J. King Saud Univ., Vol. 18, *Eng. Sci*. (1), pp. 161-180, Riyadh (1426H./2005)

A Long Horizon Neuro-fuzzy Predictor for MPEG Video Traffic

Adel Abdennour
Department of Electrical Engineering, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

(Received 20 April, 2005; accepted for publication 03 July, 2005)

Abstract. This paper investigates the long-term prediction of MPEG video traffic. Predicting such traffic over a long horizon is important for today's fast networks and internet multimedia services. In comparison with short-term prediction, long-term prediction of video traffic is yet to be explored especially for MPEG-4 coded videos despite its effectiveness in a number of important network-edge applications such as dynamic bandwidth allocation, quality of service (QoS) control, and network management and planning. The main reason for the shortage of publications in such area is the difficulty of the problem, especially when classical or widely used prediction techniques are the ones to be employed. Prediction results, in this paper, are obtained using a simple neuro-fuzzy system and are compared to the classical normalized Least Mean Squares (LMS) technique. The neuro-fuzzy predictor is capable of predicting various real MPEG-4 real-world video traffic hundreds of frames in advance with high accuracy.

1. Introduction

MPEG-4 (Moving Picture Expert Group) is a compression video standard that has recently gained a considerable attention and started to be widely used for multimedia and wireless applications over Internet protocol (IP). In addition, MPEG-4 is expected in the near future to be carried by both 3G mobile and broadband always-on-Internet connections. In this standard, each frame of an input video sequence is segmented into a number of arbitrary shaped image regions, called video object planes (VOP). Three types of VOPs are encoded: Intra VOP (I-VOP), encoded independently of any other VOP; Predicted VOP (P-VOP), predicted (using motion compensation) based on another previously decoded VOP; and Bidirectional Interpolated VOP (B-VOP), predicted based on past as well as future VOPs. The size of video frames varies drastically as the sequence is being generated, and thus results in variable-bit-rate (VBR) traffic. I-VOP frames have the highest size while P-VOP frames size is slightly lower and B-VOP frames have the lowest size among all the three types of frames. Predicting such dynamic and complex behavior requires sophisticated techniques that can capture these traffic fluctuations and provide correct control actions which maximize network utilization while maintaining the desired quality of service (QoS).

In the past, linear predictors were more widely used. Such predictors work well with a relatively smoothed traffic where motion activity level is uniform such as in videoconference and videophone applications. However, it is not always possible, within the linear modeling framework, to estimate traffic characteristics when the applied video applications have rapid traffic variation and frequent scene changes such as in entertainment and broadcast videos. Accurate prediction of such complex traffic requires the use of non-linear predictors that can capture some statistical characteristics of the traffic and the inherent non-stationarities and non-linearities associated with MPEG video traffic. In the last decade, and particularly in the last five years, much of the research in the area was geared towards neural networks and, to a less extent, fuzzy logic. The main reason for such attention is that these “intelligent” predictors are often adaptive and rely on observed data rather than on an analytical model usually difficult to obtain. The resulting scheme is, therefore, robust, efficient, and capable of reflecting changes in the traffic behavior.

Several papers dealing with the prediction of video traffic using neural networks are available in the literature [1-10]. These papers use either a single step-ahead predictor to predict the next frame or a short-term predictor to predict two to four future frames ahead. None of these papers addresses the long-term case. The short-term prediction, while useful, may not be very practical because the predictor comes with its own computation burden (no matter how small it is), and very frequent prediction tasks increase the overhead and reduce efficiency.

Nowadays, and with the expansion of high-speed networks, there is a pressing need for long-term prediction. Some of the domains that can directly benefit from these predictors are: scene prediction [11], congestion control [12-14], dynamic bandwidth allocation and admission control [15-17], wireless and network management [18-19], and quality assessment [20]. In these domains, it is useful to have an estimate of more than 100 frames ahead to be able to capture scene changes, evaluate buffer occupancy, predict delays, or monitor network congestion.

In this paper, we propose a long-term predictor based on neuro-fuzzy systems to estimate MPEG-4 video traffic. Closest to this work is that of Liang [21], who analyzes long-term prediction of video traffic using neural networks trained through multi-resolution learning. In Linag’s approach, multi-resolution learning which is based on wavelet theory is employed for decomposing the original signal and approximating it at different levels of details. A feed-forward neural network predictor is used to predict MPEG-1 video traces with good results. This paper attempts to make contributions in two aspects. First, we develop long horizon prediction for estimating MPEG-4 video traffic that has not been reported thus far. Second, we use a neuro-fuzzy system that is straight forward to design, yet powerful enough, to predict up to few hundred frames ahead. Due to the unavailability of any results to compare to, we develop in this paper a classical predictor based on least mean squares for comparison purposes. Applying both predictors to a number of entertainment MPEG-4 coded videos should reveal the potential that the neuro-fuzzy system may have in predicting long-term video traffic.

The remainder of this paper is organized as follows. Section two analyzes the MPEG-4 video traces used in this study in terms of distribution, autocorrelation, and long range dependence. Section three presents the neuro-fuzzy predictor. Simulation results are presented in section four while section five contains the conclusions.

2. Analysis of MPEG-4 Video Traces

MPEG-4 is object-based coding, different from MPEG-1 and MPEG-2 of frame-based coding. The scene of MPEG-4 video contains a number of video objects, which are coded and decoded independently. Each frame of an input video sequence is segmented into a number of arbitrary shaped image regions, called VOPs. The shape and location of a VOP vary from frame to frame. Successive VOPs belonging to the same physical object in a scene are referred to as video objects. Each video object is partitioned into units of Group of Video Object Planes (GOV) which can provide random access points into the bit stream. Each GOV encodes three types of VOPs: I-VOP, P-VOP, and B-VOP. Intra frames (I-VOP) contain information that results from encoding a still image. They are the points of reference and random access in the video stream. They can be encoded independently of any other frames. The compression ratio for these frames is the lowest of all frame types.

Predicted frames (P-VOP) require information from previous I-VOP and/or P-VOP frames for encoding and decoding. Bidirectional Interpolated frames (B-VOP) require information from past I-VOP as well as future I-VOP for encoding and decoding. They have the highest compression ratio among the three frame types. The encoding process of each VOP is very similar to other frame-based standards such as MPEG-1, MPEG-2, and H.263, with motion and texture encoding applied to each VOP. Therefore, GOV, I-VOP, P-VOP, and B-VOP of MPEG-4 are usually referred as GOP, I frames, P frames, and B frames, respectively. Figure 1 shows the GOP (group of pictures) of MPEG-4 video that is used throughout this paper. The GOP is made of 12 frames arranged in periodic nature as: IBBPBBPBBPBB.

In this paper, we use four MPEG-4 video traces that are available from the Telecom Networks Group, technical University of Berlin [22]. These four traces are carefully chosen to represent different categories of movies (e.g., talks, sports, animation, and action). These traces are coded with Momusys MPEG-4 video software, where the entire scene is one video object and the single video object is encoded into a single video object layer. All detailed sequences were encoded with the following parameters:

- A pattern of GOP is repeated continuously to create the encoded frame sequence.
- Number of frames per sequence = 65,000.
- Duration of each frame = 40 ms.
- Encoder input =352 * 288 pels.

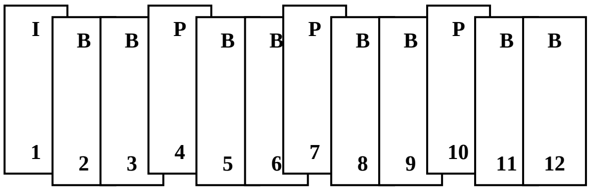
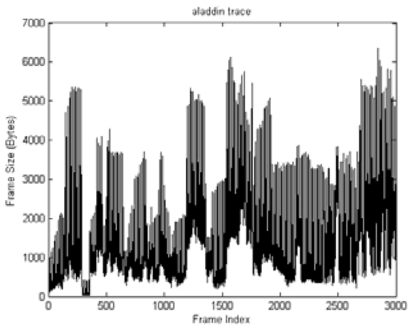


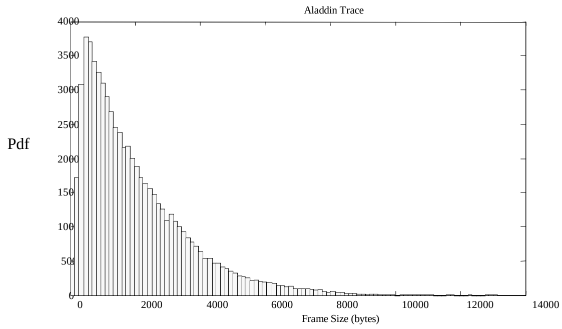
Fig. 1. Group of pictures of MPEG-4 frame sequence.

Table 1 gives an overview of the statistical properties of the generated MPEG-4 traces. These data sets represent entertainment video sequences with potential real-time applications. In addition, these sets contain frequent scene changes, camera zooming, panning, and different degree of motion activity. Figure 2 (a) shows a segment of the Aladdin video trace. The sequence is highly bursty and the maximum rate is more than six times larger than the minimum rate. Figure 2 (b) and (c) show the probability distribution function (PDF) and the autocorrelation function (ACF) of the same video trace. Analyzing these figures reveals that the MPEG-4 video data exhibits a heavy tailed PDF and an ACF with a mix of short range dependence (SRD) and long range dependence (LRD). The time dependent statistics are important for network and traffic engineering since correlations in the video traffic can have a significant impact on the performance of packet switched networks. The long range correlation is formally characterized as self-similarity or long range dependence. The cumulative effect of the correlations for large lags is significant and can cause large losses and delays.

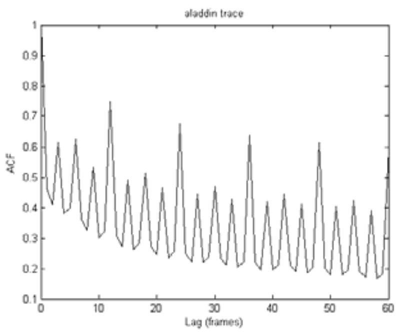
A number of video traffic studies show that video traffic is LRD or exhibits self-similar properties [23-24]. This is due to the nature of the video content (frames, sequences, scenes, and sub-scenes) and the encoding techniques employed. It is, therefore, important that any video predictor design considers these properties to allow full analysis of the complex statistics involved. Basically, LRD, also known as persistence or the Hurst effect, is the phenomenon of observations of an empirical record being significantly correlated to observations that are far removed in time or across a wide range of time scales.



(a)



(b)



(c)

Fig. 2. Statistical properties of MPEG-4 video trace: (a) Frame size, (b) PDF, (c) ACF.

Table 1. MPEG-4 statistical properties of encoded video sequences

	Frame Size			Bit rate	
	Mean (Kb)	Peak/ Mean	Coef. of variation	Mean (Mb/s)	Peak (Mb/s)
Aladdin	2.04	6.46	0.79	0.41	2.63
Star Wars	1.36	6.86	0.67	0.27	1.87
Lect. Room	1.08	6.92	0.88	0.22	1.49
Skiing	3.82	4.14	0.58	0.76	3.16

Hurst parameter (H) is commonly used to measure long-range dependence (the degree of self-similarity). Generally, time series without long-range dependence (i.e. momeryless) have a Hurst parameter of 0.5. Hurst parameters between 0.5 and 1 indicate a higher degree of long-range dependence (i.e., long memory). Table 2 shows the estimated Hurst parameters of the tested MPEG-4 frame size traces from pox plots of the R/S statistics as a function of aggregation level [22]. A larger H-value reflects a larger amount of movement in the video sequence [23, 25]. From the table, the four traces can be classified in terms of motion activity as follows:

- Low to moderate motion: Lecture room
- Moderate motion: Skiing
- Moderate to high motion: Star Wars
- High motion: Aladdin.

Therefore, we can safely assume that these movies are indeed representatives of almost all activity levels**.**

Table 2. Hurst parameters of MPEG-4 traces

	Hurst parameter (H)			
Aggregation level (frames)	**Aladdin**	**Star Wars**	**Lecture Room**	**Skiing**
1	0.919	0.903	0.571	0.848
12	0.896	0.838	0.739	0.787
50	0.891	0.808	0.713	0.734
100	0.906	0.785	0.716	0.727
200	0.909	0.776	0.722	0.696
300	0.904	0.765	0.71	0.702
400	0.882	0.756	0.714	0.697
500	0.894	0.758	0.719	0.729
600	0.928	0.752	0.721	0.751
700	0.857	0.727	0.723	0.711
800	0.845	0.722	0.718	0.764

3. Neuro-fuzzy Predictor

Fuzzy logic as a theory was formally introduced by Zadeh in the 1960's [26]. However, it was not widely applied till many years later. The first attempt in fuzzy modeling was formulated by Mamdani and, hence, called the Mamdani fuzzy model [27]. Such model imitates a human expert using a number of if-then rules. It is usually difficult to formalize this model as it is highly dependant on the particular application and heavily based on expert knowledge. In the process of designing the system, the designer should have more than adequate knowledge about the problem in order to specify the fuzzy membership functions for the inputs and the outputs, in addition to building the if-then rule-base. A typical rule in Mamdani's model has the form:

If (x is A) and (y is B) then (z is C)

where A, B, and C are all fuzzy sets. Despite the wide skepticism towards this model and the difficulties associated with it, it has been popular and its success has been indisputable.

An alternative model, called the Sugeno fuzzy model, came to overcome some of the drawbacks of Mamdani's model. Specifically, it came as an effort to provide a more systematic way to generate fuzzy rules. An important byproduct of this systemization was adaptation. A typical fuzzy rule in a Sugeno model is of the form:

If (x is A) and (y is B) then ($z=f(x,y)$)

Here, z is a crisp (non-fuzzy) number and f is a polynomial in x and y. Depending on the order of the polynomial, we have various types of Sugeno models. For example, if the output z is a constant (i.e., the polynomial is of order zero), the system becomes a zero-order Sugeno model. The structure of the Sugeno model can be viewed as a special case of gain scheduling where the gains depend on the input membership functions. Due to this quality, predictions, and modeling in general, can be easily incorporated with this system. Therefore, an efficient and simple blend that outperforms any linear modeling or prediction is reachable.

In fact, this is what Adaptive Neuro-fuzzy Inference Systems (ANFIS) were developed to do. ANFIS systems are simple fuzzy Sugeno models put in the framework of adaptive systems to facilitate learning and adaptation. The details of ANFIS can be found in [28]. The general idea, however, is simply to provide the system with only a set of input-output data along with the input membership functions. The system, then, is trained for a number of iterations (epochs) and generates the fuzzy rules and the output polynomial for each rule by minimizing an error measure between the system output and the actual output. Put in the framework of prediction, ANFIS can be implemented as shown in Fig. 3.

The size of the training data has an important effect on the generalization capabilities of the predictor. In general, the more training data is available, the better is the training. However, more training data means longer training time which is not desirable as it makes online implementation less feasible. In this paper, 1500 training points and 500 testing points are used to design the predictor. The number of points is small enough for very fast training and large enough to capture sufficient dynamics from the traces. A zero-order ANFIS predictor is designed with only three history points (three inputs). Each input is represented with three fuzzy membership functions. The structure of this predictor is shown in Fig. 4. The normalized membership functions for the three inputs and the output are depicted in Fig. 5. Note that since the model is of the zero-order Sugeno type, the output is a set of constant polynomials representing constant crisp values. Since there are three inputs with three membership functions each, the total number of rules is $3^3=27$ rules. These rules are generated during training and are listed in Table 3.

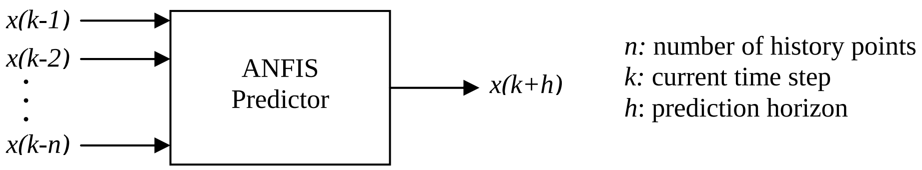


Fig. 3. Block diagram of the neuro-fuzzy predictor.

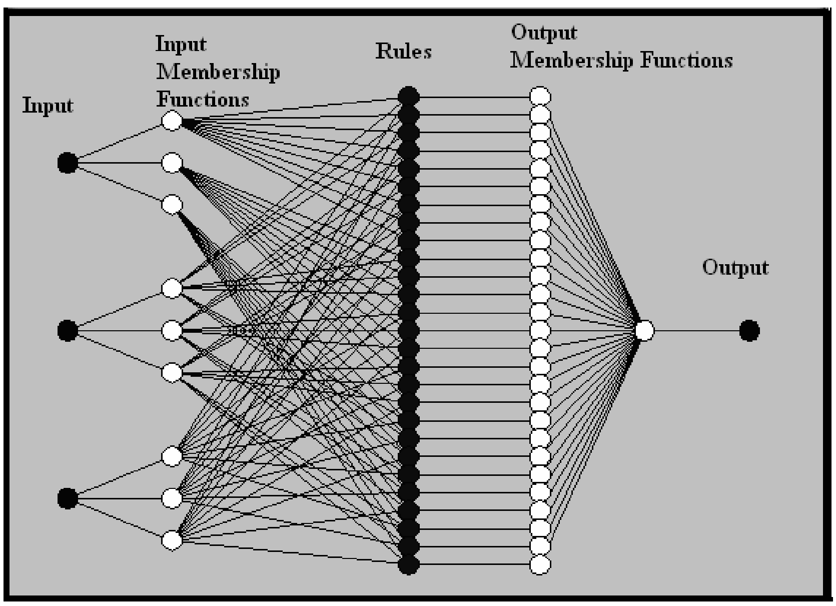


Fig. 4. Neuro-fuzzy predictor structure.

Table 3. ANFIS predictor fuzzy rules

Input 1 MF	input 2 MF	Input 3 MF	Output MF	Output value
1	1	1	1	0.242
1	1	2	2	0.282
1	1	3	3	0.271
1	2	1	4	0.555
1	2	2	5	0.369
1	2	3	6	0.363
1	3	1	7	0.437
1	3	2	8	0.514
1	3	3	9	0.000
2	1	1	10	0.259
2	1	2	11	-0.009
2	1	3	12	0.751
2	2	1	13	0.366
2	2	2	14	0.370
2	2	3	15	0.305
2	3	1	16	0.365
2	3	2	17	0.439
2	3	3	18	0.756
3	1	1	19	0.366
3	1	2	20	0.921
3	1	3	21	0.000
3	2	1	22	0.372
3	2	2	23	0.326
3	2	3	24	0.055
3	3	1	25	0.000
3	3	2	26	0.963
3	3	3	27	0.000

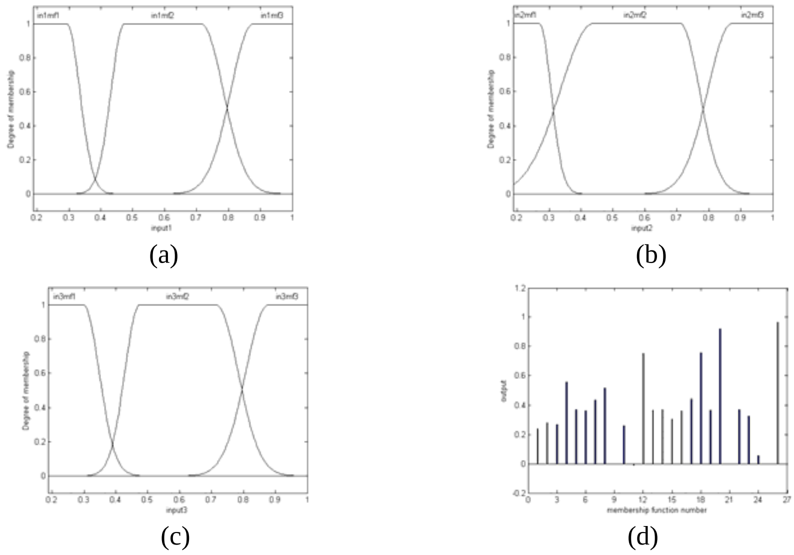


Fig. 5. Membership functions:
(a) input 1, (b) input 2, (c) input 3, (d) output.

4. Results

The ANFIS predictors are evaluated in terms of their performance in predicting MPEG-4 video traffic and compared to the normalized Least Mean Square (LMS) predictor developed in [29, 30] and redesigned in this work to be implemented with the used video traces. For fair comparison, the LMS is optimized in terms of its convergence rate. The convergence factor that gave the best performance in terms of minimum error was found to be 0.3. However, any value between zero and one could be used without significant degradation of the performance. The performance of the predictors is evaluated in terms of three measures: the Normalized Mean Square Error (NMSE), the Mean Square Error (MSE), and the Maximum Relative Error (MRE). These measures are given by:

$$NMSE = \sqrt{\frac{\sum (x-\hat{x})^2}{\sum (x-\bar{x})^2}} \qquad (1)$$

$$MSE(\%) = 100\frac{\sum (x-\hat{x})^2}{\sum x^2} \qquad (2)$$

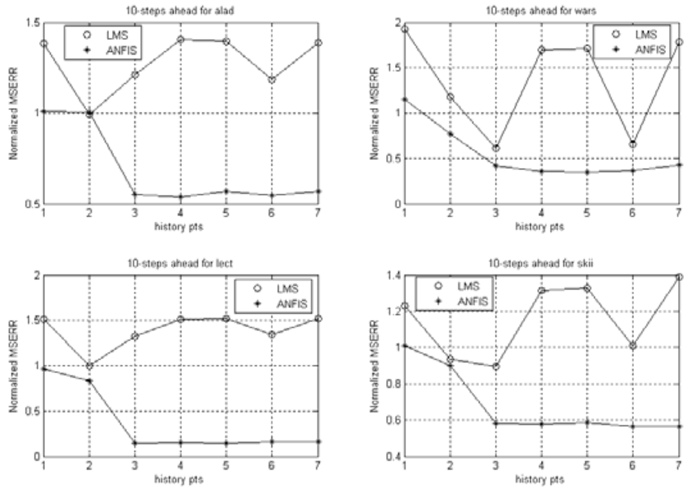
$$MRE = \max\left(\frac{|x-\hat{x}|}{|x|}\right) \tag{3}$$

where x, $\hat{x}$, and $\bar{x}$ denote the actual, predicted, and mean values, respectively. The NMSE is widely used in evaluating multistep prediction. For a perfect predictor, NMSE is zero. For a predictor predicting the statistical mean (trivial predictor), NMSE is equal to 1. Therefore, for a predictor to give better prediction than the trivial, the NMSE must be less than 1. The second performance measure (MSE) gives a quantitative measure about how close (on the average) are the outputs to the target. The MRE, on the other hand, gives a measure about the farthest point (or the largest error) which may be useful in problems related to bandwidth allocation and congestion avoidance. Needless to say that a network may give a very small MSE even if some finite data points are too far from the target; this could cause serious problems in practical applications.

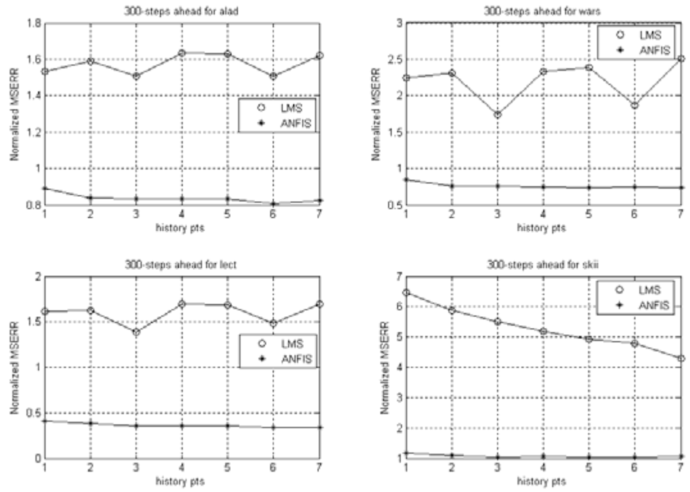
Effect of history points

Any prediction scheme relies on some sort of historical information. Less historical data normally leads to less reliable prediction; the reverse is not always true. However, if the scheme is provided with more historical data, the possibility of accurate prediction is higher. Unfortunately, the more history points are used, the more complex and computationally expensive the predictor becomes. Therefore, a typical trade-off does exist between accuracy and complexity. To examine the effect of history points on prediction results for the case of MPEG-4 videos, the normalized mean square error is examined into different scenarios: relatively short term (10 steps ahead) and relatively long term (300 steps ahead). The ANFIS performance in comparison with LMS is shown in Fig. 6. The figure reveals that the number of history points has an inconsistent effect on the LMS performance. However, having only three history points is sufficient for accurate prediction in the case of ANFIS. The LMS performance is expected to improve with a much more increase in history points because for very short term MPEG-4 traffic prediction, reference [5] reports that 10 to 15 history points are needed. For long term MPEG-1 traffic prediction, reference [21] reports the need for at least 24 history points. In light of this, ANFIS seems to have the power of capturing the trace dynamics with a much less number of inputs than reported techniques; resulting in a relatively simple predictor.

To evaluate the performance of the ANFIS, prediction is performed for horizons spanning 50 to 500 steps ahead. The NMSE as a function of prediction horizon, Fig. 7, shows that ANFIS is capable of giving excellent prediction results even for very long horizons. The LMS, on the other hand, performs worse than a trivial predictor for long term predictions. To more visually examine the performance, the 300 step-ahead predictions, in comparison with the actual traces, for Star Wars and Lecture room are shown in Fig. 8. The figure confirms the favorable results given by ANFIS. In order to quantitatively evaluate the performance, the NMSE, MSE, and MRE given by (1) to (3), are tabulated in Tables 4-6, respectively.



(a)



(b)

Fig. 6. Effect of history points on NMSE
(a) horizon=10, (b) horizon=300.

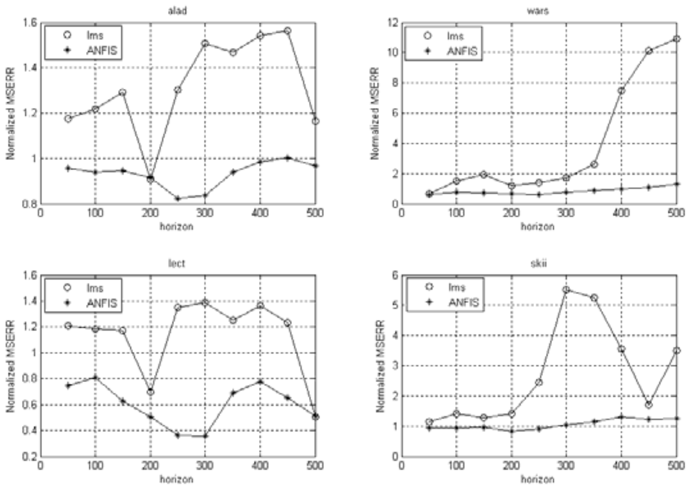
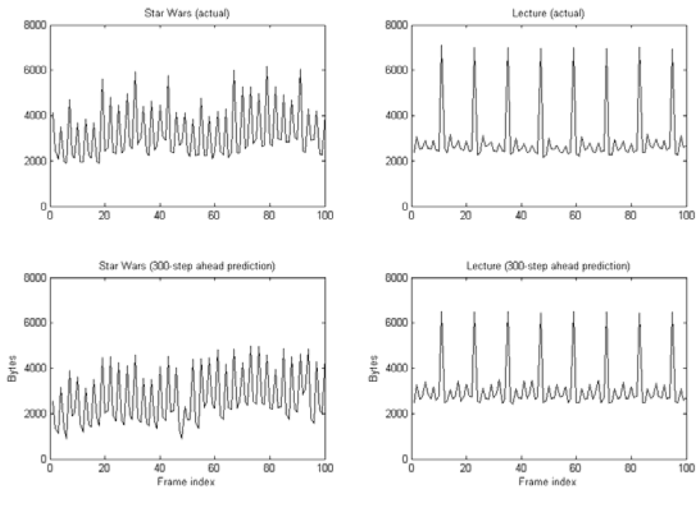


Fig. 7. NMSE for history=3.



(a) (b)

Fig. 8. A sample of prediction results
(a) Star Wars, (b) Lecture room.

Table 4. NMSE for different prediction horizons (history=3)

	100 steps		200 steps		300 steps		400 Steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	0.94	1.22	0.91	0.91	0.84	1.51	0.99	1.54	.97	1.17
Star Wars	0.75	1.48	0.65	1.19	0.76	1.73	0.98	7.48	1.29	10.92
Lect. Rm	0.81	1.18	0.51	0.7	0.36	1.39	0.78	1.36	0.51	0.51
Skiing	0.93	1.41	0.83	1.41	1.04	5.51	1.31	3.56	1.25	3.51

Table 5. MSE for different prediction horizons (history=3)

	100 steps		200 steps		300 steps		400 Steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	33	56	31	30	23	75	29	70	27	40
Star Wars	6	25	5	19	11	56	27	1547	63	4568
Lect. Rm	9	19	3	6	2	25	8	24	3	3
Skiing	7	16	5	16	15	412	29	221	25	202

Table 6. MRE for different prediction horizons (history=3)

	100 steps		200 steps		300 steps		400 Steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	3.74	3.55	4.14	3.12	2.79	2.03	5.03	1.40	4.28	11.85
Star Wars	0.66	1.70	1.12	2.36	2.26	4.66	4.72	27.95	10.23	77.22
Lect. Rm	0.79	1.36	0.55	1.00	0.33	1.81	0.72	1.43	0.83	1.00
Skiing	0.91	1.00	0.96	1.13	2.52	12.60	3.19	8.13	2.17	9.51

From these tables, few points can be drawn. It seems that low-order LMS predictors are not capable of providing meaningful results for long-term predictions. ANFIS, on the other hand, seems to have excellent results in terms of NMSE and MSE and acceptable results in terms of MRE. On the average (all four movies are taken into consideration), ANFIS performs well for less than 400 steps ahead. Beyond that, the performance is not as remarkable. The average MSE is less than 14% when the prediction horizon is less than 400 steps and increases to between 20% and 30% for longer horizons.

In terms of MRE, it is interesting to note that it is the only performance measure that is directly and consistently affected by with the motion activity in the traces. In other words, the higher the motion activity is, the higher the MRE, and vise-versa. On the average, the MRE is found to be less than 2 for horizons less than 400 steps. The error increases to about 4 beyond that. It is worth noting here that when training the ANFIS, the objective was to minimize the MSE between the ANFIS output and the actual output. No attention was given to the MRE. However, if MRE must be minimized (as could be the case for some applications), then, it should be the measure to be used during training.

Effect of smoothing

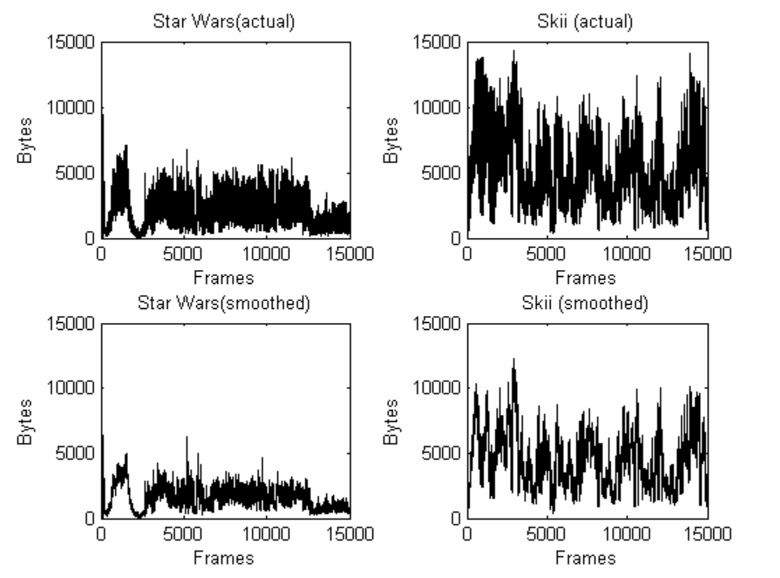
In this scenario, instead of using the actual video sequence, a smoothed version is considered. Since GOP series are very noisy, smoothing is expected to remove the undesired noise while preserving the dynamics of the traffic and the long-term dependencies. In addition, and particularly for long-term predictions, estimating the moving average frame size over a given averaging horizon has been claimed to be better for control and planning. The moving average as a function of the original series is given by [5]:

$$X(k) = \frac{1}{w} \sum_{j=kp-w+1}^{kp} x(j) \qquad (4)$$

where w is the window size, p is the amount by which the window is moved, and $x(j)$ is j^{th} GOP size of the video trace. Here, we will choose $p=1$ and $w=3$ (corresponding to quarter of the GOP length). The actual and smoothed traces for Star Wars and Skiing are shown in Fig. 9. It is clear from the figure that smoothing (even with a relatively small window size) greatly reduces the spikes in the traces.

To investigate the effect of smoothing, ANFIS is retrained with the smoothed data and the performance is evaluated in exactly the same scenarios used for the unsmoothed trace in the previous section. The NMSE's for the smoothed traces, as a function of prediction horizon, are shown in Fig. 10. From the figure and in comparison to Fig. 7, it is clear that smoothing deteriorates the performance of the predictor at least in terms of NMSE. This is the case even for the LMS. To further examine the ANFIS performance, the best and worst 300-step-ahead predictions, compared to the actual traces, are shown in Fig. 11. The figure confirms the unfavorable results caused by trace smoothing. It is worth noting here that the performance deteriorates with increasing window size.

In order to quantitatively evaluate the performance, the NMSE, MSE, and MRE for ANFIS and the LMS for the case of smoothed traces are tabulated in Tables 7-9, respectively. The tables expose the degradation of the performance caused by smoothing even in terms of MSE and MRE. The only unsurprising exception is the improvement of the MRE for the Lecture room trace. Such trace, unlike other traces, is characterized by a low to moderate level of activity as was shown in Table 2. The low level of activity results in limited and notable peaks. Smoothing such trace highlights these peaks even more as relatively smaller ones are reduced due to averaging. While this does not necessarily improve MSE and NMSE, it does improve the MRE.



(a) (b)

Fig. 9. Smoothing using moving average with window=3
(a) Star Wars, (b) Skiing.

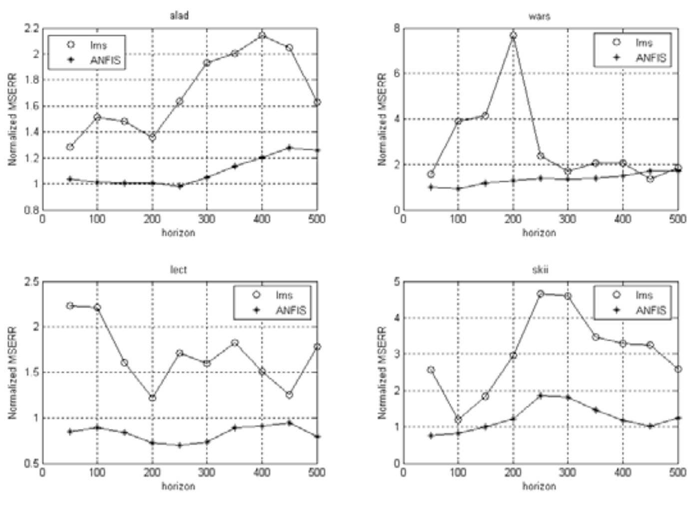


Fig. 10. NMSE for smoothed traces.

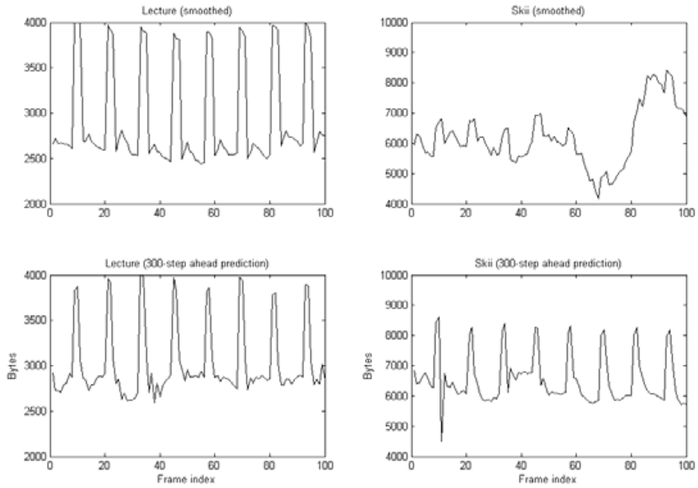


Fig. 11. Prediction results for the smoothed traces
(a) best result (Lecture Room), (b) worst result (Skiing).

Table 7. NMSE for smoothed traces

	100 steps		200 steps		300 steps		400 steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	1.02	1.51	1.01	1.36	1.05	1.93	1.20	2.14	1.26	1.63
Star Wars	0.93	3.90	1.28	7.66	1.35	1.70	1.49	2.04	1.70	1.87
Lect. Rm	0.89	2.21	0.72	1.22	0.74	1.60	0.91	1.51	0.80	1.78
Skiing	0.83	1.19	1.23	2.96	1.82	4.60	1.19	3.29	1.23	2.58

Table 8. MSE for smoothed traces

	100 steps		200 steps		300 steps		400 steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	22	49	23	42	21	71	22	71	22	36
Star Wars	16	284	25	906	28	44	33	62	61	73
Lect. Rm	3	17	2	5	2	9	3	8	2	11
Skiing	9	19	9	51	8	53	5	37	6	28

Table 9. MRE for smoothed traces

	100 steps		200 steps		300 steps		400 steps		500 steps	
	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS	ANFIS	LMS
Aladdin	1.98	1.85	2.18	3.40	1.72	1.46	1.66	1.28	1.64	6.64
Star Wars	2.51	9.55	3.38	21.27	2.73	2.78	8.92	5.44	35.78	9.04
Lect. Rm	0.57	1.26	0.38	1.00	0.40	1.00	0.43	1.00	0.43	1.20
Skiing	2.44	3.22	3.16	1.52	0.66	1.00	0.66	1.00	0.56	1.00

5. Conclusions

In this study, a neuro-fuzzy system is developed to predict MPEG-4 coded videos over a long horizon (up to few hundred frames ahead). Predicting such traffic and over such a long horizon is important for today's fast networks and internet multimedia services. Particularly, it is of importance to many network-edge applications such as dynamic bandwidth allocation, quality of service (QoS) control, and network management and planning. The results presented in this paper indicate that neuro-fuzzy systems can be very effective in long-term prediction of this type of traffic. Such promising results are achieved with a simple architecture having only three inputs and one output. The simplicity of the system does not only simplify the design but also makes the execution fast. The results are evaluated using four entertainment movies representing different activity levels ranging from low to high. When exploring the effect of smoothing the traces using moving average, it was concluded that such smoothing deteriorates the results.

References

[1] Drossu, R., Lakshman, T., Obradovic, Z. and Raghavendra, C. "Single and Multiple Frame Video Traffic Prediction Using Neural Network Model." *Computer Networks: Architecture and Applications*, Chapman Hall, Chapter 9 (1995), 146-158.

[2] Chang, P. and Hu, J. "Optimal Non-linear Adaptive Prediction and Modeling of MPEG Video in ATM Networks Using Pipelined Recurrent Neural Networks." *IEEE J. Selected. Areas. Commun.*, 15, No. 6 (1997), 1087-1100.

[3] Doulamis, A., Doulamis, N. and Kollias, S. "An Adaptable Neural Network Model for Recursive Non-linear Traffic Prediction and Modeling of MPEG Video Sources." *IEEE Trans. on Neural Networks*, 14, No. 1 (2003), 150-166.

[4] Doulamis, A., Doulamis, N. and Kollias, S. "Recursive Non-linear Models for On-line Traffic Prediction of VBR MPEG Coded Video Sources." *IEEE-INNS-ENNS Int Conf. Multimedia Networks*, Italy (2000), 114-116.

[5] Bhattacharya, A., Parlos, A. and Atiya, A. "Prediction of MPEG-coded Video Source Traffic Using Recurrent Neural Network." *IEEE Trans. on Signal Processing*, 51, No. 8 (2003), 2177-2190.

[6] Khotanzad, A. and Sadek, N. "Multi-scale High Speed Network Traffic Prediction Using Combination of Neural Networks." *International Joint Conference on Neural Networks (IJCNN03)*, Portland, Oregon (2003), 1071-1075.

[7] Sang, A. and Li, S. "A Predictability Analysis of Network Traffic." *IEEE INFOCOM* (2000), 342-351.

[8] Qiu, B., Suthaharna, S. and Wu, H. "Neuro-fuzzy Modeling and Prediction of VBR MPEG Video Sources." *ACM*, 5 (1999), 207-215.

[9] Alheraish, A. and Abdennour, A. "Assessment and Comparison of Neural Networks Techniques for MPEG-4 Video Traffic Prediction." *First IEEE International Conference on Modeling, Simulation and Applied Optimization ICMSAO*, Sharajh, UAE (2005).

[10] Habib, I., Tarraf, A. and Saadwi, T. "A Novel Neural Network Controller Using Reinforcement Learning Method." *IEEE J. Selected. Areas. Commun.*, 12, No. 6 (1994), 988-1007.

[11] Li, H., Liu, G., Zhang, Z. and Li, Y. "Adaptive Scene Detection Algorithm for VBR Video Stream." *IEEE Trans. on Multimedia*, 6, No. 4 (2004), 624-633.

[12] Aweya, J., Montuno, D.Y., Zhang, Q.J. and Orozco-Barbosa, L. "Multi-step Neural Predictive Techniques for Congestion Control - Part 1: Prediction and Control Models." *International Journal on Parallel and Distributed Systems and Networks*, 3, No. 1 (2000), 1-8.

[13] Aweya, J., Montuno, D.Y., Zhang, Q.J. and Orozco-Barbosa, L. "Multi-step Neural Predictive Techniques for Congestion Control - Part 2: Control Procedures." *International Journal on Parallel and Distributed Systems and Networks*, 3, No. 3 (2000), 139-143.

[14] Habib, I., Tarraf, A. and Saadwi, T. “A Neural Network Controller for Congestion Control in ATM Multiplexer.” *Computer Networks*, 29, No. 3 (1997), 325-334.

[15] Chong, S., Li, S.Q. and Ghosh, J. "Predictive Dynamic Bandwidth Allocation for Efficient Transport of Real-time VBR Video over ATM." *IEEE J. Selected Areas in Communication,* 13, No. 1 (1995), 12-23.

[16] Casetti, C., Kurose, J. and Towsley, D. “An Adaptive Algorithm for Measurement-based Admission Control in Integrated Services Packet Networks.” *Computer Communications*, 23, No.14-15 (2000), 363-1376.

[17] Onyiagha, C., Kransiqi, X. and Clarkson, T. “Adaptive Access Control of ATM Traffic Using Neural Network.” *IEEE Global Telecommunication Conference,* 1 (1996), 201-205.

[18] Kim, M. and Noble, B. "Mobile Network Estimation." *Proceedings of the Seventh Annual Conference on Mobile Computing and Networking,* (2001), 298-309.

[19] Bush, S. “Active Virtual Network Management Prediction: Complexity as a Framework for Prediction, Optimization, and Assurance.” *IEEE Computer Society Press, Proceedings of the 2002 DARPA Active Networks Conference and Exposition (DANCE 2002),* San Francisco, USA (2002), 534-553.

[20] Mohamed, S., Rubino, G., Afifi, H. and Cervantes, F. “Real-time Video Quality Assessment in Packet Networks: A Neural Network Model.” *Proc. PDPTA’01*, Las Vegas, NV (2001), 102-111.

[21] Liang, Y. “Real-time VBR Video Traffic Prediction For Dynamic Bandwidth Allocation.” *IEEE Trans. on Systems, Man, and Cybernetics*, 34, No. 1 (2004), 32-47.

[22] Fitzek, F. “MPEG-4 and H.263 Video Traces for Network Performance Evaluation.” Technical University, Berlin (2001), available online at: *http://www-tkn.ee.tu-berlin.de/research/trace/trace.html*.

[23] Wang, L. and McCrosky, C. “Self-similarity of Aggregated Network Traffic Models." *Technical Report*, Department of Computer Science, University of Saskatchewan, Canada (1995).

[24] Beran, J., Sherman, R., Taqqu, M.S. and Willinger, W. "Long-range Dependence in Variable-bit-rate Video Traffic.” *IEEE Transactions on Communications*, 43, No. 2/3/4 (1995), 1566-1579.

[25] Rose, O. "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems.” University of Wuerzburg institute of Computer Science Research Report Series, Report No. 101 (1995), 1-25.

[26] Zadeh, L.A. "Fuzzy Sets.” *Information and Control*, 8 (1965), 338-353.

[27] Mamdani, E.H. "Applications of Fuzzy Algorithms for Simple Dynamic Plant.” *Proceedings of IEE,* 121, No 12 (1974), 1585-1588.

[28] Jang, J.S.R. “ANFIS: Adaptive-network-based Fuzzy Inference System.” *IEEE Transactions on Systems, Man and Cybernetics*, 23 (1993), 665 – 685.

[29] Adas, A. “Using Adaptive Linear Prediction to Support Real-time VBR Video Under RCBR Network Service Model.” *IEEE/ACM Trans. Networking*, 6, No. 5 (1998), 635-644.

[30] Yoo, S. “Efficient Traffic Prediction Scheme for Real Time VBR MPEG Video Transmission over High Speed Networks.” *IEEE Trans. on Broadcasting*, 48 (2002), 10-18.

التوقع على ا دى البعيد ركة نقـل صـور الفيديو باستخدام الشبكات العصبية الاصطناعية و منطق الغموض

عادل عبدالنور بن عبدالنور

قسم الهندسة الكهربائية، كلية الهندسة، جامعة الملك سعود، ص.ب ٨٠٠،
الرياض ١١٤٢١، المملكة العربية السعودية

(قدِّم للنشر في ٢٠/٠٤/٢٠٠٥م؛ وقبل للنشر في ٠٣/٠٧/٢٠٠٥م)

ملخص البحث. يهتم هذا البحث بموضوع التوقع على المدى البعيد لحركة نقل صور الفيديو المرمزة بنظام الـ MPEG وتحديدا الترميز الجديد المعروف بـ MPEG-4. يحظى هذا النوع من التوقع بأهمية كبيرة خصوصا مع التطور المطرد لتقنيات الاتصالات الحديثة كالشبكات السريعة والإنترنت والخدمات متعددة الوسائط. من أهم أسباب الاهتمام بهذا الموضوع هو أن التوقع الدقيق وعلى مدى زمني بعيد لحركة نقل صور الفيديو يزيد من فعالية ادارة وتخطيط الشبكات، ويحسن جودة خدمات، إضافة إلى مساهمته في تقسيم النطاق الترددي بأكثر فعالية. مع هذا، تكاد تكون البحوث في هذا الميدان معدومة نظرا لصعوبته خصوصا باستعمال الطرق التقليدية للاستقراء والتوقع. تم في هذا البحث تطبيق نظام توقع باستخدام الشبكات العصبية الاصطناعية ومنطق الغموض، كما تم تقييم أداء هذا النظام وتطبيقه على أفلام فيديو حقيقية وتبين أنه أحسن أداء مقارنة بالطرق الخطية حيث تمكن هذا النظام من التوقع الدقيق وعلى مدى زمني تجاوز بضع مئات من الخطوات.