A NEW FRAMEWORK FOR NOISE-RESISTANT VIDEO COMPRESSION USING MOTION-COMPENSATED PREDICTION

P. B. Penafiel and N. M. Namazi

Department of Electrical Engineering
The Catholic University of America,
Washington D. C. 20064
penafiep@cua.edu,namazi@cua.edu

ABSTRACT

This paper introduces a new framework for video compression. The proposed method considers noise directly in the video sequence and seeks the optimal compression ratio and video quality. Compression is achieved by eliminating the spatial and temporal redundancies found in the intensity and motion fields of the video. Processing is performed in blocks of $\mathcal{N}$ frames stored in a video buffer. Encoder and decoder are synchronized prior to the transmission of a new block. A reference frame is chosen from each block and encoded before transmission. Spatial redundancies in the intensity domain are reduced by a wavelet filter. The pixel-motion field between the reference frame and other frames in a block is evaluated using a Kalman filter that estimates the pixel motion in the presence of noise. Video frames are predicted from the reference frame and the corresponding motion field. Prediction errors, motion vectors and the reference frame are compressed in wavelet domain before transmission. The compression system includes quantization and entropy coding.

1. INTRODUCTION

The objective of video-compression techniques is to reduce both the bandwidth requirements for transmission and the memory requirements for storage of video. Compression techniques are known as *lossless* if the uncompressed video can be totally recovered from the compressed version, otherwise the techniques are referred to as *lossy*. The compression ratio that lossless methods can achieve depends mainly on the entropy of the video sequence. Typical compression ratios for lossless systems is in the order of 2:1. Lossy methods, on the other hand, achieve higher compression ratios by reducing the video quality. Typical compression ratios for the MPEG-1 standard are in the order of 16:1. The final performance of lossy techniques is measured not only in terms of compression ratio but also fidelity, or equivalently, the video quality expressed in terms of the signal-to-noise ratio (SNR).

The video sequences used for this study contain 256×256 pixels per frame and the intensity field is represented by 8 bits per pixel which corresponds to 256 gray-levels. Additive White Gaussian Noise (AWGN) has been added to the intensity field of the video sequences. The noise increases the entropy of the intensity field which results in a reduction of the compression ratio and fidelity.

Standard video-compression techniques assume that the video sequences are noise-free. This paper introduces a new framework for motion-compensated video-compression in the presence of AWGN (see for example cf. [5]). The transmitter is assumed to have large memory and computational capabilities. Compression ratios of 0.15 bits per pixel are achieved in the simulations for video frames corrupted with AWGN with signal-to-noise-ratio (SNR) of 10 dB.

1.1. System Description

The block diagram for the proposed video-encoder is shown in Figure 1. The encoding cycle starts with the selection of a reference frame from the processing buffer. The relative pixel-motion of the $\mathcal{N}-1$ remaining frames with respect to the reference frame is estimated next. The cycle concludes with the computation of the motion compensated prediction error of the $\mathcal{N}-1$ frames relative to the reference frame. Intraframe wavelet-based encoding is used to transmit the intensity field of the reference frame, the motion field, and the prediction-error field of the $\mathcal{N}-1$ remaining frames.

Lossy Data Compression of the intensity, motion and prediction-error fields is described in the following sections in terms of three subsystems: a) Representation; b) Quantization; and c) Codeword assignment (cf. [5]). The performance of each module is intimately

0-7803-2431-5/95 $4.00 © 1995 IEEE

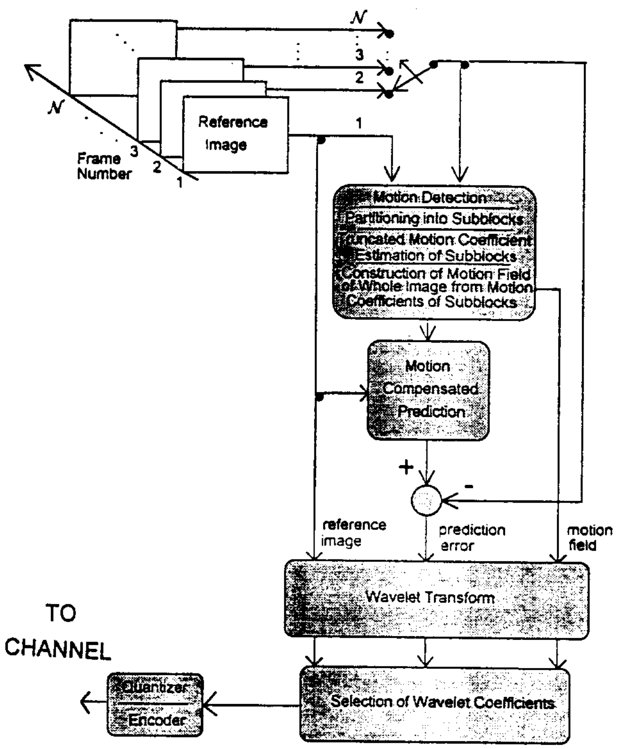

Figure 1: Block Diagram of the Motion-Compensated Video Encoder.

related to the performance of the other modules, therefore the optimization process considers the overall effect of the system rather than the particular effect of each subsystem.

2. WAVELET-BASED FILTER

Wavelet theory is used to formulate the wavelet-based filter that is included in the encoding sections of the intensity field of the reference frame, and the motion and prediction-error fields of the other frames in the processing-buffer. A comprehensive review on the theoretical foundation of wavelets is given by Daubechies (cf. [1]). Several families of orthogonal wavelets with compact support have been defined in the literature. Fast algorithms have been developed for the computation of these wavelets (cf. [4]).

The filtering scheme consists of a systematic selection of wavelet coefficients. The coefficients are sorted by amplitude. High amplitude coefficients are selected until the energy of the filtered sequence is a predefined fraction k_f of the total energy, or until the maximum quota of wavelet coefficients N_w is reached. Due to the good time-frequency localization feature of commonly used wavelets, it is expected that most of the energy contained in the two-dimensional signal field be concentrated in few wavelet components.

The parameters k_f and N_w that control the wavelet filter could be adaptively adjusted to comply with bandwidth and video-quality restrictions.

A characteristic problem associated with wavelet filtering is that the encoder must transmit both the magnitude and the location of the wavelet coefficients. Assuming an 8-bit quantizer and an $M \times N$ pixels image, $(8 + \lceil \log_2(M+N) \rceil)$ bits are needed to transmit each wavelet coefficient. This results in a significant reduction of the compression ratio. A solution to this problem is presented next with the formulation of a new code that achieves a substantial compression of the location information of the wavelet coefficients to be transmitted.

2.1. Pyramidal Encoding of Wavelet-Coefficient Position Information

The filtered signal can be typically represented using a small fraction of the total amount of wavelet coefficients. In general, the non-zero wavelet coefficients are clustered in small groups scattered about the two-dimensional array. Because of these features, a code can be formulated to efficiently transmit the position of the wavelet coefficients. The following pyramidal-code is formulated to reduce the amount of bits required for the transmission of the location information of the filtered wavelet coefficients.

The encoding is based on a two-dimensional binary tree with its root in a block associated with the two-dimensional array of wavelet coefficients. The code assigns a **0** to this block if it is partially filled with wavelet coefficients requiring transmission, or a **1** if none or all of the matrix elements need to be transmitted. The latter is totally specified by adding a **0** (code **10**) if none of the coefficients need to be transmitted, or a **1** (code **11**) if all of the coefficients are to be sent.

In most of the cases, the encoded stream will start with a **0**, meaning that some of the wavelet coefficients will be transmitted. This being the case, the two-dimensional array of wavelet coefficients is partitioned into four equally-sized subblocks. Each one of the subblocks is then recursively encoded as specified for the root block. The only requirement is that the encoder and the decoder are synchronized with respect to the order in which the subblocks are to be encoded.

2.2. Quantization and Encoding of Wavelet Coefficients

In general, the sign entropy of the wavelet coefficients is close to one which implies that one bit could be used

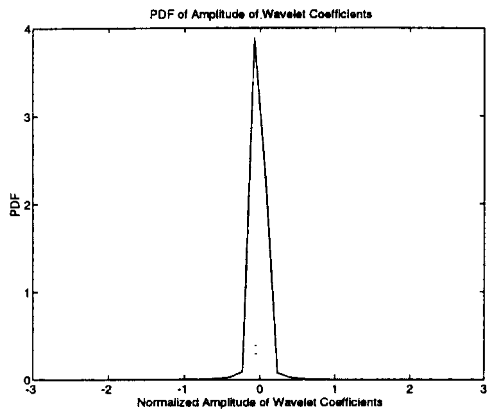

Figure 2: PDF of the normalized wavelet coefficients.

to encode the sign of the wavelet coefficients.

A typical probability density function of the amplitude of the wavelet coefficients is shown in Figure 2. Scalar quantization can be applied if the distribution is assumed Laplacian (cf. [7]). Taubman and Zhakhor (cf. [6]) propose a quantization scheme that varies with the scale of the wavelet basis functions. Instead, we encode the logarithm of the magnitude of the wavelet coefficients selected from the wavelet filter. This results in a pdf that after normalization can be modeled as exponential ($p_x(\alpha) = (1/b)e^{-\alpha/b}, \alpha >= 0, b = 0.1$) as shown in Figure 3. Lloyd iterations are then computed from this model to define the optimum codebook for quantization of the wavelet coefficients.

3. VIDEO ENCODER

The intensity field of the reference frame is encoded using the wavelet-filter previously described. The motion field is estimated from the noisy video frames using a Kalman filter. In order to utilize the Kalman filter, the motion vector is appropriately modeled in state-space. A proposed approach is to transform the motion field into another space using a known, unitary transformation before modeling the motion coefficients in the state-space. A Kalman-filter-based algorithm for estimation of the motion-coefficients is described in [3, 2].

The motion field is computed by applying the model to subblocks within the video frame. If the subblocks are sufficiently large (32×32 pixels) and the number of motion coefficients per subblock is significantly low (2×2), the reduced-dimension motion coefficients can be quantized and sent through the channel. An alternative way is to reconstruct the motion field so that motion-compression can be achieved in wavelet-domain.

Reference frame, motion vectors and prediction errors are encoded in wavelet domain as indicated in Section 2. Further compression is achieved in the outgoing bit-stream by using adaptive Lempel-Ziv encoding.

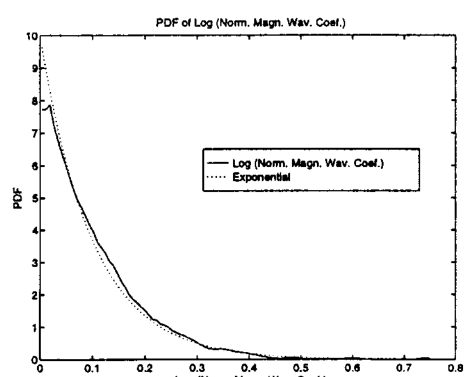

Figure 3: PDF of the natural logarithm of the magnitude of the normalized wavelet coefficients (continuous) and exponential model (dotted).

4. PERFORMANCE MEASUREMENTS

Simulation experiments were performed using the noisy sequence shown in Figure 4 (SNR = 10 dB). The images were partitioned into blocks of 32×32 pixels, and the Kalman-based motion estimator was designed to estimate the 2×2 DCT coefficients of the motion in each subblock. In the simulations we used the Daubechies wavelets with extremal phase and 20 filter coefficients (cf. [1]). We used a nonuniform quantizer, and we applied adaptive Lempel-Ziv coding to the outgoing stream.

Two video buffers have been considered at the encoder video input. The first buffer is assumed to be of infinite size (or the size of the video sequence). The second buffer, referred to as the processing buffer, is a circular buffer with capacity for $\mathcal{N}$ frames. At the decoder end, the video sequence is recovered in an 3-frame buffer. The first frame-buffer is used for the reference frame. The second and third frame-buffers can be thought of as a double-buffer where the previous frame is being displayed from one buffer while the present frame is being decoded in the other buffer. Encoder and decoder are synchronized prior to the transmission of each block of frames stored in the processing buffer.

Table 1 lists the wavelet coefficients as well as the bit rate per frame. It is noticeable from this table that a significant reduction in error coefficients is obtained using motion compensation. Figure 5 shows the decoded images. The good visual quality of these images demonstrates the resistance of our motion-compensated video compression scheme to relatively high levels of noise.

5. REFERENCES

[1] Ingrid Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[2] Nader M. Namazi. *New Algorithms for Variable Time Delay and Nonuniform Image Motion Estimation.* Ablex Publishing Corporation, Norwood, New Jersey 07648, 1994.

[3] Nader M. Namazi, Pablo B. Penafiel, and Chieh M. Fan. Nonuniform image motion estimation using kalman filtering. *IEEE Transactions on Image Processing,* 3(5):572–588, September 1994. Special Issue on Image Sequence Compression.

[4] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, New York, NY 10011-4211, second edition, 1992.

[5] M. Ibrahim Sezan and Reginald L. Lagendijk, editors. *Motion Analysis and Image Sequence Processing.* The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Norwell, MA 02061, 1993.

[6] David Taubman and Avideh Zakhor. Orientation adaptive subband coding of images. *IEEE Transactions on Image Processing,* 3(4):421–437, July 1994.

[7] Ping Wong and Steven Noyes. Space-frequency localized image compression. *IEEE Transactions on Image Processing,* 3(3):302–307, May 1994.

Table 1: Simulation Results (SNR = 10 dB).

Frame #	Error Coef.	Motion x Coef.	Motion y Coef.	Bit Rate bit/pixel
1*	4000	–	–	0.65
2	600	27	30	0.15
3	600	15	15	0.15
4	600	18	9	0.15
5	600	21	10	0.15
5	600	12	8	0.15

* Reference Frame

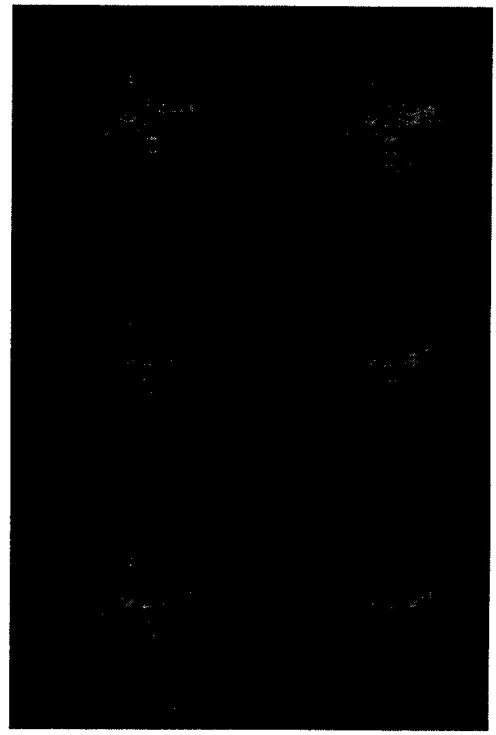

Figure 4: Noisy first six video frames of the Miss America sequence (SNR=10 dB). The sequence direction is from left to right and from top to bottom.

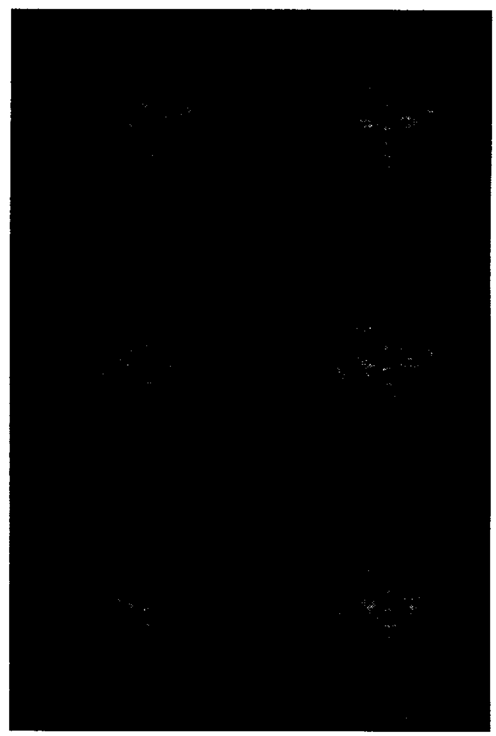

Figure 5: Video frames decoded from the 10 dB *noisy* sequence. The sequence direction is from left to right and from top to bottom.