OPTIMAL BIT ALLOCATION FOR MPEG-4 MULTIPLE VIDEO OBJECTS

Zhenzhong Chen and King Ngi Ngan

Department of Electronic Engineering
The Chinese University of Hong Kong, Shatin N. T., Hong Kong
Email: {zchen, knngan}@ee.cuhk.edu.hk

ABSTRACT

In this paper, we present an optimal bit allocation algorithm for multiple video objects (MVO's) in MPEG-4 video coding. We combine the rate-distortion models for texture and shape information to achieve the accurate rate control. To optimize the bit allocation for multiple video objects, we use the dynamic programming (DP). The simulation results demonstrate that the proposed algorithm outperforms the MPEG-4 verification model not only in terms of buffer stability but the objective quality.

1. INTRODUCTION

In MPEG-4, a scene is viewed as a composition of video objects (VO's) with intrinsic properties such as shape, motion, and texture. Instances of video object in a given time are called video object planes (VOP's). A natural video object consists of a sequence of VOP's. In general, VOP's are generated by segmentation of the scene according to some semantic meaning. For such scenes, shape information is thus binary and is referred to as an alpha plane. The binary alpha plane is coded on a macroblock basis by a coder that uses context information, motion compensation and arithmetic encoding.

Rate control plays an important role of digital video-coding standards and applications. In many applications, video sequences must be transmitted over constant bit-rate (CBR) channels. However, the amount of information in compressed video sequences is inherently variable. To solve this problem, a buffer is placed between the video encoder and the transmission channel. The rate control scheme is then responsible for adjusting the coding parameters to ensure that the buffer never underflows or overflows. While a stable buffer level is desired, the rate control scheme aims to obtain the maximum picture quality. In [1], [2] and [3], it was shown how the single video object (SVO) rate control can be extended to multiple video object (MVO) rate control algorithm where SVO is used to describe the entire frame in MPEG-4.

In this paper, we propose an optimal bit allocation algorithm for MPEG-4 multiple video objects based on our combined rate distortion models for texture and shape information. The organization of the paper is as follows. The next section formulates the optimal bit allocation problem. In Section 3, the rate-distortion models for arbitrarily shaped video objects are presented and the optimal bit allocation method is then given based on these novel models. We then give the detailed rate control algorithm in Section 4. We provide our experimental results and the concluding remarks in Sections 5 and 6, respectively.

2. PROBLEM FORMULATION

The optimal multiple video object bit allocation problem can be formulated by:

$$\min\left(\sum_{i=1}^{m} D_i(t)\right) \quad \text{subject to} \begin{cases} \sum_{i=1}^{m} R_i(t) \le R_t \\ \sum_{t=1}^{n} R(t) \le R_{\text{total}} \\ 0 \le B(t) \le B_s, \forall t = 1, \dots, n \end{cases} \tag{1}$$

where $D_i(t)$ is the distortion and $R_i(t)$ denotes the bit rate for the ith VO. m and n are the number of the VO's and the number of frames, respectively. $R(t)$ is the number of bits for current frame, R_{total} is the total number of bits for the source information, and R_t denotes the target bit rate. $B(t)$ and B_s denote the buffer level and buffer size, respectively.

3. OPTIMAL BIT ALLOCATION

3.1. Rate-Distortion Modeling for Texture

We employ a quadratic rate-quantizer (R-Q) relationship for texture information of the arbitrarily shaped VO [4]:

$$R_{\text{texture}} = S\left(\frac{X_1}{Q} + \frac{X_2}{Q^2}\right) \tag{2}$$

where R_{texture} is the bits for current VOP, S is the encoding complexity, Q denotes the quantization parameter and X_i denote the model parameters that are updated by linear regression method from previous coded parameters. We propose to use the sum of absolute difference (SAD) instead of

0-7803-8554-3/04/$20.00

the mean of absolute difference (MAD) since the SAD will increase with the size of the VOP and thereby reflects the larger number of bits required for encoding the larger VOP.

The distortion model is defined as [2]:

$$D_{\text{texture}} = Y_1 \times Q + Y_2 \times Q^2 \tag{3}$$

where D_{texture} is the distortion for current VOP and Y_i denote the model parameters that are updated by linear regression method from previous coded parameters.

3.2. Shape Bit Rate Estimation

Besides considering the amount of bits used for header and motion information, we need to consider the bits used for shape information before we allocate the target bit rate for texture information. We proposed an accurate rate-distortion model for MPEG-4 shape coding in [5]. Generally, the shape bit rate increase in proportion with the number of border blocks. Therefore, we derived a rate model as following:

$$R_{\text{shape},j} = a_j \alpha + b_j \tag{4}$$

where α is the number of the border blocks, a_j and b_j are model parameters, j denotes the conversion ratio (CR), i. e., 0 is full resolution, 1 is CR = 1/2, 2 is CR = 1/4. It should be noted that if $\alpha = 0$, then $R_{\text{shape}} = 0$. Considering the similarity between successive VOP's, we use linear regression to estimate the model parameters.

3.3. Dynamic programming

To solve the problem described in Section 2, we use the dynamic programming (DP) [6, 7] to achieve the maximum picture quality of the decoded video frame which consists of multiple video objects under the bit rate constraint.

We define (referring to Fig. 1):

- Trellis: the trellis is made of all possible paths that line the initial stage to nodes in the final stage.
- Stage: each stage corresponds to a VO to be coded.
- Node: each node is a pair (i, j), where $i \in 1, \cdots, m$ is the stage number, and $j \in Q_{\min}, \cdots, Q_{\max}$ is the quantizer.
- Branch: if quantizer j at stage i has R-D characteristics (R_{ij}, D_{ij}) then node $(i-1, j)$ will be linked by a branch of distortion d_{ij} to node (i, j) provided $\sum_{l=1}^{i} R_l \leq R_{\text{target}}$, where R_{target} is the target bit rate for the texture of all VO's.
- Path: a path is a concatenation of branches. A path from the initial node to a final stage corresponds to a feasible choice of quantizers for all VO's.

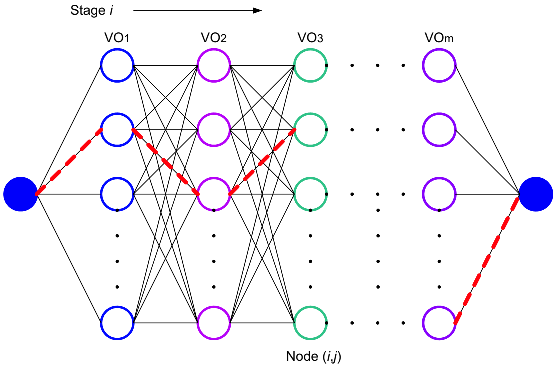

Fig. 1. Trellis diagram.

VA is to find the optimal result for trellis growth:

Step 0 : Choose an initial node and an end node as shown in Fig. 1.

Step 1 : At each stage i add permissible branches to the end nodes of all surviving paths.

Step 2 : Of all the paths arriving at a node in stage $i+1$, the one having the minimum distortion cost under the rate constraint is chosen, and all the paths exceeding the rate constraint are "pruned".

Step 3 : Increment i and go to Step 1.

3.4. Temporal Smoothing

Considering the temporal smoothing, we allow the integer quantizer of a VO varying within 25% of previous one. The number of nodes at each stage is reduced as well.

4. RATE CONTROL ALGORITHM

The proposed rate control algorithm is as follows:

Initialization: We use the same stage of MPEG-4 VM [8] to initialize the encoding parameters and assume a lossless shape coding for both VM and the proposed approach.

Buffer Control: The buffer occupancy filter can be formulated by:

$$R_t = B_d - B(t-1) + R_p \tag{5}$$

where R_t is the target bit rate for the current frame, $B(t-1)$ denotes buffer level before encoding, R_p denotes the average number of bits to be drained from the buffer per frame, and B_d denotes the desired buffer level which is normally set as half of the buffer size B_s for the tradeoff between buffer underflow and overflow. The target number of bits for the texture of VOP's is defined as:

$$R_{\text{target}} = R_t - R_{\text{header}} - R_{\text{shape}} \tag{6}$$

where R_{header} represents the amount of motion and header bits used for the previous VOP's, and R_{shape} is the bit rate for the shape information estimated by (4). Further adjustment is given by:

$$R_{\text{target}} = \max\{R_{\text{target}}, R_p/4\} \tag{7}$$

The lower bound $R_p/4$ is used to ensure the minimum quality for texture information.

Optimal Bit Allocation: Given the rate constraint R_{target} and temporal smoothing constraint, the optimal quantization parameters for each VO can be obtained using the algorithm described in Section 3.3.

Update the Model Parameters: The model parameters for both the texture rate-distortion model and shape rate-distortion model are updated based on the encoding results of the current VOP as well as the past VOP's within the slide window. All the model parameters are solved by the least mean squares (LMS) estimation. In the estimation process, the outlier data points are rejected and the monotonicity is checked.

Frame Skipping Control: After encoding a VOP, the buffer level is updated:

$$B(t) = B(t-1) + R_c - R_p \tag{8}$$

where R_c denotes the actual encoding bit rate. To avoid the buffer overflow, if the current buffer level $B(t)$ is higher than 80% of the buffer size B_s then the incoming frame will be skipped. After the frame is skipped, the buffer level is updated again.

5. SIMULATION RESULTS

The test conditions, such as the format, bit rate, and frame rate, are indicated in Table 1.

The rate control algorithm [8] aims to control the buffer level at half of the buffer size to avoid buffer overflow or underflow, hence the buffer stability is a measurement for the bit rate estimation accuracy. To examine the buffer stability, we present the buffer level for each test video sequence as shown in Figs. 2-5 . It is obvious that the proposed rate control scheme can achieve a more stable buffer than the VM rate control. Our shape model can accurately estimate the bit rate for the current shape instead of copying the information from the previous shape as in the VM. Since the shape bit rate is estimated accurately, the bit allocation for texture information is appropriate. Moreover, using our new complexity measure SAD, the estimated quantization parameter Q is more accurate since MAD cannot reflect the relationship between the texture bit rate and video object size.

To further evaluate the performance of the proposed approach, we consider the decoded picture quality as well. Table 1 provides the PSNR results. Unlike VM which uses fixed bit allocation parameters to allocate bits for MVO's, we use DP to achieve optimal bit allocation based on the R-D characteristics of the individual video object. Using our optimal bit allocation algorithm , the PSNR results were improved significantly.

6. CONCLUSIONS

In this paper, we present an optimal bit allocation algorithm for multiple video objects based on our combined the rate-distortion models for texture and shape information for the arbitrarily shaped video object. As demonstrated by the simulation results, the proposed algorithm outperforms the MPEG-4 verification model both in terms of the buffer stability and the objective quality.

7. ACKNOWLEDGMENT

This work was partly funded by Nanyang Technological University, Singapore.

8. REFERENCES

[1] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 186–199, Feb. 1999.

[2] J. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate control and bit allocation for MPEG-4," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 1243–1258, Dec. 1999.

[3] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 5, pp. 878–894, Sept. 2000.

[4] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate-distortion modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[5] Z. Chen and K. N. Ngan, "Linear rate-distortion models for MPEG-4 shape coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 869–873, June 2004.

[6] R. E. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[7] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. on Image Processing*, vol. 3, no. 1, pp. 26–40, Jan. 1994.

[8] MPEG Video Group, *MPEG-4 video verification model v18.0*, Pisa, Jan. 2001.

Table 1. Simulation results.

Test Sequence	Format	No. of VO	Bit Rate (Kbps)	Frame Rate (fps)	Average PSNR (VM) (dB)	Average PSNR (Proposed) (dB)	PSNR Gain (dB)
Akiyo	QCIF	2	128	30	40.09	41.66	+1.57
Akiyo	QCIF	2	96	15	40.69	41.56	+0.87
Akiyo	CIF	2	256	30	39.32	40.16	+0.84
Akiyo	CIF	2	128	15	38.83	38.90	+0.07
News	QCIF	4	256	30	39.71	40.11	+0.40
News	QCIF	4	128	15	35.98	37.48	+1.50
News	CIF	4	512	30	36.48	37.37	+0.89
News	CIF	4	256	15	35.52	36.13	+0.61

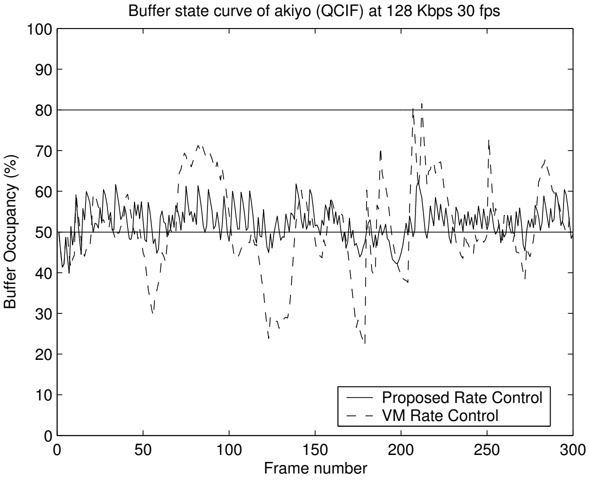

Fig. 2. Comparison of buffer levels of *Akiyo* (QCIF) encoded at 128 kbps and 30 fps.

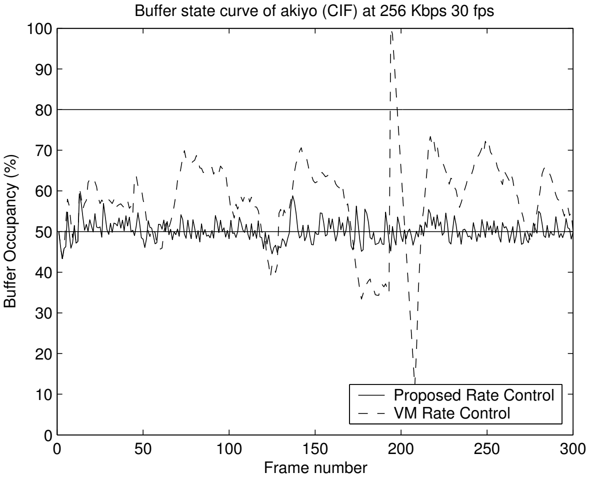

Fig. 3. Comparison of buffer levels of *Akiyo* (CIF) encoded at 256 kbps and 30 fps.

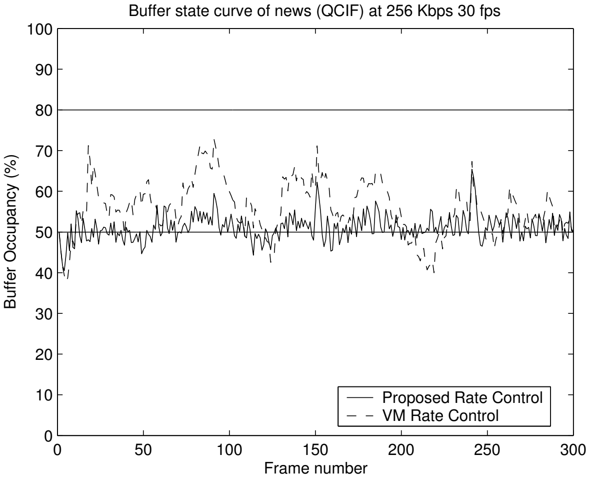

Fig. 4. Comparison of buffer levels of *News* (QCIF) encoded at 256 kbps and 30 fps.

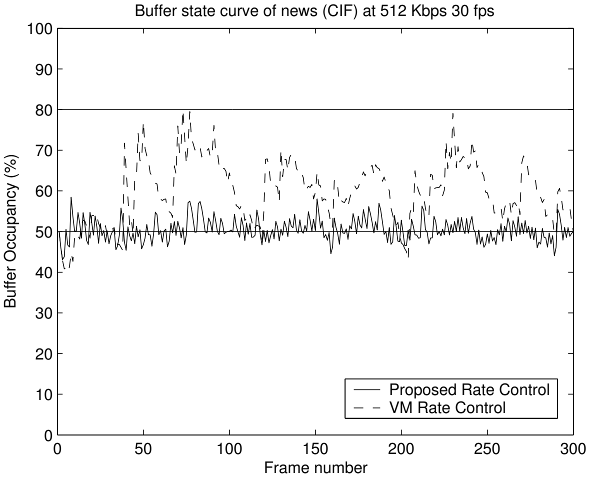

Fig. 5. Comparison of buffer levels of *News* (CIF) encoded at 512 kbps and 30 fps.