper, we present the results of our study on the statistical characterization of MPEG-2 VBR video source. The frame trace, marginal distribution of the frame size and the high autocorrelation existing in the continuous frames are studied. Based on these characteristics, an ARMA model for MPEG2 video source is proposed. The different frame types of MPEG2 sources are considered together by performing a normalization transformation during modeling. The performance of the model is examined and the results show that the model is good in fitting the histogram and preserving the property of autocorrelation. We also indicate that the proper order of the ARMA model depends on the frame pattern of the MPEG source.

References

[1] D.Le.Gall, *MPEG: A Video Compression Standard for Multimedia Applications*, Communications of the ACM, vol.34, No.4, pp.47-58, April 1991

[2] P.N.Tudor, *Tutorial: MPEG-2 Video Compression*, Electronics & Communication Engineering Journal, Dec. 1995, pp.257-264

[3] *Generic Coding of Moving Pictures and Associated Audio: Video. Recommendation ITU-T H.262, ISO/IEC 13818-2,* November 1994

[4] M.Nomura, T.Fujii, and N.Ohta, *Basic Characteristics of Variable Rate Video Coding in ATM Environment*, IEEE J. on Selected Areas in Communications, vol.7, No.5, pp.752-760, June 1989

[5] Reto Grunenfelder, John P. Cosmas, et al, *Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queuing System Performance*, IEEE J. on Selected Areas in Communications, vol.9, No.3, pp.284-292, April 1991

[6] D.P.Heyman, T.V.Lakshman, *Source Models for VBR Broadcast-Video Traffic*, IEEE/ACM Trans. Networking, pp.40-48, 4 Feb, 1996

[7] D.P.Heyman, Ali Tabatabai, and T.V.Lakshman, *Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks*, IEEE trans. on circuits and systems for video technology, vol.2, No.1, March 1992