Video Compression with Wavelets and Random Neural Network Approximations

Fan Hai, Khaled Hussain, Erol Gelenbe, Ratan Guha

School of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL 32816
Telephone: (407) 823-2341, Fax (407) 823-5419
Email: {fhai, khaled, erol, guha}@cs.ucf.edu

ABSTRACT

Modern video encoding techniques generate variable bit rates, because they take advantage of different rates of motion in scenes, in addition to using lossy compression within individual frames. We have introduced a novel method for video compression based on temporal subsampling of video frames, and for video frame reconstruction using neural network based function approximations. In this paper we describe another method using wavelets for still image compression of frames, and function approximations for the reconstruction of subsampled frames. We evaluated the performance of the method in terms of observed traffic characteristics for the resulting compressed and subsampled frames, and in terms of quality versus compression ratio curves with real video image sequences. Comparisons are presented with other standard methods.

Keywords: Wavelets, Video Compression, Neural Networks

http://www.ixueshu.com

1. INTRODUCTION

In modern real-time network traffic, video is a major source of information due to distance learning, video-conferencing and similar applications. Video consumers are faced with a variety of networks through which they will transmit and receive video, whose quality must remain satisfactory. The end-to-end quality of service (QoS) is highly sensitive to bit rates and packet rates which are available, and also to other network performance matrices. Video compression standards use motion prediction and compensation to encode the sequence of frames. Such techniques create additional performance constraints since they generate encoded frames of varying types and sizes, and place a major burden on the correct reception of certain critical frames. Network performance and characteristics compound these associated problems of video compression. In a recent publication [7] we demonstrated the feasibility of dropping entire frames from a video sequence during compression and reconstructing them upon decompression. Thus, significant reduction of the network traffic is achieved by dropping frames from the sender and constructing them at the receiver. Random neural network based techniques have been used for subsampling of frames on the sender site and interpolating the missing frames on the receiver site. In this paper, we extend our work by applying a wavelet based compression technique to subsampled frames. We evaluate the performance of the method in terms of observed traffic characteristics for the resulting compressed and subsampled frames, and in terms of quality versus compression ratio curves with real video image sequences.

In Section 2, we propose a video compression scheme based on random neural networks and fast wavelet transforms. In Section 3, we present experimental results of our method and several other techniques. Finally in Section 4, we state our conclusion and indicate our continuing work in this area.

2. A VIDEO COMPRESSION SCHEME

In this Section, we present a video compression scheme using video reconstruction with random neural networks from temporally subsampled frames, with a wavelet based image compression for still frames. The outline of the proposed method is shown in Figure 1 and the detailed treatment of the scheme is presented in the following subsections. In the first stage, a majority of frames are dropped, leaving one of the S frames for transmission. We consider the consecutive values of any pixel in a video sequence as a function of time. Because the video sequence represents some reality, time is continuous in the abstract.

The pixel value function can be computed by knowing at only a few points in a time interval. Details are presented by Cramer and Gelenbe[7]. Once a temporal subselection is made, the frame is compressed by a lifting scheme, a fast second generation wavelet technique. This compressed frame is transmitted over the network. At the receiving end, the compressed frame is decompressed and finally frames are reconstructed from the subsampled frames.

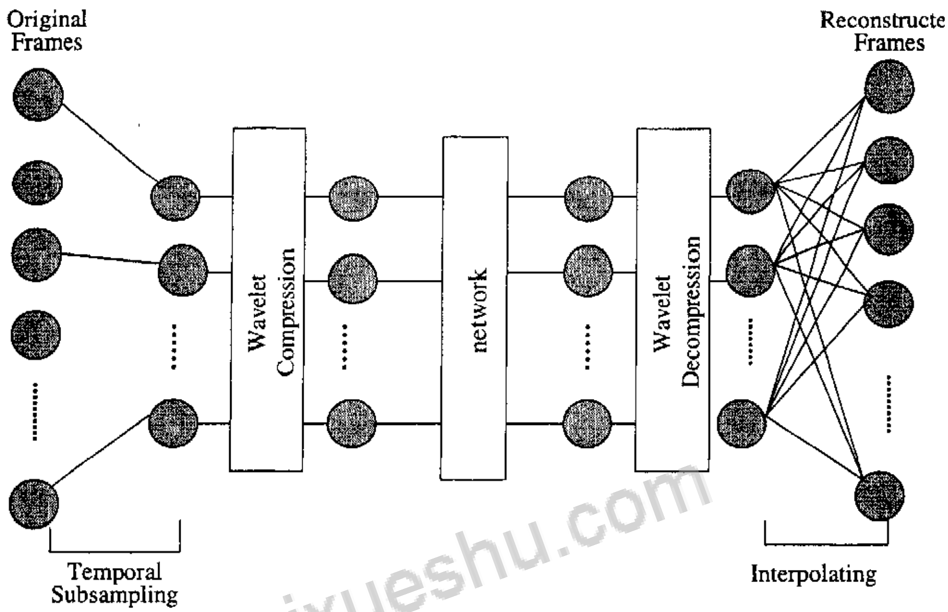


Figure 1: Video compression scheme

http://www.ixueshu.com

2.1. Random Neural Network Model

The random neural network model, introduced by Gelenbe (1989, 1993), consists of n neurons. The state of each neuron i at time t is represented by $k_i(t)$, where $k_i(t)$ is the potential of neuron i. When an excitation signal arrives to neuron i, from neuron j, the potential of neuron i will be increased by one and the potential of neuron j will be decreased by one. When an inhibitory signal arrives to neuron i from neuron j, the potential of neuron i will be reduced by one if k_i is positive, and the potential of neuron j will be reduced by one. When an exogenous excitatory signal arrives to neuron i at rate Λ_i, the potential of neuron i will be increased by one. When an exogenous inhibitory signal arrives at neuron i at rate λ_i, the potential of neuron i will be reduced by one if its potential is positive. The neuron i can fire if its potential is positive. The output spikes from neuron i will be sent out at rate r_i, with independent, identically, and exponentially distributed inter-spike intervals. These spikes will go out to neuron j with probability p_{ij}^+ as excitatory signals, or with probability p_{ij}^- as inhibitory signals. The probability that these signals are sent out of the network is d_i. Thus, we should have $d_i + \sum_{j=1}^{n}[p_{ij}^+ + p_{ij}^-] = 1$. The stationary probability distribution associated with this model is given by:

$$p(k) = \lim_{t\to\infty} p(k,t), \quad q_i = \lim_{t\to\infty} q_i(t), \; i = 1,\ldots\ldots,n. \tag{1}$$

$$\lambda_i^+ = \sum_j q_j w_{ji}^+ + \Lambda_i, \qquad \lambda_i^- = \sum_j q_j w_{ji}^- + \lambda_i, \tag{2}$$

$$q_i = \frac{\lambda_i^+}{r_i + \lambda_i^-} \tag{3}$$

$$r_i = \sum_i w_{ij}^+ + w_{ij}^- \tag{4}$$

where $q_i(t)$ is the probability that neuron i is excited at time t and p(k,t)=Pr(k(t)=k) is the probability distribution of the network state.

2.2 Video Compression and Reconstruction with RNN[7]

We consider the consecutive values of any pixel I (x, y) in a video sequence as a function f (x, y, t) of time t. Suppose that f had the property that its values for all t∈[a, b] can be computed by knowing f at only a few points in [a, b], then this would provide us with a very powerful tool for reconstructing the video form just a few reliable frames.

We will use neural network function approximation techniques which will construct an approximation techniques which will construct an approximation g (x, y, t) to the function f (x, y, t) in an interval t∈[a, b], using only a "small" finite number of values f (x, y, ti) with i=1,..., m ant t_i∈[a,b]. The implicit assumption will be that f (x, y, t) is continuous, which is reasonable in view of the fact that video sequences represent the motion of natural objects in a natural scene. The construction of the approximation function g (x, y, t) will use RNN function approximation[6] using a small number of video frames which act as training samples in supervised learning. It would obviously be impractical to train a different network for each pixel position (x, y). Fortunately, it suffices to train a distinct network for pixel blocks[7] which include 64 pixels.

http://www.ixueshu.com

2.3 A Wavelet Based Image Compression

For wavelet transform, we start from a function $\psi(t)$ as a modulation function and obtain a family of functions from ψ by varying the scale. We fix $p \geq 0$ and for all $s \in R, s \neq 0$, and define:

$$\psi_s(u) = |s|^{-p}\,\psi(\frac{u}{s}) = \frac{1}{|s|^p}\psi(\frac{u}{s}).$$

Next we localize each function ψ_s in time:

$$\psi_{s,t}(u) = \psi_s(u-t) = |s|^{-p}\,\psi(\frac{u-t}{s}) = \frac{1}{|s|^p}\psi(\frac{u-t}{s}).$$

Note that if $\psi \in L_2(R)$, then $\psi_{s,t} \in L_2(R)$, and

$$\|\psi_{s,t}\|^2 = |s|^{1-2p}\|\psi\|^2.$$

By taking p = ½, we have $\|\psi_{s,t}\| = |\psi|$.

Now we can define a transform on $L_2(R)$ using functions from the family $\psi_{s,t}$ as the modulating functions:

$$\hat{f}(s,t) = \int_{-\infty}^{+\infty} f(u)\psi_{s,t}(u)du.$$

This transform is known as *the wavelet transform.* The function ψ is called the "*mother wavelet*". The term *wavelet* corresponds to sets of functions $\psi_{s,t}$, which are derived from ψ by dilations and translations. Here we describe our wavelet based compression. A wavelet compression algorithm essentially consists of three steps: transform, quantization and encoding [10]. This is shown in Figure 2.

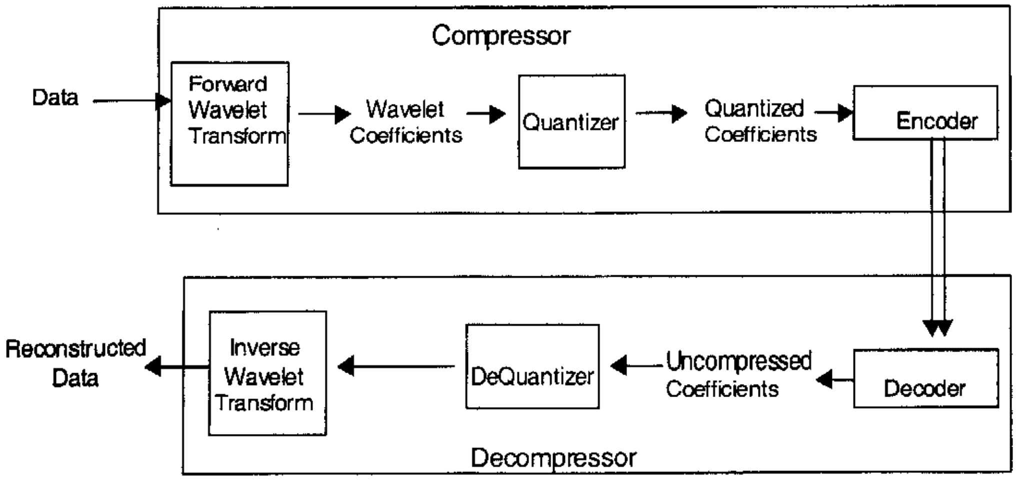


Figure 2. Scheme of Wavelet Encoding/Decoding

In choosing a particular wavelet one has to consider issues like compactness of support, smoothness, number of vanishing moments, length of filters, etc.. The transformed coefficients can have nearly arbitrary values. The purpose of quantization is to restrict the values of the coefficients to a limited number of possibilities. Wavelet thresholding can be applied in this step. The encoding step involves efficiently replacing the string of input symbols coming from the quantizer with a bit stream.

A new way of constructing a biorthogonal wavelet has been provided by the lifting scheme [9 - 11]. The lifting scheme was originally designed for the construction of families of wavelets which are independent of the Fourier transform, or the so called "*second generation wavelets.*" It is observed that this construction offers certain advantages over the classical wavelets (the first generation wavelets). These advantages include the following: a fully in-place calculation of the wavelet transform, easy inverse transform, and the construction of wavelets entirely in the spatial domains. In this paper, we use lifting scheme for wavelet transform since high compression speed is very important for video transmission.

3. EXPERIMENTAL EVALUATION

In this section, we evaluate the described schemes from an experimental standpoint. We will consider both the video quality and the resulting traffic characteristics.

3.1 Video Quality and Compression Ratios for Interpolated Sequences

The video quality, determined by the difference between the interpolated video sequence and the original sequence before it is subsampled, cannot be considered in isolation. It must be considered with the resulting compression ratio. Therefore the performance measure commonly used is the quality-versus-compression-ratios (QvsCR) characteristic, which plots the observed PSNR in decibels averaged frame by frame for all the test sequence, against the measured average compression ratio for the same video sequence. Figure 3 shows the QvsCR performance for the proposed method for S = 1, 3, and 6. A comparison between the proposed method (S=3) and the H.261 for Miss America sequence is shown in Figure 4. In Figure 5, we show three original and reconstructed frames in the Miss America sequence using our proposed method.

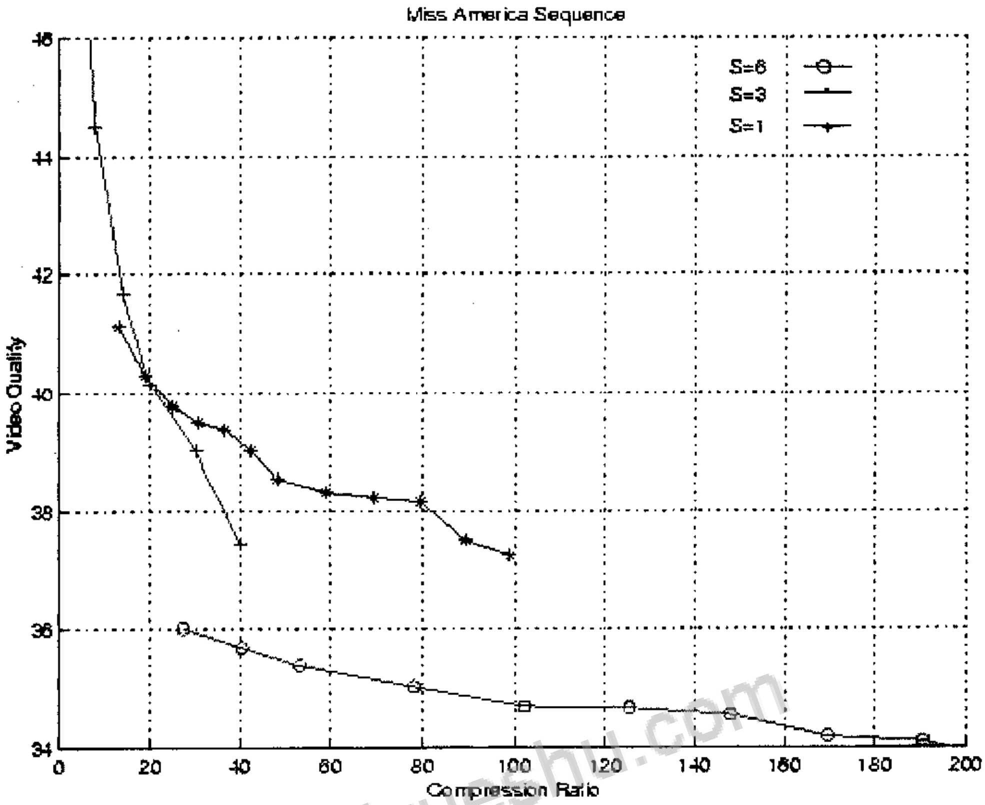


Figure 3: Video quality versus compression ratio with S=1, S=3, and S=6 for the proposed method.

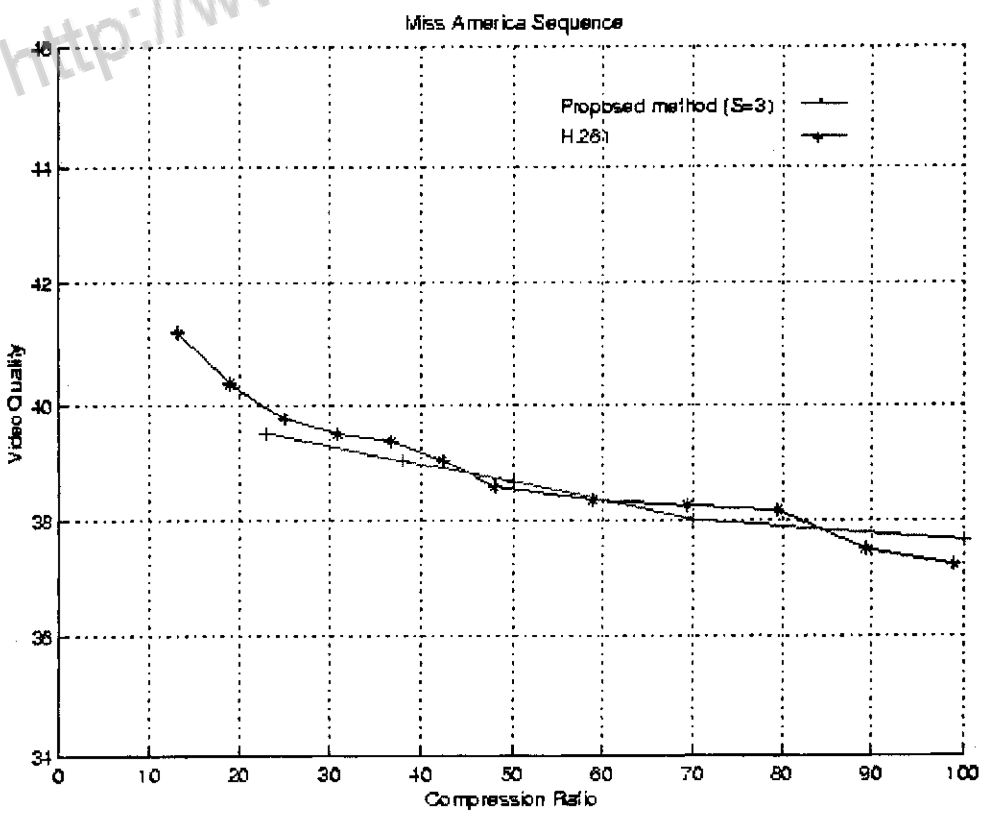
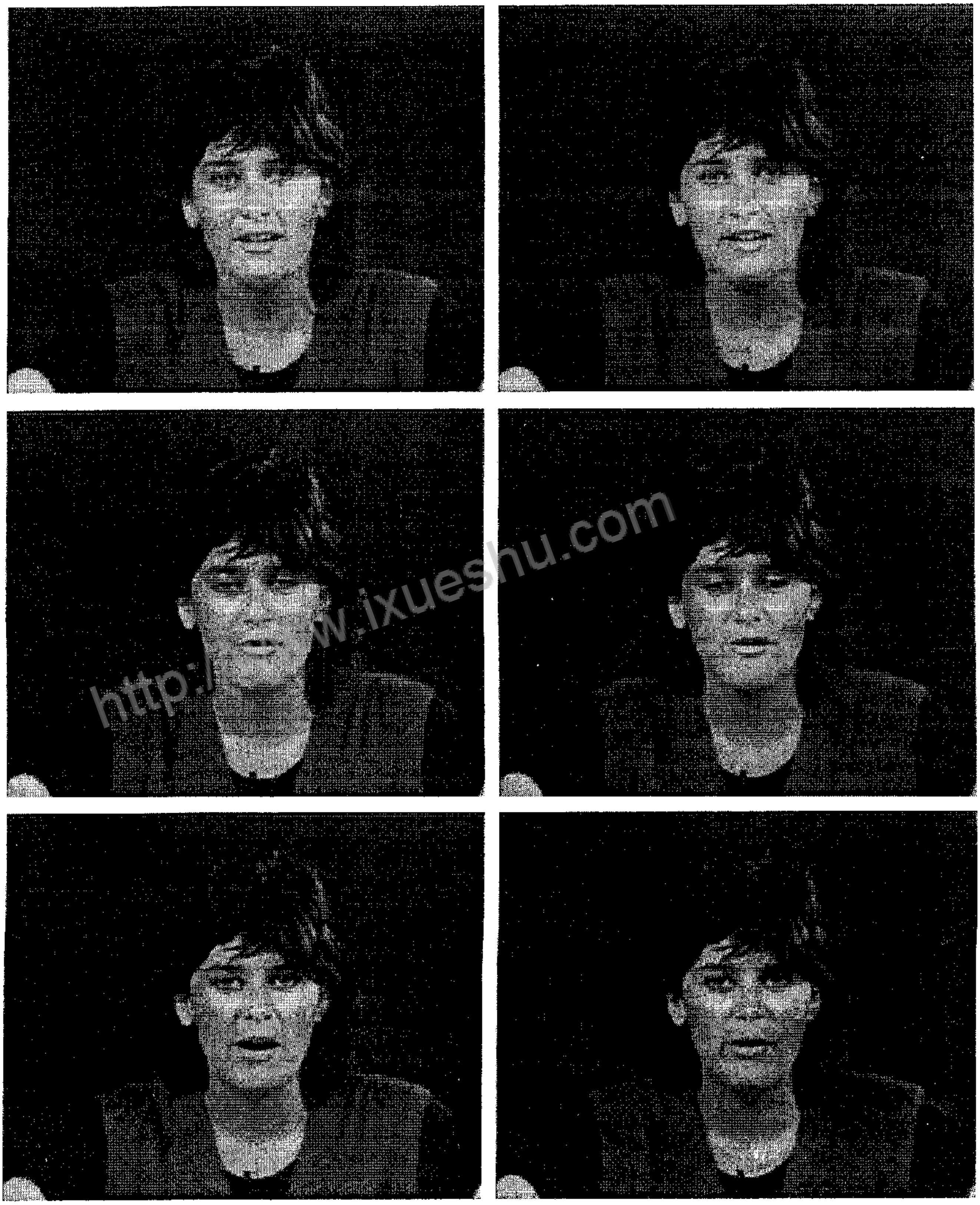


Figure 4: Comparison between the proposed method (S=3) and the H.261 for Miss America sequence.

http://www.ixueshu.com



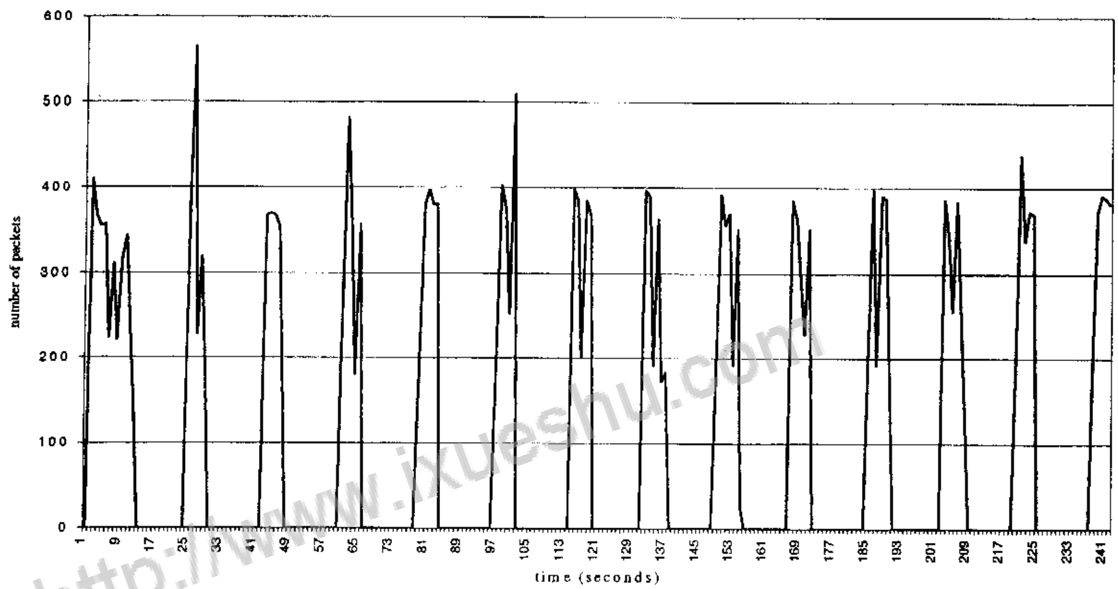
(a) (b)

Figure 5: (a) Original and (b) reconstructed frames in the Miss America (CR = 99).

http://www.ixueshu.com

3.2 Traffic Characteristics of the Video Sequence

The packet rates for the compressed Miss America sequence, where frames are initially compressed using wavelets with subsamplings S = 3 and 6, are shown in Figure 6 and 7. Figure 6 shows the traffic characteristics of the video sequence of Miss America over a 241-second time period, whereas Figure 7 shows the result over a 100-second time period. We obtain time varying packet rates in both cases with a significant compression obtained for the non–subsampled video sequence, indicating the importance of subsampling in the reduction of overall traffic.



http://www.ixueshu.com

Figure 6: Packet-rates versus time for subsampled with S = 3 compressed with wavelets

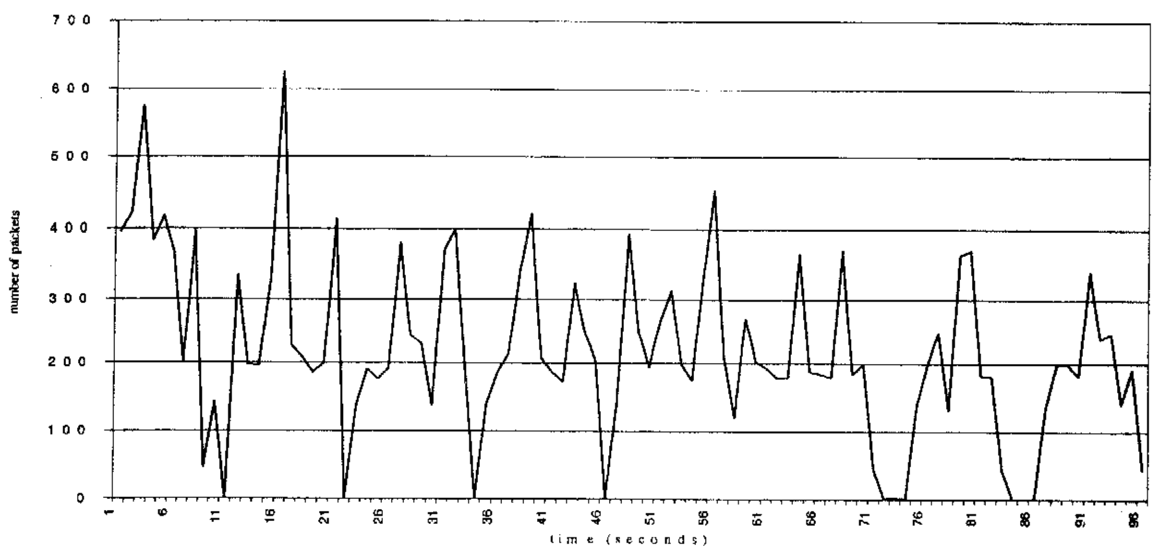


Figure 7: Packet-rates versus time for subsampled with S = 6 compressed with wavelets

These results show that subsampling can provide signicificant reductions in overall traffic. In the previous subsection, we show that this reduction in traffic does not diminish video quality.

4. CONCLUSION

This paper explores the idea of using frame interpolation with wavelets, as a means to build video receivers which reconstitute video sequences having high frame rates with good quality, even when a very significant fraction of frames may have been dropped either at the input of the network or during transmission.

We are continuing our work on video quality. We fix the compression level and allow the image quality factor (expressed as target PSNR) to be varied, resulting in variations in quality and compression ratios. For H.261, the target compression ratio is varied, resulting in variation of the image quality. In the wavelets case, we fix the wavelets used as D4, and vary the target compression ratio to obtain results in variations of quality. In order to base our results on viable video compression techniques, we compare between our method and the H.261.

ACKNOWLEDGEMENTS

The authors acknowledge the partial support provided by the Army Research Office (Contract No. DAAH04-95-1-0250). The views and conclusions herein are those of the authors and do not represent the official policies of the funding agency.

REFERENCES

1. E. Gelenbe, "Random Neural Networks with Negative and Positive Signals and Product Form Solution", *Neural Comput*, **1**, 4, pp 502-511, 1989.
2. E. Gelenbe, "Learning in the Recurrent Random Neural Network", *Neural Comput*, **5**, **1**, pp 154-164, 1993.
3. E. Gelenbe, C. Cramer, M. Sungur, P. Gelenbe "Traffic and video quality in adaptive neural compression", *Multimedia Systems* , **4**, pp. 357-369, 1996.
4. E. Gelenbe, "User oriented video: quality, traffic and computational trade-offs", Invited Plenary Paper, *Proceedings of International Council for Computer Communications* (ICCC'97), 1997.
5. C. Cramer, E. Gelenbe, P. Gelenbe "Image and video compression", *IEEE Potentials* , February/March 1998.
6. Y. Feng and E. Gelenbe "Adaptive Object Tracking and Video Compression", *Networks and Information Systems* , **1**, 1, 1998.
7. E. Gelenbe "The spiked Random Neural Network model: algebraic non-linearities, computational learning", Invited Plenary Lecture, *CNNA98: 5th IEEE International Workshop on Cellular Neural Networks*, London, South Bank University, April 14-17, 1998.
8. E. Gelenbe, Z.H. Mao, Y.D. Li, "Function approximation by spiked random networks", *IEEE Transactions on Neural Networks*, **10**, 1, pp. 3-9, 1999.
9. C. Cramer, E. Gelenbe, "Video quality and traffic QoS in learning based subsampled and receiver-interpolated video sequences"', *IEEE Journal of Selected Areas in Communications*, **18**, **2**, pp. 150-166, 2000.
10. F. Hai, *Wavelet Thresholding Algorithms for Terrain Data Compression*, M. S. Thesis, SEECS, University of Central Florida, 2000.
11. G. Fernandez, S. Periaswamy and W. Sweldens, *" LIFTPACK: A Software Package for Wavelet Transforms using Lifting"*, in M. Unser, A. Aldroubi, and A. F. Laine, editors, Wavelet Applications in Signal and Image Processing IV, pp. 396-408, *Proc. SPIE 2825*, 1996.
12. W. Sweldens, " Wavelets, Signal Compression and Image Processing", Course notes:wavelets and their applications in computer graphics, *SIGGRAPH* 1994.
13. W. Sweldens, "The Lifting Scheme: A New Philosophy in Biorthogonal Wavelet *Constructions"*, wavelet application in signal and image processing III, pp.68-79. *Proc. SPIE 2569*, 1995b.
14. I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure and Appl. Math, **41**, pp. 909-996, 1988.

http://www.ixueshu.com

免费论文查重：http://www.paperyy.com

3亿免费文献下载：http://www.ixueshu.com

超值论文自动降重：http://www.paperyy.com/reduce_repetition

PPT免费模版下载：http://ppt.ixueshu.com