# ADAPTIVE JOINT SOURCE-CHANNEL CODING USING RATE SHAPING[+]

*Trista Pei-chun Chen and Tsuhan Chen*[*]

Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, U.S.A.
Email: {peichun, tsuhan}@andrew.cmu.edu

## ABSTRACT

*We present in this paper an adaptive joint source-channel coding scheme using rate shaping on pre-coded video data. Rate shaping selectively drops portions of the video bitstream before transmitting them in order to satisfy the network bandwidth requirement. In wireless multimedia transport over heterogeneous networks, limited bandwidth is not the only issue. The high error rate of the channel should be considered as well, so channel coding is often applied. We propose a rate shaping method that drops not only the source-coding segments of the video bitstream, but also the channel-coding segments of the video bitstream, adaptively according to the network condition, in order to achieve the optimal rate-distortion performance. The proposed method is based on "discrete rate-distortion combination" to accomplish joint source-channel coding. We consider both the simulcast and multicast scenarios and show promising results.*

## 1. INTRODUCTION

Video transmission is challenging in nature because it has high data rate compared to other data types/media such as text or audio. In addition, the channel bandwidth limit and error prone characteristics also impose constraints and difficulties on video transport. A joint source-channel coding approach is needed to adapt the video bitstream to different channel conditions. The joint source-channel coding approach should be scalable as well to be applied to a multicast scenario.

We propose a joint source-channel coding scheme based on the concept of *rate shaping* to accomplish the task of video transmission. The video sequence is first source coded followed by channel coding. Popular source coding methods are H.263 [1], MPEG 2 [2], etc. Example channel coding methods are Reed-Solomon codes, BCH codes, and the recent turbo codes [3][4]. The source and channel coded video bitstream then passes through the rate shaper to fit the channel bandwidth requirement while achieving the best reconstructed video quality. As opposed to conventional rate shapers [5][6], the proposed rate shaper explicitly considers the channel error condition.

The proposed rate shaper performs *discrete rate-distortion combination* to the encoded video bitstream before it sends the video bitstream to the channel. The proposed discrete rate-distortion combination algorithm is especially computationally efficient and performs almost as well as the exhaustive search.

For the multicast scenario, we propose to use a modified path-based rate shaper. In large-scale heterogeneous networks, this is especially useful. The network cannot afford the simulcast transmission specifically tuned for each user. On the other hand, with multicast, a single adaptable video bitstream can be decoded by many users. Therefore, multicasting saves a lot of bandwidth. The proposed path-based rate shaper generates video bitstream that is decodable for both the high-end user, such as a broadband network user, as well as the wireless phone user, with limited bandwidth and error-prone channel.

This paper is organized as follows. Section 2 starts with the system description of the joint source-channel coder. Source and channel coding methods are described. In Section 3, we introduce the proposed rate-distortion optimized rate shaping, with illustration of the *discrete rate-distortion combination algorithm*, and with the experiment showing the performance of the proposed algorithm. In Section 4, we introduce *path-based rate shaping;* and the experiment shows that it utilizes the bandwidth efficiently. We conclude our work and provide future research directions in the last section.

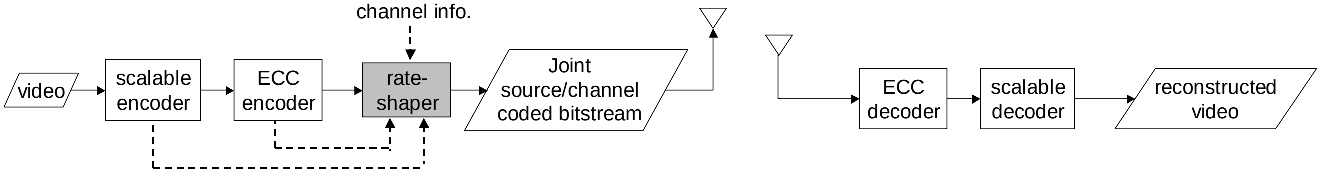
## 2. SYSTEM DESCRIPTION

To transmit the video bitstream effectively, we propose the joint source-channel coding system as shown in Figure 1. In this system, the video sequence is first source coded. We in particular use the scalable video coding [1][2]. With scalable video coding, the video sequence is encoded into several layers: the base layer and the enhancement layers with quality refinements to the base layer video bitstream. After the source coding stage, channel coding is performed to each layer of the video bitstream by an error correction code encoder (shown as ECC encoder in Figure 1). Applying channel codes with different error correction capabilities to different layers of the video bitstream is known as unequal error protection (UEP). We apply Reed-Solomon codes [3] to the source coded video bitstream. The proposed rate shaper then shapes the source and channel coded video bitstream by selectively sending portions of the video bitstream to the channel. Taking information from the source encoder, the channel encoder, and the channel, the rate shaper optimizes video transmission in the sense of rate-distortion. The decoder then decodes the received video bitstream.

---

[+] Published in Proc. of ICASSP 2002, Orlando, Florida, May 13-17, 2002

[*] Corresponding author: Tsuhan Chen

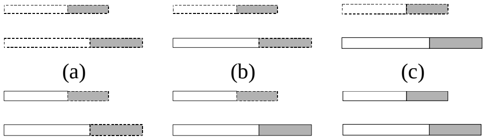


(a) (b)

**Figure 1. System diagram of the joint source-channel coder: (a) encoder; (b) decoder**

## 3. RATE-DISTORTION OPTIMIZED RATE SHAPING

After the video sequence has been source and channel coded, the rate shaper then decides which portions of the encoded video bitstream will be sent to the channel. Let us consider the case where the video sequence is scalable coded into two layers: one base layer and one enhancement layer. Each of the two layers is error correction coded with different error correction capability. Thus, there are four *segments* in the video bitstream: the source-coding segment of the base layer bitstream (lower left segment of Figure 2 (f)), the channel-coding segment of the base layer bitstream (lower right segment of Figure 2 (f)), the source-coding segment of the enhancement layer bitstream (upper left segment of Figure 2 (f)), and the channel-coding segment of the enhancement layer bitstream (upper right segment of Figure 2 (f)). The rate shaper will decide which of the four segments to send. The rate shaper cannot randomly choose any combination of these segments to send. For example, if the channel-coding segment of the base layer is chosen, the source-coding segment of the base layer should be chosen as well. In the two-layer case, there are totally six valid combinations of segments (Figure 2 (a)~(f)). The segments with the solid boundary are the ones being sent, while the segments with the dash boundary are the ones being dropped. We call each valid combination a *state*. Each state is represented by a pair of integers $(x, y)$, where $x$ is the number of source-coding segments chosen counting from the base layer and $y$ is the number of channel-coding segments chosen counting from the base layer. $x$ and $y$ satisfy the relationship of $x \geq y$.

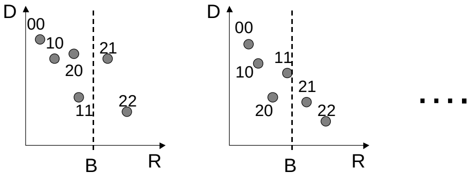


(a) (b) (c)

(d) (e) (f)

**Figure 2. Valid states: (a) State (0,0); (b) State (1,0); (c) State (1,1); (d) State (2,0); (e) State (2,1); (f) State (2,2)**

The decision of the rate shaper can be optimized given the rate-distortion map, or R-D map, of each coding unit. A coding unit can be a frame, a macroblock, etc., depending on the granularity of the decision. The R-D maps vary with different channel error conditions. Given the R-D map of each coding unit with a different constellation of states (Figure 3), the rate shaper finds the state with the minimal distortion under certain bandwidth constraint "B". In the example of Figure 3, State (1,1) of Unit 1 and State (2,0) of Unit 2 are chosen. Such decision is made on each of the coding unit given the bandwidth constraint "B" of that unit.



(a) (b) (c)

**Figure 3. R-D maps of coding units: (a) Unit 1; (b) Unit 2; (c) Unit 3 and so on**

Consider taking a frame as a coding unit. Video bitstream is typically coded with variable bit rate in order to maintain constant video quality. To minimize the overall distortion for a group of pictures/frames (GOP), it is not enough to choose the state for each frame based on the equally allocated bandwidth to every frame. We will introduce a smart rate shaping scheme that allocates different bandwidth to each frame in a GOP. The rate shaping scheme is based on the discrete rate-distortion combination algorithm.

### 3.1. Discrete Rate-Distortion Combination Algorithm

Assume there are $F$ frames in a GOP and the total bandwidth constraint for these $F$ frames is $C$. Let $x(i)$ be the state chosen for frame $i$ and let $D_{i,x(i)}$ and $R_{i,x(i)}$ be the resulting distortion and rate at frame $i$ respectively. The goal of the rate shaper is to:

$$\text{minimize} \qquad \sum_{i=1}^{F} D_{i,x(i)} \tag{1}$$

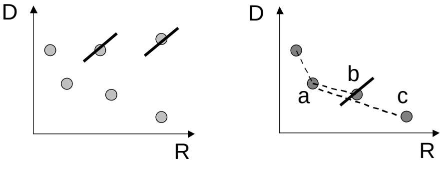
$$\text{subject to} \qquad \sum_{i=1}^{F} R_{i,x(i)} \leq C \tag{2}$$

In principle, this optimization problem can be accomplished using Dynamic Programming [7][8][11]. The trellis diagram is formed with the x-axis being the frame index $i$, y-axis being the cumulative rate at frame $i$, and the cost function of the trellis being the distortion. If there are $S$ states at each frame, the number of nodes at Frame $i = F$ will be $S^F$ (if non of the cumulative rates are the same). This method is too computationally intensive.

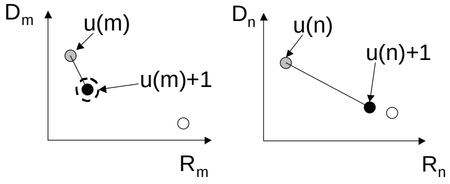
If the number of states, $S$, is large, the R-D map becomes a continuous curve. The Lagrangian Optimization [9][10][11] method can be used to solve this optimization problem. However, Lagrangian Optimization method cannot reach the states that do not reside on the convex hull of the R-D curve.

In this paper, we propose a new discrete rate-distortion combination algorithm as follows:

1. At each frame, eliminate the state in the map if there exists some other state that is smaller in rate and smaller in distortion than the one considered. This corresponds to eliminating states in the upper right corner of the map (Figure 4 (a)).
2. At each frame $i$, eliminate State $b$ if $R_{ia} < R_{ib} < R_{ic}$ and $\left|\frac{D_{ib}-D_{ia}}{R_{ib}-R_{ia}}\right| < \left|\frac{D_{ic}-D_{ib}}{R_{ic}-R_{ib}}\right|$, where State $a$ and State $c$ are two neighboring states of State $b$. This corresponds to eliminating states that are on the upper right side of any line connecting two states. For example, State $b$ is on the upper right side of the line connecting State $a$ and State $c$ (Figure 4 (b)). Thus, State $b$ is eliminated.
3. Label the remaining states in each frame from the state with the lowest rate, State 1, to the state with the highest rate. Let us denote the current decision of state at Frame $i$ as State $u(i)$. Start from $u(i)=1$ for all frames. The rate shaper examines the next state $u(i)+1$ of each frame and finds the one that gives the largest ratio of distortion decrease over rate increase compared to the current state $u(i)$. If Frame $\tau$ is chosen, increase $u(\tau)$ by one. As an example, let us look at two frames, Frame $m$ and Frame $n$ in Figure 4 (c). Current states are represented as gray dots and the next states as black dots. We can see that updating $u(m)$ gives larger ratio increase than updating $u(n)$. Thus, the rate shaper updates $u(m)$.
4. Continue Step 3 until the total rate meets $C$ or will exceed $C$ with any more update of $u(i)$. If $C$ is met, we are done.
5. If the bandwidth constraint is not yet met after Step 4, reconsider the states that were eliminated by Step 2. For each frame, re-label all the states from the state with the lowest rate to the state with the highest rate, and let $u(i)$ denote the current state. Choose the frame with the next state giving the most distortion decrease compared to the current state. If Frame $\tau$ is chosen, increase $u(\tau)$ by one.
6. Continue Step 5 until the total rate meets $C$ or will exceed $C$ with any more update of $u(i)$.



(a) (b)



(c)

**Figure 4. Discrete R-D combination: (a) Step 1; (b) Step 2; (c) Step 3**

### 3.2. Experiment

We compare four methods: (M1) transmits a single non-scalable and non-ECC coded video bitstream; (M2), proposed by Vass and Zhuang [12], switches between State (1, 1) and State (2, 0) depending on the channel error rate; (M3) allocates the same bit budget to each frame and chooses the state that gives the best R-D performance for each frame; (M4) is the proposed method that dynamically allocates the bit budget to each frame in a GOP and chooses the state that gives the best overall performance in a GOP, using the algorithm shown in Section 3.1. Each GOP has $F=5$ frames.

The test video sequence is "stefan.yuv" in QCIF (quarter common intermediate format). The bandwidth and channel error rate vary over time and are simulated as AR(1) processes. The bandwidth ranges from 4k bits/frame to 1024k bits/frame; and the channel error rate ranges from $10^{-0.5}$ to $10^{-6.0}$.

The performance is shown in mean square error (MSE) versus the GOP number as in Figure 5. In the case that all four methods satisfy the bandwidth constraint, the average MSE of all four methods are 10050, 5356, 2091, and 1946 respectively. The proposed M4 has the minimum distortion among all. In addition, let us compare M1 and M2 with M3 and M4. Since M1 and M2 do not have the R-D maps in mind, the network could randomly discard the bitstream sent by these two methods. The resulting MSE performance of M1 and M2 are bad. On the other hand, M3 and M4 are more intelligent in knowing that the bitstream could be non-decodable if the channel error rate is high and thus decide to allocate the bit budget to the channel-coding segments of the video bitstream.

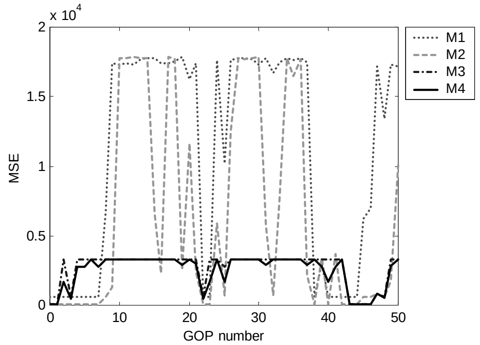


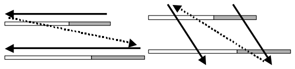
**Figure 5. MSE performance of four rate shaping methods**

## 4. RATE SHAPING FOR MULTICAST

### 4.1. Path-Based Rate Shaping

When there are multiple users, the rate shaper described so far serves the users in a *simulcast* fashion. This means that it tries to serve the request one-by-one with the best achievable video quality, given the channel error rate and the bandwidth of each client. We can modify the rate shaper to be path-based to be used in a *multicast* fashion.

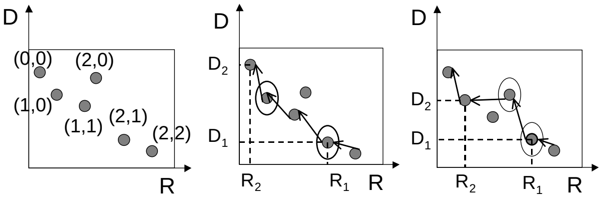
Assume that there are two users, with different channel conditions, requesting for the video bitstream. If we can transmit only one bitstream that serves both users, then the bandwidth is saved. To do so, the rate shaper has to make the partial bitstream sent to the second user decodable. That is, the bitstream represented by the state of the second user should be a subset of the bitstream represented by the state of the first user. In this case, the two states are in the same path as shown in Figure 6, where a state visited earlier in a path corresponds to a bitstream that is a subset of the bitstream corresponding to a state visited later. Consider that the video sequence is scalable coded into two layers, there are two possible paths.



(a) (b)

**Figure 6. Two paths: (a) Path A; (b) Path B**

Each path contains five states, as shown in Figure 7.



(a) (b) (c)

**Figure 7. States chosen by the rate shapers based on different paths: (a) all states; (b) Path A; (c) Path B**

When the channel error rate is low, the rate shaper tends to use Path B to have more bits allocated to the source-coding segment of the bitstream. On the other hand, if the channel error rate is high, more bits are needed to send the channel-coding segment instead of the source-coding segment of the enhancement layer. Thus, Path A is preferred. With this in mind, we want to have a rate shaper that can dynamically choose Path A or B. The rate shaper first decides which two states are the best for these two users given each path. After choosing the states with respect to each path, the rate shaper chooses between two paths to minimize the overall distortion.

### 4.2. Experiment

We show experiment results with four rate shapers: (P0) a simulcasting rate shaper sending two bitstreams to both users; (P1) the path-based rate shaper with Path A only; (P2) the path-based rate shaper with Path B only; (P3) the path-based rate shaper with the path that minimizes the overall distortion.

With all four rate shapers having similar MSE performance, the mean rates (including the bitstream sent to both users) of all four rate shapers are 74.2k bits/frame, 53.7k bits/frame, 55.0k bits/frame, and 49.4k bits/frame. We can see that P0 needs to use almost double the bandwidth. P3 has the best performance among all.

Figure 8 shows that to meet the same bandwidth requirement, P1 and P2 have larger distortions compared to the proposed P3. P1 has an average MSE of 3189; P2 has an average MSE of 3209; and P3 has an average MSE of 3122.

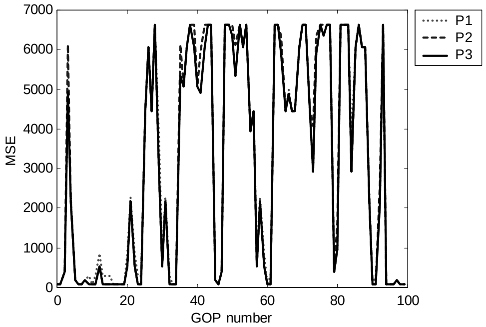


**Figure 8. Distortion performance of P1, P2, and P3**

## 5. CONCLUSION

In this paper, we proposed to use rate shaping to accomplish joint source-channel coding. The rate shaping method specifically considered the channel error in optimizing the rate-distortion performance using the *discrete distortion combination algorithm.* The experiment result showed that the proposed rate shaper outperformed conventional methods. We also proposed a modified path-based rate shaper for the multicast scenario, and showed that it utilized the bandwidth efficiently.

In practice, it is not easy to determine the R-D maps for a given video source under a certain network condition. We are currently working on modeling video statistics to estimate the R-D maps. In addition, when the decoder receives incorrect bitstream, or no information at all, error concealment is often used to improve the video quality. Work is in progress in taking error concealment into consideration for rate shaping.

## 6. REFERENCES

[1] ITU-T Recommendation H.263, January 27, 1998.

[2] B. G. Haskell, A. Puri, and A. N. Netravali, Digital Video: An Introduction to MPEG-2, Chapman & Hall 1997.

[3] Shu Lin, Daniel J. Costello, Jr., Error Control Coding: Fundamentals and Application, Prentice-Hall.

[4] S. Wicker, Error Control Systems for Digital Communication and Storage, Prentice-Hall, 1995.

[5] W. Zeng and B. Liu, "Rate Shaping by Block Dropping for Transmission of MPEG-precoded Video over Channels of Dynamic Bandwidth". ACM International Conference on Multimedia 1996.

[6] S. Jacobs and A. Eleftheriadis, "Streaming Video Using Dynamic Rate Shaping and TCP Congestion Control". Journal of Visual Communication and Image Representation, 9(3), 211-222, 1998.

[7] B. Bellman, Dynamic Programming, Prentice-Hall, 1987.

[8] G. D. Forney, "The Viterbi Algorithm". *Proc. Of the IEEE*, 268-278, March 1973.

[9] H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources". *Operations Research*, 399-417, 1963.

[10] Y. Shoham and A. Gersho, "Efficient Bit Allocation for an Arbitrary Set of Quantizers". *IEEE Trans. ASSP*, 1445-1453, Sep 1988.

[11] A. Ortega and K. Ramchandran, "Rate-Distortion Methods for Image and Video Compression". *IEEE Signal Processing Magazine*, 15(6), 23-50, Nov. 1998.

[12] Jozsef Vass and Xinhua Zhuang, "Adaptive and Integrated Video Communication System Utilizing Novel Compression, Error Control, and Packetization Strategies for Mobile Wireless Environments", Packet Video 2000.