MINIMAX DISAPPOINTMENT CRITERION FOR BROADCASTING BASED ON JOINT SOURCE-CHANNEL CODING

Leiming Qian and Douglas. L. Jones

University of Illinois at Urbana-Champaign

1308 W. Main Street, Urbana, IL 61801

Email: *lqian@uiuc.edu*

ABSTRACT

Users of media broadcasting services generally experience different grades of overall performance because of the non-uniform conditions of their channels. In order to provide all users equally satisfactory performance, we propose a new performance criterion named "minimax disappointment" (MD), based on layered service levels and the basic principle of joint source-channel coding [1]. This criterion minimizes for all users the maximum value of performance degradation between the received performance and the expected optimal performance, given each user's individual channel situation. In support of this criterion, we develop four broadcasting systems and a gradient-based optimization scheme. Our systems achieve near-maximal performance for multiple user classes. [1]

[1]This work has largely been supported by the U.S National Science Foundation through grant no. MIP-9707742 and the Motorola Center for Communications at the University of Illinois.

1. INTRODUCTION

In a broadcasting situation, the central broadcasting station transmits the same set of signals to all users of its service without regard to their individual connection to the station. Received performance for different users may vary because of different channel qualities. A typical example would be a simple wireless broadcasting scenario, where users are scattered in a pattern of co-centered circles of different radii (shown in Figure 1); depending on their respective distances to the station and other factors such as geographical environment, structure density, local weather and electromagnetic noise activity, etc. users can be categorized into different classes which observe different channel signal-to-noise ratios (SNR) and fading pro£les. In general, the farther

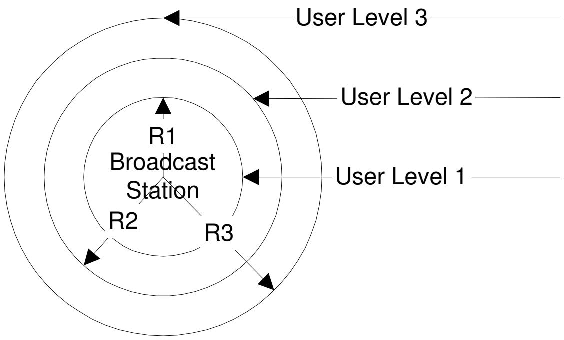

Fig. 1. A simpli£ed wireless broadcasting scenario

away the user is from the station, the worse the transmission quality because of signal energy decay over distance.

In view of the above situation, it is considered good practice to de£ne multiple user classes and provide layered levels of service. In other words, the broadcast signal is divided into multiple successively re£nable layers; each class of users can then receive a maximum number of layers within its ability and obtain a satisfactory performance. It has been shown that layered service indeed provides performance increase for all users.

However, even with layered service and the performance increase it introduces, the problem of unequal performance for different users still exists, if somewhat alleviated. How to effectively and meaningfully evaluate the performance of the overall broadcasting system, given different classes of users, has been a major concern in the development of broadcasting schemes. Traditionally people have developed heuristic and intuitive measures such as averaging the individual performance over all users, using weighted averaging schemes, or simply designing for the worst-case user. They either penalize certain class(es) of users, or are too ad-hoc to retain much theoretical signi£cance.

We believe that a desirable and practical broadcast performance criterion should possess the following properties:

- Fairness to all users: no class of users should be penalized more than the other classes; for example, greater geographical distance from the broadcasting station is not a legitimate reason for heavier penalty in terms of performance.

- Individual consideration: although no class of users should be unfairly treated, the fact that people do anticipate different service quality under different situations should also be taken into consideration (assuming the users are reasonably *informed*). For example, those users far away from the station would naturally expect the transmission quality to be somewhat inferior to that of the users next to the station. Thus, *our objective is to offer different classes of users the same level of relative satisfaction / disappointment, rather than exactly or approximately the same service quality*.

- Theoretical tractability: we should be able to obtain the solution to the optimization problem based on the performance measure using established optimization methods, and the performance measure itself should have some theoretical signi£cance.

Based on the observations above, in analogy to the *minimax regret* criterion in decision theory [2], we propose a new performance measure named "minimax disappointment", which can be explained by Figure 2 and the following formula:

$$P = \max_{i \in I} (P_i - p_i) \quad (1)$$

$$I = \{ i \mid P_{min} \leq p_i \leq P_{max}, i \in 1, \ldots, N\} \quad (2)$$

where N is the number of user classes, P_i is the ith user's *expected best performance* (Here *performance* is de£ned using any suitable metric, later we explain how this P_i is obtained using the joint source-channel matching principle), and p_i is its actually *received performance* (intuitively, for most users $P_i > p_i$ because we have to trade off performance between users. Equality can be achieved under certain circumstances but usually does not lead to the optimal con£guration.). P_{min} and P_{max} are the upper and lower bounds for p_i; performances smaller than P_{min} are considered *unacceptable* and performances greater than P_{max} are deemed *perfect*; those users whose performances satisfy these two inequality conditions form I, the *valid user set*.

We de£ne $P_i - p_i$ as the i-th user's *disappointment* and our objective is to minimize the maximum performance degradation for all users in set I; in other words; all users in this set will be "disappointed" to a certain extent because they cannot receive their expected optimal performance, and the

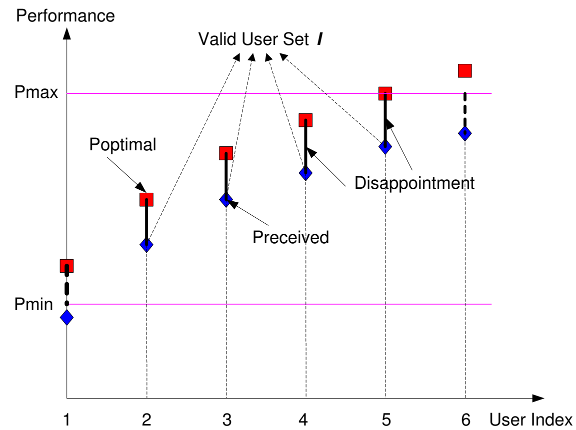

Fig. 2. The Minimax Disappointment Criterion

maximum value of their "disappointment", P, is the proposed performance measure: the smaller P is, the better the performance. We introduce the valid user set I to prevent users with extremely bad channels from having parasitic effects on the entire system, and avoid wasting resources on users whose received performances are already so good that any extra gain would not introduce perceptible effects. In this paper we make the assumption that the set I is predetermined before system optimization.

This performance measure does not penalize any class of users, and it is reasonable and fair because all users in set I know that everybody else is equally as satis£ed / disappointed as they are. It is theoretically tractable because optimizing the performance simply falls into the Minimax Optimization framework.

The minimax disappointment criterion can be extended to minimax *relative* disapppointment criterion (de£ned below):

$$P = \max_{i \in I} \left(\frac{P_i - p_i}{P_i} \right)$$

This criterion optionally scales each user's disappointment by its optimal expectation P_i. It re¤ects the fact that the user's perceived video quality is not proportional to the transmission system's performance: users who have extremely good performance might be willing to accept a relatively larger performance degradation since it will not signi£cantly affect their perceived service quality; while users who have mediocre or bad performance already might be reluctant to accept any further performance degradation. Since this is simply a scaling operation, everything we describe in the following sections still applies.

In Section 2, we give a brief introduction to the joint source-channel coding principle and the various source / channel coders we use in our simulation; in Section 3 we describe the simulation systems (we base our simulations on video broadcasting but the results can be extended to other types of broadcast media), their underlying mathematical descriptions, and our optimization algorithms; in Section 4 we elaborate on the performance and optimality of our algorithms; in Section 5 we present our simulation results in support of our minimax disappointment criterion; and £nally in Section 6 we conclude our paper.

2. BACKGROUND KNOWLEDGE

In this section we brie¤y introduce the readers to the idea of joint source-channel coding and certain source / channel coders such as Motion-SPIHT, 3D-SPIHT, RCPC, etc.

2.1. Joint Source-Channel Coding

Although the joint source-channel coding theorem of Shannon [3] implies that source coding and channel coding can be treated separately without any loss of overall performance, this separability holds only if the communication is point-to-point and allows in£nite codeword length, which is not realistic in a practical environment with £nite delay or a channel with multi-path fading. Joint source-channel coding takes advantage of this unsatis£ed prerequisite to achieve performance gain and a certain level of error resilience.

The advantages of JSCC for image and video transmission have been extensively studied; here we present an incomplete list of previous research: Davis and Danskin [4] described a joint source-channel allocation scheme for transmitting images losslessly over block erasure channels such as the Internet; Ramchandran *et al*. [5] studied multiresolution coding and transmission in a broadcasting scenario; Azami *et al*. [6] acquired performance bounds for joint source-channel coding of uniform memoryless sources using binary decomposition; Belzer *et al*. [7] developed a joint source-channel image coding method using trellis-coded quantization and convolutional codes; Sherwood and Zeger [8] in-

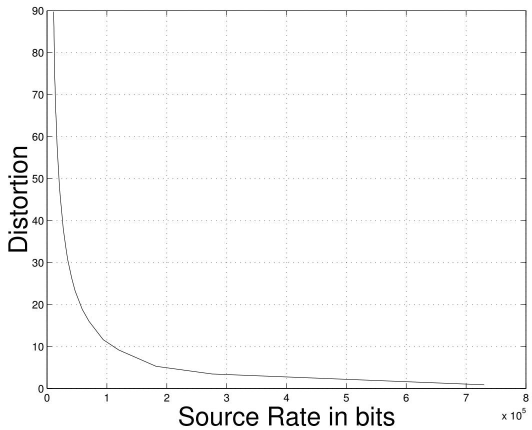

Fig. 3. Rate-distortion curve for a H.263 encoder

vestigated unequal error protection for the binary symmetric channel; Man *et al*. [9] examined unequal error protection and source quantization; Lu, Nosratinia and Aazhang developed a closed-form solution for progressive source-channel coding of images over bursty error channels [10]; Fossorier, Xiong, Cheng and Zeger studied joint source-channel image/video coding for a power-constrained noisy channel [11] [**?**]; Bystrom and Modestino investigated combined source-channel coding for video transmission over a slow-fading Rician channel [12]; Lan and Tew£k studied power-optimized mode selection for H.263 video coding in wireless communication [13], and Zheng and Liu used a subband modulation approach to transmit image and video over wireless channels [14]. A general method that can be applied to most source and channel coders can be found in [15] and [16].

Source coding, in most cases (except for lossless coders), is a trade-off between encoding rate and decoding distortion

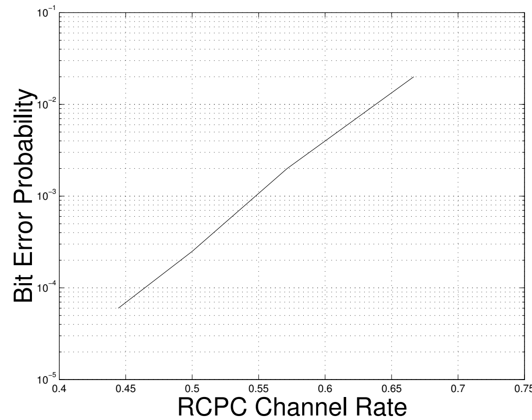

Fig. 4. Performance of an RCPC encoder

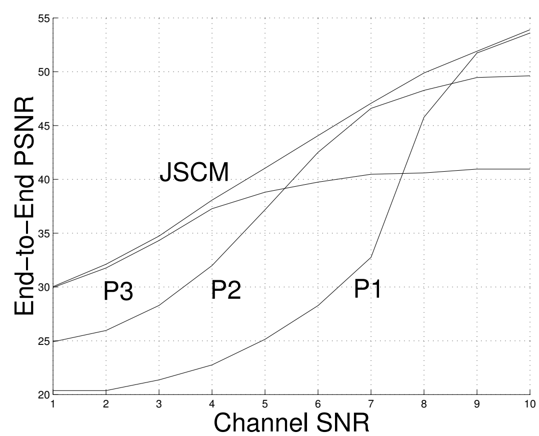

Fig. 5. Comparison of JSCC vs. Non-JSCC

(the rate-distortion curve for an H.263 encoder is shown in Figure 2.1). Thus, it is in our best interest to send the source information at the highest rate possible. However, to increase compression performance, most current video source coders assume perfect channels and proli£cally use data types such as marker bytes, which are sensitive to channel errors. To reduce the impact of channel errors, channel coding is needed (the performance of an RCPC coder is shown in Figure 2.1). There is a fundamental trade-off between source and channel coders because they both require system resources. Furthermore, most existing channel coders are designed for a speci£c channel and target bit error rate without explicit regard for the source coder characteristics, such as the importance of the bits being transmitted; a better collaboration between source and channel is bene£cial. Figure 2.1 shows the performance comparison of a JSCC-system and non-JSCC systems: while the non-JSCC systems all fail at certain ranges of the channel SNR, the JSCC system performs well across the entire SNR range by intelligently allocating system resources and achieves the optimal performance.

Based on the previous reasoning, we make the following observation:

Conjecture 1 *The expected optimal performance, P, for a certain user with regard to a speci£c JSCC-capable transmission system can be obtained by applying the JSCC principle on this system toward this particular user.*

2.2. Progressive Property and Video Source Coders

The objective of video source coding is to remove redundancies in the bitstream in the most ef£cient fashion. To achieve this, various advanced techniques such as motion-compensation are usually used. However, these techniques also introduce sensitive data types such as start codes and marker bytes, the corruption of which could lead to disastrous results. Their existence also renders layered service dif£cult.

Layered service usually means the application of unequal protection (UEP) to different sections of the source bit stream, which requires a layered source encoder. In other words, the source encoder should produce bits prioritized according to their relative importance. Thus, we can apply unequal protection in such a fashion that all users can receive at least a coarse version of the original, and when users' channel situations improve, the extra bits they receive re£ne the coarse version for better quality.

The progressive property also simpli£es the JSCC system to a great extent. With a non-progressive coder, raw video frames must be kept in memory until the JSCC system provides the source encoder the desired compression ratio to use; for a strictly progressive coder, we can compress the sequence at a reasonable rate, and later use simple truncation to get any desired rate (assuming all desired rates are no greater than the initial rate chosen.)

Based on the above reasoning and experimental results,

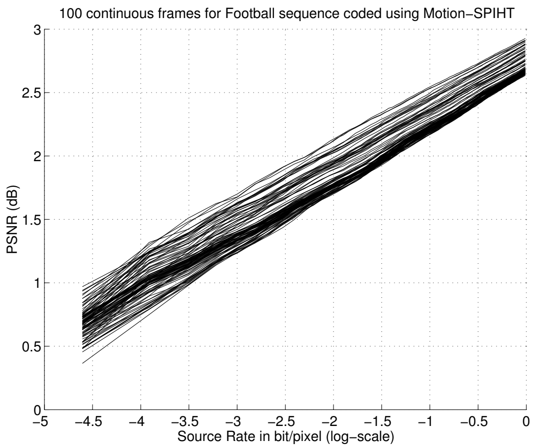

Fig. 6. The collection of rate-distortion curves for 100 consecutive frames in the football video sequence, each frame encoded using the SPIHT coder

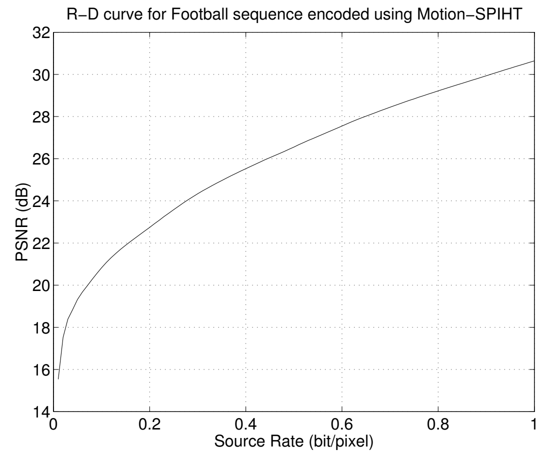

Fig. 7. Overall rate-distortion curve for the Football video sequence encoded using the Motion-SPIHT encoder

we make the following observation:

Conjecture 2 *It is necessary to have a (semi)progressive source encoder to develop a JSCC system with low disappointment ("semi" denotes the situation when the bitstream is progressive in the spatial dimension but not the temporal dimension).*

Conjecture 2 leads to our choice of two progressive video source coders used in simulation (we also use a non-progressive video source coder as comparison). Next we give a brief discussion of them and how to obtain their respective rate-distortion curves, by which the video source coders are characterized.

2.2.1. Motion-SPIHT Source Coder

Strictly speaking, Motion-SPIHT is not a high-performance video encoder, since it simply compresses every individual frame using the SPIHT image encoder [17] and does not take into consideration of the temporal redundancy between consecutive frames as do MPEG and H.263. However, we still elect to consider this coder because of its excellent progressive nature in the spatial dimensions. SPIHT produces a highly progressive output bit stream that can be truncated at any location and yet still produces a coarse version of the original.

It is observed that, when plotted in a log-log scale, R-D curves for frames in a Motion-SPIHT-coded video sequence can be approximated by a collection of converging linear functions (see Figure 6). The R-D function for any particular frame is thus determined by a single point on the curve, which is readily available when we perform the initial source compression. The overall rate-distortion curve for the Motion-SPIHT encoder is shown in Figure 7.

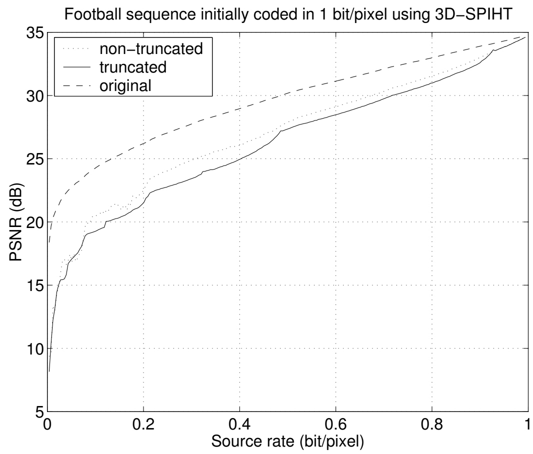

Fig. 8. Rate-distortion curve for the football video sequence encoded using the 3D-SPIHT encoder

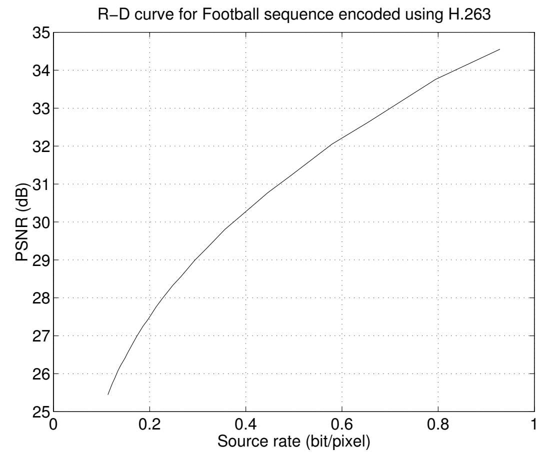

Fig. 9. Rate-distortion curve for a video sequence encoded using the H.263 encoder

2.2.2. 3D-SPIHT Source Coder

3D-SPIHT [18], an extension of the 2D-SPIHT image compression method [17], is another progressive video encoder. It takes video frame groups of size 16 and compresses them using the 3D-SPIHT algorithm. The bit stream it produces is almost-strictly progressive and has a typical rate-distortion curve as shown in Figure 8.

On Figure 8 the double-dash curve is the original 3D-SPIHT rate-distortion curve (obtained by encoding at different rates and calculating decoding distortion), the dotted curve is generated by introducing a byte error in the encoded 3D-SPIHT bitstream and the solid curve is generated by *truncating* the encoded 3D-SPIHT bitstream at the error location. It can be observed that the latter two curves are very close to each other, which again demonstrates the excellent progressive property of the 3D-SPIHT encoder. Either of these two curves is relatively easy to obtain and can be used as an approximation of the 3D-SPIHT rate-distortion curve.

However, since 3D-SPIHT achieves a high compression ratio by utilizing both the spatial and temporal redundancy in a video sequence, it gives only mediocre performance when the video sequence contains a high amount of motion information (less spatial and temporal redundancy). Thus, it is most suited for applications such as video-conferencing and video-phone.

2.2.3. Layered H.263 Source Coder

As a demonstration of the case when the source coder does not possess the progressive property, we use a layered H.263 video encoder as an example. The H.26x (H.261, H.263, H.263+) coders are based on advanced motion-compensation compression techniques and target mostly low-rate applications such as videoconferencing and videophone, where the input frames consist of mostly background information and the amount of motion is limited. It uses sensitive markers extensively to increase coding ef£ciency, so a random error can be disastrous. Figure 9 shows the rate-distortion characteristics of an H.263 encoder.

Our layered H.263 encoder encodes the same video sequences at different rates and then simply concatenates the compressed sequences together to form the layered stream. Each layer is an independent H.263-encoded sequence targeted for a speci£c performance level, typically in a descending order. Upon the reception of the signal, users decode the layer which gives least decoding distortion.

We note that this layered H.263 encoder is a highly inef£cient multi-resolution coder, with layers that are totally independent and cannot enhance each other. Furthermore, simulation results show that the inef£ciency of retransmitting coarse information at £ne layers renders anything more than three layers wasteful and impractical. However, the system is simple and exploits the increased compression ef£ciency of the advanced H.263 coder.

2.3. Channel Coders

The joint source-channel coding approach and minimax disappointment criteria do not impose any constraint on the channel coders. In principle, any channel coder, including popular coders such as Reed-Solomon, RCPC, and Turbo Codes, can be used so long as it is possible to modify the coder's bit-error-probability behavior. In our simulation we have opted to use the standard well-performing Rate-Compatible Punctured Convolutional Codes (RCPC); cases using other coders are essentially similar.

2.3.1. RCPC Channel Coder

RCPC codes, introduced by Hagenauer [19] [20], extend traditional convolutional codes by puncturing a low-rate $1/N$ code periodically with period P to obtain a family of codes with rate $P/(P+l)$, where l can be varied between 1 and $(N-1)P$. The *rate-compatible* restriction on the puncturing tables ensures that all code bits of high-rate codes are used by the lower-rate codes, thus allowing transmission of incremental redundancy in ARQ/FEC schemes and continuous rate variation to change from low to high error protection within a data frame. RCPC codes are good for source-channel coding because the basic structure of the codec remains the same as the coding rate changes.

Although there is not an analytical expression for error probability calculation for RCPC codes [21], there exist performance bounds that are relatively tight. A log-af£ne relationship has also been observed to exist between the block length and the error probability [10].

2.3.2. Adjusting Bit Transmission Power

We consider another case where we have a power-constrained transmission system that can adjust the transmission power down to bitwise granularity. Instead of using a channel coder, we allocate the total transmission power across the

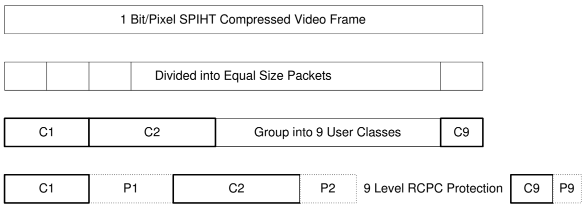

Fig. 10. The Channel Encoding Procedure

bitstream. We also make the assumption that the ef£ciency of the transmission system is a constant.

3. JSCM FRAMEWORK

In this section we explain in detail our four simulation systems and the underlying mathematical optimization problems.

3.1. Motion-SPIHT and RCPC: Rate-Constrained

First we consider the case where the channel capacity is limited. We choose Motion-SPIHT as the source encoder and RCPC as the channel encoder. The encoding procedure is illustrated in Figure 10.

1 Raw video frames (indexed by i) are compressed equally at one-bit-per-pixel and their respective rate-distortion functions are estimated.

2 Each frame is divided into £xed-size packets, of which only the £rst K_i will be sent. These K_i packets are grouped into J packet groups, where J is the number of available RCPC protection rates. The jth packet group for frame i contains m_{ij} packets, where the m_{ij} are not necessarily equal. The values for K_i, m_{ij} are determined by the system to yield the optimal performance, which, in our case, is the minimax disappointment for all the users.

3 RCPC protection of rate r_j is applied to the packet group j, which is then broadcast.

The minimax optimization problem can be shown mathematically as follows:

$$K_i = \sum_{j}^{J} m_{ij} \tag{3}$$

$$s_i = \sum_{j}^{J} m_{ij}\, r_j\, p \tag{4}$$

$$D_{i,l} = \sum_{k=1}^{K_i} \left[P_{k+1,l} \sum_{n=1}^{k} (1 - P_{n,l}) \right] d_i(k) \tag{5}$$

$$D_{0,l} = \min_{m_{ij}} \left(\frac{\sum_i^N D_{i,l}}{N} \right) \tag{6}$$

$$P = \max_{l, m_{ij}} \left(D_{0,l} - \frac{\sum_i^N D_{i,l}}{N} \right) \tag{7}$$

$$b_i = \max(b_{i-1} + s_i - R, 0) \tag{8}$$

$$b_i \leq b_{max},\ i = 1, 2, \ldots, N \tag{9}$$

where K_i, m_{ij}, r_j are de£ned as before, p is the £xed packet size in bytes, s_i is the total length (source plus protection) for the ith frame, $D_{i,l}$ is the expected received performance for the ith frame by the lth user, and $d_i(k)$ is the quality of the decoded ith frame if the kth packet is the £rst packet lost. $P_{n,l}$ is the transmission failure probability of the nth packet for the lth user (the m_{ij} packets in the jth packet group would have $P_{n,l}$ equal to the error probability for user l when using RCPC protection rate r_j), $D_{0,l}$ is the expected optimal performance for user l, which can be obtained by solving Equation 6, and N is the size of the video sequence. P is the maximum of the disappointment, the value we need to minimize. Note that we solve (6) and (7) as two separate optimization problems; the solution of (6) gives us $D_{0,l}$, while the solution of (7) gives us the optimal m_{ij} and minimax disappointment P.

A particular problem related with video transmission, called rate control [22], must also be addressed. Since video transmission usually allows only a small, £nite delay, and the decoder usually has a £nite decoding buffer size and constant decoding ¤ow, we should avoid buffer over¤ow or under¤ow. If we use b_i to denote the buffer occupancy at the ith time index and R to denote the constant decode ¤ow, we should ensure that b_i is never greater than b_{max}, the maximum decoding buffer-size. Our optimization becomes a constrained problem because of this particular requirement and the total channel capacity constraint.

One way to incorporate the rate-control constraint is to modify the cost function P by adding a penalty item associated with buffer over¤ow:

$$P' = P + C \times \sum_{i=1}^{N} \max(0, b_i - b_{max})$$

where C is a large constant. When C goes to in£nity, the non-constrained solution converges to the constrained solution; in a discrete optimization setting, they will be equal given C suf£ciently large.

To solve this optimization problem, we initialize the values for m_{ij} with the expectation that they are close to the optimal values, and employ a gradient-based procedure to derive the optimal solution. Usually at least $6 - 7$ iterations are required for the algorithm to converge, but since broadcasting parameter optimization is usually done of¤ine, we can afford the cost.

3.2. 3D-SPIHT and RCPC: Rate Constrained

In this case we substitude the Motion-SPIHT coder in the previous simulation system with the 3D-SPIHT video encoder. The same encoding procedure still applies except now instead of encoding on a frame basis, we encode and transmit the entire GOP in one step. We do not have frame-wise rate-control issues with this transmission scheme; on the other hand, a total delay of one GOP is introduced.

Similar to the previous case, the minimax optimization problem can be shown mathematically as follows:

$$K = \sum_{j}^{J} m_j \quad (10)$$

$$s = \sum_{j}^{J} m_j \, r_j \, p \quad (11)$$

$$D_l = \sum_{k=1}^{K} \left[P_{k+1,l} \sum_{n=1}^{k} (1 - P_{n,l}) \right] d(k) \quad (12)$$

$$D_{0,l} = \min_{m_j} D_l \quad (13)$$

$$P = \max_{l,m_j} (D_{0,l} - D_l) \quad (14)$$

the entire 1 bit/pixel encoded 3D-SPIHT bitstream is £rst divided into packets of size$p$, and only the £rst K would be sent; The K packets would be further divided into J groups, each containing m_j packets; each packet group is assigned a particular RCPC encoding rate r_j; $P_{n,l}$ is again the transmission failure probability of the nth packet for the lth user; $d(k)$ is the decoding distortion if the kth packet is the £rst packet lost; D_l is the expected distortion for the l-th user; $D_{0,l}$ is the expected minimal distortion for the l-th user; and P is our minimax disappointment. Equation (13) de£nes the single user JSCC optimization problem, and Equation (refeqn:md5) de£nes the minimax disappointment optimization problem for this broadcasting system. They can be solved using exactly the same technique used in the previous case.

3.3. Layered-H.263 and RCPC: Rate Constrained

In this case we again consider the situation when we have a limited channel capacity. We choose layered-H.263 as the source encoder. The encoding procedure can be described as follows:

1. Obtain all layers by performing multiple compression using predetermined layer-speci£c compression parameters.

2. Apply RCPC protection with different rates to each layer.

3. Concatenate the encoded sequences together and transmit.

This minimax problem can be mathematically expressed as follows:

$$\sum_{i}^{N} s(q_i) r_i \leq C \quad (15)$$

$$D_k = \sum_{i}^{N} \prod_{j}^{i-1} P_k(q_j, r_j) (1 - P_k(q_i, r_i)) d(q_i) \quad (16)$$

$$\log(P_k(q_i, r_i)) = A_k s(q_i) r_i + B_k \quad (17)$$

$$D_{0,k} = \min_{Q,R} D_k \quad (18)$$

$$P = \max_{k,Q,R} (D_{0,k} - D_k) \quad (19)$$

where N is the number of layers, q_i is the i-th quality factor we use to compress the sequence, $s(q_i)$ is the resulting compressed sequence length, r_j is the RCPC coding rate for the i-th layer; the sum of the length of all layers with channel protection should be less than the total channel capacity. $P_k(q_j, r_j)$ is the *layer-error probability* for the j-th layer and k-th user; we approximate its value using a log-af£ne function given in (17) (the channel SNR for the k-th user is then hidden inside parameter A_k and B_k, which need to be pre-estimated before optimization) [10]. $d(q)$ is the decoding distortion when using quality factor q. Q and R are the quality factor vector $(q_1, q_2, \ldots, q_N)$ and the RCPC coding rate vector $(r_1, r_2, \ldots, r_N)$. (18) and (19) again de£ne the joint source-channel optimization problem for a single user and the minimax optimization problem for the entire broadcasting system. They can similarly be solved using gradient-descent-based algorithms.

3.4. 3D-SPIHT: Power Constrained

Unlike the previous three cases, this time we consider the case where the total system transmission power is limited. We employ 3D-SPIHT as the source coder and adjust the transmission power for each bit in order to achieve unequal protection. The transmission bit-power pro£le is shown in Figure 11 and the transmission procedure can be described as follows:

1 Raw video frames are £rst divided into groups of size 16 and compressed using the 3D-SPIHT video encoder; its rate-distortion curve is then estimated.

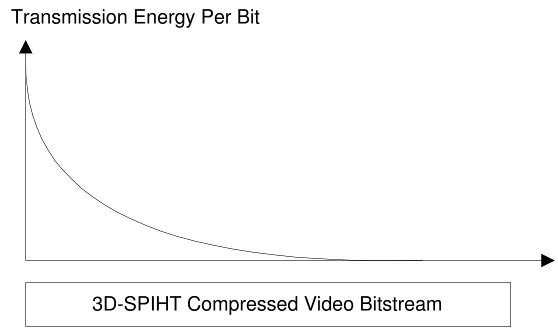

Fig. 11. Power-Constrained System

2 The compressed video source stream is transmitted with energy per bit optimized under the total energy constraint.

The minimax optimization problem can be cast mathematically as follows:

$$\sum_{j}^{L_s} e_j \leq E_{total} \tag{20}$$

$$p_k(e_j) = Q\left(\sqrt{\frac{2e_j}{N_{0,k}}}\right) \tag{21}$$

$$D_k = \prod_{j=1}^{L_s} \left(1 - p_k(e_j)\right) D_c + \sum_{i=1}^{L_s} \left[\prod_{j=1}^{i-1} \left(1 - p_k(e_j)\right)\right] p_k(e_i)\, d(i) \tag{22}$$

$$D_{0,k} = \min_{E} D_k \tag{23}$$

$$P = \max_{k,E} \left(D_{0,k} - D_k\right) \tag{24}$$

where E_{total} is the total limit on system transmission power, $p_k(e_j)$ is the transmission error probability for the jth bit for the kth user, given bit energy e_j and channel noise $N_{0,k}$; D_k is the resulting distortion for the kth user; L_s is the total number of bits and D_c is the distortion caused by source compression. $d(i)$ is essentially the rate-distortion curve for the current video stream; it is the decoding distortion if the ith bit is the £rst bit in error (we discard all bits after it); E is the bit transmission power pro£le vector $(e_1, e_2, \ldots, e_{L_s})$. $D_{0,k}$ is the expected optimal performance for the kth user, which can be obtained by solving (23), and we obtain P, the maximum disappointment value for all users, by solving (24).

We try to minimize P by £nding the optimal energy allocation vector E $(e_1, e_2, \ldots, e_{L_s})$. This minimax optimization problem can be similarly solved by a gradient-descent algorithm. Rate-control is not a concern in this case since we always encode a group of frames as an entity and do not start decoding until we receive the entire stream. This entails an initial delay of one group of frames, which is acceptable in typical video-conferencing applications.

4. MINIMAX OPTIMIZATION

In this section we elaborate on the performance and convergence issues of our algorithms used to solve the minimum disappointment minimax [23] optimization problem. We focus on the case of a power-constrained system.

The standard minimax optimization problem with constraints can be stated as follows: Let $f_i(X), i \in [0:N], X = (x_1, \ldots, x_n)$ be functions de£ned and continuously differentiable on some open set $\Omega' \subset E_n$; let Ω be a convex closed (not necessarily bounded) subset of Ω'. The objective is to £nd a point $X^* \in \Omega$ such that

$$\max_{i \in [0:N]} f_i(X^*) = \inf_{X \in \Omega} \max_{i \in [0:N]} f_i(X)$$

Consider the function de£ned on Ω'

$$\phi(X) = \max_{i \in [0:N]} f_i(X)$$

The problem we have set is precisely to minimze the function $\phi(X)$ on the set Ω. Also de£ne the following set $R(X)$

$$R(X) = \{i \in [0:N] \mid f_i(X) = \phi(X)\}$$

The necessary (and suf£cient for convex cost functions) conditions for a minimax solution is stated in the following theorem:

Theorem 1 *[23] A necessary condition for a point $X* \in \Omega$ to be a minimum point of $\phi(X)$ on Ω is that*

$$\inf_{Z \in \Omega} \max_{i \in R(X^*)} \left(\frac{\partial f_i(X^*)}{\partial X}, Z - X^* \right) = 0$$

If $\phi(X)$ is convex, this condition is also suf£cient, and the point X^* is called a stationary point of $\phi(X)$ on Ω.

Condition 1 is equivalent to the following theorem:

Theorem 2 *[23] In order that $\phi(X)$ have a minimum on E_n at a point X^*, it is necessary, and if $\phi(X)$ is convex,* *also suf£cient that*

$$\inf_{\substack{g\in\bar{\Gamma}(X^*)\\ \|g\|=1}} \max_{i\in R(X^*)} \left(\frac{\partial f_i(X^*)}{\partial X},\ g\right) \geq 0$$

Theorem (1) and (2) essentially state that we can apply gradient-descent type algorithms to solve the minimax optimization problem. Based on the above theorems, we employ the £rst method of successive approximations [23] (a direct generalization of the steepest-descent algorithm) to solve the minimax optimization problem. For details of the algorithm please see the Appendix. We discuss the convexity of our $\phi(X)$ function in the following section.

4.1. Convexity of the Cost Function

The necessary and suf£cient conditions for a minimax solution stated in the previous section requires that the cost function $\phi(X)$ be convex, which is equivalent to the convexity of $f_i(X)$, based on Theorem 3. Thus, it is of great interest to investigate the convexity of $f_i(X)$ (in the following derivations we treat concavity as equivalent to convexity since it can be corrected by a simple change of sign.) In showing $f_i(X)$ is convex, we also prove that the joint source-channel optimization problem for a single user class has a convex cost function, and thus various gradient-descent algorithms can be used with a guarantee of convergence to a global minimum.

Theorem 3 *[23] Let Ω be a convex set. If all $f_i(X)$ are convex on Ω, then $\phi(X)$ is also convex on Ω.*

We make the following observations:

1. The rate-distortion functions for both the Motion-SPIHT and the 3D-SPIHT encoders are *continuous, monotonic, and convex*. This is not strictly true in a real situation, but it is a reasonable assumption to make.

2. The error probability versus bit-energy function of an AWGN channel is expressed by the Q function.

We revisit the mathematical expression for our optimization problem and relate it with the minimax optimization theorems:

$$\begin{aligned} D_k &= \prod_{j=1}^{L_s} (1 - p_k(e_j))\, D_c + \\ &\quad \sum_{i=1}^{L_s} \left[\prod_{j=1}^{i-1} (1 - p_k(e_j))\right] p_k(e_i)\, d(i) \quad (25) \\ f_k(E) &= D_{0,k} - D_k, \quad E = (e_1, e_2, \ldots, e_{L_s}) \quad (26) \\ \phi(E) &= J = \max_{k,E} (D_{0,k} - D_k) \quad (27) \end{aligned}$$

Based on assumptions 1 and 2, both the Q function and $d(i)$ are continuous, twice differentiable functions. We show that D_k is a convex function by £rst calculating the gradient of the distortion D_k with regard to the l-th energy factor e_l:

$$\frac{\partial D_k}{\partial e_l} = -D_c \frac{\partial p_k(e_l)}{e_l} \prod_{j=1, j\neq l}^{L_s} [1 - p_k(e_j)]$$
$$+d(l) \frac{\partial p_k(e_l)}{\partial e_l} \prod_{j=1}^{l-1} [1 - p_k(e_j)]$$
$$-\sum_{i>l}^{L_s} \left\{ \frac{\partial p_k(e_l)}{\partial e_l} \prod_{j=1, j\neq l}^{i-1} [1 - p_k(e_j)] \right\} p_k(e_i) d(i) \quad (28)$$
$$= \frac{\partial p_k(e_l)}{\partial e_l} \Big\{ - \prod_{j=1, j\neq l}^{L_s} [1 - p_k(e_j)] \, D_c$$
$$+ \prod_{j=1}^{l-1} (1 - p_k(e_j)) d(l)$$
$$- \sum_{i>l}^{L_s} \left[\prod_{j=1, j\neq l}^{i-1} (1 - p_k(e_j)) \right] p_k(e_i) d(i) \Big\} \quad (29)$$

Denote the item in the braces as A (note that A does not depend on e_l) and take the second-order derivative and we get:

$$\frac{\partial^2 D_k}{\partial e_l^2} = \frac{\partial^2 p_k(e_l)}{\partial^2 e_l} A \quad (30)$$

so the convexity of D_k is determined by convexity of $p_k(e_l)$. We next observe the following:

$$p_k(e_j) = Q\left(\sqrt{\frac{2e_j}{N_{0,k}}}\right) \quad (31)$$
$$\frac{\partial p_k(e_j)}{\partial e_j} = \frac{1}{\sqrt{2\pi} N_0} e^{-\frac{e_j}{N_0}} {e_j}^{-\frac{1}{2}} \quad (32)$$
$$\frac{\partial^2 p_k(e_j)}{\partial {e_j}^2} = \frac{1}{\sqrt{2\pi} N_0} e^{-\frac{e_j}{N_0}} {e_j}^{-\frac{1}{2}} \left(-\frac{1}{N_0} - \frac{1}{2e_j}\right) \quad (33)$$

We note that the second derivative of $p_k(e_j)$ is always negative when e_j is positive, thus $p_k(e_j)$ is convex over this region, and so is D_k. Following Theorem 3, $\phi(E) = J$ is thus convex.

4.2. Simpli£cation of the Optimization Problem

In the power-constrained case, since we are optimizing with regard to each bit, the size of the optimization problem can become relatively large when we have a large number of bits to transmit. To reduce the computational complexity, we can make the following two simpli£cations.

The £rst simpli£cation comes from an observation during simulation. We observe that when we have a reasonable number of user classes (more than seven), with channel SNRs relatively spread out, the optimal energy allocation vector for the minimax optimization problem is approximately a linear combination of all the optimal energy allocation vectors for the joint source-channel coding problem for each individual user class. Thus, we can reduce the dimension of the problem down to less than 10 variables by solving for the linear scaling factors. This signi£cantly reduces computational complexity.

For the second simpli£cation, instead of adjusting the transmission power for each bit, we group bits into blocks and adjust the transmission power on a block basis. Using blocks of bits means a lesser computational load but possibly larger minimax disappointment since we have lost part of the ¤exibility to adjust. A reasonable block size is essential to the system performance (a typical value could be 8, for a byte). Furthermore, we need to show that the cost function is still convex. Assuming a block size of N, we make the following modi£cations to the original expressions:

$$
\begin{aligned}
L_B &= \left\lfloor \frac{L_s}{N} \right\rfloor && (34)\\
P_k(e_j) &= \sum_{i=1}^{N} \binom{N}{i} p_k(e_j)^i \\
&\quad (1-p_k(e_j))^{(N-i)} && (35)\\
D_k &= \prod_{j=1}^{L_B} (1-P_k(e_j))\, D_c + \\
&\quad \sum_{i=1}^{L_B} \left[\prod_{j=1}^{i-1} (1-P_k(e_j)) \right] \\
&\quad P_k(e_i)\, d(iN) && (36)
\end{aligned}
$$

where N is the bit-block size, L_B is the total number of blocks, $P_k(e_j)$ is the *block* transmission error probability if each bit inside this block is transmitted with power e_j (we assume that the total block is lost if even a single bit is in error and ignore the location of the error).

Similarly, we can show that the convexity of the cost function depends on the convexity of $P_k(e_j)$. We can calculate the second derivative of $P_k(e_j)$ as follows:

$$
\begin{aligned}
\frac{\partial P_k(e_j)}{\partial e_j} &= N \frac{\partial p_k(e_j)}{\partial e_j} (1-p_k(e_j))^{N-1} && (37)\\
\frac{\partial^2 P_k(e_j)}{\partial {e_j}^2} &= N(1-p_k(e_j))^{N-2} \Bigg(\frac{\partial^2 p_k(e_j)}{\partial {e_j}^2} \\
&\quad - (N-1)(\frac{\partial p_k(e_j)}{\partial e_j})^2 \Bigg) && (38)\\
\frac{\partial^2 p_k(e_j)}{\partial {e_j}^2} &< 0 && (39)
\end{aligned}
$$

thus we conclude that the second derivative of $P_k(e_j)$ is always negative, so the cost function remains convex.

5. SIMULATION RESULTS

In order to demonstrate the use of minimax disappointment criteria and compare systerm performances, we choose to use the same set of simulation parameters for all four simulation systems:

1. Simulation sequence: fast-motion football test sequence cropped to QCIF format.
2. Simulation channel: AWGN channels with SNRs ranging from 1dB to 9dB.
3. Number of user classes: 5 classes evenly spread out in the entire SNR range.

The last example is different from the previous three because it uses an energy-varying transmission with a power constraint instead of a rate-constraint. For this case we cannot directly compare its performance with the other three cases.

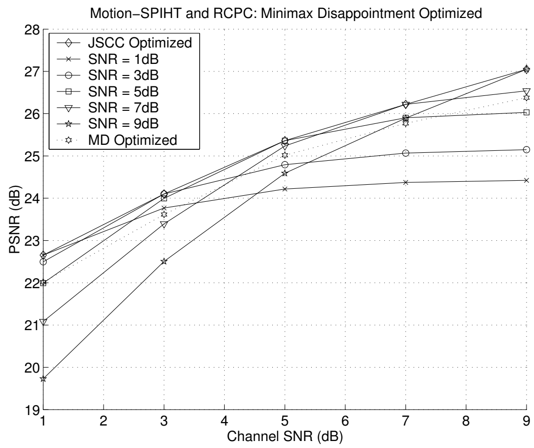

Fig. 12. Simulation results for a Minimax-Disappointment optimized system with £ve user classes in a rate-constrained situation, using Motion-SPIHT source coder and RCPC channel coder with 9 available protection rates

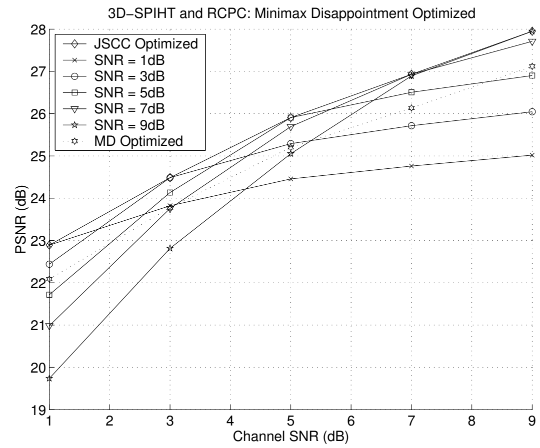

Fig. 13. Simulation results for a Minimax-Disappointment optimized system for £ve user classes in a channel-capacity-constrained situation, using 3D-SPIHT source coder and RCPC channel coder with 9 available rates.

In order to adapt to the human visual system characteristics, we also elect to calculate minimax disappointment using the log-scale PSNR (Peak Signal-to-Noise Ratio).

5.1. Motion-SPIHT and RCPC

To test the Motion-SPIHT-RCPC simulation system, the broadcasting station transmits the test sequence using the optimized system parameters with nine available RCPC protection rates. At the decoder end, a constant decode buffer size of approximately 1 bit/pixel is used. Simulation results are shown in Figure 12.

From Figure 12 we observe that the received performance for all users of our MD-optimized system is roughly the same distance from their individual expected best performance; in other words, they experience the same level of disappointment, whereas in the other system con£gurations optimized toward one singel user class, one or more users are invariably highly disappointed; i.e, their received performance is far below their best expectation. Thus, our system does give the optimal performance in terms of the minimax disappointment criterion, achieving a minimax disappointment of 0.6703 dB.

5.2. 3D-SPIHT and RCPC

In the case of 3D-SPIHT with RCPC simulation, the result (shown in Figure 13) is quite similar with the previous case. We again observe that systems designed for individual user class suffer from great disappointment values for other users, and the minimax disappointment optimized system gives each user class approximately the same level of performance degradation (0.82 dB).

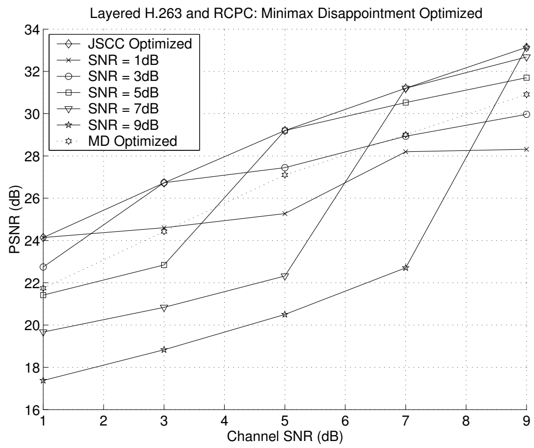

Fig. 14. Simulation results for a Minimax-Disappointment optimized system for £ve user classes in a channel-capacity-constrained situation, using layered-H.263 source coder and RCPC channel coder with 9 available rates.

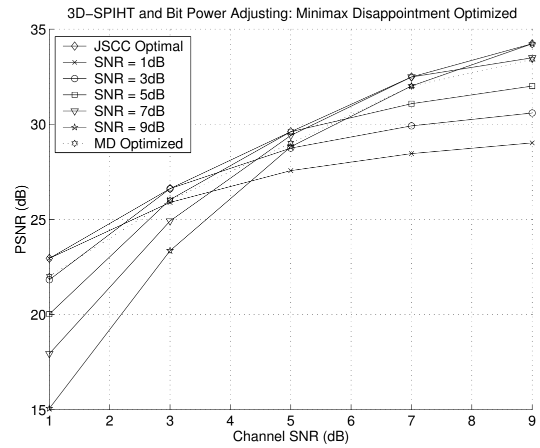

Fig. 15. Simulation results for a Minimax-Disappointment optimized system toward £ve user classes in a transmission-power-constrained situation, using 3d-SPIHT source coder and a bit-power-adjustable transmission scheme

5.3. Layered-H.263 and RCPC

In the case of layered-H.263 with RCPC, we use the same simulation sequence and RCPC codes as before; the result is shown in Figure 14. In this case, we observe that not only the amount of disappointment experienced by each user varies, but also the level of maximum disappointment in the system is much greater (almost 2.4 dB) than the previous cases. This behavior is expected because of the inef£ciency of retransmitting the coarse information in every layer.

5.4. 3D-SPIHT and Energy Adjustment

To demonstrate the generality of the minimax disappointment criterion, we take an alternative approach by using the 3D-SPIHT-Energy simulation system. We similarly make the assumption that we have £ve classes of users mapped to different SNR levels of AWGN channels, and the transmission system can adjust the power for each bit. A total power constraint of E_{total} is also assumed. Simulation results are shown in Figure 15.

From the plot we again observe that all users experience approximately the same level of disappointment in our MD-optimized system, while using a transmission bit-energy pro£le optimized for any single user leads to one or more highly disappointed users. Our system achieves a minimax disappointment of 0.9831 dB.

5.5. Result Comparison

Comparing the result for Motion-SPIHT, 3D-SPIHT and layered-H.263 under the capacity constrained situation, we can see that for any individual user, using layered-H.263 with JSCC actually gives a higher absolute performance because of the H.263 encoder's much greater coding ef£ciency. However, the minimax disappointment for layered-H.263 is much higher because its lack of progressive property makes unequal-error-protection dif£cult and inef£cient. This example illustrates that the minimax disappointment approach can be applied with success both to progressive and non-progressive video source coders, but that a highly-progressive coder is needed to maximally exploit the potential of minimax disappointment broadcasting.

6. CONCLUSIONS

Video broadcasting systems have always been in need of an intelligent and meaningful performance criterion that supports a desirable trade-off between multiple classes of users and remains tractable. The maximum value of the deviation between received performance and expected best performance for each user introduced in this paper largely satis£es this need. It is a fair criterion toward all users, and theoretically appealing because it falls into the minimax optimization framework. We observe that little penalty is paid for layered service levels using progressive source coders; our simulation showed less than one dB of performance loss for £ve user classes. Although we proposed the minimax disappointment criterion based on video transmission and the joint source-channel matching principle, we believe that it is limited to neither and can be used as a universal broadcasting performance criterion.

The minimax disappointment optimization achieves almost equal disappointment for multiple classes of users, each corresponding to a particular received SNR. However, minimax disappointment is not guaranteed for SNRs between these levels. A smoother distribution of disappointment *vs.* SNR can be achieved by increasing the number of user classes, and arbitrary precision can be obtained by introducing a suf£ciently large number of user classes. A gradient-based algorithm effectively £nds the optimal operating parameters.

7. APPENDIX: THE FIRST METHOD OF SUCCESSIVE APPROXIMATIONS

In this section we discuss very brie¤y the £rst method of successive approximation for solving discrete minimax optimization problems. Instead of explaining the details of how the algorithm is applied to our problem, we introduce the fundamental algorithm itself to be conceptually clearer. For more details, please refer to [23].

We start by £rst introducing the following notations (definition of $\phi(X)$, $f_i(X)$ and g are as before):

$$\begin{aligned} R_\epsilon(X) &= \{i \mid \phi(X) - f_i(X) \le \epsilon\}, \quad \epsilon \ge 0 \\ \psi_\epsilon(X) &= \min_{\|g\|=1} \max_{i \in R_\epsilon(X)} \left(\frac{\partial f_i(X)}{\partial X},\ g \right) \end{aligned}$$

We also de£ne ϵ-stationary point X_ϵ^* and ϵ-steepest descent direction $g_\epsilon(X)$ as follows:

De£nition 1 *A point $X^* \in E_n$ for which*

$$\psi_\epsilon(X^*) \geq 0$$

is known as an ϵ-stationary point of $\psi(X)$.

De£nition 2 *A vector $g_\epsilon(\bar{X})$, $\|g_\epsilon(\bar{X})\| = 1$, is known as a direction of ϵ-steepest descent of $\phi(X)$ at the point $\bar{X}$ if*

$$\max_{i \in R_\epsilon(\bar{X})} \left(\frac{\partial f_i(\bar{X})}{\partial X}, g_\epsilon(\bar{X}) \right) = \psi_\epsilon(\bar{X})$$

The £rst method of successive approximations can now be described by the following procedure:

1. We £x two parameters $\epsilon_0 > 0$ and $\rho_0 > 0$, take random initial approximation X_0.

2. Suppose we found k-th approximation X_k, if $\psi(X_k) < 0$, then we check sequence $\epsilon_\nu = \epsilon_0 / 2^\nu$ and £nd the £rst ν such that

$$\psi_{\epsilon_\nu}(X_k) \leq -\frac{\rho_0}{\epsilon_0}\epsilon_\nu$$

3. Set $g_k = g_{\epsilon_k}(X_k)$, the ϵ-steepest descent direction and update X_k as

$$\begin{aligned} \phi(X_k(\alpha_k)) &= \min_{\alpha \in [0:\infty]} \phi(X_k(\alpha)) \qquad (40) \\ X_{k+1} &= X_k(\alpha_k) = X_k + \alpha_k g_k \qquad (41) \end{aligned}$$

4. Iterate until convergence occurs.

Note that this algorithm guarantees convergence to a local minimum if $\phi(X)$ is convex.

8. REFERENCES

[1] S. B. Z. Azami, P. Duhamel, and O. Rioul, "Joint source-channel coding: Panorama of methods," in *Proceedings of CNES Workshop on Data Compression*, Nov. 1996.

[2] Xie Jun and A. R. Barron, "Asymptotic minimax regret for data compression, gambling, and prediction," *IEEE Transacations on Information Theory*, vol. 46, pp. 431–445, March 2000.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[4] G. M. Davis and J. M. Danskin, "Joint source and channel coding for image transmission over lossy packet networks," in *SPIE Conference on Wavelet Applications of Digital Image Processing XIX*, Aug. 1996, vol. 2847, pp. 376–387.

[5] K. Ramchandran, A. Ortega, K. Uz, and M. Vetterli, "Mutilresolution broadcast for digital HDTV using

joint source/channel coding," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 6–23, Jan. 1993.

[6] S. B. Z. Azami, O. Rioul, and P. Duhamel, "Performance bounds for joint source-channel coding of uniform memoryless sources using a binary decomposition," in *Proceedings of European Workshop on Emerging Techniques for Communication Terminals*, Jul. 1997, pp. 259–263.

[7] B. Belzer, J. D. Villasenor, and B. Girod, "Joint source-channel coding of image with trellis coded quantization and convolutional codes," in *Proceedings of ICIP'95*, Aug. 1995, vol. 2, pp. 85–88.

[8] G. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Letters*, vol. 4, pp. 72–81, Jul. 1997.

[9] H. Man, F. Kossentini, and M. Smith, "Progressive image coding for noisy channels," *IEEE Signal Processing Letters*, vol. 4, pp. 8–11, Aug. 1997.

[10] Jin Lu, Aria Nosratinia, and Behnaam Aazhang, "Progressive source-channel coding of images over bursty error channels," in *Proceedings of ICIP98*, 1998, vol. 2, pp. 127–131.

[11] M. P. C. Fossorier, Z. Xiong, and K. Zeger, "Joint source-channel image coding for a power constrained noisy channel," in *Proceedings of ICIP98*, 1998, vol. 2, pp. 122–126.

[12] M. Brystrom and J. W. Modestino, "Combined source channel coding for transmission of video over a slow-fading Rician channel," in *Proceedings of ICIP98*, 1998, vol. 2, pp. 147–151.

[13] Tse-Hua Lan and Ahmed H. Tew£k, "Power optimized mode selection for H.263 video coding and wireless communications," in *Proceedings of ICIP98*, 1998.

[14] H. Zheng and K. J. R. Liu, "Image and video transmission over wireless channel: A subband modulation approach," in *Proceedings of ICIP98*, 1998, vol. 2, pp. 132–136.

[15] Leiming Qian, "Joint source-channel matching for wireless video transmission," M.S. thesis, University of Illinois at Urbana-Champaign, http://www.ifp.uiuc.edu/˜ lqian/Research/JSCC/jscc.html, 1998.

[16] S. Appadwedula, D. L. Jones, K. Ramchandran, and I. Konzintsev, "Joint source-channel matching for a wireless communications link," in *Proceedings of ICC 98*, Jun. 1998, vol. 1, pp. 482–486.

[17] A. Said and W. A. Pearlman, "A new fast and ef£cient image codec based on set partitioning in hierarchical

trees," in *IEEE Trans, Circuits and Systems for Video Technology*, Jun. 1996, vol. 6, pp. 243–250.

[18] B.-J Kim and W. A. Peralman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proceedings of DCC'97*, Mar. 1997, pp. 251–260.

[19] Joachim Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Tranaction on Communications*, vol. 36, pp. 389–400, Apr. 1988.

[20] J. Hagenauer and T. Stockhammer, "Channel coding and transmission aspects for wireless multimedia," in *Proceedings of IEEE*, Oct. 1999, vol. 87, pp. 1764–1777.

[21] J. Hagenauer, N. Seshadri, and C-E. W .Sundberg, "The performance of rate-compatible punctured convolutional codes for digital mobile radio," *IEEE Transaction on Communications*, vol. 38, pp. 966–980, Jul. 1990.

[22] L-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 3, pp. 665–676, Feb. 1998.

[23] V. F. Demyanov and V. N. Malozemov, *Introduction to Minimax*, Dover Publications, Inc, New York, 1990.