Rate Control for MPEG-4 FGS Coded Video Using Piecewise Rate Distortion Model

Li Zhuo¹,²  Kin-Man Lam¹  Lansun Shen²
siplsvr@bjut.edu.cn  enkmlam@polyu.edu.hk  slx@bjut.edu.cn
¹ Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong
² Signal & Information Processing Laboratory, Beijing University of Technology, Beijing, China

ABSTRACT

Three efficient rate allocation schemes for the MPEG-4 fine granular scalability (FGS) enhancement layer with constant quality are proposed in this paper. With the distortion measured using the mean squared error (MSE) or peak signal noise ratio (PSNR) metric, the rate-distortion (R-D) curve of the MPEG-4 FGS enhancement layer exhibits piecewise characteristics. For each piecewise region, a first-order or a second-order model is adopted to estimate the actual R-D curve. Based on these piecewise models, three optimal rate allocation schemes are proposed to allocate an appropriate number of bits to the enhancement layer of each frame. Experimental results show that, compared to the average rate allocation method and the exponential model method, our proposed algorithms can achieve a much constant quality in the reconstructed video.

1. INTRODUCTION

Compare with other scalable video coding schemes, the fine granular scalability (FGS) coding method is outstanding due to its ability of adapting to the changing network conditions more accurately [1][2]. In the MPEG-4 FGS standard, video is encoded into two bit-streams: the base layer and the enhancement layer. The base layer is encoded with a non-scalable MPEG-4 compliant encoder. The enhancement layer is encoded with the bit-plane coding method, thus providing fine-grained scalability. The video data will be over-coded and saved at the video server. While transmission, the entire base layer bit-stream will be transmitted to the clients and the enhancement layer bit-stream will be truncated to meet the network bandwidth requirement. The human visual system (HVS) is very sensitive to the quality variations between adjacent frames. Therefore, how to allocate the rate among enhancement layers in order to achieve the constant reconstructed quality becomes very important.

In the MPEG-4 FGS streaming profile [2], the enhancement layer bits are equally allocated for every frame. This method is very simple and easily implemented, but will cause the quality of the reconstructed images to fluctuate greatly. Wang et al. [3] set up a R-D exponential model for the enhancement layer of the progressive fine granularity scalable video coding in [4]. And then an optimal rate allocation scheme is proposed to minimize the overall distortion. But there is still significant quality fluctuation because the closed form of the model cannot estimate the actual rate distortion characteristics of the MPEG-4 FGS enhancement layer accurately. Zhang et al. [5][6] used linear interpolation to estimate the actual R-D curve of the enhancement layer. Experimental results showed that the method can effectively minimize the quality variation of the reconstructed frames.

It is observed that, no matter whether the distortion is measured with the MSE metric or the PSNR metric, the R-D curve of the MPEG-4 FGS enhancement layer also exhibits a piecewise property, where each piecewise curve corresponds to a different bit plane of the enhancement layer. Therefore, in this paper, a second-order model and a first-order or linear model corresponding to each bit plane are firstly set up to describe the actual R-D characteristics. Based on the piecewise models, three rate allocation methods for
each metric are proposed to optimally allocate the enhancement layer bits among the frames. Experimental results validate that, compared with the average rate allocation and closed form exponential model, the proposed rate allocation schemes can achieve much more constant reconstructed video quality.

2. RATE ALLOCATION FOR THE ENHANCEMENT LAYER

2.1. R-D Curve of the Enhancement layer

To illustrate the actual R-D characteristics for the MPEG-4 FGS enhancement layer, a number of R-D data points are extracted during the encoding process. The actual R-D data points of the enhancement layer for the 2nd frame of the “Mobile” video sequence measured with the MSE and PSNR metrics are illustrated in Figure 1. It can be seen that each of the R-D curves exhibits piecewise characteristics, where each piecewise region corresponds to a different bit plane of the enhancement layer.

2.2. Piecewise Models

In this paper, to estimate the actual R-D curve for each frame, we propose to use a first-order model and a second-order model for each bit plane region to describe the actual piecewise R-D relationship. The first-order or linear model $R^{(1)}$ and the second-order model $R^{(2)}$ used are given below:

$$R^{(1)} = AD + B$$  \hspace{1cm} (1)

$$R^{(2)} = AD^2 + BD + C$$  \hspace{1cm} (2)

where $D$ represents the distortion of the frame under consideration, measured with either the MSE metric or the PSNR metric, and $(A, B)$ or $(A, B, C)$ are the parameters of the model. In this paper, the R-D relationship is represented as $R(D)$ instead of $D(R)$, which will facilitate solving the following optimal problems. The R-D data points used to calculate the parameters of the model can be obtained during the encoding process. As the DCT transform is a kind of unitary transform, after encoding a bit plane, the corresponding distortion can be computed directly.

A comparison of the actual R-D curves with the first-order and the second-order piecewise models is shown in Figure 2. The R-D curves of the 22nd frame of the “Mobile” sequence with the MSE and PSNR metrics are illustrated. It is obvious that the proposed second-order model can describe the actual R-D characteristics of the MPEG-4 FGS enhancement layer more accurately than that of the first-order model.

2.3. Optimal Rate Allocation Method

The purpose of our rate allocation method is to achieve reconstructed image quality as constant as possible through allocating an appropriate number of bits to the enhancement layer of each frame subject to the given target bandwidth constraint [5][6]. The optimal rate allocation problem can be formulated as follows:

$$\text{Min } \delta = \sum_{i=1}^{N-1} |D_i(R_i) - D_{i+1}(R_{i+1})| \quad s.t \quad \sum_{i=1}^{N} R_i = R$$  \hspace{1cm} (3)

where $R_i$ denotes the number of bits allocated to the enhancement layer of the $i$-th frame, $D_i$ the corresponding distortion, $N$ the total number of frames in the rate allocation group, and $R$ the given target bit budget. Ideally, if $\delta = 0$, there is no variation of the reconstructed quality between adjacent frames, that is $D_i(R_i) = D_{i+1}(R_{i+1}) = \ldots = D_{N}(R_N) = D_c$, where $D_c$ is a constant value.

Based on the use of different models, we propose three schemes to solve the optimal rate allocation problem. They are called Method I, II, and III, respectively, in the following.

**Method I:**

This method uses the second-order model for each of the piecewise R-D regions. With (2), (3) can be rewritten as follows:

$$\sum_{i=1}^{N} (A_i D_i^2 + B_i D_i + C_i) = R$$  \hspace{1cm} (4)

Suppose that $D_c$ is of a constant value. Then, we have

$$D_c^2 \sum_{i=1}^{N} A_i + D_c \sum_{i=1}^{N} B_i + C_i = R$$  \hspace{1cm} (5)

where $R$ is given and $A_i, B_i, C_i$ have been computed during encoding. Therefore, the single unknown, $D_c$, can be computed. Then, the number of bits to be allocated, $R_i$, for the enhancement layer of the $i$-th frame can be calculated by using (2).

**Method II:**

This method uses the first-order or linear model for each of the piecewise R-D regions. Similarly, with (1), (3) can be rewritten as follows:

$$\sum_{i=1}^{N} (A_i D_i + B_i) = R$$  \hspace{1cm} (6)

Suppose that $D_c$ is of a constant value. Then, we have

$$D_c = \frac{R - \sum_{i=1}^{N} B_i}{\sum_{i=1}^{N} A_i}$$  \hspace{1cm} (7)
The number of bits to be allocated for the enhancement layer of the $i$-th frame can be calculated by using (1).

**Method III:**

In this method, a linear model is used to represent the R-D curves, as in Method II, but adjacent piecewise curves which have a similar slope will be merged to form a single curve. Suppose that the slope of the R-D piecewise curve for the $p$-th bit plane is denoted as $g_p$. Two adjacent R-D curves will be merged if $|g_p - g_{p+1}| < \tau$, where $\tau$ is a pre-set threshold. $k$ consecutive adjacent curves will be combined if $|g_p - g_{p+i}| < \tau$, for $i = 1, 2, ..., k-1$. This method can further decrease the amount of side information needed and the computations required for computing $D_c$, as compared to Method II. The optimal solution can be achieved by the same way as Method II.

3. EXPERIMENTAL RESULTS

Six standard video sequences with CIF and QCIF formats were used to test our proposed rate allocation algorithms. The encoding frame rate is 10 frames per second. The number of frames used for each video sequence is 100. The encoding format is IPPPPP.... The bit rate of the base layer is 128kbps for CIF and 32kbps for QCIF. The bit rates of the enhancement layer are 480kbps for CIF format and 128kbps for QCIF format. The proposed methods are compared with two other rate allocation methods, the average rate allocation method [2] and the exponential model method [3].

To show the rate allocation performance, we define two terms: mean quality fluctuation, $\epsilon_{\text{mean}}$, and maximum quality fluctuation, $\epsilon_{\text{max}}$, to represent the degree of quality fluctuation between adjacent frames. These two terms are defined as follows:

**Mean Quality Fluctuation:**

$$\epsilon_{\text{mean}} = \frac{1}{N-1} \sum_{i=1}^{N-1} |Q_i - Q_{i+1}|$$  \hspace{1cm} (9)

**Maximum Quality Fluctuation:**

$$\epsilon_{\text{max}} = \max_{0 \leq i \leq N-1} |Q_i - Q_{i+1}|$$  \hspace{1cm} (10)

where $Q_i$ represents the measured quality of frame $i$. Table 1 tabulates the maximum and mean fluctuations in terms of the MSE and PSNR metrics of the different rate allocation algorithms with different video sequences and formats. Figure 3 show the changes in the Foreman sequences based on the different allocation schemes.

It can be seen that our proposed three rate allocation schemes can achieve lower fluctuation levels than that of the average allocation method and the exponential model method. Method I can achieve the most constant quality, while the performance of Method II is better than that of Method III. This is because the second-order model can represent the actual R-D curves more accurately than the first-order model can, but at the expense of a higher computational complexity.

4. CONCLUSIONS

Three efficient rate allocation schemes for constant quality are proposed in this paper. We propose to use a piecewise second-order and a piecewise first-order model to describe the R-D characteristics. We also propose to merge adjacent linear R-D curves if their slopes are similar. Experimental results show that these piecewise models can describe the actual R-D curves of the MPEG-4 FGS enhancement layer more accurately than that of the exponential model. Compared to two other methods, our proposed schemes can achieve a reconstructed quality with much less fluctuation, irrespective of the distortion metric being used.

5. REFERENCES


Figure 1. Actual R-D data points of the enhancement layer with (a) the MSE, and (b) the PSNR metrics.

Figure 2. Comparison of the actual R-D curve with the second-order and first-order piecewise model.

Figure 3. Reconstructed quality of the Foreman Sequence based on the different rate control algorithms.

Table 1. Maximum and average fluctuations for different rate control algorithms and video sequences

<table>
<thead>
<tr>
<th>Methods</th>
<th>CIF</th>
<th>QCIF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Akiyo</td>
<td>Mobile</td>
</tr>
<tr>
<td>Average Rate Allocation</td>
<td>0.232</td>
<td>0.066</td>
</tr>
<tr>
<td>Exponential Model</td>
<td>0.191</td>
<td>0.058</td>
</tr>
<tr>
<td>MSE Metric</td>
<td>Ⅰ</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>Ⅱ</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Ⅲ</td>
<td>0.198</td>
</tr>
<tr>
<td>PSNR Metric</td>
<td>Ⅰ</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>Ⅱ</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>Ⅲ</td>
<td>0.197</td>
</tr>
</tbody>
</table>