Scene Change Detection-Based Discrete Autoregressive Modeling for MPEG-4 Video Traffic

I. Spanou*, A. Lazaris†, and P. Koutsakis‡
* Department of Electrical and Computer Engineering, University of California, Davis, USA.
†Department of Electrical Engineering, University of Southern California, USA.
‡Department of Electronic and Computer Engineering, Technical University of Crete, Greece.
Email: espanou@ucdavis.edu, alazaris@usc.edu, polk@telecom.tuc.gr

***Abstract*—With more than 16 billion videos streamed on YouTube during last May and recent estimates by Cisco that mobile video traffic will increase 25-fold between 2011 and 2016, there is a pressing need to adequately serve large numbers of simultaneous online video transmissions. Network providers need to be able to guarantee the strict Quality-of-Service (QoS) requirements of real-time variable bit rate (VBR) video users, and a good statistical model for multiplexed video traffic can help significantly to evaluate and enhance network performance under various video loads.**

In this paper, we propose and evaluate a new hybrid video traffic model for MPEG-4 video. The genres of the videos considered in our study include lectures, cartoons, talk shows, action movies, sci-fi and sports. In the first part of our work, we build a Discrete Autoregressive model of order one (DAR(1)) and discuss its efficiency on capturing the behavior of real MPEG-4 video traces. We then proceed to build and evaluate a hybrid model which combines the DAR(1) model with a *scene-change detection and classification* algorithm, in order to provide us with higher modeling accuracy. The classification is performed based on the average number of bits generated during the scenes and the scene activity is modeled by a Markov chain where each state represents the degree of activity (high/low). Our extensive experimental evaluation study shows that the proposed hybrid method significantly improves the efficiency of pure DAR(1) schemes.

I. Introduction

The popularity of online Web video streaming has increased dramatically during the last decade. Video has strict Quality-of-Service (QoS) requirements, two of which are the maximum allowable video packet delay and the maximum allowable video packet dropping probability. Delays in video streaming can be annoying to the viewers and whenever the delay experienced by a video packet exceeds the corresponding maximum delay threshold, the packet is dropped. As a consequence, a very challenging problem for the research community is the design of wired and wireless networks that are able to guarantee the strict QoS requirements of video traffic, despite the tremendous increase in the number of multimedia users.

MPEG-4 is a video encoding standard that was introduced in 1998 for more efficient compression of audio and visual (AV) digital data. Standard MPEG encoders generate three types of video frames: I (intracoded), P (predictive) and B (bidirectionally predictive). An I frame uses only transform coding and provides a point of access to the compressed video data. On the other hand, a P frame uses motion-compensated prediction from the most recent previous I or P frame. In terms of size, as P frames use information already transmitted in previous anchor frames, their size can be significantly smaller than that of I frames. B frames are coded based on both past and future I or P frames, and thus their size is about an order of magnitude less than the size of I frames. So, I frames are usually the largest in size, followed by P and then B frames.

A very important feature of common MPEG encoders is the manner in which frame types are generated. Typical encoders use a fixed Group-of-Pictures (GoP) pattern when compressing a video sequence. The GoP pattern specifies the number and temporal order of P and B frames between two successive I frames and it is defined by the distance N between I frames and the distance M between P frames. In our work, the GOP pattern of the traces under study has M=3 and N=12.

The problem of modeling traffic generated by standard MPEG encoders is not new. Several methodologies have been proposed in the relevant literature for modeling Variable Bit Rate (VBR) video traffic. Some examples include: first-order autoregressive (AR) models [10] [11], discrete AR (DAR) models [9] [10], Markov renewal processes (MRP) [12], MRP transform-expand-sample (TES) [13], finite-state Markov chain [14], Gamma-beta-auto-regression (GBAR) models [15] [16] (which capture data-rate dynamics of VBR video conferences well but were found in [16] to not be suitable for general MPEG video sources), and wavelets [17] [19]. However, most of the approaches in the literature have either been proposed and evaluated for past video encoding standards, or incur high complexity (e.g., [19]) in order to achieve accurate modeling results. For these reasons, we revisit the problem and focus on modeling a large number of MPEG-4 movies. In this paper, we propose a hybrid model comprising two components: a) a DAR(1) model, and b) a simple two state Markov chain model. The reason for this selection is that both models come with significant advantages such as low computational complexity and easy deployment, since they are based on four traffic parameters, namely the mean, the peak, the variance and the autocorrelation coefficient of the traffic.

In the first part of this work (Section II), we follow the methodology presented in [2] for modeling MPEG-4 *video-conference* traces. We initially try to model the video traces using well known probability distributions and we use the best fit results in order to build a DAR(1) model that models the traffic more accurately. However, due to the bursty nature of the traces under study, a simple DAR(1) model is not always enough. For this reason, we propose a hybrid scheme (Section III) that combines the DAR(1) model with a scene-

978-1-4673-3122-7/13/$31.00 ©2013 IEEE

change detection scheme and classification technique (taken from [3]), in order to further improve the modeling accuracy. The proposed work differs from our work in [2] in that the traffic considered here is much different in nature (although it is encoded using the same standard, i.e. MPEG-4), and also that in [2] we do not combine DAR(1) with scene-change detection. In addition, in [2], we use only 4 videoconference traces of low or moderate motion. On the other hand, in this work, we use 63 traces (21 movies in 3 different formats, each), much more bursty in nature, as it will be described below.

II. Video Traffic Modeling

In this work, we analyze 21 MPEG-4 movies from [1], each in 3 different versions (characterized by a triplet x-y-z, for the quantization parameter used), yielding a total of 63 traces. The list of the movies used is given in Table I. Throughout this paper, we use the word "movie", to refer to the actual movie, and the word "trace" to refer to a particular version of this movie (in terms of the quality, quantization parameters, etc).

Movie Title	**Movie Title**
1) Aladdin	12) Lecture - Reisslein
2) Baseball with Commercials	13) O Brother DVD
3) Charlie's Angels DVD	14) Oprah without commercials
4) Cinderella	15) Silence of the Lambs
5) Citizen Kane	16) Star Wars IV
6) Die Another Day DVD	17) Terminator One
7) Die Hard One	18) The Transporter DVD
8) Friends vol.4 DVD	19) Tokyo Olympics DVD
9) Ice Age DVD	20) Tonight Show with Commercials
10) Lecture-Gupta	21) Tonight Show without Commercials
11) Lecture HQ - Reisslein	

TABLE I: MPEG-4 movies under study. Three versions of each movie were used (with different quantization parameters) yielding to a total of 63 traces.

The first step of this work is to attempt to model the various MPEG-4 traces of different genres (i.e., drama, action, sci-fi, cartoon, lecture, talk show and sports) using well-known distributions, and then use this distribution to build a DAR(1) model. For this purpose, we try to model the MPEG-4 traces in two ways: 1) by using one trace at a time (as a whole trace, or separately for each frame type (i.e. I,P,B)), and 2) by multiplexing (i.e., superposing) several sources.

A. Statistical Tests

Throughout this work, we use three kinds of statistical tests in order to assess the capabilities of each model to capture the behavior of MPEG-4 video traffic. Specifically, we use 1) Q-Q plots [4] [9], 2) the Kolmogorov-Smirnov [4] test, and 3) the Kullback-Leibler divergence test [5]. The Q-Q plot is a powerful goodness-of-fit test which graphically compares two data sets in order to determine whether the data sets come from populations with a common distribution (if they do, the points of the plot should fall approximately along a 45-degree reference line). More specifically, a Q-Q plot is a plot of the quantiles of the data versus the quantiles of the fitted distribution (a z-quantile of X is any value x such that $P(X \leq x) = z$). The KolmogorovSmirnov test (KS-test) tries to determine if two datasets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data, i.e., it is non-parametric and distribution free. The KS-test uses the maximum vertical deviation between the two curves as its statistic D defined as: $D = \sup_x \{|F(x) - G(x)|\}$, where $F(x)$ is the empirical distribution of the data, and $G(x)$ is the Cumulative Density Function (CDF) of the model. The Kullback-Leibler divergence test (KL-test) is a measure of the difference between two probability distributions and is defined as: $I(f,g) = \sum_{i=1}^{k} p_i \log \frac{p_i}{\pi_i}$, where k is the number of possible outcomes of the underlying random variable, p_i is the true probability of the i^{th} outcome, and $\pi_1, \pi_2, \ldots, \pi_k$ constitute the approximating probability distribution. The notation $I(f,g)$ denotes the information lost when g is used to approximate f or the distance from g to f. If the quantity $0log0$ appears in the formula, it is interpreted as zero.

In this work, we deploy seven widely used distributions as candidates for modeling the video traffic. Specifically, the distributions used are: 1) Exponential, 2) Gamma, 3) Lognormal, 4) Weibull 5) Pearson-V, 6) Geometric, and 7) Negative Binomial. The above distributions were chosen as fitting candidates, since almost all of them (with the exception of exponential) have often been used in video traffic modeling studies.

Initially, when handling traces of whole movies, our efforts focused on whether a best fit existed or not. More specifically, if the Q-Q plots were unable to indicate a distribution that could serve as a best fit, the corresponding traces were split into their respective I, P and B frames and each sequence of frame types was modeled separately, and tested with the same statistical tests, in order for us to finally decide upon the best fit in each case. If, on the other hand, the Q-Q plots indicated a best fit for the whole trace, we then took the results of the KL and the KS tests into consideration, in the way described below. In the cases where all three statistical tests or two out of the three tests denoted the same distribution as the best fit, we considered that distribution to provide the best fitting result. However, in the cases where each out of the three tests denoted a different distribution as the best fit, we compared the result of the KL-test with the result of the respective Q-Q plot. If the distribution being the best fit according to the KL test was a close second best according to the Q-Q plot, that distribution was considered to be the best fit for the real data. Throughout our work, we encountered this in several of the traces under study. Still, the most frequent case was that of two statistical tests denoting the same distribution as the best fit. Finally, if all the tests offered different results and the KL-test best fit did not happen to be a Q-Q plot good fit, then the specific trace was not included in our DAR(1) modeling.

B. Single-Source MPEG-4 Traffic Modeling

Similarly to [2], our first attempt here is to capture the statistical behavior of the 63 aforementioned single-source MPEG-4 video traces. This can be achieved by finding the distribution providing the best fitting results (i.e., capturing well the video frame sizes).

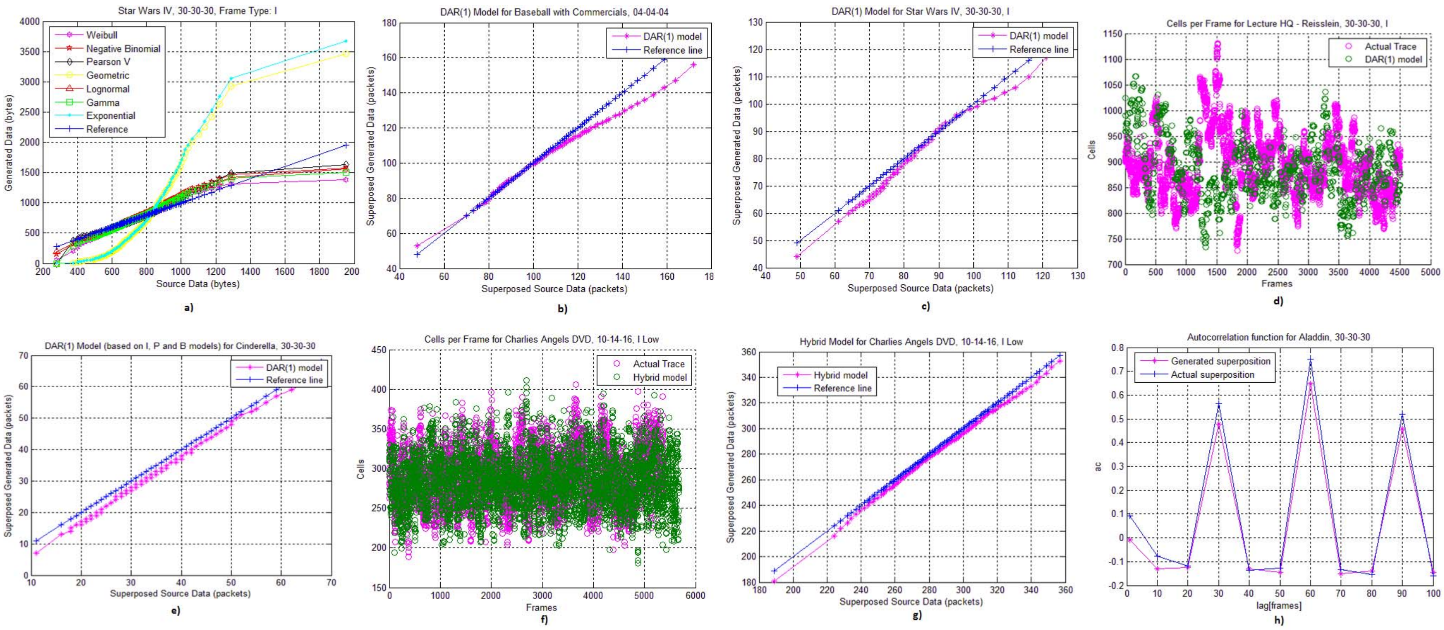

Fig. 1: Figs. a), b), and c) represent the Q-Q plots for an I frame sequence using several distributions, the DAR(1) model for the whole trace (5 sources), and the DAR(1) model of an I frame sequence (5 sources), respectively. Fig. d) shows an example of a DAR(1) time series (15 sources). Fig. e) shows a Q-Q for the synthetic trace generated using the DAR(1) model (5 sources). Fig. f) illustrates the DAR(1) results for the hybrid model (10 sources), and g) the respective Q-Q plot (10 sources). Finally, Fig. h) plots the autocorrelation coefficients of the generated superposition of traces using the hybrid model, vs the actual trace.

Our statistical test results showed that, in the majority of cases, the three tests used did not unanimously agree on which distribution provided the best fit. This is intuitively expected, because of the different nature of the tests. More specifically, the Q-Q plot is of a qualitative nature and is based on visual inspection. The KS-test does provide a quantitative result, but this refers to the maximum vertical deviation between two curves, which is not a very representative metric. On the other hand, the KL-test provides the information lost during a fitting attempt, by comparing the probabilities of every respective element of the real and the distribution data, and thus giving a more accurate metric of the goodness of the fit. Therefore, among the three tests, KL is the one to provide the most accurate results. For all the traces under study, we proceed to model the whole traces with each of the seven distributions (an example is shown in Fig. 1 a)), and if the modeling results are not adequate, we split the traces into I, P, and B frames. Despite the fact that the majority of the tests provided good to mediocre fitting results, no distribution was proven capable of perfectly "capturing" the behavior of a trace under study. The reason is that the high autocorrelation of the traffic, especially in the cases of individual frame types, cannot be "captured" by a distribution generating frame sizes independently, according to a declared mean and standard deviation.

C. Multiplexed MPEG-4 Traffic Modeling

In the next step of our work, we proceed to implement a DAR(1) [7] model to capture the behavior of multiplexed videoconference traffic. Our goal is to build a model based only on parameters which are either known at call set-up time or can be measured without introducing much complexity in the network. In addition, since in the case of multiplexed traffic the resulting frame sizes (in Bytes) can be quite big, we choose to express the frame sizes in *packets* of 48 [2] or 200 [8] Bytes, depending on the resulting size of the multiplexed traces. This way, we can control the number of states needed in order to represent the frame sizes generated by our model, as it is described below.

Autoregressive models have been used in the past to model the output bit-rate of VBR encoders. A Discrete Autoregressive model of order p, denoted as DAR(p), generates a stationary sequence of discrete random variables with an arbitrary probability distribution and with an autocorrelation structure similar to that of an autoregressive model [6] [7]. DAR(1) is a special case of a DAR(p) process and is defined as follows: Let $\{V_n\}$ and $\{Y_n\}$ be two sequences of independent random variables. The random variable V_n can take two values, 0 and 1, with probabilities $1-p$ and p, respectively. The random variable Y_n has a discrete state space S and $P(Y_n = i) = \pi(i)$. The sequence of random variables $\{X_n\}$ which is formed according to the linear model: $X_n = V_n X_{n-1} + (1-V_n)Y_n$ is a DAR(1) process. A DAR(1) process is a Markov chain with discrete state space S and transition matrix: $\mathbf{P} = \rho\mathbf{I} + (1-\rho)\mathbf{Q}$, where ρ is the autocorrelation coefficient, $\mathbf{I}$ is the identity matrix, and $\mathbf{Q}$ is a matrix with $Q_{i,j} = \pi(i)$, for $i, j \in S$. Autocorrelations are usually plotted for a range of lags. The term "lag" is used to refer to the distance between the frames examined. The autocorrelation can be calculated by the following formula: $\rho(W) = E[(X_i - \mu)(X_{i+W} - \mu)]/\sigma^2$, where μ is the mean and σ^2 is the variance of the frame size for a specific video trace. The rows of the $\mathbf{Q}$ matrix above consist of the probabilities of the distribution found to be the best fit $f_0, f_1, \ldots, f_k, F_K$, where $F_K = \sum_{k>K} f_k$, and K is the peak rate. Each k, for $k < K$, corresponds to possible source rates less than the peak rate of K.

As explained in the Introduction, the advantage of this model is its simplicity and the fact that it is based only on four physically meaningful parameters, i.e., the mean, the peak, the variance and the lag-1 autocorrelation coefficient ρ. This allows the model to be easily used for on-the-fly computations in the network, e.g., for call admission control purposes.

In order to simulate N multiplexed sources of a specific type, we generate N copies of a particular trace by shuffling the trace using a different start point each time. Then, we add (superpose) the resulting traces to generate the multiplexed traffic. In our study, we have used 5, 10, and 15 superposed traces in order to investigate the behavior of DAR(1) vs. the number of multiplexed sources. After running our experiments for all the traces under study, we observe that out of the four categories examined (whole traces and I, P and B frames), the number of good modeling results in I frames was the highest (e.g. Fig. 1 c), d)), with P and B frames following, leaving whole traces last (e.g. Fig. 1 b)). More importantly, although there were a few cases of high accuracy modeling results, perfect accuracy was never achieved. In addition, as the number of the superposed sources increases from 5 to 10 and then to 15, the majority of the results, regardless of frame type, remain unaffected or deteriorate. Improvements can be observed in a limited number of cases. Also, most of the bursty movies (burstiness ratio (i.e. peak/mean) > 10) that were tagged as bad modeling results in the case of 5 traces tend to worsen with the increase in the number of superposed sources.

In order to better assess the ability of DAR(1) to model real multiplexed MPEG-4 traces, we proceed to generate the MPEG-4 GoP patterns using separate DAR(1) models for each frame type and then running our statistical tests in order to quantify the modeling accuracy. The derived results illustrate once again that when the number of superposed sources increases (from 5 to 10 and from 10 to 15), the majority of the initial results for 5 sources remain unaffected or deteriorate (e.g. Fig. 1 e)). Interestingly, this result does not coincide with [2], due to the different types of studied traces (video vs. videoconference).

III. Scene Change Detection Based Modeling

Due to the unsatisfactory modeling results of the DAR(1) model, we proceed to design a hybrid model by combining the DAR(1) model with a scene-change based Markov chain model, similar to the one proposed in [3]. The main concept behind this model lies on the idea of dividing each video trace into scenes, and then classifying the detected scenes into low or high-activity ones. Therefore, each video trace produces two new traces, one consisting of all the low-activity scenes and one consisting of all the high-activity ones. For each movie out of the initial 63 ones, we use a 2-state (High, Low) Markov chain model, as shown in Fig. 2, in order to determine the number of low and high scenes of our modeled source. For every scene in both categories, we then determine the number of I, P and B frames in it and finally, run the DAR(1) model, for every case.

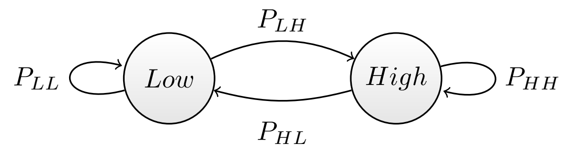

Fig. 2: The 2-state Markov chain model for characterizing the video frame size activity.

A. Scene Detection

A scene typically consists of a few tens or hundreds of frames that depict a real-world scene and can also be interpreted as a collection of related GoPs. Every new scene starts with an I frame, so the I frames of the video streams under study are used to detect scene changes. More specifically: Let $X(i)$ denote the bits generated during the i^{th} I frame. The i^{th} frame of the sequence of I frames is said to be the start of a new scene if the following equations are satisfied [3]:

$$\frac{|X(i+1)-X(i)|}{(\sum_{j=\text{start of scene}}^{i} X(j))/(i-j+1)} \geq \text{first\_threshold} \quad (1)$$

$$\frac{|X(i+2)-X(i)|}{(\sum_{j=\text{start of scene}}^{i} X(j))/(i-j+1)} \geq \text{second\_threshold} \quad (2)$$

It can be seen that the i^{th} I frame size is compared against both the $i+1^{th}$ and the $i+2^{th}$ I frames, therefore scenes are identified only when the change is persistent. The numerators of the fractions above represent the difference in bits between the size of the i^{th} and $i{+}1^{th}$ or the $i{+}2^{th}$ I frame, respectively. The denominators represent the sum in bits of all the I frames from the last I frame identified as a scene start, up to the current i^{th} I frame, divided by the number of those I frames. A typical range of thresholds, as suggested in [3], is [0.15, 0.25] for the first and second threshold. In our approach, we have used the value 0.15 for both of them.

B. Scene Classification

The next step, after the scene identification, is the scene classification. A scene can fall into one of two categories: low-activity scenes or high-activity scenes, depending on the average bit rate of the video trace; if a scene has an average bit rate that is greater than the average bit rate of the whole movie, then it is classified as a high-activity scene, else it is classified as a low-activity scene. The average bit rate of each movie is provided in [1] along with other trace statistics, whereas the average bit rate of every scene is calculated as the number of bits transmitted during the scene divided by the scene duration. We divide each low and high activity trace into their respective I, P and B frames, hence, every movie is "split" into 6 traces: the low-activity and high-activity I frames, the low-activity and high-activity P frames and the low-activity and high-activity B frames.

In the next step, we calculate the mean and variance of the frame sizes of each of the 63*6=378 traces under study and determine all the distribution parameters. We then run KL and KS tests and generate Q-Q plots, in order to determine the best fit for all cases, similarly to our work in Section II. Then, we calculate the transition probabilities of the 2-state

Markov chain of Fig. 2 (Low-High Activity scenes) using the following equation for each of the 63 video traces.

$$p_{s_1,s_2} = \frac{\textit{number of transitions from } s_1 \textit{ to } s_2}{\textit{total number } (>0) \textit{ of transitions from } s_1} \quad (3)$$
$$s_1 = H, \text{ and} s_2 \in \{H, L\}$$
$$p_{s_1,s_2} = \frac{\textit{number of transitions from } s_1 \textit{ to } s_2}{\textit{total number } (>0) \textit{ of transitions from } s_1}$$
$$s_1 = L, \text{ and} s_2 \in \{H, L\}$$

In the next step, we generate the model traces by generating one scene per each state reached. A low-activity scene is generated after reaching state ”Low” whereas a high-activity scene is generated after reaching state ”High”. Once the number of the generated scenes equals the number of the real trace scenes, the procedure ends. A different approach to ours would be to work deterministically and thus, to ignore the Markov chain model and run the DAR(1) model for as many low and high-activity scenes as detected in the real video trace. However, this last approach defies the concept of modeling: we attempt to model video traffic in order to be able to estimate its required bandwidth even for traces where we have only very basic knowledge of their traffic characteristics. The alternative approach mentioned above would need a perfect knowledge of every trace that would, e.g., attempt to be transmitted over a network; the exact sequence of high and low activity scenes of the trace would be required.

	DAR(1) Model	**Hybrid Model**
Good Results	15.91%	32.69%
Mediocre Results	56.82%	61.54%
Bad Results	27.27%	5.77%

TABLE II: DAR(1) vs. Hybrid modeling results (the percentage is over the number of traces).

After an extensive experimental evaluation study, we observe that the hybrid model clearly outperforms the simple DAR(1) model in most of the cases under study. A summary of the comparison of the two schemes is given in Table II. For about one third of the movies we achieve very high modeling accuracy, while bad modeling results occur very infrequently. Additionally, the results classified as ”mediocre” in Table II for the hybrid model are shown via the three statistical tests to provide much smaller modeling inaccuracy than the ”mediocre” results of the DAR(1) model. These results show that the proposed hybrid model is a good candidate for a low-complexity video traffic modeling solution. An example of a video trace that is generated using the hybrid model is given in Fig. 1 f) and the respective Q-Q plot is given in Fig. 1 g). As we can observe, the two curves are almost identical. Finally, to further investigate the properties of our hybrid model, we proceed to plot the autocorrelation coefficients for several lags. As we can observe in Fig. 1h, the autocorrelation is captured very well by our model. This result shows the importance of combining the model proposed in [3] with DAR(1) into a hybrid model: as noted in [18], the work in [3] suffered from an inability to capture the autocorrelation function, a weakness that is alleviated with the use of our hybrid model.

IV. Conclusion

In this paper we initially investigated the possibility of modeling 21 MPEG-4 video traces, in 3 different quantization scales, by building a simple DAR(1) model and studying the I, P and B frames of the traces separately. The lack of sufficient modeling accuracy led us to propose a new hybrid model which combines the DAR(1) model with scene-change detection and scene classification techniques. The hybrid model produced results of much higher accuracy and appears to be a promising solution for low-complexity, on-the-fly computations of future video traffic behavior.

References

[1] P. Seeling, M. Reisslein and B. Kulapala, ”Network Performance Evaluation Using Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial”, IEEE Communications Surveys and Tutorials, Vol. 6, No. 2, Pages 58-78, 2004.

[2] A. Lazaris, P. Koutsakis and M. Paterakis, ”A new model for video traffic originating from multiplexed MPEG-4 videoconference streams, Performance Evaluation”, Vol 65, 51-70, 2008.

[3] G. Chiruvolu, T. K. Das, R. Sankar and N. Ranganathan, ”A Scene-based Generalized Markov Chain Model for VBR Video Traffic”, in Proc. of the IEEE Int. Conference on Communications (ICC), Atlanta, USA, 1998.

[4] A. M. Law and W. D. Kelton, ”Simulation Modeling & Analysis”, 2nd Ed., McGraw Hill Inc., 1991.

[5] K. P. Burnham and D. R. Anderson, ”Model Selection and Multi-Model Inference”, Springer-Verlag, New York, 2002.

[6] A.Adas, ”Traffic models in broadband networks”, IEEE Communications Magazine Vol. 35, No. 7, 1997, pp. 82-89.

[7] P.A. Jacobs, P.A.W. Lewis, ”Time series generated be mixtures”, Journal of Time Series Analysis Vol. 4, No. 1, 1983, pp. 19-36.

[8] F. Xue, V. Markovski, and L. Trajkovic, ”Packet loss in video transfers over IP networks”, in Proc. of the IEEE Int. Symp. Circuits and Systems, Sydney, Australia, 2001, vol. II, pp. 345-348.

[9] D. P. Heyman, A. Tabatabai and T. V. Lakshman, ”Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 2, No. 1, 1992, pp. 49-59.

[10] D. M. Lucantoni, M. F. Neuts and A. R. Reibman, ”Methods for Performance Evaluation of VBR Video Traffic Models”, IEEE/ACM Trans. Networking, Vol. 2, 1994, pp. 176-180.

[11] M. Nomura, T. Fuji and N. Ohta,”Basic Characteristics of Variable Rate Video Coding in ATM Environment”, IEEE Journal on Selected Areas in Communications, Vol. 7, No. 5, 1989, pp. 752-760.

[12] A. M. Dawood and M. Ghanbari, ”Content-based MPEG Video Traffic Modeling”, IEEE Trans. on Multimedia, Vol. 1, No. 1, 1999.

[13] B. Melamed and D. E. Pendarakis, ”Modeling Full-Length VBR Video Using Markov-Renewal Modulated TES Models”, IEEE Journal on Selected Areas in Communications, Vol. 16, No. 5, 1998, pp. 600-611.

[14] K. Chandra and A. R. Reibman, ”Modeling One- and Two-Layer Variable Bit Rate Video”, IEEE/ACM Trans. on Networking, Vol. 7, No. 3, 1999, pp. 398-413.

[15] D. P. Heyman, ”The GBAR Source Model for VBR Videoconferences”, IEEE/ACM Trans. on Networking, Vol. 5, No. 4, 1997.

[16] M. Frey and S. Ngyuyen-Quang, ”A Gamma-Based Framework for Modeling Variable-Rate Video Sources: The GOP GBAR Model”, IEEE/ACM Trans. on Networking, Vol. 8, No. 6, 2000, pp. 710-719.

[17] D. Arifler and B. L. Evans, ”Modeling the Self-Similar Behavior of Packetized MPEG-4 Video Using Wavelet-Based Methods”, in Proc. of the IEEE Int. Conference on Image Processing, NY, USA, 2002.

[18] Y. Sun, J. N. Daigle, ”A source model of video traffic based on full-length VBR MPEG4 video traces”, in Proceedings of the IEEE GLOBECOM 2005, St. Louis, USA.

[19] M. Dai, Y. Zhang, D. Loguinov, ”A Unified Traffic model for MPEG-4 and H.264” video traces, IEEE Transactions on Multimedia, Vol. 11, No. 5, 2009.