NASA-CR-204472

A Robust Model-Based Coding Technique for Ultrasound Video *

Alen Docef and Mark J. T. Smith

Georgia Institute of Technology, School of Electrical and Computer Engineering
Atlanta, Georgia 30332-0250

ABSTRACT

This paper introduces a new approach to coding ultrasound video, the intended application being very low bit rate coding for transmission over low cost phone lines. The method exploits both the characteristic noise and the quasi- periodic nature of the signal. Data compression ratios between 250:1 and 1000:1 are shown to be possible, which is sufficient for transmission over ISDN and conventional phone lines. Preliminary results show this approach to be promising for remote ultrasound examinations.

Keywords: telemedicine, ultrasound, video coding, model-based coding, subband coding.

1 INTRODUCTION

Ultrasound video is a very cost effective diagnostic modality, and thus is widely used throughout this country and the world. Although ultrasound equipment is often available in rural and remote corners of the country, specialists to interpret data are typically in short supply in these locations. With the interest and support in telemedicine, the notion of having specialists perform ultrasound examinations at remote locations via electronic data exchange is very attractive. In the absence of channel bandwidth constraints, such an approach is straightforward, with high potential benefits related to providing immediate care and lowering overall expense. Unfortunately, many of these remote locations do not have access to or cannot afford to use high capacity channels (such as T1 lines) to interface with large well-staffed urban medical centers where such specialists reside.

In the presence of channel bandwidth constraints, this approach is encumbered by the large volume of data associated with digital video. Effective compression of the ultrasound prior to transmission will allow this data transfer to occur. The key is to achieve sufficient compression with acceptable reconstruction quality at rates compatible with telephone and ISDN lines. In this work we consider ultrasound video of the heart, where the remote examination involves a specialist at a remote location guiding the attending practitioner by telephone. A critical part of this examination is obtaining proper positioning of the ultrasound probe, so that a diagnosis can be made. The transmitted video quality standards for positioning purposes are clearly not as stringent as those for diagnosis. If positioning quality can be achieved, then higher quality video can be transmitted in a non-real time mode for diagnosis. Of course we hope to eventually be able to transmit diagnostic quality ultrasound in real time, but this not yet in reach. Regardless, the approach outlined above is a marked improvement in terms of accuracy and speed over sending video tapes by courier.

*This work was supported in part by the National Science Foundation under contract MIP-9116113 and by NASA.

The target goals imply compression ratios in the range from 250:1 to 1000:1. An obvious first line of attack on this problem is to investigate to what extent spatial and temporal sampling (i.e. frame size and frame rate) can be decimated without significantly impairing the quality. This has the advantage of being attractive computationally. Based on feedback from the Medical College of Georgia, a 4:1 reduction in spatial resolution to a size of 256×256 was judged to be acceptable. However, the full 30 f/s frame rate was recommended, particularly for pediatric cardiology where the heart rates are often very high.

Conventional coding methods such as H.261 and MPEG are not well suited to ultrasound video. The data rates tend to be too high and they have difficulty representing the high frequency information in the input. Model-based methods on the other hand are known for high compression ratios but suffer typically from variegated performance behavior over a wide variety of inputs.

In this paper, we introduce a model-based method that provides both high compression and robust behavior. To meet the difficult compression requirements imposed by the telephone bandwidth, it is important to identify and exploit all available properties of the signal and preserve with fidelity those parts of the signal that are important for expert analysis. In the case of ultrasound video in cardiology, for example, cardiologists must be able to see the shape of the walls, the shape and thickness of the valves and the tissue texture. By taking into account the nature of the noise/texture associated with the ultrasound images and identifying the important components (wall boundaries, valves, etc.), we formulated a visual model that can be used for very low bit rate coding.

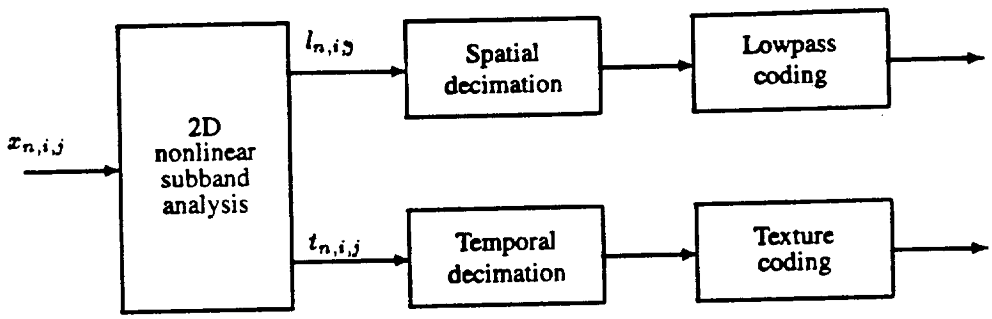
2 SYSTEM DESCRIPTION

The components of the proposed coding system are outlined in Figure 1. First, each input frame $x_{n,i,j}$ is decomposed nonlinearly into two components: a lowpass component, which is denoted by $l_{n,i,j}$; and a highpass or texture component, $t_{n,i,j}$, where n is the frame number, i is the row number, and j is the column number. The decomposition is based on a signal model and is optimized empirically such that the lowpass component contains most of the information needed for diagnosis, such as the contours of walls and valves. The highpass component contains information about the texture of the tissue being examined. The non-linear subband decomposition (upon which we elaborate later) is shown as the first block in Figure 1. After the decomposition, the lowpass component is then decimated in i and j to the Nyquist rate. Signal coding is then performed using an optimized subband coding method recently developed in the digital signal processing laboratory at Georgia Tech.[1] Some details of this method are presented in a later section.

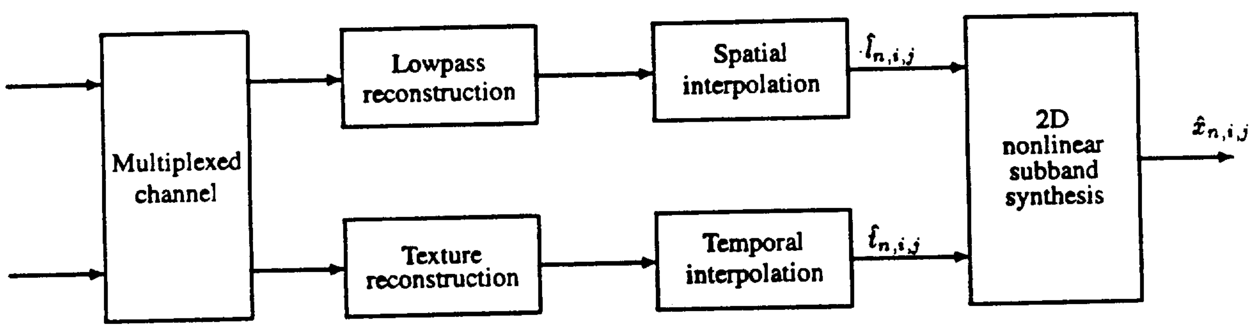
The lower branch of the system contains the texture information. It is decimated in the *temporal* domain and encoded using an in-house version of entropy-constrained residual scalar quantization.[2] The two encoded components are then time-multiplexed into the narrowband telephone channel for transmission to the remote location. At the receiver, the channel signals are demultiplexed and the individual components are decoded. The lowpass component is then upsampled and interpolated spatially to restore it to is proper size and the texture component is upsampled temporally to the original frame rate. After the components are restored, they are combined the 2-D nonlinear synthesis section to form the reconstructed video. In the next sections, we take a closer look at the individual operations shown in the block diagram in Figure 1.

3 NONLINEAR SUBBAND DECOMPOSITION

Two particular characteristics of the ultrasound video signal support the idea of using the model-based decomposition. First, if a static tissue is examined, the ultrasound image can be interpreted as the product of a luminance lowpass component, representing the intensity of the ultrasonic wave in the vicinity of the examined tissue, and a constant reflectance component, representing the reflection coefficients associated with the tissue. Second, ultrasound images are typically very noisy. Usually, additive noise models are used to describe the effect of noise in images. Filtering out the noise could enhance the



(a)



(b)

Figure 1: The components of the coding system: (a) sender, (b) receiver

images, but more important it makes the image easier to code. If the noise has a gaussian distribution then a linear filter is optimal for maximizing the signal-to-noise ratio. In our case, however, the goal is to maximize the subjective quality of the lowpass component.

Thus two approaches can be considered: an additive model and a multiplicative model. A model formulation that covers both additive and multiplicative variates is depicted in Figure 2. It is similar in nature to the homomorphic model pioneered by Stockham[3] for the purpose of image enhancement.

The filter $H(\omega)$, shown in Figure 2, is a lowpass filter with a cutoff frequency of $\omega_c = \pi/D_1$. The nonlinear decomposition is then described by the equation

$$x_{n,i,j} = \Psi^{-1}(\Psi(l_{n,i,j}) + \Psi(t_{n,i,j})).$$

This decomposition is equivalently a nonlinear subband decomposition. The nonlinearity $\Psi(\cdot)$ is chosen to be of the form

$$\Psi(x) = \beta x^{\alpha}.$$

For $\beta = 1$ and $\alpha = 1$, $\Psi(\cdot)$ is the identity mapping and we obtain the additive model. For $\alpha = 0.231$, $\Psi(x) \simeq \beta \log(x)$ in the range 0 to 255, and we obtain the multiplicative model.

The parameter β was chosen so that x and $\Psi(x)$ have the same dynamic range, i.e. from 0 to 255. The parameter α was chosen empirically to optimize the subjective performance. Qualitatively, we want the lowpass component to contain as much useful detail as possible, while keeping constant the cutoff frequency of the filter $H(\omega)$. To quantify this criterion, we could try to minimize the difference between $l_{n,i,j}$ and $x_{n,i,j}$ to address the aforementioned goal. Similarly, we could try to minimize the energy in the texture $t_{n,i,j}$. This ensures that the amount of information contained in the texture is not significant. We have measured these quantities for values of α in the range $0.1 \leq \alpha \leq 2$ for a sample set of ultrasound images and the results are summarized in Figure 3. The graph (a) shows the dependency of the mean square difference between $l_{n,i,j}$ and $x_{n,i,j}$ and the graph (b) shows the dependency of the energy of $t_{n,i,j}$ on the parameter α.

We can see that the two criteria are conflicting, and a compromise between them is needed. The value $\alpha = 0.231$ that implies an approximately logarithmic mapping is in the range of values that provide a good tradeoff between the two criteria. Therefore, the multiplicative model is a reasonable model to use for the encoding of ultrasound images.

The additive and multiplicative models are compared in Figure 6. A sample original ultrasound image is presented together with the reconstructed images obtained by using the additive ($\alpha = 1$) and multiplicative ($\alpha = 0.231$) models. We can see that the multiplicative model has improved subjective appearance.

Because this decomposition is similar to the homomorphic luminance-reflectance decomposition introduced by Stockham[3] in the context of image enhancement, we can also hope to be able to introduce some image enhancement capability to the ultrasound images. In fact, the system is constructed with this feature. Unlike Stockham's approach where different gains are imposed on the two components, we perform histogram modification of the lowpass component. This provides greater flexibility for enhancement. The histogram transformation used in this paper is nonlinear and has the profile shown in Figure 4. It was observed experimentally that the features most difficult to preserve during encoding are represented in the low and medium amplitude range of the lowpass component. Thus, contrast modification in this region is expected to enhance perceived quality. Preliminary results indicate that this is true. At this point enhancement results have not been evaluated by medical specialists, but hopefully will be by the time of the conference presentation.

4 LOWPASS COMPONENT CODING

Taking into account the way the lowpass component $l_{n,i,j}$ was obtained, it can be represented by $\Psi(l_{n,i,j})$, which is a bandlimited signal with a cutoff frequency of π/D_1 in both horizontal and vertical directions. Therefore, $\Psi(l_{n,i,j})$ can be

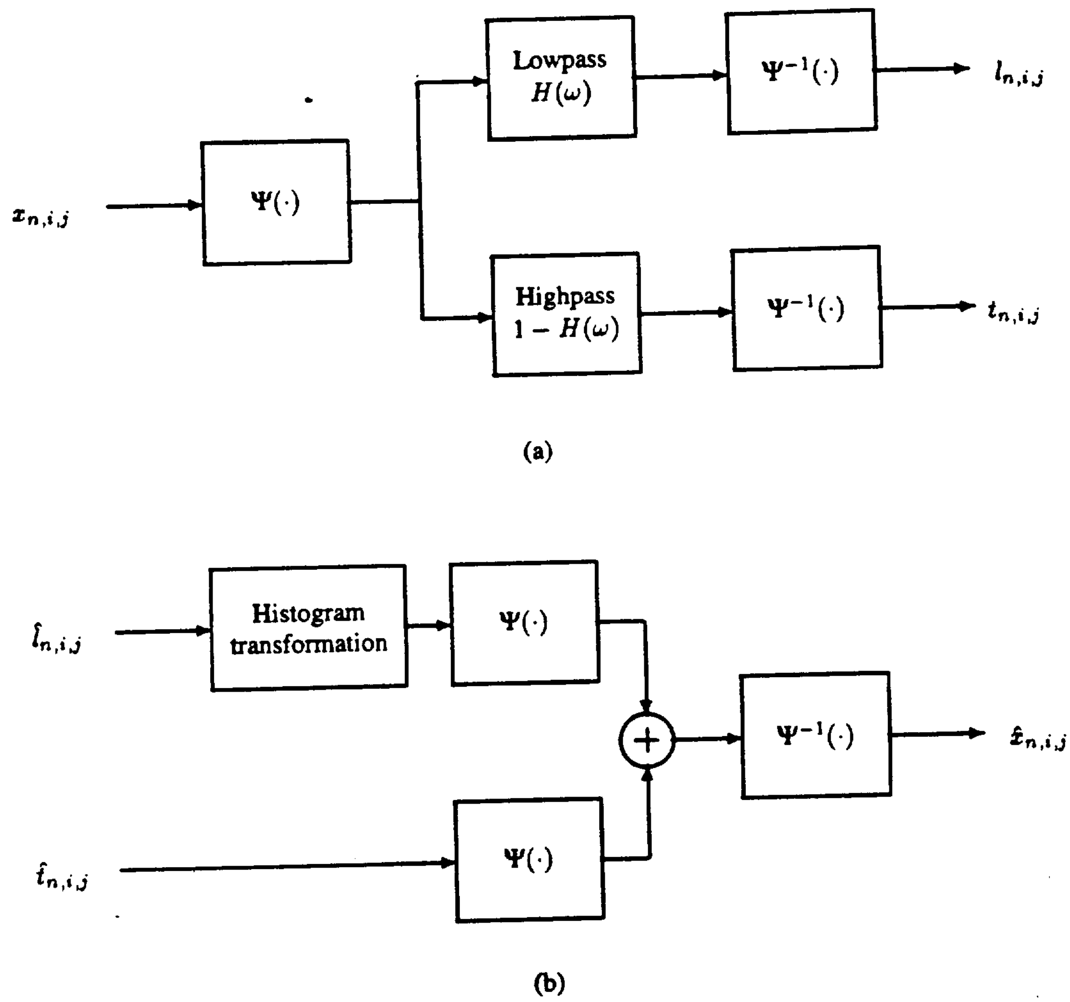


Figure 2: The (a) analysis and (b) synthesis subsystems

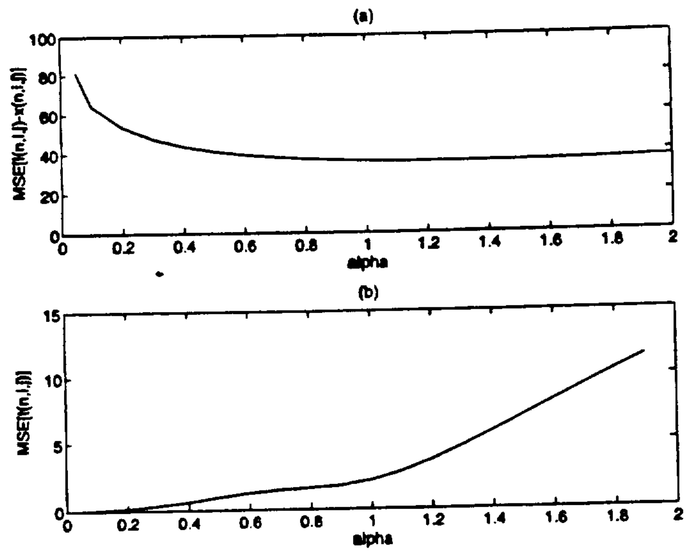


Figure 3: Dependency on α of the two optimality criteria. (a) MSE of $(l_{n,i,j} - t_{n,i,j})$ as a function of α. (b) MSE of $t_{n,i,j}$ as a function of α.

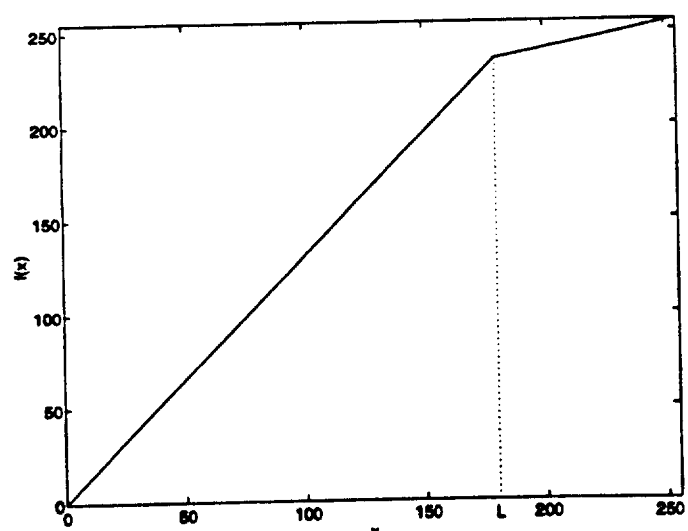


Figure 4: Example of nonlinear histogram transformation for image enhancement.

decimated by D_1 in both horizontal and vertical directions without loss of information due to aliasing. Since the mapping $\Psi(\cdot)$ is one-to-one and onto, $l_{n,i,j}$ can be decimated and reconstructed. By coding the decimated version of $l_{n,i,j}$, the net bit rate can be reduced dramatically.

One of the front-running image coding techniques is subband coding.[4,5] It refers to a broad class of systems where the input is decomposed into subband images and the subband images are coded for transmission or storage. In this work, an new optimized subband image coder is employed.[1] This particular subband coding system consists of decomposing the lowpass component into 16 uniform subbands using the A11 two-band analysis filters introduced in reference [6]. The implementation can be made very efficient computationally by using specially designed recursive filters that require no multiplication operations.[1] The subbands are then quantized using entropy-constrained multistage quantizers with intra-band and inter-band conditioning.

This subband coding system is described in detail in references [1] and [2]. Hence our discussion of this part of the coder is brief. Let it suffice to say that the subband coder is based on encoding each subband pixel (one quantization stage at a time) using conditional entropy coding. The conditioning is based on the quantized symbol values in the local neighborhood of the pixel and in corresponding locations across the subbands. Conditional entropy coding of this form allows statistical dependencies within and across subbands to be used to our advantage.

In addition however, we also extend the conditioning to include corresponding pixels in previous frames. Implementation complexity limits the number of conditioning symbols that can be used practically, which is unfortunate. Therefore only the most statistically important conditioning symbols are used (the precise number being fixed a priori by implementation constraints). For a fixed number of conditioning symbols, an algorithm that finds the location of conditioning symbols such that the overall entropy is minimized is described in.[7]

Conditioning on previous frames is reasonable since there is a lot of correlation between consecutive frames, especially after the noise has been filtered out in the nonlinear subband decomposition stage. This conditioning scheme is described in Figure 5, where only spatial conditioning is depicted. Inter-subband and inter-stage conditioning are not shown in the figure for clarity resons. Solid lines represent intra-frame conditioning and dashed lines represent inter-frame conditioning. Note that this conditioning scheme requires a large number of previous frames to be buffered. However, this is not a big problem in our case, because the frames are small (64 by 64 pixels).

This type of conditioning for cardiology ultrasound video can be used to exploit the fact that the image sequence is quasiperiodic, with a period given by the heartbeat rate. Therefore, we can use conditioning based on symbols from the frame located one heartbeat period before the current frame. A couple of techniques can be used for estimating the heartbeat period. Ideally, we would choose the value that minimizes the average codeword in the current frame. This method is computationally intensive. A simpler method is to use for conditioning the frame that minimizes the difference between itself and the current frame. However, ultrasound machine outputs often provide the EKG signal explicitly. Thus the simplest way is to extract the period directly from the accompanying EKG.

5 TEXTURE CODING

The texture component is a valuable part of the coded signal in the sense that it contributes to the natural appearance of the reconstructed image. However, much of this texture component is just random noise. One can postulate that the texture of the tissue in the examined region is the same for a relatively long period of time, except when the sensor device is in motion, and additive noise contributes most to the rapid time variations in the signal statistics. A simple approach to encode the texture component is to decimate it in the temporal dimension. A texture frame is then only encoded and transmitted once every D_t frames. At the receiver, the same decoded texture frame is used for the synthesis of D_t consecutive frames.

For large values of D_t, this method may produce an unpleasant effect of static texture. In order to reduce this effect and to have a more subjectively realistic decoded video sequence, two consecutively transmitted texture frames may be used

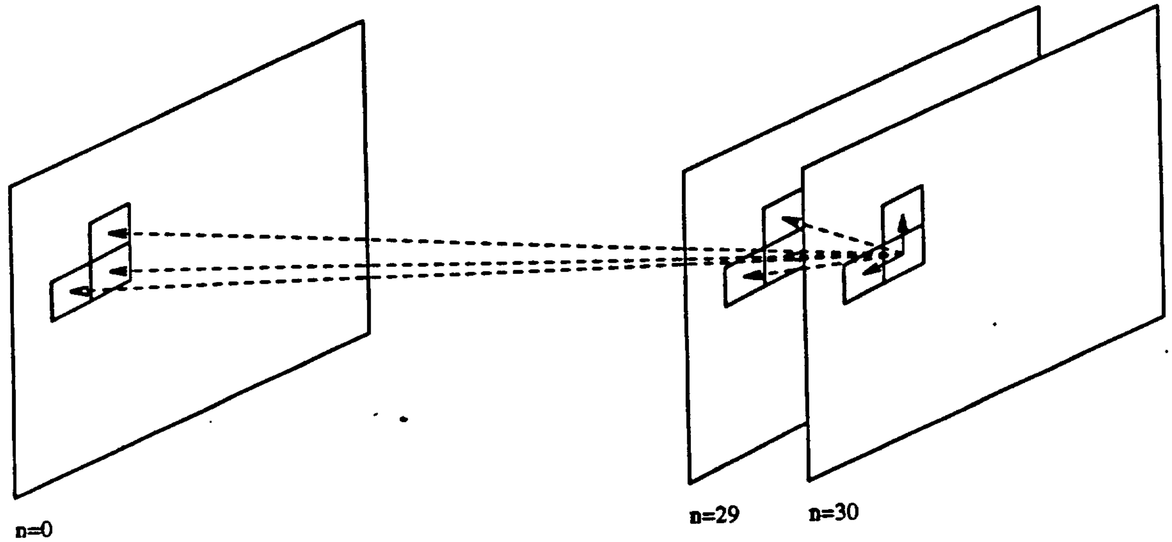


Figure 5: The inter-frame and intra-frame conditioning scheme

alternatively. Alternating between these two texture frames every $1/30$ seconds is an improvement but the flicker effects are too strong. Further subjective improvement is achieved by switching between them every two or three $1/30$ second periods. Experiments have shown that the best subjective quality is obtained when we switch texture frames every three lowpass component frames.

The same entropy-constrained quantization technique is used to encode the texture. The only difference is that conditioning is realized only with respect to neighboring pixels and neighboring subbands. This is because there is no significant correlation between textures D_2 frames apart.

6 EXPERIMENTS AND RESULTS

If the lowpass component is encoded at R_1 bits per pixel and the texture at R_2 bits per pixel, the overall bit rate in bits per second is given by

$$R = 30\left(\frac{256 \times 256}{D_1^2}R_1 + \frac{256 \times 256}{D_2}R_2\right).$$

The lowpass component spatial decimation factor D_1 can be equal to 8 if the coding system is used for positioning only, 4 if we need diagnostic quality, and 2 or even 1 if we implement a multiresolution system allowing zooming in the area of interest. The texture component temporal decimation factor is in the range 25 to 40.

In Figure 7 we present an original ultrasound image (a), the corresponding lowpass component (b), the reconstructed image (c), and the reconstructed image with contrast enhancement (d). Enhancement has been performed using the histogram transformation depicted in Figure 4 with the parameters $L = 170$ and $f(L) = 210$.

The parameters used for encoding are $D_1 = 4$, $D_2 = 60$, $R_1 = 0.45$, $R_2 = 0.25$. The overall bit rate is then $R = 55kbps + 8kbps = 63kbps$, so this example can be used for transmission over an ISDN line. We have used a uniform 64-subband decomposition, and a quantizer having six stages and two code vectors per stage.

Encoded video segments will be presented at the conference.

7 ACKNOWLEDGMENTS

The authors would like to thank Dr. Casimir Eubig and Dr. William A. Lutin at the Medical College of Georgia for their general input, providing us data, and their evaluation of the coded ultrasound images.

8 REFERENCES

[1] A. Docef, F. Kossentini, W. Chung, and M. Smith, "Multiplication-free subband coding of color images," in *Data Compression Conference*, (Snowbird, UT, USA), Mar. 1995. Submitted for publication.

[2] F. Kossentini, M. Smith, and C. Barnes, "Entropy-constrained residual vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. V, (Minneapolis, MN, USA), pp. 598–601, Apr. 1993.

[3] T. G. Stockham, Jr., "Image processing in the context of a visual model," *Proc. IEEE*, vol. 60, pp. 828–842, July 1972.

[4] J. W. Woods, ed., *Subband Image Coding*. Norwell, MA: Kluwer Academic Publishers, 1991.

[5] P. H. Westerink, *Subband Coding of Images*. PhD thesis, T. U. Delft, 1989.

[6] M. Smith and S. Eddins, "Analysis/synthesis techniques for subband image coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1446–1456, Aug. 1991.

[7] F. Kossentini, W. Chung, and M. Smith, "A jointly optimized subband coder," *Submitted to Transactions on Image Processing*, July 1994.

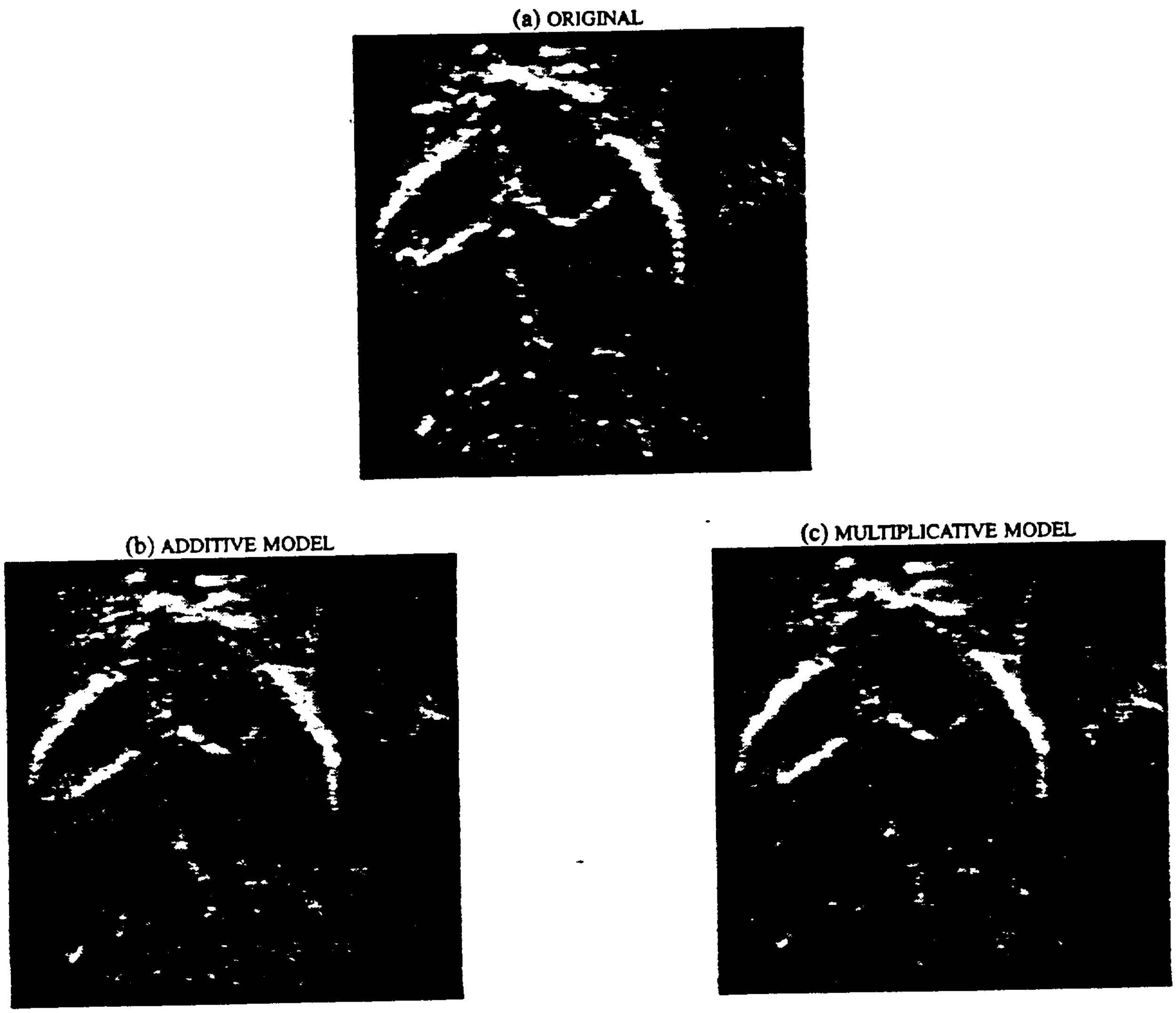


Figure 6: (a) original ultrasound image (b) reconstructed image with additive model (c) reconstructed image with multiplicative model.

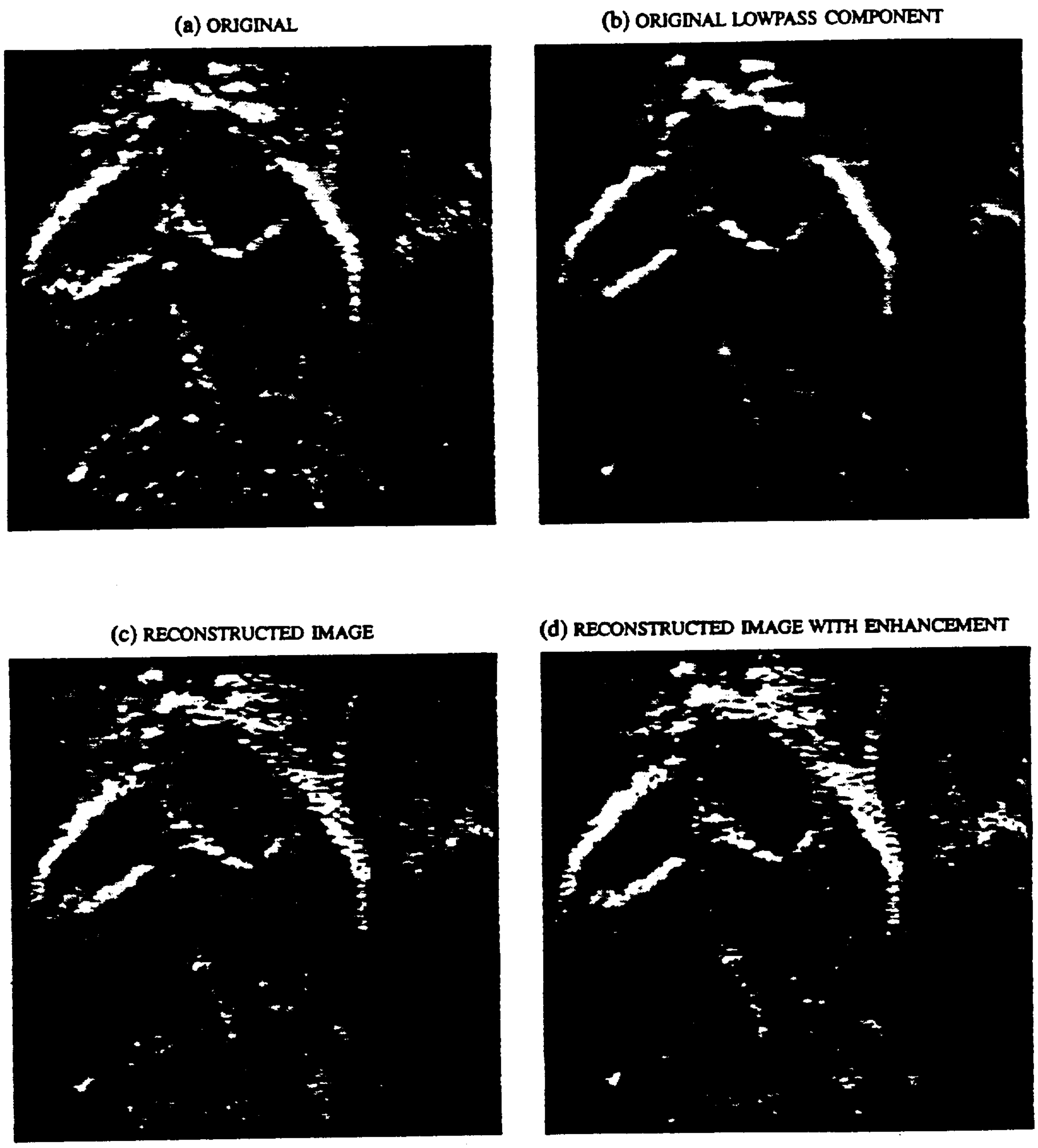


Figure 7: Encoded images at $R_1 = 0.45$bpp and $R_2 = 0.25$bpp.