Activity-Adaptive Modeling of Dynamic Multimedia Traffic

Deepak Turaga and Tsuhan Chen

Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213
{dturaga, tsuhan}@andrew.cmu.edu

ABSTRACT

Most of the existing work on modeling variable bit rate (VBR) video sources does not either explicitly take into account group-of-pictures (GOP) or assumes a fixed GOP structure. Real video data inherently possesses a variable GOP structure. We propose a number of doubly Markov models for such real data. These models outperform presently proposed models and have reasonable complexity in terms of the number of parameters.

1. INTRODUCTION

Variable bit rate video coding allows for great flexibility in terms of video coding, efficient compression ratios and can maintain desired video quality. Bitstreams from various VBR video sources can also be efficiently multiplexed over the network using statistical multiplexing techniques. All the above factors have led to VBR video encoders being the preferred mode of coding video streams and the focus of this paper is on modeling such video sources. Modeling video sources is important as it allows for network designers to estimate the parameters of networks like packet loss probabilities and end-to-end delays so that they can guarantee a desired quality of service (QoS).

Modeling VBR video traffic poses difficulties as the bit rate for a given video sequence is determined by a large number of factors. Different compression schemes can lead to different bit rates for the same video sequence. Models for VBR traffic are dependent on the choice of the compression scheme. The popular standards defining compression schemes today are ISO MPEG series and ITU H.26x series with MPEG-4 and H.263 being some of the latest versions. More information about these standards can be obtained from [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 often 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. Frames between two successive I frames, including the leading I frame, 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, and can be easily extended to B frames.

Several models for VBR video traffic have been proposed in literature. 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. Izquierdo and Reeves [6] have performed a survey of different statistical models proposed to model VBR video traffic.

Most of the work does not explicitly take into account differences between I and P frames. Some work done by Doulamis et al [7] models I, P and B frames explicitly with an additional layer corresponding to the activity level of the video scene. This is a good model for video traffic. However, they impose a constraint of a fixed GOP structure. They assume that every GOP consists of an I frame followed by a fixed number of P and B frames in a fixed pattern. This model may not be appropriate for all video sequences, as video content does not necessarily follow any regular pattern. Chandra and Reibman [8] model I and P frames explicitly and allow for a variable GOP structure. However, their model requires a large number of parameters and they do not allow for any temporal correlation or different activity levels for I frames.

In this paper we describe a simple two state model that can model I and P frame data and allow for flexibility in GOP structure. We then extend the model to account for different activity levels in the video bit rate. We try four different approaches and these are described and evaluated.

This paper is organized as follows. Section 2 describes the simple two-state model. Section 3 describes the extensions to this model to allow for different activity levels in bit rate. Section 4 includes analysis of experimental results from using these models to model real video traffic. Section 5 consists of the conclusion and future work.

2. TWO STATE I AND P MODEL

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 a Markov chain having two states, one corresponding to I frames and the other corresponding to P frames. The model transitions between these states with probabilities based on the training data,

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

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, AR(1) processes with Gaussian distributions are used to generate I and the P frames. The I frame AR(1) process is never restarted, while the P frame AR(1) process is restarted every time the model transitions from an I state to a P state. Therefore the AR(1) process for I frames captures long term correlations while that for P frames captures short term correlations. The parameters for the AR processes are estimated from the training data. Our model may be pictorially represented as in Figure 1.

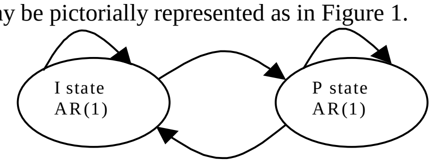

Figure 1. Two-state I and P model

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. Clearly, the parameters needed to specify the model are the four transition probabilities and three parameters for each AR(1) process (mean, variance and parameter ρ). The starting probabilities can be ignored as we always start with the I state. Hence this model needs a total of ten parameters.

3. ACTIVITY MODELS

The model proposed in the previous section is very simple and it performs well in modeling bitstreams that have a reasonably fixed activity level. If the video sequence has large differences in action levels between scenes, this leads to large variations of the bit rate within I or P frames, corresponding to different activity levels. It is difficult to capture this variation with a simplistic model and hence another level of complexity needs to be introduced in the model. We propose a number of doubly stochastic processes to model both the activity level changes and I and P frames corresponding to a certain activity level. As before, the temporal correlation between I frames or P frames corresponding to an activity level is captured using AR(1) processes with Gaussian distributions. The AR(1) processes are restarted as described for the Two-State I and P model. Again, we impose no constraint on the GOP structure or on the activity levels. The models we propose are described in the following subsections.

3.1 Type I Models

We describe two models in this sub-section. In each of these models we first decide whether to generate an I or a P frame (based on a Markov chain) and after that decide which activity level the frame should belong to.

3.1.1 Doubly Markov Model

This model has two Markov chains, one corresponding to the I and P selection and the other corresponding to the activity level selection. The outer Markov chain corresponds to the I and P frame selection. Within each state of this Markov chain we decide whether the frame corresponds to low (L), medium (M) or high (H) activity level. As we need to be flexible to allow a random GOP structure, we need to allow for transitions between the two states of the I and P Markov chain at every frame. Every time we enter an I or P state we reinitialize the activity Markov chain to obtain a frame corresponding to an activity level. However, if we re-enter an I or P state i.e. if the previous state was the same as the current state, we do not re-initialize the activity Markov chain and remember the state it was in previously. Once we decide on an I or P frame and the activity level, this frame is generated using an AR(1) process to obtain the temporal correlation between frames. This model is shown in Figure 2.

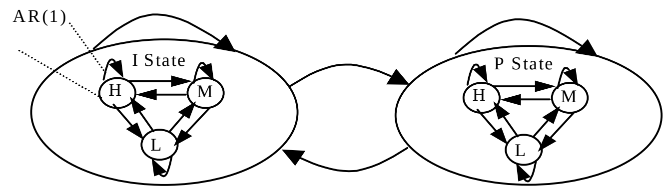

Figure 2. Type I Doubly Markov model

As can be seen from the figure, first we decide whether a frame is I or P, after which the I or P frame is classified into an activity state and then the output is generated using an AR(1) process. To specify this model we need four transition probabilities for the I and P Markov chain, nine transition probabilities for each activity Markov chain, three parameters per AR(1) process and three starting probabilities per inner Markov chain (the outer chain is always started in the I state), totaling to 46 parameters.

3.1.2 Simplified Model

The model described in the previous section has a large number of parameters. On evaluation of the probabilities within an I or a P state we found that the transition probabilities between the activity states were very close to the unconditional probabilities $P(S(n) = S_i \mid S(n-1) = S_j) \approx P(S(n) = S_i)$ where S_i is a state of the activity Markov chain and $S(n)$ is the state of the model at time instant n. This means that the inner Markov chain can be replaced with a set of probabilities for selecting L, M or H activity state. So we can modify that model to obtain the one shown in Figure 3.

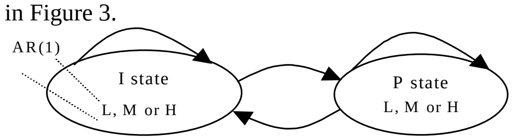

Figure 3. Type I Simplified model

As can be seen from the above model, we first decide on whether a frame is I or P and after that we decide the activity model based on the probability of a state being in a certain activity model, given that it is an I or P frame. Once we decide this then the output is generated using an AR(1) process. The total number of parameters for this model is 28.

3.2 Type II Models

The models described in the previous section pick an I or P frame first and then choose an activity level. As against this the models described in this section pick the activity level first and then choose between I and P frames.

3.2.1 Doubly Markov Model

In this model the outer Markov chain corresponds to the activity level selection, while the inner one corresponds to the I and P frame selection. The outer chain has three states, for low, medium and high activity and the inner chain has two states, corresponding to I and P frames. Each time we transition into an activity state, we initialize the inner Markov chain to start with an I frame and then let the Markov chain generate data. This is allowed till this inner Markov chain transitions back to the I state, which indicates the completion of a GOP. Once this happens we test the outer Markov chain to decide which activity state the next GOP belongs to. The assumption underlying this model is that all frames within a GOP belong to one activity level. As mentioned before, within each activity and I or P state, data is generated using an AR(1) process to model the long-term temporal correlations. This model is shown in the following figure.

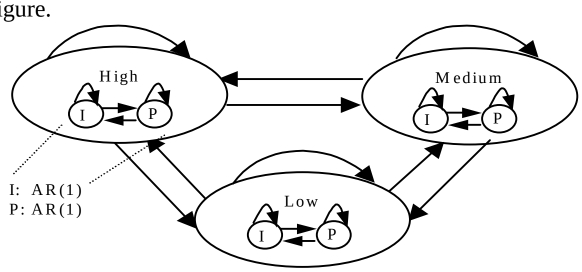

Figure 4. Type II Doubly Markov model

As can be seen from the figure, the outer Markov chain corresponds to the activity level while the inner one corresponds to the I and P frames. We need nine transition probabilities for the activity Markov chain, four transition probabilities per I and P Markov chain, three parameters for each AR(1) process and three starting probabilities for the outer Markov chain, making a total of 42 parameters. It can be noticed that this model is similar to the Doulamis model, but our model is free from the constraint of a fixed GOP structure.

3.2.2 Simplified Model

We can relax some constraints from the previous model to reduce complexity. One of the constraints that we can relax is the assumption that all frames in a GOP belong to the same activity level. By relaxing this constraint, we can make our model more flexible and also reduce its complexity. The simplified model is shown in Figure 5.

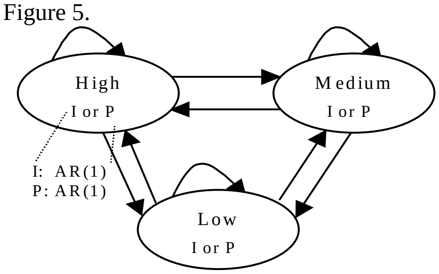

Figure 5. Type II Simplified model

The Markov chain helps in selecting the activity level of the frame, after which we decide to generate either an I or a P frame with a certain probability. Once we pick the type of the frame, it is generated using an AR(1) process, as before. Hence, the decision regarding the activity levels has to be taken for every frame, unlike in the previous case, when the decision regarding the activity level is taken once for a GOP. The total number of parameters for this model is 36.

4. RESULTS AND DISCUSSION

All the models were trained on the same data and characteristics of the generated bit rate were compared with those of the real data. The training data was from two different sequences. The first was a high motion video sequence made up of advertisements. We call this sequence Ads. This sequence had frequent scene changes, camera zooms and pans and a lot of motion. The second sequence was a news clip and we call it News. This sequence contained news reports from different locations and hence it contained a moderate amount of motion and some scene changes. Sample frames from both the sequences are shown in Figure 6.

Figure 6. Sample frames from Ads (left) and News

Both sequences consisted of five minutes of data sampled at 15 Hz, making a total of 4500 frames. Each 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 an empirical threshold (70% of blocks in a frame), 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 as high as an I frame.

Each model is trained independently on each of the sequences and then used to generate data. The autocorrelation function (R_{xx}) of the generated data ($X(k)$) is used as a measure of the performance of a model. R_{xx} is estimated from the data using $R_{xx}(i) = \frac{1}{4500}\sum_{k=i}^{4499} X(k)X(k-i)\,; i = 0 \ldots 4499$. This estimate for the generated data is then compared to that for the real data and the squared error between the two is computed. Some examples of the bit rate trace of real data and data generated by the models are shown in Figure 7.

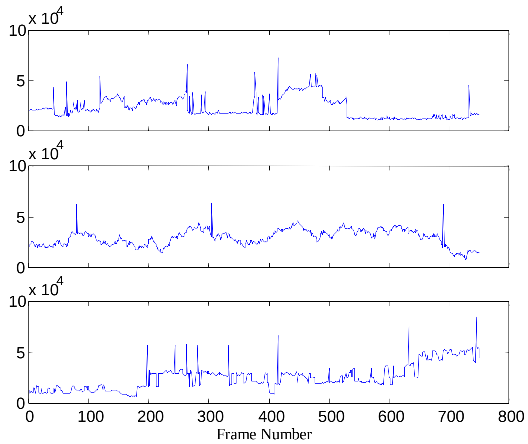

Figure 7. Trace for News: Real trace (top), Two-state I and P model (middle) and Type I Doubly Markov model

As can be seen from the figure, the real data and data generated by the Type I Doubly Markov model have three distinct activity levels. As against this, data generated by the Two-state I and P model does not show such activity levels. Hence models with activity levels can approximate the real data better.

The following table contains some experimental results for the autocorrelation function error for generated data. The number in each column represents the improvement (reduction) in error for a given model over the two state I and P model, which is taken as the base.

Table 1. Improvement in error in autocorrelation function w.r.t. Two-state I and P model

Video data	**Doulamis Model**	**Type I Models**		**Type II Models**	
		Doubly Markov	Simplified	Doubly Markov	Simplified
Ads	–54%	75.43%	69.85%	27.8%	63.69%
News	–49.3%	87.45%	86.71%	34.32%	85.03%

As can be seen from the above table, the Doulamis model performs much worse than all the other models, for both sequences. This is due to the fact that the Doulamis model assumes a fixed GOP structure, which is inconsistent with real data generated by a video encoder that selects I frames based on video content. Besides this, the Doulamis model also assumes that all frames in a given GOP belong to the same activity level. Both these assumptions place rigid constraints on the model and hence lead to poor performance. In terms of the complexity of the model, it requires 19 parameters.

It can also be seen that the addition of activity levels does lead to an improvement in performance over the two-state I and P model. All models have a smaller error in modeling data from News than from Ads. This is because the News data has a smoother autocorrelation function than the Ads data, due to the relatively smaller amount of motion. Among the models with added activity level, it can be seen that the Type I models perform better than the Type II models. This is because activity level of a bitstream of real data is well predicted by the choice of I and P frames, while the choice of I and P frames in an activity level is not as well predicted. The Type II Doubly Markov model performs the worst among all these models as it assumes that all the data within a GOP belongs to one activity level, which is too rigid a constraint. The Type I Doubly Markov model shows the best performance for both sequences. It can also be seen that the Type I Simplified model has a comparable performance. This is because in the test data the conditional probabilities for transitions within the inner Markov chain for the Type I Doubly Markov model are within 5% of the unconditional probabilities.

5. CONCLUSION

In this paper we propose some models for VBR video that allow for a flexible GOP structure. We start by introducing a simple two-state I and P model and then extend it by adding a level corresponding to the bit rate activity. These models are used to model real data and the results are evaluated in terms of the error between the generated autocorrelation function and the autocorrelation function of the real data. All the models proposed perform better than the model proposed for a fixed GOP structure. Among the models proposed, the Type I Doubly Markov model performs best, and the Type I Simplified model also gives reasonable error performance with a much smaller number of parameters. Future work includes additional tests to show that the data generated by these models also provides a better estimate of the packet loss probabilities and the delays over networks.

Acknowledgements

The authors would like to thank Dr. Amy Reibman and AT&T Labs for the fruitful discussions and insightful comments.

6. 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] M. Izquierdo and D. Reeves, "A survey of statistical source models for variable-bit-rate compressed video," Multimedia Systems, vol.7, no.3, pp. 199-213.

[7] 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.

[8] K. Chandra and A. Reibman, "Modeling One- And Two-Layer Variable Bit Rate Video," *IEEE/ACM Trans. Networking*, vol. 7, no. 3, pp. 398-413.