Combined multiresolution source coding and modulation for digital broadcast of HDTV*

K.M. Uz,* K. Ramchandran† and M. Vetterli‡

Department of Electrical Engineering and Center for Telecommunications Research, Columbia University, New York, NY 10027-6699, USA

Abstract. A practical end-to-end all-digital multiresolution system is demonstrated that employs joint source-channel coding and modulation in order to achieve efficient broadcast of digital HDTV. The threshold effect plaguing single resolution systems is softened by a stepwise graceful degradation. This can be used to increase the coverage and robustness of the digital broadcast system. This approach is seen as an alternative to traditional single resolution digital transmission systems which are not designed for broadcast situations, and which suffer from the threshold effect. This paper highlights the benefits of using an embedded multiresolution modulation constellation over a modulation scheme that resorts to time or frequency multiplexing of the broadcast resolutions. Besides showing coding results and simulations of transmission effects, the paper discusses the trade-offs between low and high resolution coverage.

Keywords. Digital broadcast, HDTV coding and transmission, multiresolution, joint source-channel coding.

1. Introduction

Recent advances in video compression technology have ushered in the era of digital television, with the advent of digital HDTV expected to make as much of an impact on the video industry as CDs have made on the audio world. Even the most demanding delivery mechanism, namely terrestrial broadcasting, might turn digital. However, all current proposals for digital terrestrial broadcasting in the US approach the problem as a point to point transmission problem, namely from the emitter to the fringe. That is, the system is geared at the poorest channel, a fact termed in [10, 11] as 'wasted capacity' near the transmitter.

This single resolution (SR) approach of catering to the broadcast fringe is known from information theory to be suboptimal: when dealing with different channels, one can do better than to transmit at the capacity of the worst channel. Cover [4] showed that one could trade capacity from the poor channels for excess capacity in the better ones, and that the trade-off can in theory be worthwhile. However, to the best of the authors' knowledge, no real system has been designed using these results.

In this work, we demonstrate that using these ideas for channel modulation, together with an appropriate multiresolution (MR) source coding, leads to a flexible way to design digital broadcast systems. A more comprehensive treatment of this work, that includes an analysis of the role of error correction codes (ECCs) in the joint source channel coding problem, is presented in [9]. We show, in particular, that the threshold effect typical of SR digital transmission can be replaced by a graceful degradation. We also demonstrate how the coverage area over SR schemes can be increased considerably, at the cost of some mid-region

suboptimality. An important benefit of employing an MR joint source channel coding scheme is shown to be realized when the scheme is coupled with state-of-the-art motion compensation techniques to increase the full-resolution broadcast area. In light of the ongoing debates on standards and compatibility, it is topical to point out that MR schemes offer a natural and practical solution to the compatibility problem, as was proposed in [2].

The outline of the paper is as follows. Section 2 reviews an MR video coding scheme [12, 13] used in this paper for the HDTV source coding. It decomposes the source into coarse versions and added refinements or details. Section 3 discusses the idea of MR transmission for broadcast channels. It reviews the classic idea of Cover [4] and how it can be applied to design a practical MR modulation constellations. The issue of joint source channel coding in this MR context is addressed next. Finally, Section 4 presents experimental results along with possible coverages using multiresolution QAM (or MR QAM) modulation.

2. Multiresolution source coding

Multiresolution (MR) source coding schemes can be seen as successive approximation methods, and under certain conditions [5] they can achieve optimality. Practical MR coders may be slightly suboptimal to SR coders for point to point communications, in terms of compression achievable at the same full resolution quality. However, a broadcast scenario involves a multiuser environment, where the MR decomposition affords a hierarchy of resolutions that are both natural and useful for the compatibility and broadcast problems. We review a specific MR video coding scheme [12], namely a three-dimensional pyramidal decomposition based on spatiotemporal interpolation, forming a hierarchy of video signals at increasing temporal and spatial resolutions (see Fig. 1(b)). The structure is formed in a bottom-up manner, starting from the finest resolution, and obtaining a hierarchy of

lower resolution versions. Spatially, images are subsampled after anti-aliasing filtering. Temporally, the reduction is achieved by simple frame skipping.

The encoding is done in a stepwise fashion, starting at the top layer and working down the pyramid in a series of successive refinement steps. The coarse-to-fine scale change step is illustrated in Fig. 1(a). At each step, first the spatial resolution is increased by linear interpolation, then the temporal motion based interpolation is done based on these new frames at the finer scale. We describe the interpolation procedure only briefly, and refer to [12] for more detail. See Fig. 3 for the different resolutions of a three-layer pyramid.

The unshaded frames shown in Fig. 1(b) are interpolated in time. For these frames, the encoder computes a set of motion vectors that are transmitted along with the temporal residual. The motion vectors are computed in a multiresolution fashion, using a hierarchical blockmatching algorithm [12] somewhat similar to [3]. For each block in the interpolated frame, three different motion vector candidates for the following interpolation modes are considered:

– Backward interpolation: the motion vector that yields the best replacement from the previous frame.
– Forward interpolation: the motion vector that yields the best replacement from the next frame.
– Motion averaged interpolation: the motion vector d that yields the best replacement by averaging the block displayed by d in the previous frame and displaced by $-d$ in the next frame.

The mode that results in the best interpolated block (in the MSE sense) is selected, and the mode selection information is also encoded and transmitted to the receiver.

A discrete cosine transform (DCT) based entropy coder is used to encode the top layer and the subsequent bandpass difference images. Quantizer steps, and consequently bit allocation at different levels in the hierarchy, is determined to obtain good perceptual quality. Another major consideration in the bit allocation scheme is the
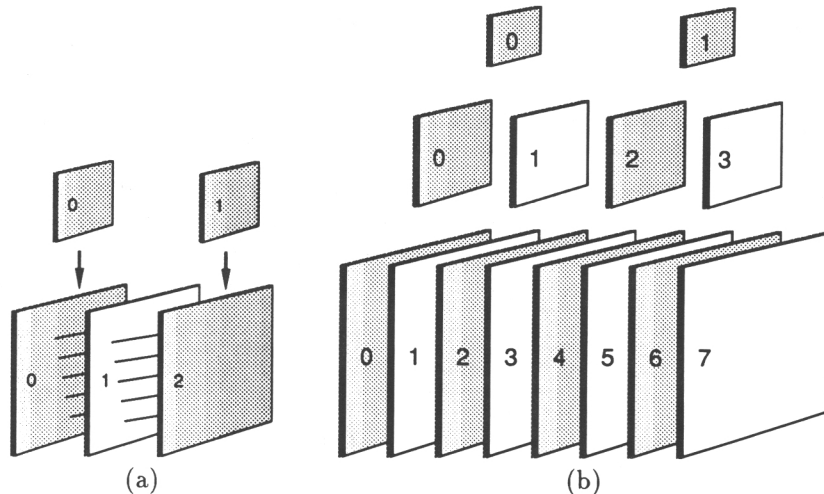
Fig. 1. Reconstruction of the pyramid. (a) One step of coarse-to-fine scale change. (b) The reconstructed pyramid. Note that approximately one half of the frames in the structure (shown as shaded) are spatially coded/interpolated.

subsequent channel coding, as described in the next section.

This MR decomposition of the video source has several advantages:
- it is suited for a digital broadcast system with graceful degradation;
- it is robust to channel errors when the coarse resolution is well protected and error concealment is employed at the decoder;
- it offers a solution to compatibility with lower resolution standards.

It was determined empirically that for the 3 layer MR source coder we consider in our system, resorting to a two-resolution hierarchy comprising the two coarsest layers of the spatio-temporal pyramid in the coarse resolution source channel, and the fine layer in the fine resolution channel resulted in a bit ratio of coarse to fine information of roughly 1:2 at SNRs of interest for typical sequences. This ratio is used to formulate a 'matched' source channel coding design, to be described in Section 4.

3. Multiresolution transmission

Efficient communication of digital information from one source to multiple receivers with varying carrier-to-noise ratios (CNRs) is the fundamental problem of digital broadcast of HDTV. The need to use the broadcast channel capacity efficiently has been pointed out by Schreiber [10] in his proposed hybrid analog-over-digital coding scheme, although no quantitative advantage in terms of received signal quality (SNR or the like) over all-digital schemes has been shown. In our work, we restrict ourselves to a digital solution that is multi-resolution (MR) in nature, and we quantify the performance of various MR transmission schemes for the broadcast range of CNRs in terms of received SNR.

3.1. Efficiency of using MR for digital broadcast

In his classical paper of 1972 [4], Cover supplied the intuition of using an MR scheme for digital broadcast channels. He demonstrated that, for a binary symmetric channel, a source wishing to send information to two receivers of different qualities could optimize its deliverable bit-rate to the receivers, in the Shannon sense, by superimposing the (refinement) information meant for the less noisy receiver within the (coarse) information intended for both. This method of superimposing or embedding information, i.e. broadcasting in an

MR fashion, where the 'detailed information' meant for the stronger receiver *necessarily includes the 'coarse information' meant for the noisier receiver*, is more efficient than 'naive independent sharing' of the broadcast channel resources in time or frequency among the receivers. An efficient end-to-end broadcast system should have its *transmission (modulation) constellation matched to its source coding scheme*, and this is the crux of our work, which we undertake in an MR environment.

3.2. Joint MR source coding/MR modulation

Though the problem of joint source and channel coding has been addressed previously in various coding contexts [6–8], in this work we propose the idea of designing an MR transmission constellation that is matched to the MR source coding scheme described earlier for a broadcast environment. We match the analog MR modulation (QAM) constellation (or MR QAM) with the digital (possibly joint source/channel coded) bitstream output by the MR source via a matched modulation design parameter λ to be described shortly.

Cover's idea of embedded transmission justifies the choice of an MR transmission scheme, the power of which is further reinforced by the MR nature of the source coder which we seek to match. Thus, while embedded transmission for broadcast makes information theoretic sense even for a non-MR source, it is even more natural when the source coder is hierarchical in nature, e.g. the various resolution layers of the spatio-temporal pyramid scheme described earlier. To retain the robustness of digital schemes while catering to the graceful degradation nature of analog schemes, we propose an MR all-digital scheme that combines the two features. The experimental results of Section 4 (see Figs. 4 and 5) highlight the importance of having multiple resolutions for digital broadcast. The presence of a robust coarse resolution channel, accompanied by error concealment techniques like motion-compensated previous-frame replacement at the decoder, can lead to a substantial increase in the high quality reception area over that of

'thresholded' SR digital schemes, which have no reliable coarse information to aid in the full-resolution reconstruction. Although the simulations we present are for a two-resolution system, in theory, the principle holds for any number of hierarchical levels desired, and would result in a 'fractal' modulation constellation, although at increased complexity and decreased practicality. A practical way of introducing more resolutions, with the aid of embedded ECCs is described in [9].

4. Experimental results

4.1. MR 64-QAM

In order to match the MR transmission constellation to the MR source coder, we simulated an MR source coder with 2 resolutions, coarse and fine, which emitted digital bits in the ratio of $1:2$. The basic idea (see Fig. 2) is that the 64-QAM is clustered into 4 'clouds' of 16-QAM constellations. For every 6 bits emitted by the $1:2$ source (of which 2 are coarse and 4 detail), the 2 coarse bits select one of the 4 clouds, while the 4 detail bits select one of the 16 points within the selected cloud. By 'matching' the relative distances between intra-cloud constellation points (D_1) and inter-cloud points (D_2), whose ratio is a design parameter λ, to the relative 'information contents' of the two bitstreams (one measure is the SNRs associated with the two bitstreams), one obtains an efficiently designed joint MR source/MR transmission system, and for a given receiver CNR, one can find the optimal λ that achieves a minimum expected distortion. Further, for a given 'broadcast cost function' encompassing the entire range of broadcast CNRs, which no doubt includes such factors as population density, one can find an 'optimal broadcast λ' for the constellation in question.

For our experiment, the bitstreams comprising the two resolutions were arrived at by clustering the two coarser layers of the 3-layer spatio-temporal pyramid of Section 2 along with indispensable information (like motion-vectors and synchroniza-
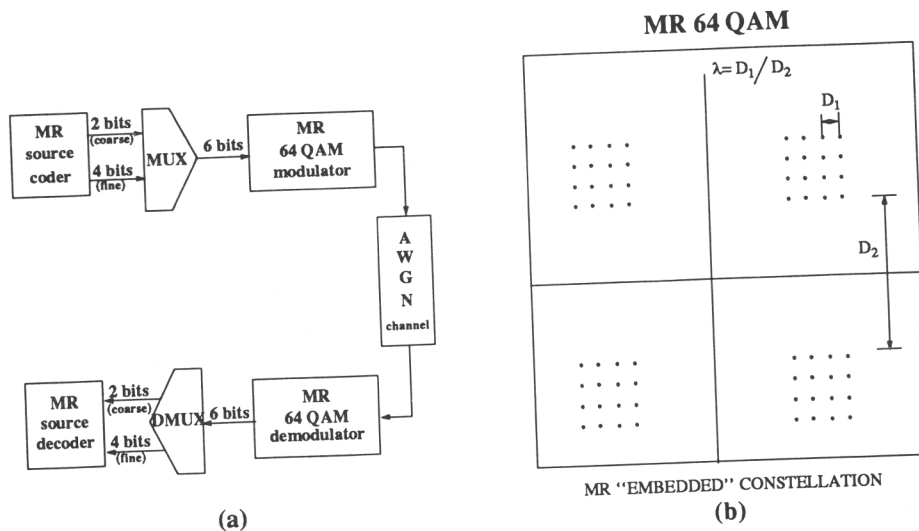
Fig. 2. (a) End-to-end MR system. (b) MR 64-QAM of parameter λ. Note that $\lambda = 1$ corresponds to 64 QAM and $\lambda = 0$ to 4 QAM.

tion bits) as the coarse virtual channel, and the bottom pyramidal layer as the fine virtual channel. Since our efficient MR source coder presents redundancy-stripped Huffman encoded bits to the modulator, isolated channel bit errors could be potentially catastrophic unless blocks of bits are decoupled as packets. To balance the tradeoff between error propagation and packet header overhead, a packet size of 1080, comprising 360 bits of coarse data and 720 bits of detail data, was picked for the simulations. A performance comparison was done on the basis of probability of packet loss, with individual bit errors assumed to cause their entire 'host' packets to be corrupted. The coarse and fine packetized channels formed by the clustering method described above were adjusted to be in the ratio of roughly 1:2 to match the channel constellation of 2 level MR 64-QAM, while preserving high SNR and perceptual quality. The two streams could be interpreted as entering virtual independent buffers with throughputs in the ratio of 1:2, with instantaneous temporal mismatches in the input channel rates being absorbed by the buffers and, if necessary, to prevent overflow or underflow, resolved by exchange of data between the buffers, resulting in minimal degradation for slight mismatches.

Error concealment

Due to the nature of the broadcast communication, it is impossible (or perhaps impractical) to achieve error-free transmission. Bitstreams are often packetized to speed up resynchronization in case of a channel error, but a single bit error still renders the whole packet unusable. Recursive systems (motion-compensated hybrid DCT being the typical example) take much longer to recover, specifically until the next restart of the prediction loop. An error concealment scheme is often required to mask those errors and provide a gracefully degrading picture. The source coder we have used is based on a finite memory structure, and errors would not accumulate but die out within a few time samples. The structure used in conjunction with the MR modulation also allows very successful error concealment.

Simulations show that for typical values of λ, at the same CNRs for which the fine channel packet error rate is greater than 10^{-1}, *the coarse channel is almost perfect* (packet error rate less than 10^{-9}). Therefore, most of the errors will occur in the fine detail, and a coarse version and motion vectors will be available for concealment. The concealment strategy differs slightly for the frames that are interpolated spatially or temporally, and assumes that

Fig. 3. Resolutions of the pyramid. (a) Coarsest layer. (b) Intermediate layer. (c) Full-resolution layer.

the motion vectors and the selected interpolation mode for each block, which are transmitted in the robustly protected coarse channel, are intact. Thus, a concealment strategy based on motion compensated interpolation gives excellent results even in extreme cases of fine channel packet loss. Complete loss of a frame can be tolerated, and sustained 15% packet loss rate causes no visible loss in quality. Figure 4 shows the effect of 15% fine-packet loss (obtained for $\lambda = 0.5$, CNR = 25.5 dB/symbol) on the spatial residual of the sequence, with Fig. 4(c) showing the reconstructed quality, while Fig. 5 illustrates the power of error concealment in an MR environment.

4.2. Comparison of MR, independent and SR constellations

Simulations were carried out for an Additive White Gaussian Noise (AWGN) channel for the 3 cases under conditions of equal power and equal average spectral efficiency. The independent case refers to separate transmission of the coarse and fine channels using 'naive multiplexing' of the frequency spectrum. To compare the MR versus independent constellations, an MR 64-QAM (of free

parameter λ), and a 16/256 QAM (coarse/fine) independent constellation pair were picked. The independent channels have a spectral efficiency of 4 b/s/Hz and 8 b/s/Hz, or an average spectral efficiency (6 b/s/Hz) identical to that of the MR 64-QAM. While a detailed description of the simulation results is provided in [9], a 'threshold' version is shown in Fig. 6(a).

Consistent with perceptual effects, thresholding for the fine channel has been applied at fine packet and coarse packet loss rates of 10^{-1} and 10^{-3}, respectively. The seemingly high fine channel packet loss rate still results in nearly flawless full resolution quality due to the incorporation of error concealment techniques at the MR decoder, as mentioned earlier. To be fair, the SR scheme has been thresholded at the same error rate as the coarse packet channel, as they both represent transitions from the region of no signal to the region of discernible signal.

As can be seen from Fig. 6, the MR constellation outperforms the independent one *over all ranges of CNRs* for a certain range of values of λ (e.g. $\lambda = 0.2$). To perform a comparison of the MR and SR cases, we included a source coding penalty of
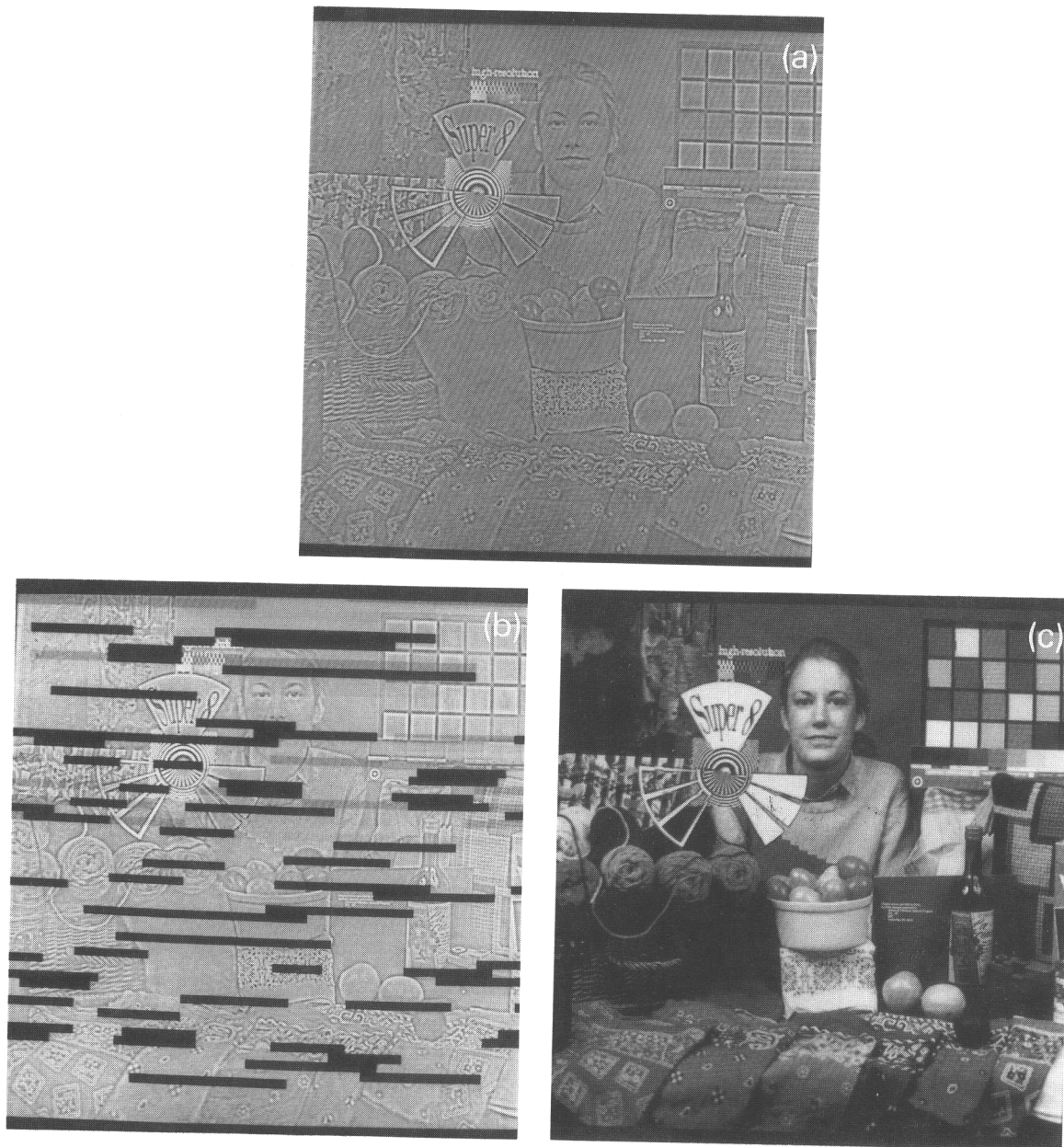
Fig. 4. Effect of channel noise for $\lambda = 0.5$, CNR $= 25.5$ dB/symbol (15% fine-channel packet loss). (a) Spatial residual frame. (b) Packets corrupted by channel. (c) Full resolution reconstruction after error concealment. Image is 512×512.

roughly 16% for resorting to an **MR** source coding scheme, as a worst case analysis from the **MR** viewpoint. This result was derived from empirical results based on the popular 'Lenna' image and other typical images, using an **MR**-unfriendly JPEG [1] coding framework. Under these conditions, the SR channel could afford a 32-QAM modulation scheme for the same transmitter power as the MR 64-QAM scheme due to a source compression advantage of 5/6. The results shown in

Fig. 5. Effect of error concealment for 15% fine-channel packet loss (blow up of Fig. 4). (a) Corrupted spatial residual frame. (b) Reconstruction without error concealment. (c) Reconstruction with error concealment.
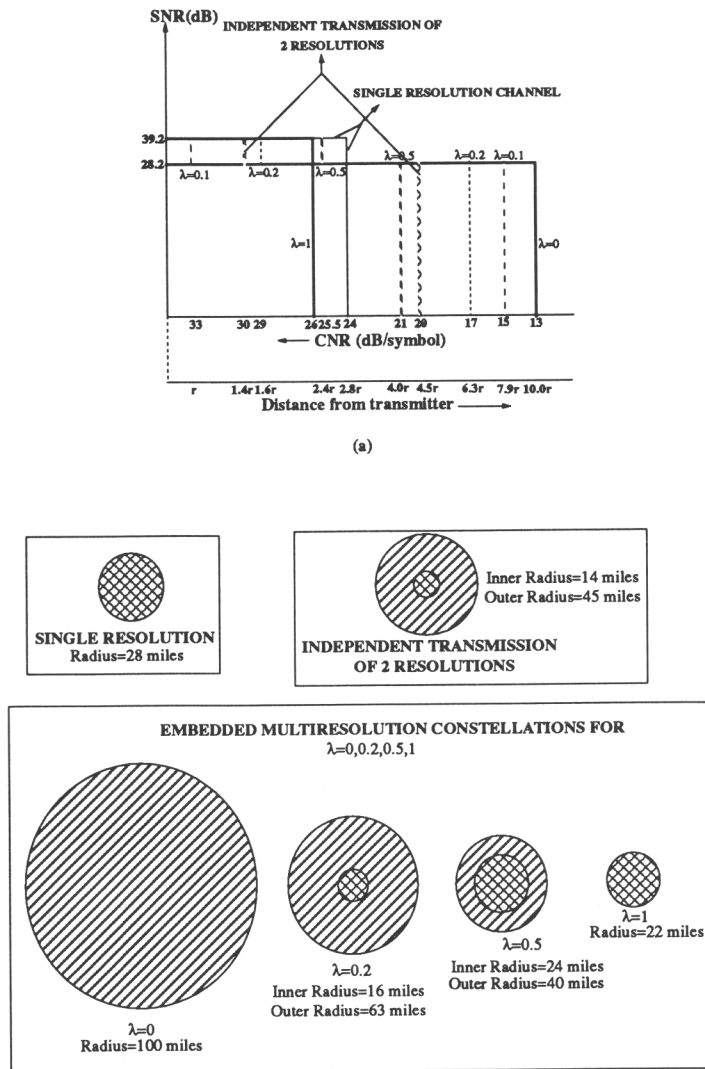
Fig. 6. Typical broadcast environment. (a) SNR versus Receiver CNR. (b) Broadcast ranges for the different constellations.

Fig. 6 indicate the tradeoffs involved. Even under these worst case assumptions, it can be seen that the MR scheme is attractive. From the broadcast area coverage plot of Fig. 6(b), it can be seen that the embedded MR scheme outperforms the multiplexed MR scheme, as promised by information theory; moreover, the tradeoffs between the embedded MR schemes and the SR scheme reveal how, at the cost of some mid-region suboptimality (see for example $\lambda = 0.5$), the MR schemes increase the broadcast coverage area considerably.

5. Conclusion

We have demonstrated a multiresolution (MR) joint source channel coding including an MR transmission modulation, in order to achieve efficient broadcast of digital HDTV. We note that joint MR source channel coding using a source coder matched to an error correction code and/or a modulation constellation would provide an efficient end-to-end MR system. The threshold effect plaguing single resolution (SR) systems is softened by

a stepwise graceful degradation similar to analog systems, without sacrificing the source coding advantage of digital schemes. We establish the superiority, as promised by information theory, of an embedded MR transmission scheme over independent transmissions of the MR source resolutions, and point out the tradeoffs in robustness and broadcast area coverage of low and high resolutions between embedded MR and SR digital systems for QAM constellations. We describe the benefits of resorting to an all-digital end-to-end MR design in exploiting the state-of-the-art digital compression and error concealment techniques to increase the coverage area and robustness for high quality HDTV digital broadcast.

Acknowledgment

The authors would like to thank Prof. W. Schreiber for pointing out the importance of spectrum efficiency for television broadcast.

References

[1] JPEG technical specification, Revision (DRAFT), joint photographic experts group, ISO/IEC JTC1/SC2/WG8, CCITT SGVIII, August 1990.

[2] D. Anastassiou and M. Vetterli, "All digital multiresolution coding of HDTV", *Proc. National Association of Broadcasting (NAB)*, Las Vegas, NV, April 1991, pp. 210–216.

[3] M. Bierling and R. Thoma, "Motion compensating field interpolation using a hierarchically structured displacement estimator", *Signal Processing*, Vol. 11, No. 4, December 1986, pp. 387–404.

[4] T. Cover, "Broadcast channels", *IEEE Trans. Inform. Theory*, Vol. IT-18, No. 1, January 1972, pp. 2–14.

[5] W.H. Equitz and T.M. Cover, "Successive refinement of information", *IEEE Trans. Inform. Theory*, Vol. IT-37, No. 2, March 1991, pp. 269–275.

[6] K. Fazel and J.J. L'Huillier, "Application of unequal error protection codes on combined source-channel coding", *IEEE Internat. Conf. Communications, ICC'90*, Atlanta, April 1990, pp. 320.5–6.

[7] G. Karlsson and M. Vetterli, "Sub-band coding of video for packet networks", *Optical Engrg.*, Vol. 27, No. 7, July 1988, pp. 574–586.

[8] J.W. Modestino, D.G. Daut and A.L. Vickers, "Combined source-channel coding of images using the block cosine transform", *IEEE Trans. Comm.*, Vol. COM-29, No. 9, September 1981, pp. 1261–1274.

[9] K. Ramchandran, A. Ortega, K.M. Uz and M. Vetterli, "Multiresolution joint source and channel coding for digital broadcast of HDTV", *IEEE J. Selected Areas Comm.*, 1991, submitted.

[10] W.F. Schreiber, "Considerations in the design of HDTV systems for terrestrial broadcasting", *Electronic Imaging*, October 1990.

[11] W.F. Schreiber, "All-digital HDTV terrestrial broadcasting in the U.S.: Some problems and possible solutions", *Workshop on Advanced Television, ENST*, Paris, May 1991.

[12] K.M. Uz, M. Vetterli and D. LeGall, "Interpolative multiresolution coding of advanced television with compatible subchannels", *IEEE Trans. CAS for Video Technology, Special Issue on Signal Processing for Advanced Television*, Vol. 1, No. 1, March 1991, pp. 86–99.

[13] M. Vetterli and K.M. Uz, "Multiresolution coding techniques for digital video: A review", *Special Issue on Multidimensional Processing of Video Signals, Multidimensional Systems and Signal Processing*, March 1992, to appear.