ON THE ENCODING OF THE ANCHOR FRAME IN VIDEO CODING

Lisimachos P. Kondi and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering,
Northwestern University, Evanston, IL 60208, USA

In this paper we determine the number of bits to be used for the encoding of the anchor frame in low bit rate video coding in order to improve the quality of the next frame to be encoded and, subsequently, the quality of the next frames to be encoded. We use a progressive method for the transmission of the anchor frame and we develop two methods for determining the optimal number of bits to be allocated to the first frame in real time.

INTRODUCTION

In most approaches known today for low bit rate real time video coding, no particular attention is given to the effect of the number of bits allocated to the first (intra, anchor) frame of the image sequence on the overall quality of the reconstructed sequence. Assuming that we are using a constant bit rate channel, the number of bits used for the anchor frame corresponds to a certain time delay for the encoding of the next frame. The time delay depends on the capacity of the constant bit rate channel and the number of bits used for the anchor frame.

Our goal in this paper is to improve the quality of the image sequence at the beginning of the transmission by using a variable number of bits for the anchor frame and exploiting the tradeoff involved. The results of this work can be used in Consumer Electronics applications, such as, videophone applications.

If we spend a lot of bits for the anchor frame, we will get a better quality of the reconstructed anchor frame at the decoder. In principle, better quality of the reconstructed anchor frame will lead to better quality of the next reconstructed (inter) frame, as long as the time delay is not that great that will cause the correlation between the two frames to be too low. Thus, the time used for the anchor frame should be large enough to yield a good quality reconstructed anchor frame but not too large so that the first frame and the frame to be coded have low correlation. Therefore, there is an optimal point where we must stop encoding the anchor frame and start encoding the next frame of the image sequence. This point will yield the best quality for the second coded frame. We expect that this will lead to improved quality for the rest of the frames since all of them are coded using prediction based on the previous reconstructed frame.

The decision on when we should stop coding the anchor frame and start coding the next frame should be done in real time, after the transmission of the anchor frame has started. In order to be able to stop at any time, we need to use a progressive method for the transmission of the anchor frame. In this work, we modify the H.263 codec to use the EZW (Embedded Zerotree Wavelet) [1, 2] method for the transmission of the anchor frame.

FORMULATION OF THE PROBLEM

Assume that we want to code an image sequence with a given frame rate in real time over a constant bit rate channel. Let SNR1 be the PSNR (Peak Signal-to-Noise Ratio) of the recostructed intra frame. Let SNR2 be the PSNR of the next reconstructed frame. Our goal is to maximize SNR2 provided that the time delay is acceptable and about the same number of bits is used for the next (second encoded) frame at all times to have a fair comparison.

The solution we propose is to treat SNR1 and SNR2 as time series SNR1[n] and SNR2[n] with index the number of source frames that correspond to the bits alloted to the first frame. We assume that we are interested in the first maximum of SNR2[n].

Past Values of SNR2[n] are Known

As noted previously, we are interested in the first maximum of SNR2[n]. Experimental results show that SNR2[n] is virtually always monotonically increasing up to its first maximum. Thus, we can wait until SNR2[n] starts to decrease and stop at that point. The index n we get using this method is the "optimal" n plus one.

For this method, we need to calculate the value of SNR2[n] for every n in real time. This means that we need to perform the motion compensation, block transform and reconstruction of the second frame for each value of n.

SNR2[n] is not Known

Let us suppose that we cannot calculate SNR2[n] in real time, but we have SNR1[n] instead. Evaluation of SNR1[n] requires far less computation, thus it is more realistic to assume that we have SNR1[n] but not SNR2[n]. However, it is clear that we need a second piece of data in order to estimate the index n which results in the maximum SNR2[n].

We expect that SNR2[n] is a function of SNR1[n] and the correlation between the first frame and frame n. We propose the use of the energy of the DFD (Displaced Frame Difference) as a measure of correlation. The energy of the DFD between images $x(i,j)$ and $y(i,j)$ of size $M \times N$ is defined as

$$DFD = \frac{1}{N \cdot M} \sum_{i=1}^{M} \sum_{j=1}^{N} [\tilde{x}(i,j) - y(i,j)]^2, \qquad (1)$$

where $\tilde{x}(i,j)$ is the displaced original frame, after motion compensation is performed.

As mentioned earlier, we assume that

$$SNR2[n] = f(SNR1[n], DFD[n]). \qquad (2)$$

Thus, our problem reduces to estimating SNR2[n] given SNR1[n] and DFD[n] for every n. Again, we stop when the estimated SNR2[n] starts to decrease. The method we use for this estimation is the Biharmonic Spline Interpolation [3].

EXPERIMENTAL RESULTS

Fig. 1 shows the actual and the estimated SNR2[n] for the "Mother and Daughter" sequence starting at frame 100. It can be seen that the first maximum of the actual SNR2[n] occurs at index 22. This is the point where we should stop encoding the anchor frame and start encoding the rest of the sequence. If SNR2[n] is not available, we can obtain estimated values as mentioned earlier. We see that the first maximum of the estimated SNR2[n] occurs very close to index 22.

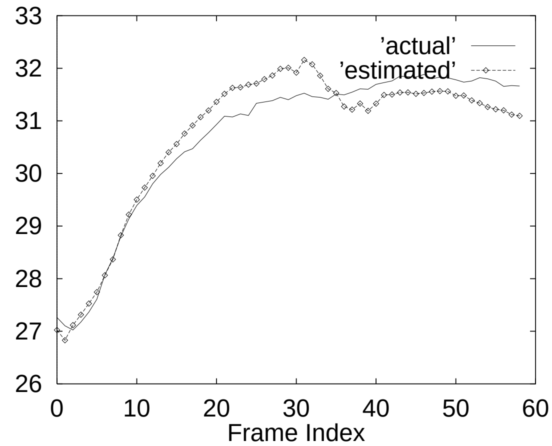

Figure 1: Actual and predicted SNR2[n]

CONCLUSIONS

In this work, we found that it is worthwhile to send the anchor frame progressively and employ methods of determining the best stopping point. The methods described here proved to estimate the stopping point quite accurately.

REFERENCES

[1] J. M. Shapiro, "An embedded wavelet hierarchical image coder," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (San Francisco, CA), Mar. 1992.

[2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, pp. 3445–3462, Dec. 1993.

[3] D. T. Sandwell, "Biharmonic spline interpolation of geos-3 and seasat altimeter data," *Geophysical Research Letters*, no. 2, pp. 139–142, 1987.