Subband coding of video for packet networks

Gunnar Karlsson
Martin Vetterli
Columbia University
Department of Electrical Engineering
and
Center for Telecommunications Research
New York, New York 10027

Abstract. Video coding has been investigated for the novel application of video transmission over packet-switched networks. The underlying design goals are presented, together with a software implementation of a coding scheme based on the technique of subband coding. The coding scheme, which divides the input signal into frequency bands in all three dimensions, seems promising in that it lends itself to parallel implementation, is robust enough to handle errors due to lost packets, and yields high compression with sustained good quality. Moreover, it may be well integrated with the network to handle issues such as congestion control and error handling. A discussion of these issues is given, together with the results obtained from the simulated system.

Subject terms: video coding; subband coding; packet switching; integrated services.

Optical Engineering 27(7), 574-586 (July 1988).

CONTENTS

1. INTRODUCTION

Traditionally, voice and video services have been provided by circuit-switched technology, while packet-switched networks have been used primarily for data transmission.[1] Since packet-switched networks are likely to dominate the communications world in the future, the need arises to provide real-time services over such networks as well. Furthermore, strong arguments exist to justify the use of packet switching for the transmission of variable rate, real-time signals. First, signals with varying rate are hard to accommodate in circuit-switched channels, resulting in wasted capacity. Second—and this is the strongest point in favor of packet transmission of voice and video—the integration of services in a network will be greatly facilitated if all of them are dealt with in a common format.[2] Of course, a price has to be paid for these advantages. The problems are related to a fundamental characteristic of packet networks, namely, that the guaranteed and timely delivery (within a bounded delay, that is) of a packet is difficult to obtain.

In this paper, we investigate a coding scheme for video signals that seems well suited for a packet-switched environment. The method is an extension of subband coding[3] to handle a three-dimensional signal. Once the original signal has been frequency divided into subsampled bands, these bands can be handled separately for both coding and transmission purposes; however, residual correlation between bands is neglected in this discussion. Because of the unequal importance of the bands with regard to the quality of the reconstructed sequence, congestion control and cost/quality tuning can be solved quite naturally with the subband coding approach. As it turns out, the video session will be possible, independently of the congestion of the network, as long as the lower frequency bands have guaranteed transmission (typically, the band corresponding to low pass filtering over all dimensions). When more capacity is available, high frequency bands will be transmitted as well and will enhance the quality. Built-in robustness features are such that packet loss in higher bands will result in nearly unnoticeable degradation, while packet loss in the most important band will lead to degradation of quality but no disruption of the session. From an implementation point of view, the subband decomposition leads to a parallel flow of data at reduced rates. Therefore, coding of the bands, packetization, and protocols can run in parallel and at the lowest possible rate.

Section 2 of this paper indicates the strong interaction that exists between signal processing and networking when a video coder is designed for a packet network and lists the key issues of such a design. Section 3 describes subband analysis and synthesis of video signals, and Sec. 4 discusses the encoding of the subbands. Section 5 addresses the interaction of the coding with the network in issues such as packetization, error recovery, congestion control, and protocols. The results

Paper 2518 received Jan. 18, 1988; revised manuscript received March 7, 1988; accepted for publication March 7, 1988; received by Managing Editor March 16, 1988. This paper is a revision of Paper 845-63, presented at the SPIE conference Visual Communications and Image Processing II, Oct. 27-29, 1987, Cambridge, Mass. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 845.

holographic plates, the orientation is not of general concern. Nevertheless, particularly when films are used, alignment errors during the reconstruction are present, and emulsion shrinkage effects can be serious. For these reasons it is better for the film to face the bisector between the subject and reference beams.[11]

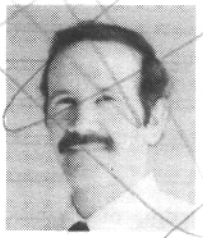
6. ACKNOWLEDGMENTS

This work was performed at the Pennsylvania State University Applied Research Laboratory under the sponsorship of the Naval Sea Systems Command. The authors would also like to thank Michael L. Billet of the Applied Research Laboratory, who helped arrange this effort.

7. REFERENCES

1. B. J. Thompson, "Holographic methods of dynamic particulate measurements—current status," in *High Speed Photography and Photonics,* Part Two, L. L. Endelman, ed., Proc. SPIE 348, 626–633 (1983).
2. J. D. Trolinger, "Particle field holography," Opt. Eng. 14(5), 383–392 (1975).
3. R. E. Brooks, L. O. Heflinger, R. F. Wuerker, and R. A. Briones, "Holographic photography of high-speed phenomena with conventional and Q-switched ruby lasers," Appl. Phys. Lett. 7(4), 92–94 (1965).
4. W. K. Witherow, "High resolution holographic particle sizing system," Opt. Eng. 18(3), 249–255 (1979).
5. G. Haussmann and W. Lauterborn, "Determination of size and position of fast moving gas bubbles in liquids by digital 3-D image processing of hologram reconstructions," Appl. Opt. 19(20), 3529–3535 (1980).
6. W. Lauterborn and K. J. Ebeling, "High-speed holography of laser-induced breakdown in liquids," Appl. Phys. Lett. 31(10), 663–664 (1977).
7. W. Hentschel, H. Zarchizky, and W. Lauterborn, "Recording and automatical analysis of pulsed off-axis holograms for determination of cavitation nuclei size spectra," Opt. Commun. 53(2), 69–73 (1985).
8. J. Katz, T. J. O'Hern, and A. J. Acosta, "An underwater holographic camera system for detection of microparticulates," in *Cavitation and Multiphase Flow Forum—1984,* J. W. Hoyt, ed., American Society of Mechanical Engineers, FED-Vol. 9, 22–25 (1984).
9. B. J. Thompson, "Droplet characteristics with conventional and holographic imaging techniques," in *Liquid Particle Size Measurement Techniques,* J. M. Tishoff, R. D. Ingebo, and J. B. Kennedy, eds., American Society for Testing and Materials ASTM-STP-848, 111–122 (1984).
10. C. S. Vikram and M. L. Billet, "Volume magnification in particle field holography," Appl. Opt. 26(6), 1147–1150 (1987).
11. J. D. Trolinger, "Analysis of holographic diagnostics systems," Opt. Eng. 19(5), 722–726 (1980).

Chandra S. Vikram is a senior research associate with both the Applied Research Laboratory and the Materials Research Laboratory of The Pennsylvania State University. He received his education in India, leading to the Ph.D. in optics from the Indian Institute of Technology-Delhi in 1973. Dr. Vikram's research interests include holographic interferometry, particle field holography, speckle metrology, and geometrical optics. He is author of more than 90 papers in these areas. He is listed in *American Men and Women of Science, Who's Who in Technology Today,* and *Marquis International Who's Who in Optical Sciences and Engineering.* He is a Fellow of OSA and the Optical Society of India and a member of the Materials Research Society and SPIE.



Timothy E. McDevitt received his MS degree in civil engineering from the Pennsylvania State University in 1978, after which he worked for two years at the Allis Chalmers Nuclear Components Division in York, Pa. Since 1980 he has been employed at the Applied Research Laboratory at the Pennsylvania State University, where he is currently working toward his Ph.D. degree in mechanical engineering. His interests include holography and laser Doppler techniques, particularly as applied to mechanical vibrations.

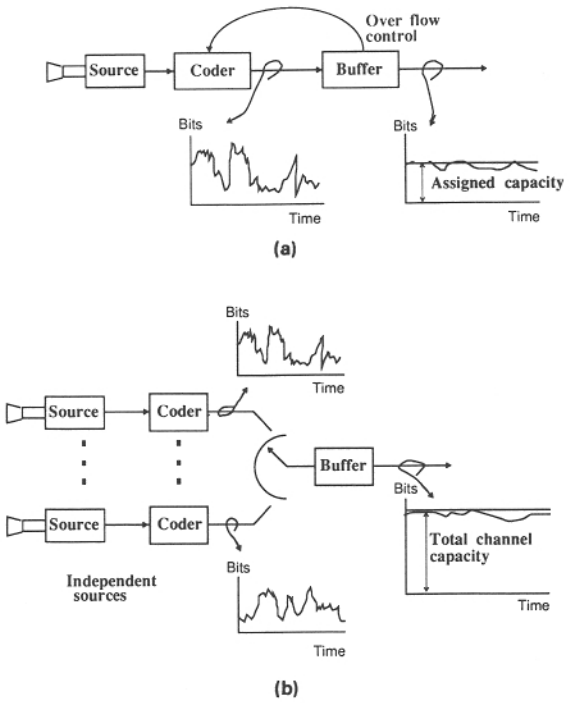


Fig. 1. Multiplexing in circuit switching and packet switching. (a) Fixed rate coder: in a circuit-switched system the multiplexing is done by assigning a fixed amount of capacity for each transmitter. (b) Variable rate coder: for a packet-switched system, there is no predetermined capacity assignment for the transmitter, and with independent sources the transmitted signals are statistically multiplexed.

obtained with the coder under lossless conditions as well as with simulated packet loss are presented in Sec. 6. Finally, future work is outlined in Sec. 7, together with conclusions.

2. INTERACTION OF SIGNAL PROCESSING AND NETWORKING

Video transmission over packet-switched networks, or "packet video" for short, poses the following general problem: a signal with a high and greatly varying bit rate has to be delivered under a given time constraint.

From a signal processing point of view, most coders take into account that the signal they deal with is nonstationary, and thus they yield a variable rate when the signal is encoded: a simple example is the silence detection in voice coding. When such coders are aimed at circuit-switched transmission, in which the channel capacity is fixed, a buffer is required between the coder and the transmission link to smooth out the variations in the rate. To avoid unacceptable delays, the buffer has to be limited in size. This leads to nonnegligible probabilities of buffer overflow. To reduce the occurrence of overflow, a feedback mechanism is put in place that enforces a higher compression when the amount of data in the buffer exceeds some prescribed threshold. In a video coding context this means that the overflow control degrades the image quality during times when the coder generates a high output rate, as during periods of motion in the video scene. This system is depicted in Fig. 1(a).

When the transmission link is packet switched, however, the averaging effect is obtained by statistically multiplexing

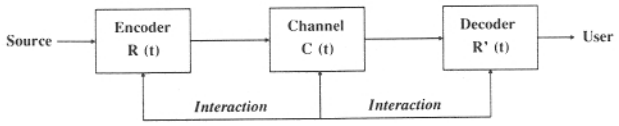


Fig. 2. Transmission of variable rate signals over a packet-switched network. Information on required and available capacity is exchanged, thus not allowing the classical separation of coder and channel. R(t) is the coder output rate, C(t) is the available capacity, and R'(t) is the rate of the decoder. R(t) and R'(t) might differ since there is no common time reference in the system.

several sources over the same link or network,[4,5] as shown in Fig. 1(b). There is a fundamental difference between the two averagings that are taking place. In the buffering for circuit switching, the bit rate of the signal has high correlation over time, whereas in the statistical multiplexing the sources are independent. Therefore, the averaging is expected to be more effective in the packet-switched case. Additionally, the perceptual quality is expected to remain constant[5,6] since the bit rate of the transmitted signal is allowed to vary.

From a networking point of view, delivery of packets within a bounded time delay, thus providing real-time service, is a difficult resource allocation and control problem,[5] especially when the users provide high and varying rates. Fixed allocation of resources, which could eliminate packet loss, would waste channel capacity. Packet switching, on the other hand, cannot guarantee the absence of packet loss, which depends on the load of the network. However, by not allocating fixed channels, packet switching may yield a higher utilization of the total channel capacity.

We have seen that the video coder will require various amounts of capacity over time but that the packet network will provide a channel whose capacity changes depending on the total load of the network. In addition, the decoder has to function together with the channel, in which packets are lost or delayed depending on the total network load. The decoding rate might differ from that of the encoder since there is no common time reference in the system.[7] Therefore, the separation between source, channel, and receiver[8] of the classical communication theory no longer holds. An optimal solution to the transmission of variable rate signals over a variable capacity channel requires an interaction, or information exchange, between the entities involved. Hence, it is necessary to look at the global system. Figure 2 depicts this interaction.

Because the coder and the network are interacting during transmission, we have considered the goals that should be met in designing a video coder for a packet-switched network. The issues of importance are as follows:

(a) Adaptibility of the coding scheme—The signals we are dealing with have changing amounts of information over time. It is expected that the coder will take these variations into account, which, through the redundancy reduction, will give a varying output rate. Typically, when there is no movement in a video sequence, the output rate should go to zero or to a small residual rate.

(b) Robustness to packet loss—Packet loss is unavoidable in a packet-switched environment. Thus, the coding scheme has to be robust so that the video session is never seriously disrupted when packets are lost; suitable error recovery schemes must exist, and redundancy should be built in so as to minimize the quality degradation of the reconstructed image

sequences from the incomplete data. Note that retransmission should be avoided for two main reasons: First, the tight timing requirements make retransmission difficult. Second, there is a risk of positive feedback: assume a congested network in which all users experience packet loss; if they all rely on retransmission for error recovery, the congestion will be aggravated and might lead to more severe performance degradation.

(c) Resynchronization of the video—The transmitted signal is presented to the decoder without any synchronization information and, possibly, with bursts of data lost. From the received data the decoder not only has to reconstruct the image information in the best possible manner but also must present it synchronously to the terminal equipment. The complications relate to variable rate, variable transmission delays, and packet loss. Variable coding does not yield a predetermined number of packets per video frame, and hence the packet arrival rate cannot be used for synchronization. Since there is no notion about the number of packets transmitted, it could be difficult to determine whether a packet is delayed, lost, or simply not transmitted. Further complication is caused by the absence of a common time reference for the encoder and the decoder.[7,9–11] According to its clock, the decoder might thus expect data at a higher or lower rate than is being transmitted.

(d) Control of the coding rate—We extend the notion of congestion control to the whole system (i.e., the coder and the network). Thus, a mechanism for congestion control is incorporated into the coder itself. In the case of a congested network, congestion control has to be enforced at the coder in a graceful manner in terms of image quality degradation.

(e) Interaction with the protocols—The coding scheme should allow for various degrees of quality, as required by the user, both in normal operation and in a congested network. Typically, a lower quality of service would mean a higher probability of packet loss in a congested network (i.e., the transmitted data get a lower priority) and a higher compression, which yields a lower transmission cost but thereby also a lower image quality.

(f) Parallel architecture—The coder should be implementable in parallel because of the high data rate. This will enable the encoding, packetization, and protocols to run at the lowest rate possible, which simplifies their implementation. To clarify, protocols, which are often implemented in software, may not be executable at the speed necessary to handle encoded video as a single bit stream.

Among these various requirements, most new and interesting issues appear at the intersection of coding and networking. Having defined our design goals, we now explore a particular video coding scheme and its suitability for packet video.

3. SUBBAND ANALYSIS AND SYNTHESIS OF VIDEO SIGNALS

A digital video signal is three-dimensional, with two dimensions in space and the third over time. We wish to remove redundancy in each dimension. At the same time, the coding has to incorporate the networking aspects outlined in Sec. 2. We have found that multidimensional subband coding lends itself to packet video through good integration with the network and through high compression. Owing to its successful use in compression of still images[12–15] and of video,[16,17] we believe it to be a powerful video coding method comparable

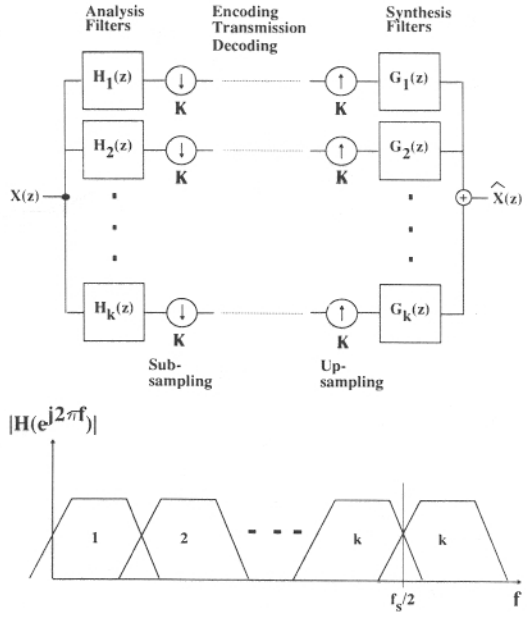


Fig. 3. A subband coding system with K bands. The frequency responses shown are for the analysis filters, H_1 to H_K.

with other methods such as transform coding, vector quantization, and predictive coding. Such methods are commonly applied on subblocks of the image, typically of size 8×8 pixels or 16×16 pixels for transform coding, or 4×4 pixels for vector quantization. When the block of transform coefficients or pixels is coarsely quantized, the subblock structure may become visible in the decoded image. This is often referred to as "blocking effects." Moreover, with the block processing, redundancy due to correlation across block boundaries is not removed. In subband coding, the images are processed in their entirety and therefore the aforementioned problems do not exist. Subband coding also allows the compression to be adjusted according to perceptual criteria, which may reduce the visibility of the introduced distortion. Thus, the compression can be made highest in those frequency bands in which the distortion becomes least visible. The underlying theory of multidimensional subband coding can be found in various references, for example, Ref. 18.

Note that throughout this paper we denote the process of achieving compression as "encoding" and the reverse process as "decoding."

3.1. Description of the method

The technique of subband coding is depicted in Fig. 3. A signal is passed through a bank of bandpass filters, the analysis filters. Owing to the reduced bandwidth, each resulting component may be subsampled to its new Nyquist frequency, thus yielding the subband signals. Following that, each subband would be encoded, transmitted, and, at the destination, decoded. To finally reconstruct the signal, each subband is up-sampled to the sampling rate of the input. All up-sampled components are passed through the synthesis filter bank, where they are interpolated, and are added to form the reconstructed signal.

In the following, we focus on an analysis into two bands only. Assume low and high pass analysis filters with z-transforms, $H_\ell(z)$ and $H_h(z)$, and the corresponding synthesis fil-

ters $G_\ell(z)$ and $G_h(z)$. The subsampling and the following up-sampling, both by a factor of 2, yield the analysis-filtered signal modulated by $\frac{1}{2}[1 + (-1)^n]$, where n is the sample index. In other words, every second sample has been discarded by the subsampling and has been reinserted as a zero-valued sample by the up-sampling. For the two-band case we thus have the input/output relation

$$\hat{X}(z) = \frac{1}{2}[H_\ell(z)X(z) + H_\ell(-z)X(-z)]G_\ell(z) + \frac{1}{2}[H_h(z)X(z) + H_h(-z)X(-z)]G_h(z) \ . \quad (1)$$

The aliasing components, given by $H_\ell(-z)X(-z)$ and $H_h(-z)X(-z)$, are nonzero when the frequency responses of the analysis filters overlap, as depicted in Fig. 3. Aliasing distortion is highly visible but can be canceled by the following constraints on the synthesis filters:

$$G_\ell(z) = 2H_h(-z) \ , \quad G_h(z) = -2H_\ell(-z) \ . \quad (2)$$

The aliasing is now perfectly canceled regardless of the amount of overlap of the passbands of the filters. Hence, we are left with

$$\hat{X}(z) = [H_\ell(z)H_h(-z) - H_h(z)H_\ell(-z)]X(z) \ . \quad (3)$$

It is desirable to choose the filters $H_\ell(z)$ and $H_h(z)$ so that $\hat{X}(z)$ is a perfect, but possibly delayed, replica of the input signal $X(z)$, in the absence of coding and transmission loss. This can be written as

$$\hat{X}(z) = z^{-k}X(z) \ , \quad k \in \mathcal{N} \ . \quad (4)$$

It is beyond the scope of this paper to investigate the design of perfect reconstruction filters. However, a review of the entire topic can be found, for example, in Refs. 19 through 21.

3.2. Implementation of three-dimensional subband analysis and synthesis

The implementation of a three-dimensional subband analysis system is shown in Fig. 4. The system consists of temporal, horizontal, and vertical filtering, with further spatial analysis of the subband that has been obtained through low pass filtering in all of the three first stages. Note that owing to the subsampling, the bit rate in each of the output branches from an analysis stage is half of its input rate. Consequently, the first stage with temporal filters has the highest computational burden since there is no parallelism, while each of the spatial filters operates in parallel at a lower rate. Hence, we chose the temporal filters to be the shortest possible, which also minimizes the number of frames needed to be stored as well as the delay. In the z-transform domain the filters are given by

$$H_\ell(z) = \frac{1}{2}(1 + z^{-1}) \ , \quad H_h(z) = \frac{1}{2}(1 - z^{-1}) \ . \quad (5)$$

Together with synthesis filters given by Eq. (2), these filters yield perfect reconstruction, as in Eq. (4), with the delay $k = 1$.

The spatial filters operate in parallel at a lower rate, and the filter lengths do not affect the storage requirements. Consequently, we can allow higher-order filters. LeGall[22] derived pairs of perfect reconstruction filters that are well suited for image processing. These filters have linear phase, low computational complexity, and relatively good characteristics for frequency selection and interpolation. From among them we chose the following filters for the spatial subband analysis and synthesis:

$$H_\ell(z) = \frac{1}{4}(-1 + 2z^{-1} + 6z^{-2} + 2z^{-3} - z^{-4}) \ , \quad (6)$$

$$H_h(z) = \frac{1}{4}(1 - 2z^{-1} + z^{-2}) \ , \quad (7)$$

$$G_\ell(z) = \frac{1}{4}(1 + 2z^{-1} + z^{-2}) \ , \quad (8)$$

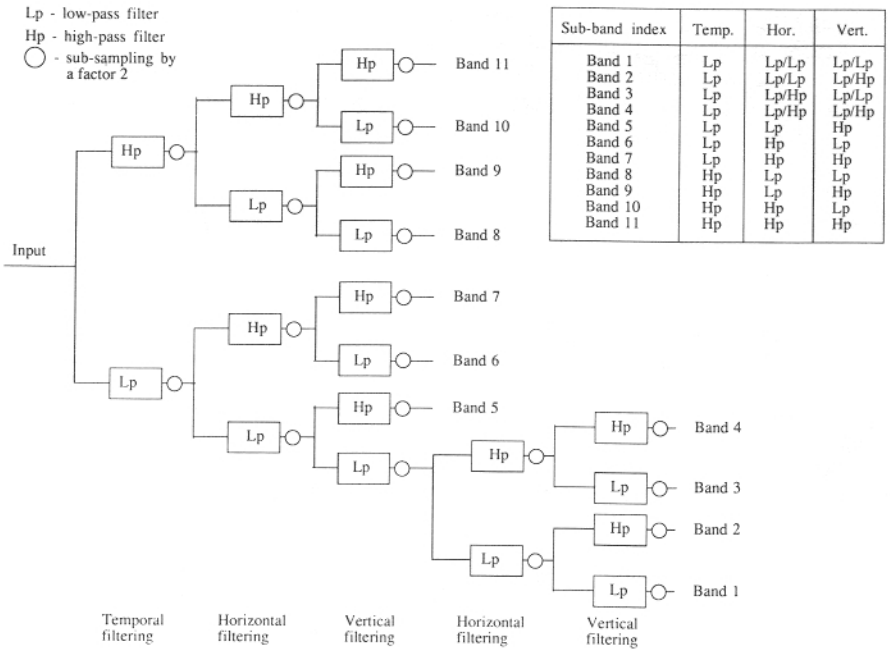
$$G_h(z) = \frac{1}{4}(1 + 2z^{-1} - 6z^{-2} + 2z^{-3} + z^{-4}) \ . \quad (9)$$

As can be verified, for the given filters the input/output relationship [Eq. (1)] reduces to Eq. (4), with the delay $k = 3$. The frequency responses of the four spatial filters H_ℓ, H_h, G_ℓ, and G_h are shown in Fig. 5. The low pass filter of shorter length, G_ℓ, given in Eq. (8), was chosen as the synthesis low pass filter because of its better interpolation characteristics.

Note that all filters are chosen with regard to their low computational complexity, which is tabulated in Table I for the analysis filters given in Eqs. (5) through (7). The synthesis filters have the same complexity. The complexity of the overall coding scheme is studied and compared with other methods in Ref. 23.

Through the subband analysis we have achieved a desirable separation of the input data into 11 parallel bands. Each band has a known importance in terms of image features and quality, to which the encoding can be tailored. Thus, masking properties of the human visual system may be incorporated to yield good image quality even though the distortion may be high in a mean square error sense. The base band (band 1), which has been low pass filtered through all stages, is the most vital, and the importance of the other bands decreases with increasing band index, loosely speaking (see Fig. 4 for the band indices). This can be seen in Fig. 6, in which the subband analysis of Fig. 4 is performed on an image sequence. In the center of the top row is the input, with the temporal low pass component on the left and the temporal high pass component on the right. In the second row, from left to right, are the subsampled horizontal low and high pass components of the respective components. In the third row, the results of vertical analysis and subsampling applied to the components is shown. The components are alternately low and high pass, starting with a low pass filtered component at the left. The four pictures in the last row represent bands 1 to 4, obtained by horizontal and vertical analysis of the leftmost picture in the third row.

The decomposition over time yields one component with an essentially constant bit rate when encoded, while the other component exhibits a bursty behavior after encoding. The encoding can be designed to handle the two cases of constant and bursty rates separately. Compare this with interframe differential pulse-code modulation (DPCM), which corresponds to the temporal high pass component alone. It is expected that a nonadaptive quantization would perform bet-



Sub-band index	Temp.	Hor.	Vert.
Band 1	Lp	Lp/Lp	Lp/Lp
Band 2	Lp	Lp/Lp	Lp/Hp
Band 3	Lp	Lp/Hp	Lp/Lp
Band 4	Lp	Lp/Hp	Lp/Hp
Band 5	Lp	Lp	Hp
Band 6	Lp	Hp	Lp
Band 7	Lp	Hp	Hp
Band 8	Hp	Lp	Lp
Band 9	Hp	Lp	Hp
Band 10	Hp	Hp	Lp
Band 11	Hp	Hp	Hp

Fig. 4. A three-dimensional subband analysis system. The subsampling is performed in the same direction as the filtering operation that it follows.

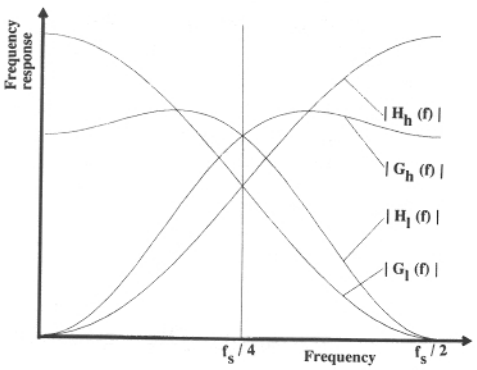


Fig. 5. Frequency responses of the spatial filters given by Eqs. (6) through (9). The filters have been normalized to a common gain factor. (The vertical and horizontal scales are linear.)

TABLE I. Number of operations for the filters given in Eqs. (5) through (7).

Filter	No. of adds	No. of shifts
Temporal		
Low pass	1	1
High pass	1	1
Spatial		
Low pass	5	3
High pass	2	2

ter on the subband analyzed data than on the prediction error in the DPCM case. Consider that the high frequency component, as well as the entire DPCM error signal, would be suboptimally encoded by a nonadaptive method, owing to lack of stationarity of the signal. On the other hand, the low frequency component, in the subband case, is nearly stationary, for which adaptivity is not required for good performance. Thus, the temporal subband analysis may, in a nonadaptive case, perform better than interframe prediction by partly obtaining near optimality. In addition, temporal filtering is more robust to transmission loss than interframe prediction.

The second stage of spatial analysis filtering, which yields bands 1 to 4, is necessary to achieve high compression. The component resulting from the low pass analysis filterings of the first three stages requires sophisticated encoding for its redundancy reduction, analogously to the input sequence, which it resembles. Thus, in accordance with the supporting arguments given for subband coding of the original, the solution for this band's coding is further spatial analysis into four more bands.

Note that seven of the bands each have only 1/8 of the input data rate, and the remaining four have only 1/32 of the input rate. Moreover, for the purpose of parallel implementation, we treat the subbands as being independent. Thus, all subsequent processing, such as packetization, protocols, and network access, may be done in parallel at more tractable data rates.

Next we present how the 11 subbands are compressed and how the subband analysis may be well integrated with the networking issues under consideration.

4. ENCODING OF THE SUBBANDS

The subband analysis has not resulted in any compression: the sum of data in the subbands equals that of the input. However, it has yielded a desirable separation of the data. The filtering operations yield one band (band 1) with an intensity distribution similar to that of the input [compare Figs. 7(a) and 7(b)]. The other 10 bands (bands 2 to 11) have distributions highly concentrated at or around zero and variance highly reduced compared with that of the input distribution, as shown for some of the bands in Figs. 8(a) through 8(d). They may therefore be quantized to a reduced number of intensity levels without introducing high amounts of visible distortion. Band 1 is still highly correlated in all dimensions since it is similar to the input. This correlation is partially removed by DPCM encoding. Note that any method of image coding could be used for this purpose, for example, discrete cosine transform.[24] (Figure 4 gives the band indices.)

Quantization of the pixel values and the prediction error, respectively, results in large connected areas of zero-valued pixels in each band, which run-length encoding exploits to further the compression.

4.1. DPCM encoding of the base band

Subband 1, which is obtained through spatial low pass filtering of the temporal low pass component, is the subband that carries most of the information about the original image. Because of its importance it is transmitted with a higher priority than the other subbands. Although this band contains only 1/32 the amount of data in the input, it still has too high a data rate for requesting high priority transmission. We encode this band with a simple one-dimensional DPCM scheme. In our case we found vertical prediction slightly superior to horizontal prediction in terms of compression. Although the signal exhibits strong correlation over all three dimensions, the issue of robustness and error recovery is too important to

Fig. 6. Three-dimensional subband analysis of a video sequence.

allow the introduction of dependencies in more than one dimension. Again for the sake of robustness, spatial prediction is preferred over temporal prediction; thereby, the effects of transmission loss are confined to the frame in which the loss occurs. The interframe redundancy may be well exploited for error recovery (see Sec. 5), so the prediction loop does not have to be made leaky, which would reduce the compression.

The quantization of the prediction error signal has to be done with caution since, through the synthesis, each pixel will be up-sampled and interpolated to a 4×4 block of pixels in each of two frames. This may accentuate the visibility of artifacts, such as contouring. This subband offers the smooth form from which the other bands carve out features and detail. Consequently, contouring is the foremost artifact to be guarded against, followed by a lowered dynamic range; both are due to improper quantization. Henceforth, the quantizer has to have enough steps to allow fine quantization while still covering the dynamic range of the prediction error.

In our case, the prediction errors are quantized by a symmetric, uniform quantizer. The quantizer has been designed with virtually unlimited range, whereby only the width of the zero level and the step size have to be adjusted to fit the quantizer to the band's properties. The representative levels are taken as the midpoint between two quantization levels. These output levels are assigned variable length codewords that are fitted to an exponential distribution. Thus, the outer limit of the quantizer is set to the point at which the length of the codeword would be unpracticably long. It has been shown to be beneficial to allow a quantizer with a wide range, which relieves the problem of choosing between step size and number of quantization levels. The perceptual importance of the outer levels is higher than is expected from their rare occurrence. The locations of the stretches of zero values and nonzero values, respectively, are run-length encoded, as described in Sec. 4.3.

4.2. Pulse-code modulation encoded bands

Through the subband division we have obtained 11 bands that we may treat as independent, of which 10 have a much lower variance in the intensity distribution than does the original signal. Hence, for these bands little can be gained through predictive coding by exploiting the already low amounts of correlation, and therefore these 10 bands are pulse-code modulation (PCM) encoded. The distributions of some bands are shown in Fig. 8. Note that the components that have been spatially low pass filtered in any direction have a larger spread of the distribution. This is not necessarily owing to a higher information content, but to the higher gain factor of the spatial low pass filter compared with the spatial high pass filter [see Eqs. (6) and (7)]. The quantizers for these components can be designed to offset this without requiring more levels.

We have attempted to fit probability density functions to the histograms of intensity values for each subband. The candidate functions are[25] Gaussian distribution,

$$f(x) = \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha(x-\mu)^2}\ , \tag{10}$$

Laplacian distribution,

$$f(x) = \frac{\alpha}{2}\, e^{-\alpha|x-\mu|}\ , \tag{11}$$

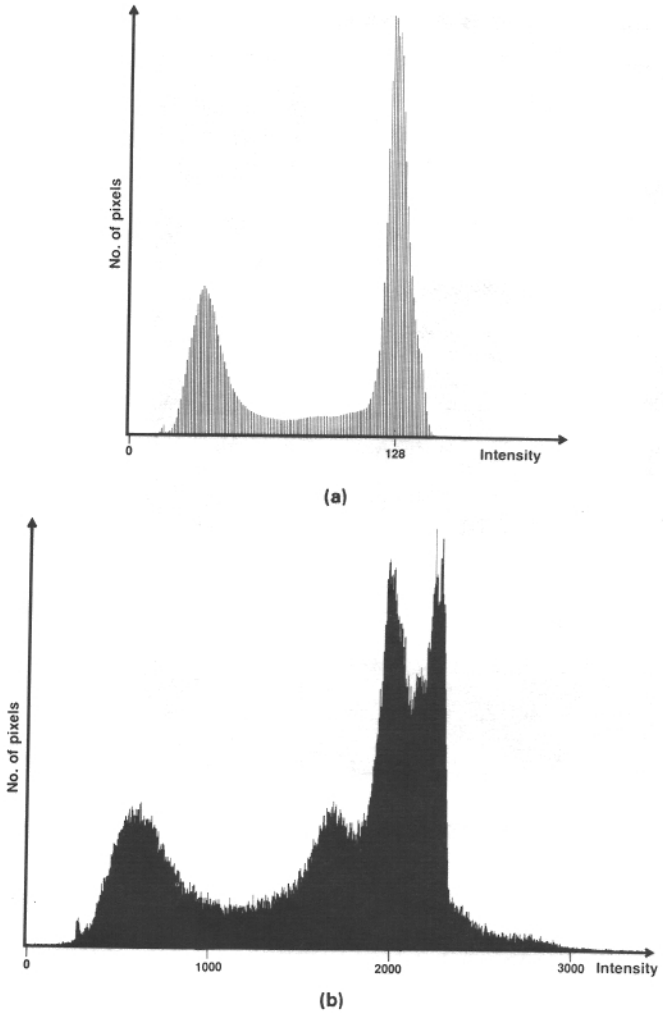


Fig. 7. Histograms of the intensity distribution of (a) an input sequence and (b) band 1. Note that band 1 has an expanded range of intensity values compared with the 256 permissible levels of the input. This is owing to the gain factor of the spatial analysis low pass filter [see Eq. (6)]. The figure is only intended to show the shapes of the histograms; therefore, the vertical scales are different for each one: (a) maximum value 278,232 and (b) maximum value 633.

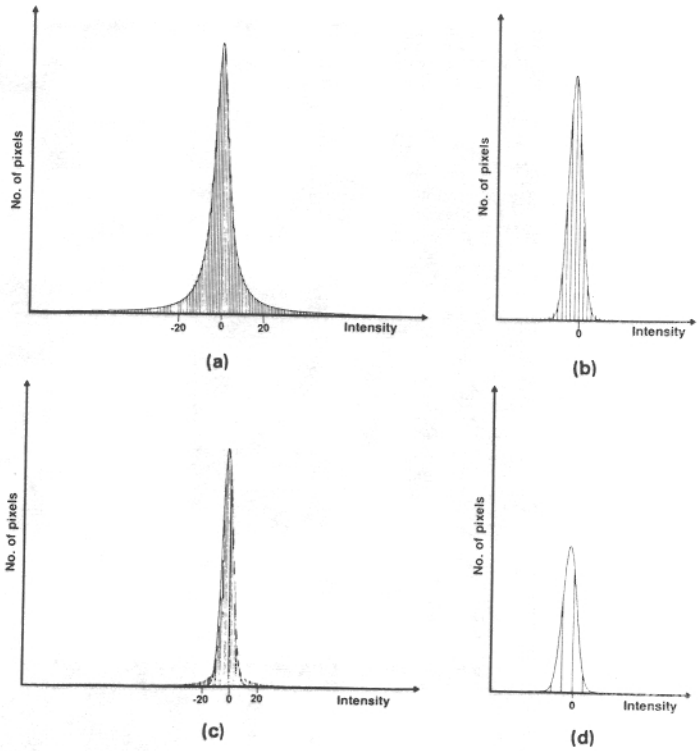


Fig. 8. Histograms of the intensity distribution for four subbands: (a) band 2, (b) band 6, (c) band 8, and (d) band 11. The best fit of the probability density functions given by Eqs. (10) through (12) is overlaid on each histogram. For band 1, the best fit is achieved by a Cauchy function, Eq. (12); for the other three bands, the Gaussian function, Eq. (10), is the best approximation to the probability density distribution of the data. (Note that the figure is intended to show only the shapes of the histograms; therefore, the vertical scales are different for each one.)

and Cauchy distribution,

$$f(x) = \frac{\alpha}{\pi} \frac{1}{\alpha^2 + (x - \mu)^2}, \tag{12}$$

where μ is the mean value. Note that the Cauchy probability density function has infinite variance, but it would be used over a finite range of intensity values, which gives a finite variance. The functions were fit to each normalized histogram (i.e., the sum of values equals 1) by adjusting the parameter α and computing the mean square error of the deviation. The best fitting function is overlaid on each histogram in Fig. 8.

The quantizers could be designed according to the assumed underlying probability density function. However, subband coding allows adapting the quantization to perceptual criteria, a feature that ought to be utilized. Consequently, the analytic design would be based on the fitted probability density functions but adjusted according to perceptual criteria when such are contrary to the dictates obtained through formal analysis. For example, Gharavi and Tabatabai[14] suggest the use of quantizers with a "dead zone" for high pass filtered components. This means that the quantizer has a wider zero level than a quantizer designed analytically, based on the signal's assumed probability density function. The reason for using a dead zone is to remove picture noise, which appears as a low level signal in the higher bands.

The current implementation has symmetric, uniform quantization of all of the bands from 2 to 11. For each band, we adjust only the width of the dead zone and the step size. Analogously to the quantizer for band 1, the quantizers have a wide range with variable length codewords for the output levels, which are taken as the midpoint between two quantization levels. The codewords are designed for an exponential distribution and have a limit on the maximal length. This limit constitutes the outer bound of a particular quantizer. Only the nonzero values are transmitted through run-length encoding of their locations.

4.3. Run-length encoding

The encoded subbands now have large connected areas of zero-valued samples in each frame. This of course suggests that run-length encoding could be fruitfully employed to indicate the stretches of zero- and nonzero-valued samples.

Run-length encoding is a technique most easily implemented as a one-dimensional encoding along the rows or columns of a picture. Our findings are intuitively clear: to necessitate the fewest runs, each subband frame should be

encoded along the spatial direction in which it has been low pass filtered. Encoding is complicated by noise since isolated values have to be encoded as runs on their own, with a large resultant overhead. To eliminate these values, which contribute little to the quality of the reconstructed image, we employ a simple "noise cleaning" by truncating each value with two zero-valued samples on each side in the direction of the run. A study of sequences of reconstructed images shows no visible degradation due to the truncation of these singular values. It does, however, reduce the number of samples to be encoded. Moreover, short gaps of one zero-valued pixel are bridged to get longer and fewer data runs. This decreases the overhead associated with the run-length encoding but slightly increases the number of data values to be sent. Therefore, only one-pixel-wide gaps are bridged.

The codewords could be any set of fixed or variable length numerals, but the B_1 codes of Meyr et al.[26] are very suitable and easy to implement. They have also been used successfully by Gharavi and Tabatabai in subband coding of still images.[12,14]

5. INTERACTION OF THE CODER AND THE NETWORK

To obtain the best possible performance under nonideal transmission conditions, when packet loss occurs, the signal processing aspects of the video coding have to be merged with the networking considerations. Thus, the packetization of the data has to be integrated with the encoding of the subbands. In addition, a suitable error recovery scheme is needed to avoid breakdown of a video session when a packet is lost in the DPCM-encoded band and to reduce the visibility of such an error. Furthermore, when the network is congested, graceful congestion control that maintains the highest possible image quality has to be enforced at the encoder rather than in the network. Finally, the protocol aspects of video sessions must be considered. This issue pertains to the type and quality of service to be established for a video session.

5.1. Packetization

To limit the propagation of the error that results from a lost packet, all packets are made independent of one another. Thus, there is no mixing of data from different subbands, and neither data runs nor runs of zeroes may be broken into different packets. Each individual packet contains the address of the starting location of the first run that it carries (see Fig. 9). This means that when a packet is lost, the data not received can be treated as erasures, i.e., the locations of the lost data are known. This simplifies the error recovery since there is no error propagation and the erased values may be substituted to alleviate the visibility of the error. Note that although the data from different bands are not multiplexed into one packet, one packetizer may still serve more than one subband.

5.2. Error recovery

Error recovery has been shown to be of importance mainly in the DPCM-encoded band.[27] By virtue of the packetization, lost packets become erasures. For the PCM-encoded subbands, such erasures have been found to be masked by the data in other bands, as we show in Sec. 6. A lost packet of PCM values is thus replaced by zero-valued samples. This appears as lost texture and detail in the reconstructed image.

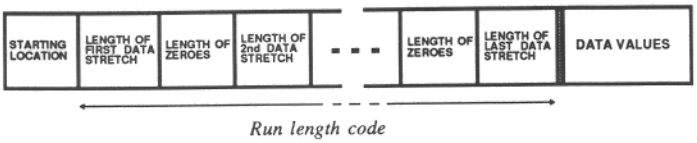


Fig. 9. Format of the packetized data. The data values at the end of the packet are to be distributed according to the run-length codewords. (The illustrated size of a block does not represent its actual size in bits.)

The visibility of this distortion is low when viewing is at video rate primarily because the effect is subtle and the erased areas are scattered spatially and over time. Hence, all further discussion in this section pertains to only the DPCM-encoded subband.

The information contained in the DPCM-encoded subband is the most vital and will be transmitted at a high priority. However, this does not guarantee that packets will not be lost or unduly delayed, even though such loss or delay is extremely rare. Therefore, some method of error recovery has to be considered to avoid a complete restart of the video session if a packet is lost and to reduce the visibility of the event. As pointed out in Sec. 2, notifying the transmitter of the lost packet and initiating an appropriate handling is not considered feasible owing to the tight time constraints involved. It may also lead to more severe congestion if all affected video providers flood the already congested network with retransmitted data. Thus, error recovery has to be handled at the receiver alone.

To achieve error recovery of the signal, the signal is packetized according to Sec. 5.1, with the addition that each packet starts with one PCM value and the error signal is packetized along the direction of the prediction. Thereby, we confine the error to the lost values; no further propagation is possible. Note that if we had used two- or three-dimensional prediction for the base band compression, this erasure property would have been obtained only by including one- or two-dimensional arrays of PCM values, respectively, with each packet, which is clearly unfeasible.

In a first attempt to cover the erased area, the values that could not be properly reconstructed were taken from the corresponding area in the previous frame.[27] The difference between the values from the previous frame, which have replaced the erasures, and the neighboring correct values was expected to be smoothed out during the subsequent synthesis low pass filtering. This proved to be an unsatisfactory remedy even when there were small amounts of motion in the video scene. To improve the performance, we will use a simple motion compensation scheme to locate the most proper area in the previous frame to be used for the replacement. We deem the scheme possible to be implemented with low complexity. As justification, consider that the error recovery has to be invoked only when a packet is lost, that the frame rate is only half the input rate, and that the motion estimation may be done only for a small number of pixels, for example, the corner pixels of a rectangular area circumscribing the erased area. However, the method is still under investigation.

Error-correcting codes[28] may offer an alternative for reducing the visibility of transmission errors in the DPCM-encoded band. By application of such codes to the data in the direction perpendicular to the packetization,[29] lost packets can be corrected as long as their number does not exceed the

error-correcting capability of the code. It is unfeasible to choose a code to be able to recover from any number of errors that might occur during a video session. Since it is not possible to put a deterministic bound on the number of lost packets within each video frame, such codes would have to have very powerful error-correcting capabilities to handle a worst-case scenario. Inevitably, this yields high amounts of overhead even during normal operation of the network, when no or few packets are lost. Rather, the code should be chosen to reduce the probability of a visible error to some acceptable low level, while yet some error handling has to be invoked at the receiver to avoid termination of the video session when the level of packet loss has exceeded the correctable. Note that use of error-correcting codes will increase the complexity of the encoder since the introduced redundancy will have to be offset by higher compression, which adds to the computations introduced by the error-correcting encoding.

5.3. Congestion control at the encoder

There are good reasons why congestion control should be enforced at the encoder rather than in the network. Assume that there is congestion: the network will enforce congestion control by blindly retaining and disregarding packets,[1] whereas the encoder will reduce its flow with regard to the importance of the data. In this manner, the encoder can shield the viewer from noticing all but the most severe network congestion.

Subband encoding yields a desirable decomposition of the image sequence into components with known properties. When we need to enforce a higher compression of the input signal, we have a good notion of which bands contain vital data and in which bands the less important data lie. The increase of compression can therefore be aimed at the least important bands, starting with a change of quantization to coarser and fewer steps, spatial subsampling, and, eventually, omission of transmission of entire bands.

Provided that no new set of codewords need be introduced for the data, congestion control can be handled at the transmitter without notifying the receiver. This avoids the problem of lost messages to indicate the change of compression mode. Spatial subsampling is easily detected since the run-length encoding indicates twice the number of data values that are actually supplied in the packet (see Fig. 9). This indicates that each sample must be repeated twice, or interpolated in a more sophisticated way. Since untransmitted data are handled at the receiver in the same way as lost data, even the frame rate of the DPCM-encoded band can be controlled at the transmitter. If no data of an entire frame are sent, the previous frame will, by virtue of the error handling, be repeated at the receiver. Hence, the output rate from the encoder may be dramatically cut, theoretically to zero, when the video has slowed down to become freeze-frame at the receiver.

If information about network congestion cannot be made available to the encoder, an alternative, less sophisticated technique is to use priorities for the transmitted bands. This has been studied for voice transmission with two priority classes,[30] but the scheme could be extended to more classes by assigning the least important band the lowest priority and increasing the priorities with the importance of the subbands. The network would then be allowed to handle the congestion control by discarding packets in order of increasing priority, whereby the least important data are lost first. The smooth degradation that is obtainable by the first scheme is here traded for the simplicity of using the network's existing mechanism for congestion control.

It is worth noting that with the first method, only data that can be successfully transmitted enter the network, while in the latter case all data enter, some only to be discarded downstream in the network.

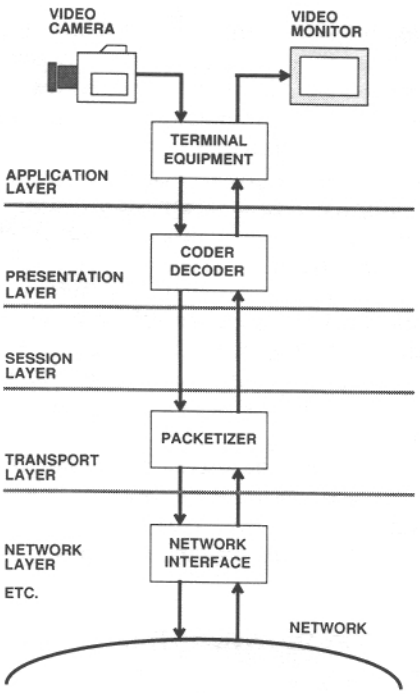


Fig. 10. The location of coding and packetization within a layered protocol model. The terminal equipment, typically a control device for initiating and terminating video sessions, would contain the A/D and D/A circuitry.

5.4. Interaction with protocols

Figure 10 shows how the coding and the associated issues fit into a layered protocol model.[5] The terminology is taken from the open systems interconnection model[1] but solely as an indication of functionality. In particular, at the application layer it is desired to offer a flexibility in the setup of a video session. For example, when the session is initiated, the user may want to decide between the quality and the cost of the session. The quality depends on two parts: encoding and transmission. The encoding may be adapted to achieve a desirable quality and output bit rate through transmission of an appropriate set of subbands. The greater the number of bands, the higher the quality and the output rate. For the transmission, the issue is the degree to which the session may be affected by the total load of the network. A more consistent quality is obtained if all transmitted bands are given high priority, rather than only the base band. Thus, through encoding, the quality without transmission loss is set, and through the priorities of the transmitted bands, their allowed deterioration under network congestion is decided. Moreover, the application layer should allow a multicast session to be arranged. The coder also lends itself to set up sessions in which, for example, four smaller images showing different speakers replace a larger image showing only one speaker. Any such setup is easily handled since the received bands are

(a)

(b)

Fig. 11. Pictures representing the video sequence used in the simulation. (a) Segment with little motion. (b) Segment with substantial motion.

TABLE II. Bit rates, in kbits/s, of the subbands and the total output. The input sequence contained 50 frames of 512×480 pixels with 256 gray levels. The given bit rates are based on a frame rate of 15 frames/s. The minimum and maximum values are the instantaneous rates, which correspond to the respective mimimum and maximum numbers of bits needed to encode a particular frame in the sequence.

Band	Mean	Standard deviation	Minimum rate	Maximum rate
1	300.0	7.0	279.4	308.3
2	98.6	15.8	62.6	123.4
3	117.2	20.5	59.7	144.8
4	29.5	8.3	12.6	42.9
5	192.8	105.4	115.3	603.2
6	87.5	22.2	38.4	124.2
7	10.3	3.1	6.7	21.7
8	539.4	265.1	160.9	1435.9
9	147.1	115.6	35.2	547.1
10	62.9	18.2	28.5	90.9
11	11.8	4.6	4.9	22.0
Total	1597.3	415.8	1077.7	3101.5

(a)

(b)

Fig. 12. The pictures of Fig. 11 shown after compression.

subsampled; thus, up-sampling and interpolation are simply bypassed.

6. RESULTS

We have obtained encouraging results pertaining to compression and associated quality possible with subband coding of video. The results include both lossless transmission and transmission with packet loss.

Figure 11 shows two representative images of the monochrome sequence used in the simulation; Fig. 11(a) is taken from a segment with little motion, and Fig. 11(b) is taken from a segment in which there is substantial motion. The total sequence contains 50 frames, each of size 512×480 pixels, which corresponds to a data rate of 29.5 Mbits/s, given a video rate of 15 frames/s. The data rates for each subband are given in Table II, together with standard deviation and minimum and maximum rates. The minimum rate is calculated as the frame rate times the minimum number of bits needed to encode any single frame, and the maximum rate is calculated analogously. The overall data rate is 1.6 Mbits/s, which is roughly 1/18 of the input rate.

The minimum signal-to-noise ratio (SNR) for any single frame is 35.9 dB, calculated as

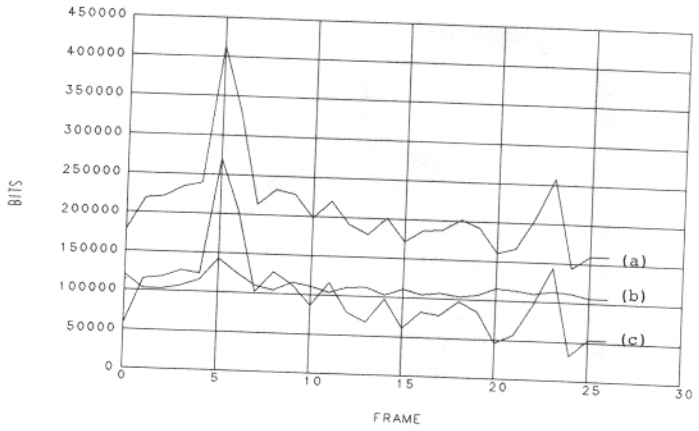


Fig. 13. Bits per frame of the sequence. (a) Total rate from all bands. (b) Sum of the output rates from the temporal low pass filtered bands (bands 1 through 7). (c) Total rate from the bands with temporal high pass frequencies (bands 8 through 11).

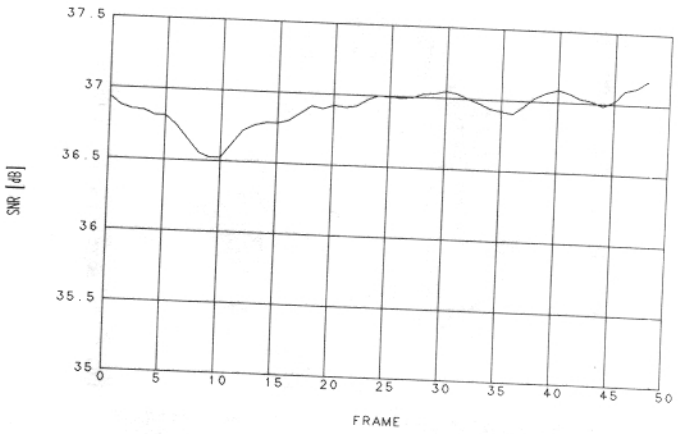


Fig. 14. SNR for each frame in the sequence, following a moving average of length 4, which is performed to yield a measure corresponding more closely to perceived quality.

$$\text{SNR} = 10\log_{10}\left(\frac{256^2}{\sigma^2_{\text{diff}}}\right), \tag{13}$$

where 256 is the intensity range of the input and σ^2_{diff} is the variance of the difference between input and output frames (the mean is 0). The mean SNR is 36.9 dB, and the standard deviation is 0.64 dB. Two frames from the reconstructed sequence, corresponding to the pictures in Fig. 11, are shown in Fig. 12.

In Sec. 2 we discussed how allowing the transmission rate to vary permits the quality to remain constant. In Fig. 13(a) the bits per frame are plotted for the whole sequence. Note that the sequence length is halved because of the temporal subsampling during analysis. The output rate from the encoder varies greatly, which is also apparent from Table II. To assess the quality, we computed the SNR for each frame and filtered it with a moving average filter of length 4, with the result shown in Fig. 14. The reason for performing the moving average is to measure an SNR over a time window that represents more closely what the eye might catch.[31] Note how the fluctuations in the SNR are contained within 1 dB, which

(a)

(b)

Fig. 15. The pictures of Fig. 11 after compression and lossy transmission with five lost packets per frame in bands 2 through 11.

for all practical purposes may be considered constant. Comparing the two plots, the statement of justification for packet video seems to hold this far; varying bit rate transmission can yield constant quality video.

Figure 13 also includes plots representing the total rates from the temporal low frequency bands (1 through 7), Fig. 13(b), and the temporal high pass bands (8 through 11), Fig. 13(c). These two curves show what we wished to obtain with the temporal filtering: one constant component and one bursty. Thus, any encoding following the time analysis can be designed to operate on a signal with either a nearly time-constant information rate or a bursty one. This is to be compared with, for example, interframe prediction, in which only one encoding mechanism has to handle both of these cases.

Throughout this presentation we have pointed to the issue of robustness to transmission loss. Although the visibility of errors is diminished when the sequence is viewed at video rate, compared with the scrutiny of individual frames, we will try to make our point with the images in Fig. 11. We corrupted the image sequence represented by Figs. 11(a) and 11(b) by five lost packets per video frame in the high frequency bands (2 through 11); the results are shown in Fig. 15. The affected

(a)

(b)

Fig. 16. Lossy transmission with one packet lost per frame in band 1 and five packets per frame in bands 2 through 11. Note that the lost data appear in different locations in (a) and (b).

bands, as well as the locations for the loss, were selected completely at random. Figure 16 is similar, but this time the more error-sensitive base band is corrupted as well, by one lost packet per frame. Again, when the sequence is viewed at real video rate, the errors are less visible. This is because the errors are randomly scattered spatially between the frames and in different combinations of the bands, whereby the nature of the error changes.

The single packet lost per frame in the base band corresponds to 2.6% packet loss for that band, and the corresponding ratio for the other bands is 3.0% (assuming equal priority transmission of these bands), which is considered high for most networks. The packets were taken to be 1024 bits, which conforms with the MAGNET standard,[32] and the data included the overhead of the run-length encoding (approximately 40% of the pixel data).

In summary, we have found that the presented results justify many of our assumptions regarding packet video coding. The quality of the reconstructed video, as conveyed by the pictures presented, is representative for the entire sequence: no noticeable impairment seems present. Thus, the SNR plot (Fig. 14), which indicates very little fluctuation in quality, corresponds to the perceived quality. This is what we hoped for by allowing the transmission rate to vary.

The handling of problematic issues of packet transmission has received encouraging results. In particular, packet loss, which was feared as a major complication, appears to be alleviated by the methods discussed in Sec. 5.

It is noteworthy that band 1 has a data rate that is close to constant (see Table II). Recall from Sec. 5 that because of its importance, this band will be transmitted at high priority (i.e., with a fixed amount of capacity allocated). Clearly, the low variance of the band's output rate will yield only a negligible waste of the allocated capacity. The concept of high priority transmission of band 1 is thereby deemed practicable.

7. FUTURE WORK AND CONCLUSIONS

The extension of the work presented comprises the network response to the encoded video signals for the target network, MAGNET.[32] Network statistics will be collected that are necessary for the future evolution of the coding scheme. These include transmission delays and losses, evaluated for different priorities and amounts of network load, and estimates of the time jitter between received packets. These statistics will be used for deciding on resynchronization methods at the receiver, based on the techniques developed for "packet voice."[9-11] Also, statistics of the coder output rate will be used for network performance analysis.[33] The analysis of voice transmission systems with two priority classes for congestion control[30] may be extended to the larger set of priorities needed for the subbands. Finally, this work can be applied to the development of a protocol model for integrated networks,[5,34] to include video in the integrated services.

In terms of refining the coding scheme, the PCM quantizers may be adapted to the intensity level in the base band, which may increase the coding gain as well as lower the perceptible coding noise. Since the base band is reliably transmitted, such a dependency is not expected to worsen the robustness of the coding. The subbands are also tractable for doing motion-dependent encoding,[17] which could include lowering the spatial resolution of high temporal frequency bands during ample motion.

In conclusion, we have described the issue of transmitting video over packet-switched networks. This environment gives a flexibility not possible with circuit switching in terms of services integration and transmission of signals with varying bit rates. There is, however, a new set of problems to be tackled related to variable delays and packet loss. We have found the following issues to be of importance in the design of a coder for packet video: adaptability, robustness to packet loss, resynchronization, control of coding rate, protocol interaction, and parallel architecture.

With these points in mind, we have investigated a video coding scheme based on the technique of subband coding. In an initial study it has revealed appealing properties, such as high compression with good perceptual quality, robustness to packet loss, tractable integration with networking issues, and an architectural simplicity suitable for parallel implementation.

The networking issues discussed in the context of this coding method include packetization, error recovery, congestion control, and protocols. For packetization, we have found it important to supply the starting location of the data in the packet and thereby enable the lost data to be treated as

erasures. Error recovery is applied only in the base band, which has been predictively encoded. Initially, the corresponding area of the previous frame was used for replacement, but this proved to be insufficient during motion. Hence, we have begun a study of more sophisticated algorithms for substitution. To resolve congestion, the data flow from the coder is decreased by affecting the least important subbands first, which leads to a smooth degradation in received image quality. Protocols to set up a video session should allow the user to define a desired level of transmission cost and reconstructed quality.

Finally, initial results from a simulation of the system show that three-dimensional subband coding is a promising way of achieving video compression, especially in connection with packet-switched transmission.

8. ACKNOWLEDGMENTS

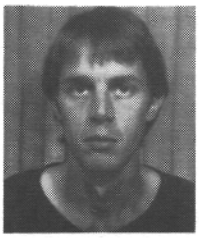
The authors thank Bell Communications Research for providing the image sequence; D. LeGall, A. Tabatabai, and S. Ericsson for helpful discussions; and A. Lazar for originally motivating the work in packet video.

This work was supported by the National Science Foundation under grant CDR-84-21402.

9. REFERENCES

1. M. Schwartz, *Telecommunication Networks,* Addison-Wesley, New York (1987).
2. P. Gonet et al., "Asynchronous time-division switching: the way to flexible broadband communication networks," in *Proc. Int. Zurich Seminar on Digital Communications,* pp. 741–748 (1986).
3. R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing,* Prentice-Hall, Englewood Cliffs, N.J. (1983).
4. W. Verbiest and M. Duponcheel, "Video coding in an ATD environment," in *Proc. Third Int. Conf. on New Systems and Services in Telecommunications* (Liege, Belgium) (1986).
5. A. A. Lazar and J. S. White, "Packetized video on MAGNET," Opt. Eng. 26(7), 596–602 (1987).
6. P. Adam and J. P. Coudreuse, "Variable bit rate coding and asynchronous time division technique," in *Proc. Picture Coding Symposium* (Stockholm) (1987).
7. J.-Y. Cochennec et al., "Asynchronous time-division networks: terminal synchronization for video and sound signals," in *Proc. GLOBECOM '85,* pp. 791–794, IEEE (1985).
8. A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding,* McGraw-Hill, New York (1979).
9. G. Barberis and D. Pazzaglia, "Analysis and optimal design of a packet-voice receiver," IEEE Trans. Commun. COM-28(2), 217–227 (1980).
10. J. G. Gruber, "Delay related issues in integrated voice and data networks," IEEE Trans. Commun. COM-29(6), 786–800 (1981).
11. P. M. Gopal et al., "Analysis of playout strategies for voice transmission using packet switching techniques," Perform. Eval. 4, 11–18 (1984).
12. H. Gharavi and A. Tabatabai, "Sub-band coding of digital images using two-dimensional quadrature mirror filtering," in *Visual Communications and Image Processing,* T. R. Hsing, ed., Proc. SPIE 707, 51–61 (1986).
13. J. W. Woods and S. D. O'Neil, "Sub-band coding of images," IEEE Trans. Acoust. Speech Sig. Proc. ASSP-34(5), 1278–1288 (1986).
14. H. Gharavi and A. Tabatabai, "Application of quadrature mirror filtering to the coding of monochrome and color images," in *Proc. ICASSP-87,* pp. 2384–2387, IEEE (1987).
15. P. H. Westerink et al., "Sub-band coding of images using predictive vector quantization," in *Proc. ICASSP-87,* pp. 1378–1381, IEEE (1987).
16. A. von Brandt, "Teilbandcodierung von Bewegtbild-Sequenzen mit 2 Mbit/s," Frequenz 40(8), 190–197 (1986).
17. A. von Brandt, "Motion estimation and subband coding using quadrature mirror filters," in *Signal Processing III: Theories and Applications,* I. T. Young et al., eds., pp. 829–832, Elsevier, New York (1986).
18. M. Vetterli, "Multi-dimensional sub-band coding: some theory and algorithms," Sig. Proc. 6(2), 97–112 (1984).
19. M. Vetterli, "Filter banks allowing perfect reconstruction," Sig. Proc. 10(3), 219–244 (1986).
20. M. Vetterli, "A theory of multirate filter banks," IEEE Trans. Acoust. Speech Sig. Proc. ASSP-35(3), 356–372 (1987).
21. P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction technique," IEEE ASSP Magazine 4(3), 4–20 (1987).
22. D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric kernel filters and arithmetic coding techniques," in *Proc. ICASSP 88,* pp. 761–763, IEEE (1988).
23. G. Karlsson and M. Vetterli, "Three-dimensional sub-band coding of video," in *Proc. ICASSP-88,* pp. 1100–1103, IEEE (1988).
24. D. J. LeGall and H. Gaggioni, "Transmission of HDTV signals under 140 Mbit/s using subband decomposition and discrete transform coding," in *Proc. Second Int. Workshop on Signal Processing of HDTV* (L'Aquila, Italy), paper No. 64 (1988).
25. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, New York (1984).
26. H. Meyr et al., "Optimal run length codes," IEEE Trans. Commun. COM-22(6), 826–835 (1974).
27. G. Karlsson and M. Vetterli, "Sub-band coding of video signals for packet-switched networks," in *Visual Communications and Image Processing II,* T. R. Hsing, ed., Proc. SPIE 845, 446–456 (1987).
28. R. E. Blahut, *Theory and Practice of Error Control Codes,* Addison-Wesley, Reading, Mass. (1983).
29. P. J. Lee, "Forward error correction coding for packet loss protection," presented at the First International Packet Video Workshop (New York) (May 1987).
30. N. Yin et al., "Performance analysis of a priority-oriented packet voice system," in *Proc. INFOCOM '87,* pp. 856–863, IEEE (1987).
31. N. Ohta et al., "Variable rate video coding using motion compensated DCT for asynchronous transfer mode network," to be published in *Proc. Intl. Conf. on Communications* (Philadelphia), IEEE (June 1988).
32. A. A. Lazar et al., "MAGNET: Columbia's integrated network testbed," IEEE J. Selected areas Commun. SAC-3(6), 859–871 (1985).
33. B. Maglaris et al., "Performance analysis of statistical multiplexing for packet video sources," in *Proc. GLOBECOM '87,* pp. 1890–1899, IEEE (1987).
34. A. A. Lazar et al., "A reference model for integrated local area networks," in *Proc. Int. Conf. on Communications* (Toronto), pp. 531–536, IEEE (1986).

Gunnar Karlsson was born in Jönköping, Sweden. He received the MS degree in electrical engineering from Chalmers University of Technology, Gothenburg, Sweden, in 1983. He is currently working towards the Ph.D. degree in electrical engineering at Columbia University. During the academic year 1982–83 he studied at the University of Massachusetts on a Fulbright scholarship. He did his master's thesis at Telefonaktiebolaget LM Ericsson, Stockholm, during the summer of 1983. In 1984 he worked as a hardware designer at Saab Training Systems in Huskvarna, Sweden. During the summers of 1985 and 1986 he did research at Bell Communications Research in Morristown, N.J. Since 1985 he has held positions as a teaching assistant in the Department of Electrical Engineering and a research assistant in the Center for Telecommunications Research at Columbia.



Martin Vetterli was born in Switzerland in 1957. He received the Dipl. El.-Ing. degree from the Eidgenössische Technische Hochschule Zürich in 1981, the master of science degree from Stanford University in 1982, and the doctorat ès science degree from the Ecole Polytechnique Fédérale de Lausanne in 1986. In 1982, he was a research assistant with the Computer Science Department of Stanford University, and from 1983 to 1986 he was a researcher at the Ecole Polytechnique. He has worked for Siemens in Switzerland and AT&T Bell Laboratories in Holmdel, N.J. Since 1986, he has been at Columbia University, first with the Center for Telecommunications Research and now with the Department of Electrical Engineering, where he is currently an assistant professor. Dr. Vetterli is a member of the editorial board of *Signal Processing.* He was recipient of the Best Paper Award of the European Association for Signal Processing (EURASIP) in 1984 and of the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986. His research interests include multirate signal processing, computational complexity, algorithm design for VLSI, and signal processing for telecommunications.