Modeling of Dynamic Video Traffic

Deepak Turaga and Tsuhan Chen
Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Abstract— We propose a simple two-state Markov modulated AR process to model the traffic created by a variable bit rate video coder. This model performs better than currently proposed models, especially for video coded with a variable group-of-pictures (GOP) structure.

I. Introduction

Many advances in visual media, network technologies and signal processing techniques have made multimedia communication possible. In particular, packet video refers to digitized and packetized video transmitted over networks in real time. Packet video sources may be classified into Constant Bit Rate (CBR) and Variable Bit Rate (VBR) sources. CBR video coding involves generating a fixed bit rate over time, while VBR video coding allows for generating a variable bit rate over time. VBR coding is more desirable for the users of video applications as it can guarantee constant video quality over time. Hence the focus of this paper is on VBR video sources.

Estimating the network resources for different types of traffic is necessary to provide a certain QoS guarantee to the users. These parameters may include cell loss probabilities and end-to-end delays. It is more difficult for a network to guarantee a certain QoS with VBR video than it is with CBR video. Hence in order for the network designers to estimate the network parameters to guarantee a QoS for VBR video, it is essential that there exist accurate models for VBR video traffic. These models need to approximate the statistical properties and traffic characteristics for different kinds of video sequences.

Modeling VBR traffic is difficult because the bit rate is determined by a large number of factors. Different compression schemes can lead to different output bit rates for the same video sequence. It is necessary to choose a compression scheme before attempting to model the VBR traffic. The popular standards defining compression schemes today are the ISO MPEG Series and the ITU H.26x Series, with MPEG-4 and H.263 Version 2 being the latest versions [1,2]. These standards allow for three different kinds of coding schemes for a video frame in order to improve coding efficiency. A frame may be Intra (I), Predictive (P) or Bidirectionally-predictive (B). An I frame is coded in isolation from other frames using transform coding, quantization and entropy coding. A P frame is predictively coded, which means that a prediction is formed using a previously coded frame and only the difference between the prediction and the actual frame is coded. A B frame is predicted bidirectionally, which means that the prediction is formed using both its previous frame as well as the successive frame. An I frame is used to efficiently code frames corresponding to scene changes, i.e. frames that are different from preceding frames and cannot be easily predicted. Frames within a scene are similar to preceding frames and hence may be coded predictively as P or B for increased efficiency. Groups of frames between two successive I frames are collectively called a group of pictures (GOP). The work in this paper focuses on modeling explicitly video traffic consisting of I and P frames.

Several models for VBR video traffic have been proposed in literature. These models do not model the I and P frames explicitly, but deal with the modeling of the interframe predictive coded video traffic, which are like a sequence of P frames only. Maglaris et al [3] have proposed a model for the coding bit rate of a single video source using interframe predictive coding. Sen et al [4] propose models for different activity levels using correlated Markov models and use queuing analysis to estimate the packet loss and delay. Yegenoglu et al [5] propose a model for VBR video using a time dependent Autoregressive (AR) model to represent data from different activity levels.

The work by Doulamis et al [6] tries to explicitly model the I, P and B frames with an additional layer corresponding to the activity level of a video scene. In this paper we refer to this work as the Doulamis model. This work, however assumes a fixed GOP structure, i.e., each GOP consists of a fixed number of P and B frames in a fixed pattern following an I frame. This model is not appropriate for most video sequences, as the video content does not necessarily follow such a pattern. For instance scene changes or large changes in video content do not occur regularly, and hence the need for I frames in most video sequences is not at regular intervals. Our proposed model can model variable GOP structures, as against the Doulamis model.

This paper is organized as follows. Section II describes the Doulamis. Section III describes our proposed model, also

Work supported in part by Institute of Information Industry and by Texas Instruments University Research Program.

called the Two-State I and P model. Section IV deals with some experimental results when both the models are used to model data generated using a variable GOP structure. Section V consists of the Conclusion and future work.

II. Doulamis Model

This model assumes a fixed GOP structure where each GOP may be represented as IBBPBBPBBPBBP with each I frame followed by twelve B and P frames. Each GOP is classified as belonging to one of three classes, high activity, medium activity and low activity. This classification is made based on the average bit rate per frame during the GOP. If the total number of frames in the sequence is N_F and the number of frames in a GOP is L the average bit rate for the n_G th GOP x^G may be written as

$$x^G(n_G) = \frac{1}{L}\sum_{i=0}^{L-1} x^F(n_G L + i),\ n_G = 0..\frac{N_F}{L}$$ where

$x^F(i)$ is the bit rate for frame *i*.

The signal x^G is used to classify the GOP into one of the three activity bands, high, medium or low. GOPs with x^G greater than a threshold T_H are classified as high activity, those with x^G less than a threshold T_L are classified as low activity and the rest are classified as medium activity. The thresholds T_H and T_L are obtained using the mean and the standard deviation of x^G. For example $T_H = \mu^G + \lambda_H \sigma^G$ where λ_H is an empirically determined parameter. The thresholds are chosen such that an autoregressive AR(1) process models the temporal behavior of the autocorrelation function of the frames inside the GOP. The video activity level of a GOP can be modeled as a Markov chain whose states correspond to high, medium and low activity states.

Once a GOP is classified as low, medium or high activity, then frames inside a GOP are generated using an AR(1) model. Hence for each of the three states three AR(1) models are required, one each for the I, P and B frames. They further maintain that the exact behavior of P and B frames in medium and low activity states does not play a significant role to traffic behavior and hence each of these may be represented by their means. This helps in reducing the complexity of the model and the total number of parameters required for this model is 21. Their model may be pictorially represented as in Figure 1.

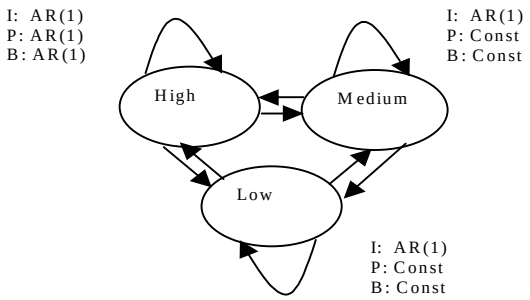

Figure 1. Doulamis model

The figure shows the two-stage model with the first comprising three activity levels and then within each level a fixed GOP structure is assumed following which each frame in the GOP is modeled using either an AR(1) process or replaced by its mean value.

III. Two-State I and P Model

The Doulamis model is useful in modeling different activity levels, but it imposes a limit on its flexibility by assuming a fixed GOP structure in every activity level state. Besides this, in order to reduce the complexity of the model, the P and B frames in the low and medium activity states are replaced by their mean values.

We propose a model for video sequences that consist of only I and P frames that is extremely simple, but still flexible enough to allow for variable GOP structure. Our model consists of only two states, one corresponding to I pictures and the other corresponding to P frames. The model transitions between these states with probabilities based on the training data, with no constraint imposed on a fixed GOP structure. So we can, in effect model data from a variable GOP structure.

Inside each state, to model the long-term temporal correlation between frames, the I frames are generated using an AR(1) process and the P frames are generated similarly. The parameters for the AR processes are estimated from the training data. Clearly, the parameters needed to specify the model are the four transition probabilities and parameters for each AR(1) process (mean, variance and parameter ρ). Thus the total number of parameters for this model is ten. If we remove B frames from the previous model, the total number of parameters reduces to 16. Hence our model is indeed simpler than the previously proposed model. Our model may be pictorially represented as in Figure 2.

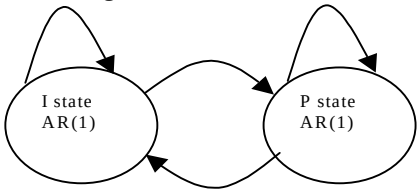

Figure 2. Two-state I and P model

The figure shows the proposed simple two-state model, with one state corresponding to I frames and the other state corresponding to the P frames. No constraints are imposed on the GOP structure and inside each state the frames are generated using an AR(1) process.

IV. Experimental Results

In order to compare the two models we modified the Doulamis model to remove B frames from it. The modified model still has activity states and a fixed GOP structure, but each GOP consists only of I and P frames.

The data we used for training was from a video sequence made up of advertisements. We call this sequence Ads. Hence there were frequent scene changes, camera zooms and pans and large motion. A sample frame from the sequence is shown in Figure 3.

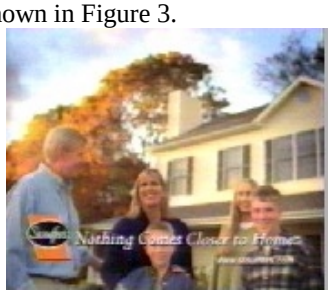

Figure 3. Sample frame from Ads

This sequence consisted of five minutes of data sampled at 15 Hz, making a total of 4500 frames. The sequence was converted to bits using a H.263 standard compliant video codec. A random GOP was achieved by inserting I frames whenever there was a great change in video content. Predictive coding in H.263 allows for individuals blocks (also called macroblocks) in a P frame to be intra coded. This happens when a good prediction for the block cannot be found. If the number of such blocks in a P frame is bigger than a threshold, it indicates a great change in video content and hence the frame is labeled as an I frame. This labeling is appropriate as a P frame with a large number of intra coded blocks will have a bit rate corresponding to an I frame.

Both models were trained using this data with the variable GOP. In order for a fair comparison, the transition probabilities between activity states were computed using the actual GOP structure and not a fixed GOP structure. Similarly parameters for I and P frames in each activity level were estimated using the actual GOP structure. Once the training was complete, a fixed GOP structure constraint was applied on the Doulamis model and data was generated. The GOP size was chosen to be the mean of the variable GOP size for the data. A sample trace of the actual video data looks as in Figure 4.

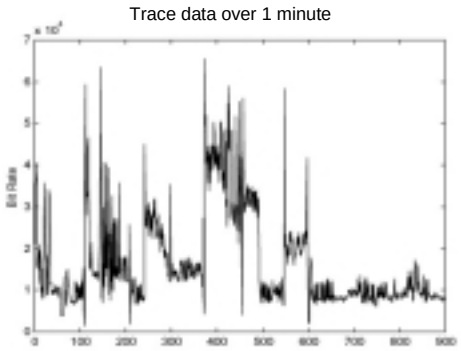

Figure 4. Trace of actual bit rate

The estimated autocorrelation function (R_{xx}) and the Power Spectral Density (PSD) are shown in Figure 5.

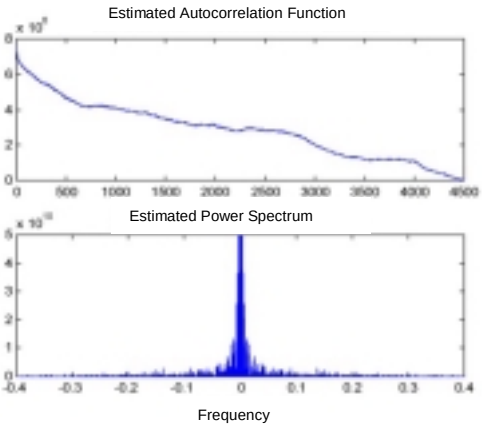

Figure 5. R_{xx} and PSD for real data

The trace generated by the Doulamis model shows signs of periodicity, because I frames are repeated regularly. This is evident when we look at the estimated R_{xx} and PSD as in Figure 6.

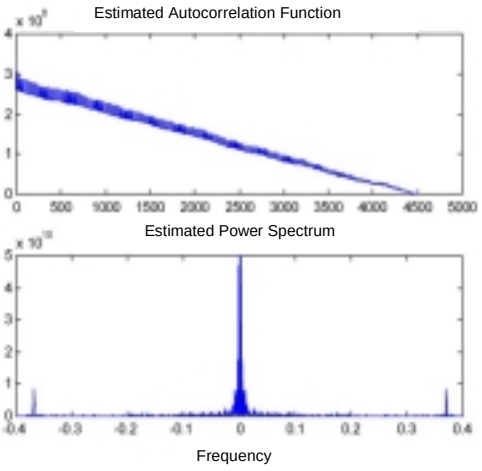

Figure 6. R_{xx} and PSD for Doulamis model

We can look at both the R_{xx} and the PSD for the data generated by the Two-State I and P model and this is shown in Figure 7.

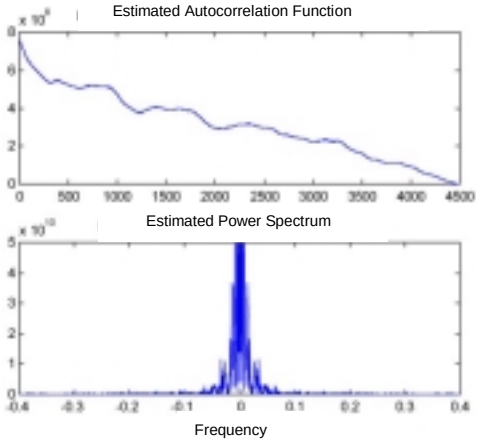

Figure 7. R_{xx} and PSD for Two-state model

In all of the plots shown, PSD is plotted only between $\left[-\frac{\pi}{8}, \frac{\pi}{8}\right]$ and the magnitude is restricted to at most 5×10^{10}. This is done to highlight the spike in the Doulamis model PSD, that occurs around $2\pi/17$, corresponding to a fixed GOP size of 17. The mean squared error between the estimated autocorrelation function for the Two-state model and the real data is smaller than the corresponding error for the Doulamis model on the average by 54%. From the plots and the mean squared error in estimating the R_{xx} we can see that that the Two-state I and P model is better when we wish to model data generated with a variable GOP size.

V. Conclusion and Future Work

The requirement of a variable GOP size is felt in typical video sequences. Most models proposed so far, however, have either not explicitly considered a GOP structure or modeled only data with constant GOP structure. Here we propose a simple Two-state I and P model that is flexible enough to model data with a variable GOP structure. The model performs better than a fixed GOP model in terms of modeling the autocorrelation function and the power spectral density of the generated data. Further tests are needed to show that the data also provides a more accurate estimate of delay and packet loss over networks. Some additional work for the future includes extending the model to add activity states, so that the model can be applied over a range of video sequences.

During the course of this research, we discovered that similar work was performed by Chandra and Riebman [7]. In their work I frames are assumed to be uncorrelated and are generated from a Gaussian distribution, as opposed to our approach where I frames are generated using an AR(1) process to model the long term correlations in the video sequence.

References

[1] *Video Coding for Low Bit Rate Communication*, ITU-T Recommendation H.263 Version 2, Jan. 1998.

[2] Motion Pictures Experts Group, "Overview of the MPEG-4 standard", ISO/IEC JTC1/SC29/WG11 N2459, 1998.

[3] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J.D. Robbins, "Performance models of Statistical Multiplexing in Packet Video Communication," *IEEE Trans. Comm.*, vol. 36, pp. 834-43, 1988.

[4] P. Sen, B. Maglaris, N, Rikli and D. Anastassiou, "Models for Packet Switching of Variable-Bit-Rate Video Sources," *IEEE J. on Select. Areas in Comm.*, vol. 7, no. 5, June 1989.

[5] F. Yegenoglu, B. Jabbari and Ya-Qin Zhang, "Motion-classified Autoregressive Modeling of Variable Bit Rate Video," *IEEE Trans. CSVT*. vol. 3, no. 1, Feb 1993.

[6] N. Doulamis, A. Doulamis and S. Kollias, "Modeling and Adaptive Prediction of VBR MPEG Video Sources," pp. 27-32, 1999 IEEE Third Workshop on Multimedia Signal Processing, September 1999, Copenhagen, Denmark.

[7] Kavitha Chandra and Amy Reibman, "Modeling One- and Two-Layer Variable Bit Rate Video," *IEEE/ACM Trans. Networking*, vol. 7, no. 3, pp. 398-413, June 1999.