JOINT RATE-DISTORTION CODING OF MULTIPLE VIDEOS

Wei-Cheng Gu and David W. Lin
Department of Electronics Engineering and Center for Telecommunications Research
National Chiao Tung University
Hsinchu, Taiwan 300, R.O.C.
Email: dwlin@cc.nctu.edu.tw

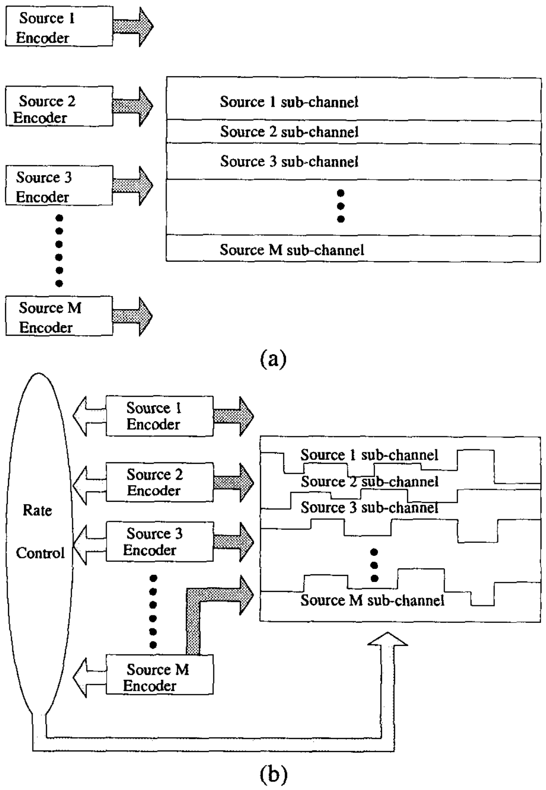
ABSTRACT

Some communications applications involve the simultaneous transmission of multiple videos from one source location. We consider the optimal bandwidth allocation and coding of such co-located multiple videos for transmission over one constant bit rate channel. It turns out that a major issue consists in proper setting of target buffer levels, which arises due to the possibly large variation in the rate-distortion relations of the aggregate video material in a short time period. In fact, this is also an issue in MPEGx coding whose I, P, and B frames possess different rate-distortion characteristics. We outline an approach to optimal solutions under several common types of distortion measures. And we present two simplified techniques and some preliminary simulation results. The simulation results show that such joint bit allocation can significantly enhance the average coding performance over multiple videos.

1. INTRODUCTION

Some communications applications involve the simultaneous transmission of multiple videos from one source location. One example is multiprogram digital video transmission over a shared satellite or cable channel. Another is videoconferencing where one location may send out multiple video streams, generated by multiple cameras and/or taken from stored video sources. In the case of MPEG4, one video may be divided into a number of video object planes (VOPs) for coding and processing. Thus the joint coding and transmission of the VOPs can also be phrased in the framework of simultaneous coding and transmission of multiple videos. Although there have been some recent publications on joint coding of multiple videos [1]–[5], a thorough study from the rate-distortion (R-D) perspective, including buffer management methods, appears lacking.

For simultaneous transmission of multiple videos from one source location, there are two basic approaches to bandwidth allocation: static and dynamic, as illustrated in Fig. 1. Intuitively, one can expect a better average performance with dynamic bandwidth allocation, if the allocated bandwidths match the time variation in the relative complexity of video sources. The question is exactly how we can do the allocation in an optimal way in the R-D sense. Complicating the picture is that the different videos may need be transmitted at different frame rates due to user preferences or other reasons.



(a)

(b)

Fig. 1: Methods of bandwidth allocation. (a) Static. (b) Dynamic.

In what follows, Sec. 2 describes the problem in more detail and outlines an approach to optimal solutions. Sec. 3 presents two related simplified techniques and some associated simulation results. And Sec. 4 contains the conclusion.

Work supported by National Science Council of ROC under grant NSC 87-2213-E-009-150.

Manuscript received November 30, 1998

0099 3063/99 $10.00 © 1999 IEEE

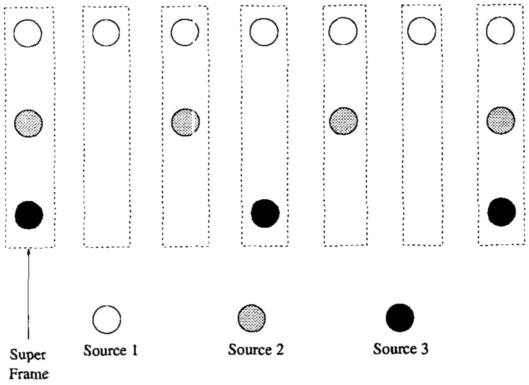


Fig. 2: An example with three video sources.

2. JOINT RATE CONTROL FOR MULTIPLE VIDEOS

2.1. The Issue

Let there be M videos to be simultaneously coded and transmitted. An example with three videos, each having a different frame rate, is shown in Fig. 2. For convenience, the collection of video frames from different video sources that are co-located in time are called a super frame. Aside from sharing one channel, let the videos share one transmitter buffer at the encoder side and one receiver buffer at the decoder side. With this system structure, the problem of simultaneous multiple-video coding becomes similar to that of single-video coding, except that now we have a super frame in the place of a regular video frame. A key difference, however, is that, due to the possibly different frame rates for the different videos, successive super frames may have vastly different complexity as opposed to the relatively constant complexity one would typically expect to encounter (at least over a short time period) in the case of a single video. This poses some challenge to buffer management. And a key in joint R-D coding consists in proper setting of either the target buffer levels at each time instant or the target number of bits for video frames over each time segment.

To appreciate the issue, assume that the channel rate is T per super frame and that the transmitter and the receiver buffers have the same size K. That it is appropriate to let the two buffers have the same size can be shown from a data-flow consideration, such as that presented in [6]. Assume that the coding control mechanism does rate allocation to N consecutive super frames, say, super frames n to $n+N-1$, at a time. (One occasion for such rate allocation is in delayed coding [6], [7].) And let $b(i)$ denote the resulting rate for super frame i. Let $x(i)$ denote the transmitter buffer level after the encoding and transmission of super frame i. Then we have

$$x(i) = x(i-1) + b(i) - T, \qquad (1)$$

for $i = n, \cdots, n+N-1$, where we must also have

$$0 \leq x(i) \leq K \qquad (2)$$

$\forall i$ to avoid buffer under- and overflows.

Consider the example shown in Fig. 2. Assume that all the frames from the different sources have the same complexity. And assume that the coding scheme employs a time-independent coding approach, rather than a time-dependent one such as that in the MPEGx standards where the I, P, and B frames have inherently different R-D characteristics. Let the first super frame be at time n. Then it is expected that $b(n)$ and $b(n+6)$, i.e., bits allocated to the first and the last super frames, would be greater than $b(n+2)$, $b(n+3)$, and $b(n+4)$, which in turn would be greater than $b(n+1)$ and $b(n+5)$. The result is that $x(n)$ and $(n+6)$, i.e., the encoder buffer levels at the end of the first and the last super frames, would be higher than $x(n+2)$, $x(n+3)$, and $x(n+4)$, which in turn would be higher than $x(n+1)$ and $x(n+5)$. Consequently, too high (resp. too low) a setting of the beginning buffer level $x(n-1)$ may cause buffer overflow at times n and $n+6$ (resp. underflow at times $n+1$ and $n+5$) while appropriate settings may avoid both. Indeed, from the above discussion, we see that the issue not only exists in mutliple-video coding, but also in MPEGx coding employing I, P, and B frames.

Optimal R-D coding under multiple rate constraints have attracted the attention of some researchers. Some representative publications are [6] and [7] for minimum sum-distortion criteria and [6] and [8] for minimum maximum-distortion or minimum lexicographic-distortion criteria. However, these studies often assume that the way to determine the initial and the final buffer levels $x(n-1)$ and $x(n+N-1)$ is given. Hence, the choice of these buffer levels remains an issue to be addressed.

2.2. Approach to Solutions

To begin, consider the case where each video source has a fixed frame rate. Then the super frame structure will exhibit periodic variation. For instance, for the case shown in Fig. 2, a super frame cycle spans six time instants, as shown in Fig. 3. In the figure, $C_k(n)$ represents the complexity of frame n of video k, whose meaning need not concern us now. If all the super frame cycles have the same R-D characteristics, then it is plain that, in the steady state, we should set the buffer level at the end of each super frame cycle (or equivalently, the buffer level at the beginning of each super frame cycle) to be the same. By the buffer and channel dynamics (1), this will determine the total number of bits for each super frame cycle. With any choice of the beginning buffer level $x(0)$, a bit allocation can be obtained using the methods in [6]–[8].

Consider minimum sum-distortion criteria first. Optimal bit allocations under such criteria are usually obtained via Lagrange-multiplier minimizations of the form

$$\min_{b(i)} D(i) + \lambda_i b(i) \qquad (3)$$

where $D(i)$ is the distortion measure for frame i and λ_i is the Lagrange multiplier whose value is determined to make the optimal $b(i)$ satisfy the rate constraints (2) [7]. Under minimum sum-distortion criteria, we have the following.

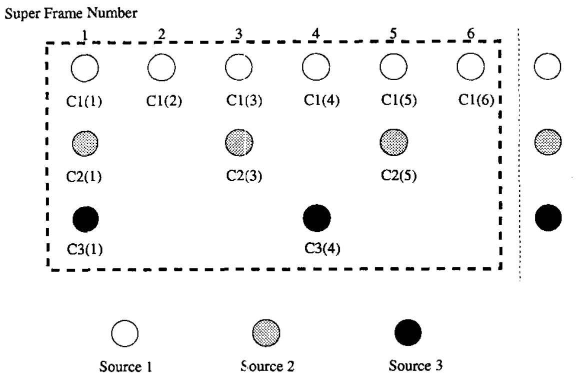


Fig. 3: An example of super frame cycle.

Proposition 1 *Under minimum sum-distortion criteria, the optimal setting of the beginning buffer level for each super frame cycle, in the steady state, is such that the optimal Lagrange multipliers for the first and the last frames in the super frame cycle are the same.*

Justification of the proposition involves a look into the change in the achievable minimum distortion over the super frame cycle between the situation when the two Lagrange multipliers in question are the same and that when they are different, as in Lemma 1 of [7]. The details are omitted.

Concerning minimum maximum-distortion or minimum lexicographic-distortion criteria, we have the following.

Proposition 2 *Under minimum maximum-distortion or minimum lexicographic-distortion criteria, the optimal setting of the beginning buffer level for each super frame cycle, in the steady state, is such that one cannot reduce the maximum or the lexicographic distortion by moving some bits from the first few frames in the super frame cycle to the last few frames or vice versa.*

The optimality condition given in this proposition is less precise than that in the last one. But a more precise statement would require the introduction of some additional notions and theoretical results, which we avoid here.

Together with the algorithms for optimal bit allocation under multiple rate constraints with given buffer-level settings [6]–[8], these propositions provide a way to find the (steady-state) optimal buffer-level settings and the overall optimal bit allocation for simultaneous multiple-video coding.

Now consider the case where successive super frame cycles may have different R-D relations, or the video sources may have time-varying frame rates that the super frames may not have a periodic structure. To handle this case, note first that we usually have little knowledge of the video content for time far into the future. Thus we may let the target buffer level at some enoughly remote future time be half full, to give it maximum capacity to go either way from that point on. As a matter of fact, if this time is reasonably far away, then it should not matter much if we know the video content after that time and hence could obtain the truly optimal buffer setting at that time, for the effects of any suboptimality in the buffer setting at that time would presumably be distributed over the long time period to yield little impact on coding of current video. Prior to that reference future time, we can solve for optimal bit allocations employing the known R-D relations of the video frames that are already captured and the expected R-D relations of the frames that are yet to come in. The expected R-D relations of the future frames need not be extensively modeled, as their function is merely to facilitate the determination of a proper ending buffer level to target at for the coding of the video frames that are already captured.

3. SOME SIMPLIFIED METHODS

The preceding section outlines an approach to jointly optimal rate-distortion coding of multiple videos whose computation is relatively complicated. In this section, we present two simplified methods and employ them in a simulation study. As we shall see, the first method is related to the optimal solution of the last section, while the second can be viewed as a further simplification of the first method.

3.1. Method 1

Consider first the case where each of the multiple sources has a fixed frame rate so that the super frame structure is periodic. In addition, let the super frame cycles have the same R-D relations. The detailed method of bit allocation is more easily illustrated with an example. Thus consider again the example shown in Fig. 3. With T being the channel rate per super frame, the number of bits to allocate over a super frame cycle is $6T$.

This method allocates bits to different video frames according to the relative complexity of these frames. Let C be the total complexity of the video frames in a super frame cycle. Then frame n in video k is given $6TC_k(n)/C$ bits. Therefore, the transmitter buffer levels progress as follows, provided no over- or underflow occurs:

$$x(1) = x(0) + \frac{C_1(1) + C_2(1) + C_3(1)}{C} \cdot 6T - T,$$

$$x(2) = x(1) + \frac{C_1(2)}{C} \cdot 6T - T,$$

$$x(3) = x(2) + \frac{C_1(3) + C_2(3)}{C} \cdot 6T - T,$$

$$x(4) = x(3) + \frac{C_1(4) + C_3(4)}{C} \cdot 6T - T,$$

$$x(5) = x(4) + \frac{C_1(5) + C_2(5)}{C} \cdot 6T - T,$$

$$x(6) = x(5) + \frac{C_1(6)}{C} \cdot 6T - T = x(0).$$

The highest and lowest buffer levels can be found from the above equations. Note that, if the buffer size is at least as big as the difference between the highest and the lowest buffer levels, then we can always find an initial buffer level $x(0)$ to accommodate the buffer level variation without over- and underflows. In the following simulation, we assume that this is the case. When this condition does not hold, the R-D coding methods of [6]–[8] can be invoked.

When successive super frame cycles may have different complexity structures, or when the video sources may have time-varying frame rates that the super frames may not have a periodic structure, we can again (like in the previous section) fix a reference time sufficiently remote in the future and set the target buffer level at that time to half full. Then we employ the complexity figures of the video frames that are already captured and that expected of the future frames up to the reference time to conduct the bit allocation.

This simple method of bit allocation finds similar solutions to that of the previous section if the R-D relations of the video frames possess some special structures. In particular, under minimum maximum-distortion or minimum lexicographic-distortion criteria, this happens if the slopes of the R-D curves for the different video frames maintain the same proportional relationship at all distortion values. Under minimum sum-distortion criteria, it happens if, at where the R-D curves of the video frames have an equal slope, there is a fixed proportional relationship among the rates which is independent of the value of the slope. In the above statements, we have also tacitly assumed that, for all the video frames, the complexity figures enjoy a fixed and equal proportional relationship with the rates of these frames.

3.2. Method 2

This method considers only two successive super frames at a time. And it is also predicated on a complexity measure for the video frames. For notational convenience, if video k skips frame n, then let $C_k(n) = 0$. Then the target transmitter buffer level at time n (after the encoding of super frame n) is decided as follows:

$$x(n) = \frac{\sum_k C_k(n)}{\sum_k C_k(n) + \sum_k C_k(n+1)} \cdot K,$$

where K denotes the buffer size as defined earlier. The idea is to take the next super frame's complexity into consideration (thus performing a one-super-frame delayed coding). If the next super frame has a higher complexity than the current super frame, then we allocate more buffer space for the next time instant; and vice versa.

3.3. Simulation Results

We consider simultaneous coding of three QCIF videos: Salesman, Miss America, and Swing, at frame rates of 10, 5, and 10/3 frames/sec, respectively. Let the total channel rate be 90 kbps and the buffer size be 15 kb. For each video, we employ an H.263 coding framework and do R-D optimal coding [9] with one quantization parameter for each GOB (group-of-blocks) under the two simplified methods of bit allocation described above. For comparison, we also simulate coding under static bandwidth allocation at a 30-kbps transmission rate for each video and with each video associated with a 5-kb transmitter buffer, since $15/3 = 5$. One-frame delayed R-D optimal coding [6], [7], also with one quantization parameter per GOB, is executed.

Table 1: PSNR from Simultaneous Coding of Three Videos

	Static	Method 1	Method 2
Salesman	30.83	33.77	32.91
Miss America	39.25	36.24	37.42
Swing	33.92	31.62	32.63
Average	32.54	33.79	33.69

Table 1 shows the PSNR results. These results were obtained *under the unrealistic assumption* that $C_k(n)$ were equal and hence did not fully exploit the potential of the approach. The suboptimality is perhaps especially acute for Method 1. Nevertheless, the results show that joint R-D coding can signicantly enhance the average coding performance over multiple videos. In addition, this enhancement in average performance is attained by favoring some component video while penalizing others in the multiple video aggregate. This effect can be seen rather clearly from the PSNR and the rate curves for the different videos in Figs. 4 and 5.

4. CONCLUSION

We studied the problem of simultaneous R-D coding of multiple videos where the video sources can have different frame rates. This problem emerges in a number of communications applications. It turned out that a major issue consists in proper setting of target buffer levels, which arises due to the possibly large variation in the rate-distortion relations of the aggregate video material in a short time period. We noted that, in fact, this is also an issue in MPEGx coding whose I, P, and B frames possess different R-D characteristics. We outlined an approach to R-D optimal solutions under several common types of distortion measures. And we presented two simplified techniques. Preliminary simulation results with the simplified techniques show that such joint bit allocation can significantly enhance the average coding performance over multiple videos.

5. REFERENCES

[1] A. Guha and D. J. Reininger, "Multichannel joint rate control of VBR MPEG encoded video for DBS applications," *IEEE Trans. Consumer Electron.*, vol. 40, no. 3, pp. 616–623, Aug. 1994.

[2] S. Lee, S. P. Park, and S.-H. Lee, "A rate control algorithm for co-located variable bit-rate MPEG-2 video encoders," in *SPIE vol. 2727, Visual Commun. Image Processing*, pt. 3, pp. 1290–1301, Mar. 1996.

[3] L. Wang and A. Vincent, "Joint rate control for multiprogram video coding," *IEEE Trans. Consumer Electron.*, vol. 42, no. 3, pp. 300–305, Aug. 1996.

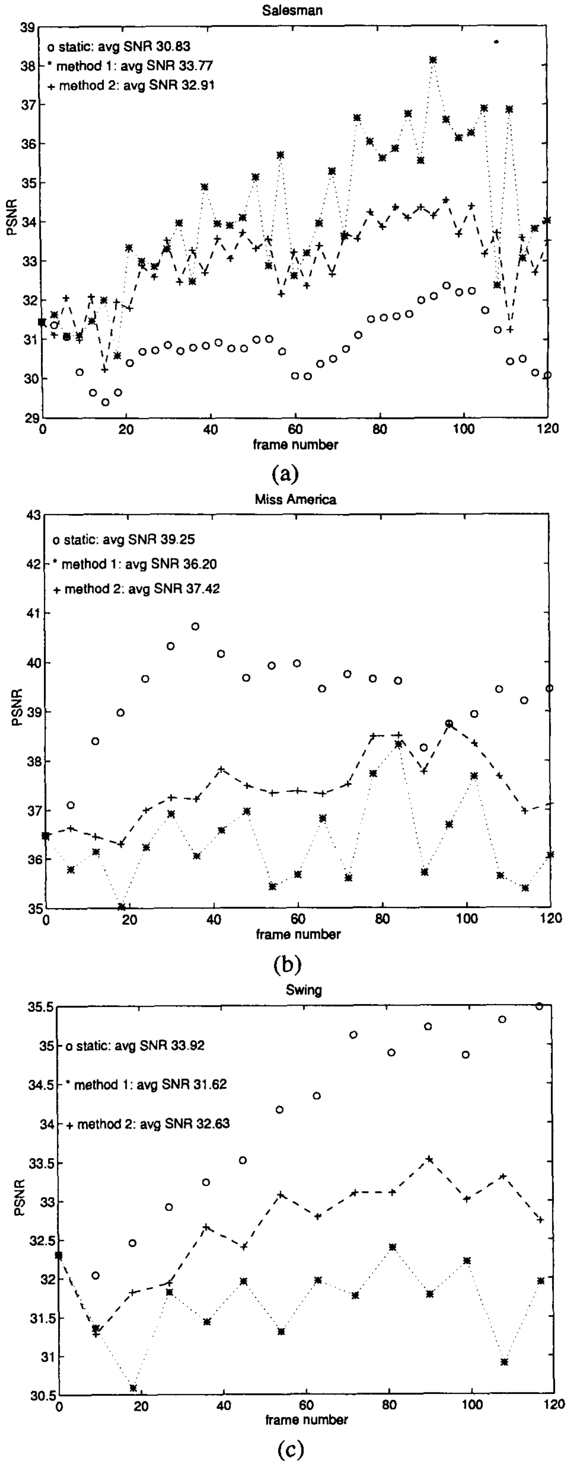


Fig. 4: PSNR curves from multiple-source coding. (a) Salesman. (b) Miss America. (c) Swing.

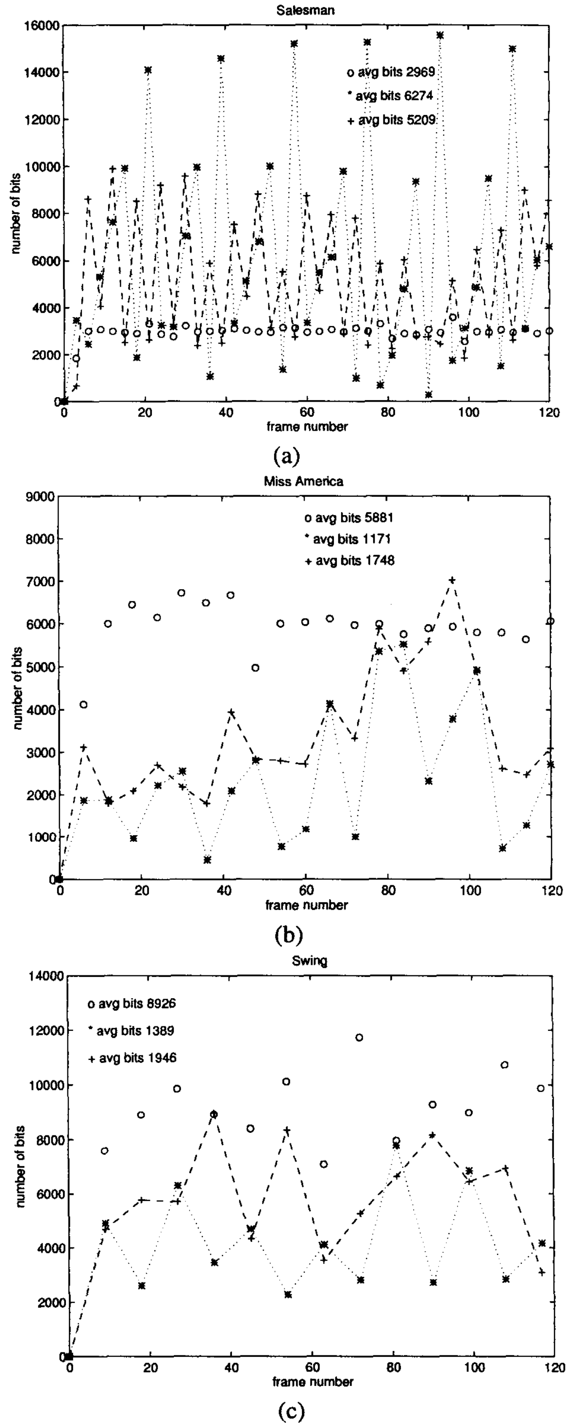


Fig. 5: Rate curves from multiple-source coding. (a) Salesman. (b) Miss America. (c) Swing.

[4] L. Wang and A. Vincent, "Joint coding for multi-program transmission," in *Proc. IEEE Conf. Image Processing*, vol. II, pp. 425–428, 1996.

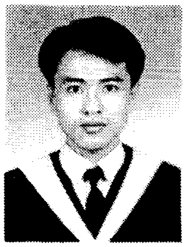
[5] L. Wang and A. Vincent, "Bit allocation for joint coding of multiple video programs," in *SPIE vol. 3024, Visual Commun. Image Processing*, pt. 1, pp. 149–158, Feb. 1997.

[6] D. W. Lin, M.-H. Wang, and J.-J. Chen, "Optimal delayed-coding of video sequences subject to a buffer-size constraint," in *SPIE vol. 2094, Visual Commun. Image Processing*, pt. 1, pp. 223-234, Nov. 1993.

[7] J.-J. Chen and D. W. Lin, "Optimal bit allocation for coding of video signals over ATM networks," *IEEE J. Select. Areas Commun.*, vol. 15, no. 6, pp. 1002–1015, Aug. 1997.

[8] D. T. Hoang, E. L. Linzer, and J. S. Vitter, "Lexicographic bit allocation for MPEG video," *J. Visual Commun. Image Represent.*, vol. 8, no. 4, pp. 384–404, Dec. 1997.

[9] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.



Wei-Cheng Gu received the B.S. degree from National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1996 and the M.S. degree from National Chiao Tung University in Hsinchu, Taiwan, R.O.C., in 1998, both in electrical engineering. His research interests include visual and mobile communication. He is presently in military service.

David W. Lin received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1975, and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1979 and 1981, respectively, all in electrical engineering. He was with Bell Laboratories during 1981–1983, and with Bellcore during 1984–1994. Since 1990, he has been a Professor in the Department of Electronics Engineering and the Center for Telecommunications Research, National Chiao Tung University. He has conducted research in digital adaptive filtering and telephone echo cancellation, digital subscriber line and coaxial network transmission, speech and video coding, and wireless communication. His research interests include various topics in signal processing and communication engineering.