http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 8 (2): 188-194, 2009
ISSN 1812-5638

Analysis and Application on Rate-Distortion Model Oriented Scalable Video Sequences

Mande Xie, Guiyi Wei and Yun Ling
College of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang, 310018, People's Republic of China

Abstract: According to the characteristic of the existing main R-D models, these R-D models are firstly classified into three categories: analytic, empirical and semi-analytic. And then their quantitative performances are evaluated. After the performances of the existing main R-D models are in-depth analyzed and compared, some general rules are proposed to select the best model for a target system. On this basis, an algorithm for establishing the piece linear R-D model and a method of property analysis and modification are proposed according to the feature of Peer to Peer (P2P) video streaming media system. The experience results show the piece linear R-D model is accuracy and the method of property analysis and modification is validated.

Key words: Fine granular scalable code, peer to peer, rate-distortion model, video streaming media

INTRODUCTION

With the increasing popularity of video service over internet, scalable-coded streams are widely adapted to optimize the perceived quality. Because the layered-coded streams and the multiple description coded streams have the low fault tolerance and limited rate scalability, the Fine Granularity Scalability (FGS) coding technique becomes the most popular code technique employed by internet video streams and has been proposed as part of the MPEG-4 standard. The most difference between FGS-encoded sequence and other scalable-coded sequence is the enhancement layer employs bit-plane coding technique. This technique makes an enhancement layer of each frame truncated in any position, which makes the quality of reconstructed video is proportional to the number of bits received. Because it makes the important parts of each coefficient are priorly decoded, the important information can also be gotten when the bit streams are truncated in any position.

Rate-distortion models are functions that describe the relationship between the bitrate and expected level of distortion in the reconstructed video stream and are a key method by which the video quality is optimized in the heterogeneous network. So, the research on R-D models has attracted more attention from researcher in a university and an enterprise. At present, several R-D models have been proposed for, the increasingly becoming popular, fine-grained scalable video sequences. However, the proposers of these models only compare the performance cared by they, the models' relative performance has not been thoroughly analyzed. Moreover, the time complexity of each model is not known. The lack of quantitative performance analysis makes it difficult to select the model that best-suits a target system.

ANALYSIS AND COMPARE OF THE EXISTING MAIN R-D MODELS

Several R-D models have been proposed for, the increasingly becoming popular, fine-grained scalable video sequences. We classify these R-D models into three categories: analytic, empirical and semi-analytic.

The analytic R-D models: According to the FGS coding enhancement, three analytic R-D models such as square root model, logarithm model and Generalized Gaussian Function (GGF) model are successively proposed. Base on the lecture (Hang and Chen, 1997). Dai *et al*. (2006) proposed a square root model and they assumes DCT coefficients follow a two-component Laplacian Mixture (LM) distribution, which is a linear combination of several Laplacian density functions and can be formally described by:

$$f_{LM}(x) = p_1\frac{\lambda_1}{2}e^{-\lambda_1|x|} + p_2\frac{\lambda_2}{2}e^{-\lambda_2|x|}$$

where p_1, p_2 and λ_1, λ_2 are parameters that need to be estimated.

Dai and Dmitri (2003) and Park and Lee (2006) proposed an effective parameter estimation method. Dai (2004) propose the logarithm model which employs a

Corresponding Author: Mande Xie, College of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang, 310018, People's Republic of China

LM source model same as the square root model. This model let C_0 represent the coefficients in $(-\Delta, \Delta)$ and let $C_{\bar{0}}$ represent the other coefficients. Note that the bitplane coding quantizes all coefficients in C_0 to the quantization index with zero reconstruction level and that the coefficients in C_0 are responsible for the majority of distortion at low and medium bitrates. Therefore, the logarithm model strives to accurately compute the distortion caused by C_0 and to roughly approximate the distortion related to $C_{\bar{0}}$. The GGF model is proposed by Sun *et al*. (2006) and it uses 64 distributions which belong to the same family but with different parameters as the source model. Each distribution models the set of coefficients of the same frequency component and leads to a R-Q and a D-Q function for that frequency component. Combining 64 R-Q functions gives a frame-level R-Q functions and aggregating 64 D-Q functions results in a frame-level D-Q function.

Though all of these three R-D models employ the analytic method, they are considerable differences in the performance and time complexity.

Accuracy: In general, the GGF model is the least accurate in all cases and it has the very high deviations at low bitrates and even it may produce the invalid results. Except for the medium and high complexity sequences coded at low base layer bitrate, where the square root model is the most accurate, the logarithm model is the most accurate. But the square root model overestimates the video quality at the high bitrate. Generally, if the base layer bitrate is low or the complexity of video sequences is low and then there are few bit planes in the enhancement layer. Because, at this case, the logarithm model properly describes the linear relationship between the bitrate and the percentage of non-zero quantized coefficients, it is the most accurate. On the other hand, the square root model has the inaccuracy estimation in the least significant bitplanes and if there are few bit planes in the enhancement layer, the inaccuracy estimation in the lease significant bitplanes dominates the overall performance and significantly affects its accuracy. In contrast, for high complexity sequences coded at low base layer rates, this linearity starts to break and significantly affects the accuracy of the logarithm model. However, with the increase of bitplanes, the effect of the inaccuracy estimation in the lease significant bitplanes reduces. Consequently, the square root model is most accurate in this case.

Time complexity: The GGF model is most efficient and square root model is slightly faster than the logarithm model, because the latter requires a time-intensive estimation of its R-Q model parameter.

Range of applicability: In fact, because the GGF model is inaccurate, it is almost not employed. The logarithm model is valid at all bitrates and it can be used to all bitrates of all sequences. The square root overestimates the video quality when the quantization step is less, namely, the bitrate is high. Hence, it is not used to low complexity sequences with high base layer bitrate.

The empirical R-D models: The empirical R-D models are classified into a pure empirical R-D model and a piece linear R-D model. The pure empirical model is derived by a high-intensity computation. Firstly, it chooses lots of sampling bitrates to encode the video sequence and decodes the video sequences at each of them. Then, for each sample bitrate, it computes the distortion as the difference between the original and reconstructed video sequences and consequently, the R-D table is generated. Interpolation is used to extend the discrete R-D table into a continuous function. The pure empirical R-D model requires tremendous amount of computational power and storage space. Nevertheless, it produces accurate R-D curve which is used as the baseline for comparing various R-D models. Traditional empirical R-D models often employ exponential interpolation between sampling points. In fact, it has been found that the exponential interpolation does not accurately track the actual R-D curves of FGS-encoded sequences (Zhang *et al*., 2003). Zhang *et al*. (2003) further revealed that when sampling bitrates are located in the same bitplane, using linear interpolation produces smaller deviation than using exponential interpolation. However, if the sampling bitrates are located in the different bitplanes, and then applying exponential interpolation is more accurate. Based on the revelation, the piece linear R-D model chooses all samplings at bitplane boundaries and computes the distortion at these points. At last, the accurate piece linear R-D curve is produced by connecting all these points.

The semi-analytic R-D models: The semi-analytic R-D models are based on the analytic R-D model and the empirical R-D model. Similarly, they also are classified into three categories: the semi-analytic square root model, the semi-analytic logarithm model and the semi-analytic GGF model. The semi-analytic square root model is based on the analytic square root model. However, instead of using the complicated D-Q function, it uses a heuristic functions:

$$\mathrm{PSNR}(z) = d_1 z^2 + d_2 z + d_3 \quad (1)$$

where, PSNR represents the quality instead of the distortion and d_1, d_2 and d_3 are the parameters required to be estimated.

This function captures the fact that the quality is a monotonically increasing function of the number of transmitted bitplanes. Furthermore, Dai *et al.* (2006) pointed out this function does not change its convexity more than once. Hence, a quadratic function is a good approximate for it. Based on the square root model and simplified D-Q function, the semi-analytic square root model can be formally described as:

$$\mathrm{PSNR}(R) = c_1 R + c_2 \sqrt{R} + c_3$$

where, c_1 and c_2 are required to be estimated by empirical R-D samples and $c_3 = 10\log_{10}(255^2/\sigma^2)$ indicates a source with variance σ^2.

The semi-analytic logarithm model is based on the logarithm model. Instead of using parameters of the Laplacian mixture density function, it employs curve fitting to compute α_1, α_2, α_3 and α_4. The semi-analytic GGF model is based on the GGF model. Instead of using numerically computing the integration, it uses a heuristic function:

$$\mathrm{PSNR}(R) = g_1 R + g_2 - \frac{g_2 - D_b}{1 + g_3 R}$$

where, PSNR is the quality, D_b is the base layer only quality and g_1, g_2 and g_3 are parameters required to be estimated.

In general, the empirical models and semi-analytic models are more accurate against the analytic models. The semi-square root model is most accurate among all semi-analytic models and the semi-logarithm is the least accurate. The semi-analytic R-D models compute the model parameters by curve fitting. This method is very time-consuming and furthermore, some inaccurate parameters are occasionally gotten. The piece linear R-D model is the most accurate and the most efficient in all empirical and semi-analytic R-D models.

THE R-D MODEL BEST- SUI- TED P2P STREAM MEDIA SYSTEM

The experiment building method of R-D model curve: The P2P streaming media system has some real time requirements, where peers' bandwidth is heterogeneous. The piece linear R-D model is very accurate and efficient and can be used to all base layer bitrates. Hence, it best-suits this system. In our P2P streaming media system, the piece linear R-D model is built by the experiment means. The chrominance components are usually ignored, because the human visual system is less sensitive to them compared to the luminance components and they only occupy about 10% of the bitrate (Hsu and Hefeeda, 2007). Hence, in this experiment, the Mean Square Error (MSE) between the luminance value of pixels in the original and reconstructed frames and the related quality measure is the Peak Signal-to-Noise Ratio (PSNR), which can be computed by the Eq. 2:

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}\,\mathrm{dB} \quad (2)$$

The key idea of the piece linear R-D model is that within each bitplane the R-D curve can be approximated by a line segment and the line segment of different bit-planes has different slope in the piece linear R-D model. Hence, in this experiment, we only encode and decode the sequence at bit rates that correspond to bitplane boundaries and compute the distortion at these points. At last, we connect these points by segments and consequently, the whole curve of this model is generated. Obviously, the curve is piece-wise continuous and smooth and all discontinuous points lie at the bit-plane boundaries. Figure 1 shows the pseudocode of the algorithm BuildTheR_Dmodel, which is the building method of R-D curve of the whole sequence. In Fig. 1, the line (1) compresses a raw video file into a base layer and an enhancement layer bitstream by the MPEG-4 reference software encoder.exe. At the same time, it trims the enhancement layer according to a given target bit rate and each truncated bitstream is then saved in a new file. Because the number of frames and the resolution of each video file are different, we customize a parameter file for each video file. The line (2) decodes all incomplete enhancement layers and produces all corresponding raw video files with proportional quality improvement by the MPEG-4 reference software decoder.exe. The line (3-9) computes the MSE of each frame between the original and reconstructed frames and the line (10) computes PSNR. The line (11-18) computes the distortions at different bitrates and builds the discrete R-D table of frame-level, where the line (14) models the bitrate difference by truncating the bitstream in different bitplane and the line (16) computes the slope of the corresponding bit-plane.

```
BuildTheR_DModel(char* origYUVFile)
{
(1) Call encoder .exe to encode the origYUVFile file ;
(2)  Call decoder .exe to decode all kinds of files truncating the enhancement file at the different bitplane ;
(3)  for (iFrameNo = 0; iFrameNo < iMaxFnum; iFrameNo++)
(4)              read the data of the frame iFrameNo from the origYUVFile file ;
(5)  for (i =0; i < iMaxbp; I++)
       {
(6)              Open the file truncating the enchancement file at the bitplane i ;
(7)              for (iFrameNo = 0; iFrameNo < iMaxFnum; frameNo++)
(8)                  read the data of the frame iFrameNo from the truncating file ;
(9)               Compute the mse frame by frame and generate the array arMse [MAXBP][MAXFNUM];
        }
(10)  arPSNR = 10 * log10((255^2)/arMSE );
(11)  for (iFrameNo = 0; iFrameNo < iMaxFnum; frameNo++)
        {
(12)           arLinear _d[i] = arMse(1,iFrameNo);
(13)          for (i=0; i < iMaxbp; I++)
              {
(14)           arLinear_rdiff(iFrameNo, i) = arFrame_size (i+1,iFrameNo) - arFrame_size (i , iFrameNo);
(15)           if (arLinear_rdiff(iFrameNo, i) != 0)
(16)                  arLinear _Slope(iFrameNo, i ) = (arMse(i+1, iFrameNo) - arMse (i,iFrameNo)) /arLinear _rdiff (iFrameNo, i );
(17)            else
(18)                    arLinear _SlopeiFrameNo, i) = arMse(i +1, iFrameNo) - arMse(i ,iFrameNo);
            }
        }
}
```

Fig. 1: The pseudocode of the building algorithm for the piece linear R-D model

Table 1: The linear R-D model parameters for the first ten frames

	Mobile sequence						Coastguard sequence					
Frame	bp1	bp2	bp3	bp4	bp5	bp6	bp1	bp2	bp3	bp4	bp5	bp6
1	-0.4369	-3.2151	-1.4935	-0.39623	-0.070545	-0.019124	-0.71482	-0.34571	-0.094614	-0.017955	0	0
2	-6.8319	-4.4523	-1.5475	-0.39226	-0.070651	-0.020451	-0.15233	-3.0570	-1.31760	-0.456750	-0.091066	-0.019973
3	-3.4044	-3.9157	-1.5179	-0.39382	-0.071601	-0.020263	-2.69150	-1.4197	-0.45747	-0.091902	-0.019372	NULL
4	-2.9818	-3.8853	-1.5112	-0.38765	-0.072714	-0.020128	-2.56910	-1.4272	-0.48045	-0.090273	-0.019765	NULL
5	-5.3699	-4.2719	-1.5510	-0.38894	-0.072068	-0.020398	-0.75660	-2.7764	-1.55780	-0.480670	-0.089939	-0.019722
6	-2.8999	-4.0344	-1.5253	-0.38538	-0.072727	-0.020786	-0.39833	-2.8112	-1.50210	-0.490550	-0.089010	-0.019895
7	-0.6186	-3.3730	-1.4428	-0.39144	-0.073572	-0.019907	-2.47380	-1.2581	-0.47202	-0.092396	-0.019205	NULL
8	-4.1599	-3.9912	-1.5291	-0.39229	-0.072090	-0.020724	-2.91850	-1.4873	-0.48511	-0.090543	-0.019568	NULL
9	-4.3098	-3.9145	-1.5034	-0.39435	-0.072989	-0.020737	-2.33490	-1.5064	-0.47633	-0.090360	-0.020026	NULL
10	-0.1687	-3.1457	-1.3675	-0.38979	-0.075373	-0.019849	-1.31620	-1.2091	-0.46435	-0.092882	-0.019326	NULL

R-D models are the basis of lots of optimizing algorithms employed P2P video streaming media system. Hence, the method of building the R-D model must be efficient. Here, we extract only once the R-D model from the video sequence and store it in a meta file which any new arrival peer in P2P streaming media system can get from a uploading peer. In the meta file, we need to store only the size of each bitplane and the slope of the line segment in that bitplane. Consequently, the meta file adds a negligible storage overhead, up to 64 bytes per frame.

The properties analysis of R-D models: By examining a large set of R-D parameters for many frames, one can find that the curves of the piece linear model have the following property: the slopes of all line segments form a monotonically increasing series, that is $g_1<g_2<...g_z<0$, where g_z is the slope of line segment in bitplane z. Table 1 shows the slopes of all bitplanes belonged to the fore ten frames which come from mobile sequence and Coastguard sequence and it alsovalidates the property. This is intuitive because, more significant bitplanes carry higher significant bits and thus the per-bit reduction in distortion is higher. This means that the most significant bitplanes will have steeper and more negative slopes.

However, we have found some abnormal points which don't follow the property in experiment datum. For example, there are some abnormal points in the frame 1 and 10 of mobile sequence and the frame 2 of coastguard in Table 1. After analyzing the datum in Table 1 and other experiment datum, we have found these abnormal points

```
MergeTheSlope(const double* arLinear_Slope, const double * arLinear _rdiff ,
              double * arLinear_a_Slope, double * arLinear_a_rdiff)
Input: arLinear_Slope , arLinear_rdiff ;
Output: arLinear _a_Slope, arLinear_a_rdiff;
{
(1)  arLinear_a_Slope =arLinear_Slope;
(2)  arLinear_a_rdiff = arLinear_rdiff ;
(3)  for (i = 1; i < iMaxfnum; I++)
     {
(4)        for (j = 1; j < iMax_bp-1; j ++)
           {
(5)             temp_sum_rdiff = arLinear _a_rdiff(i, j );
(6)             temp _sum_slope = arLinear_a_Slope(i, j);
(7)             k = j +1;
(8)             while (arLinear_a_Slope(i, k) != 0 && arLinear_a_Slope(i, k) <= temp_sum_slope)
                {
(9)                     temp_sum_slope = (temp_sum_slope * temp_sum_rdiff + arLinear _a_Slope(i,k) *
                                        arLinear_a_rdiff(i, k)) / ( temp_sum_rdiff + arLinear_a_rdiff (i,k));
(10)                    temp_sum_rdiff = temp_sum_rdiff + arLinear _a_rdiff (i , k);
(11)                    for (l = k+1; l < iMax_bp; l++)
                        {
(12)                         arLinear_a_Slope(i, l-1) = arLinear_a_Slope(i, l);
(13)                         arLinear_a_rdiff(i, l-1) = arLinear_a_rdiff(i, l);
                        }
(14)                  arLinear_a_Slope(i, iMax_bp) = 0;
(15)                  arLinear_a_rdiff (i, iMax_bp) = 0;
                }
(16)           arLinear_a_Slope(i, j) = temp_sum_slope;
(17)           arLinear_a_rdiff (i, j) = temp_sum_rdiff;
            }
       }
}
```

Fig. 2: The pseudocode of the amending algorithm

almost occurred between the bitplane 1 and 2, where the size of the first bitplane is very small. Obviously, at that small size, the overhead of the bitplane header becomes significant, which negatively impacts the effective per-bit reduction in distortion. Hence, the corresponding slope is large.

The monotonically increasing property of the piece linear R-D model is often employed by lots of optimizing algorithms used in P2P streaming media system. Hence, an amending algorithm is proposed to make the slopes of all segments follow the monotonically increasing property. Figure 2 shows the pseudocode of amending algorithm. The input of this algorithm comes from the algorithm BuildTheR_DModel and the key codes are the while loop containing the line (8-15). The line (8) validates whether a point is abnormal. If a point is abnormal, the point is merged with the next point by the line (9-10). The line (11-15) left shifts all remainder points to cover the abnormal point and sets the last item 0. Because the datum is shifted, need not to be increased. Obviously, all original successive abnormal points can be removed and the strictly monotonically increasing slope series are produced by our amending algorithm.

EXPERIMENT RESULTS

All algorithms have been implemented and tested. Table 1 shows the linear R-D model parameters for the first ten frames of mobile sequence and coastguard sequence. For the sake of clarity, slopes are amplified 1000 times and the symbol NULL indicates the non-existing bitplanes. Obviously, almost all slopes in Table 1 follow the monotonically increasing property except for the individual slopes. Figure 3 shows the piece linear R-D curve of the frame 112 of flower sequence and that of the frame 120 of coastguard sequence. In Fig. 3, the piece linear R-D curve is compared with the pure R-D model curve often employed as the baseline to validate its accuracy. The results show the piece linear R-D model is very accurate.

Figure 4 shows some amended slope curves by present amending algorithm and the first two sub-figures indicate the slope curves of two random frames of Flower sequence (Fig. 4a, b) and the last two sub-figures indicate the slope curves of two random frames of coastguard

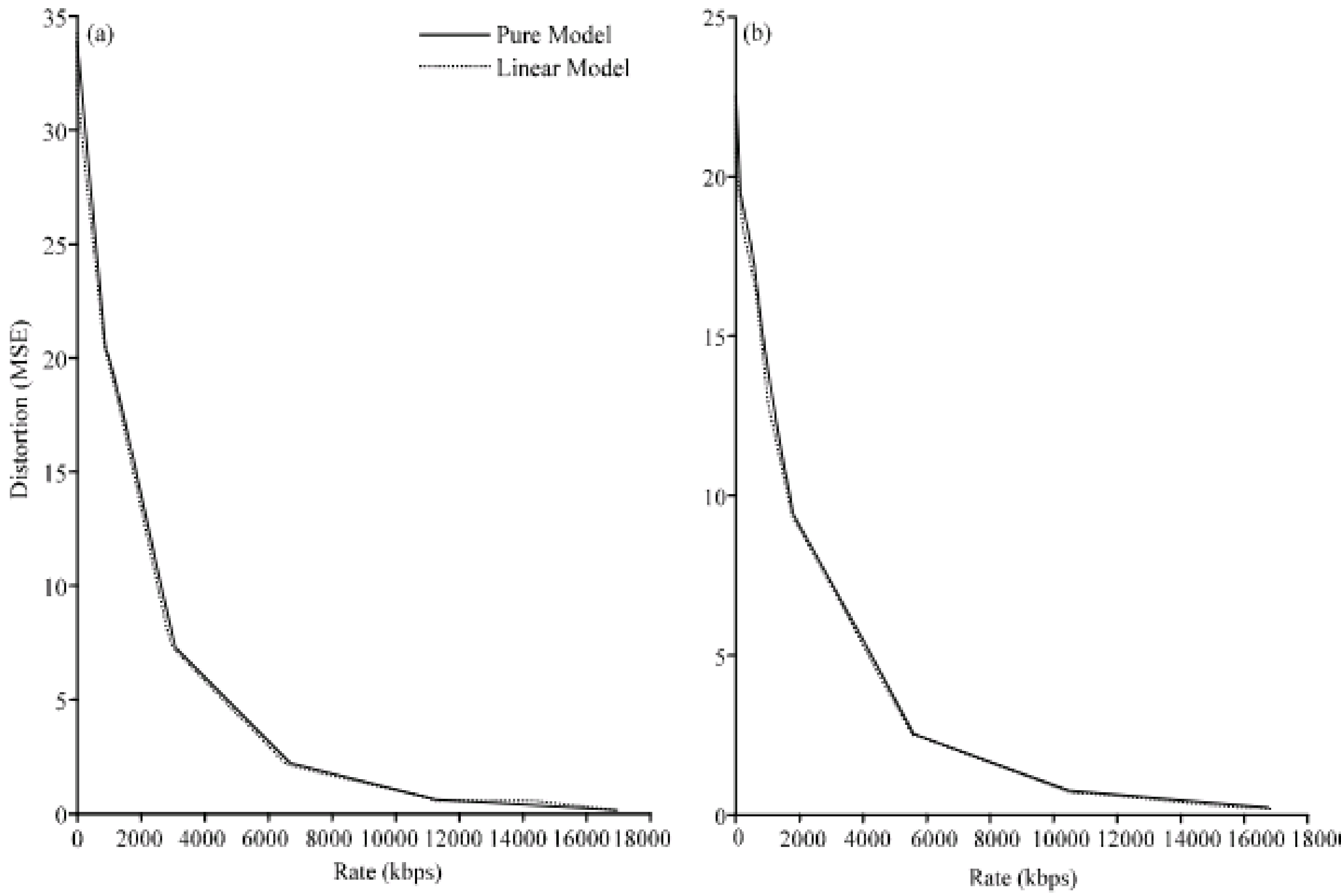

Fig. 3: The curves of the piece linear R-D model, (a) The frame 112 of flower sequence and (b) The frame 120 of coastguard sequence

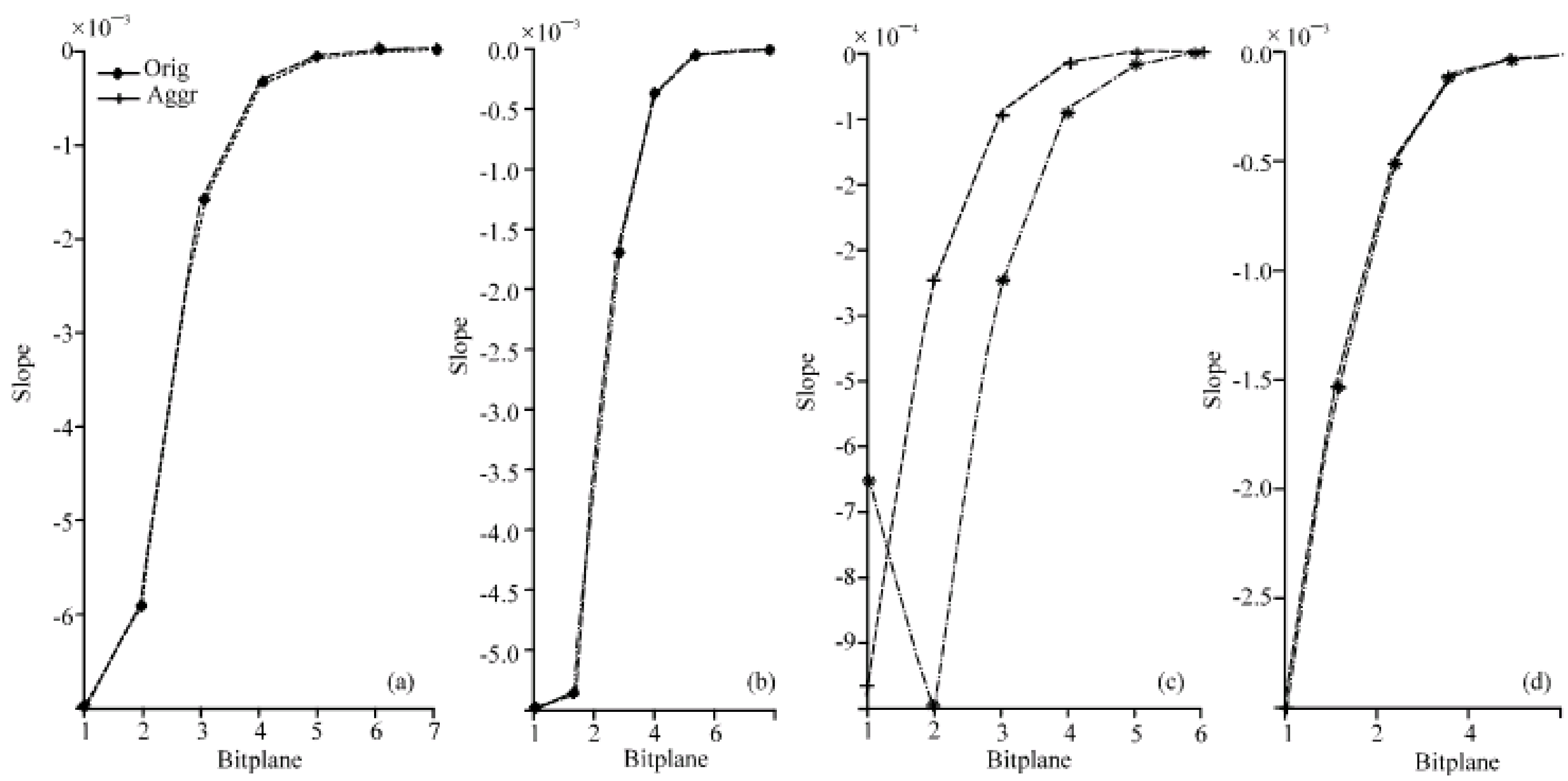

Fig. 4: The amended curve examples of Flower sequence (a, b) Frame 117, 105 and Coastguard sequence (c, d) Frame 211, 131

sequence (Fig. 4c, d), where horizontal axis indicates bitplane No. and vertical axis indicates the slope. As shown in the figure, the original curve with abnormal points becomes a strictly monotonically increasing curve after our amending algorithm ran, such as the curve of the frame 211 from coastguard sequence. Figure 5a, b compare the mean MSE between pre-and-post algorithm running. As shown in the Fig. 5a, b, the maximum distortion percent between pre-and post algorithm running does not exceed 2.5%. The results show our amending algorithm does not affect the accuracy of the piece linear R-D model and is valid.

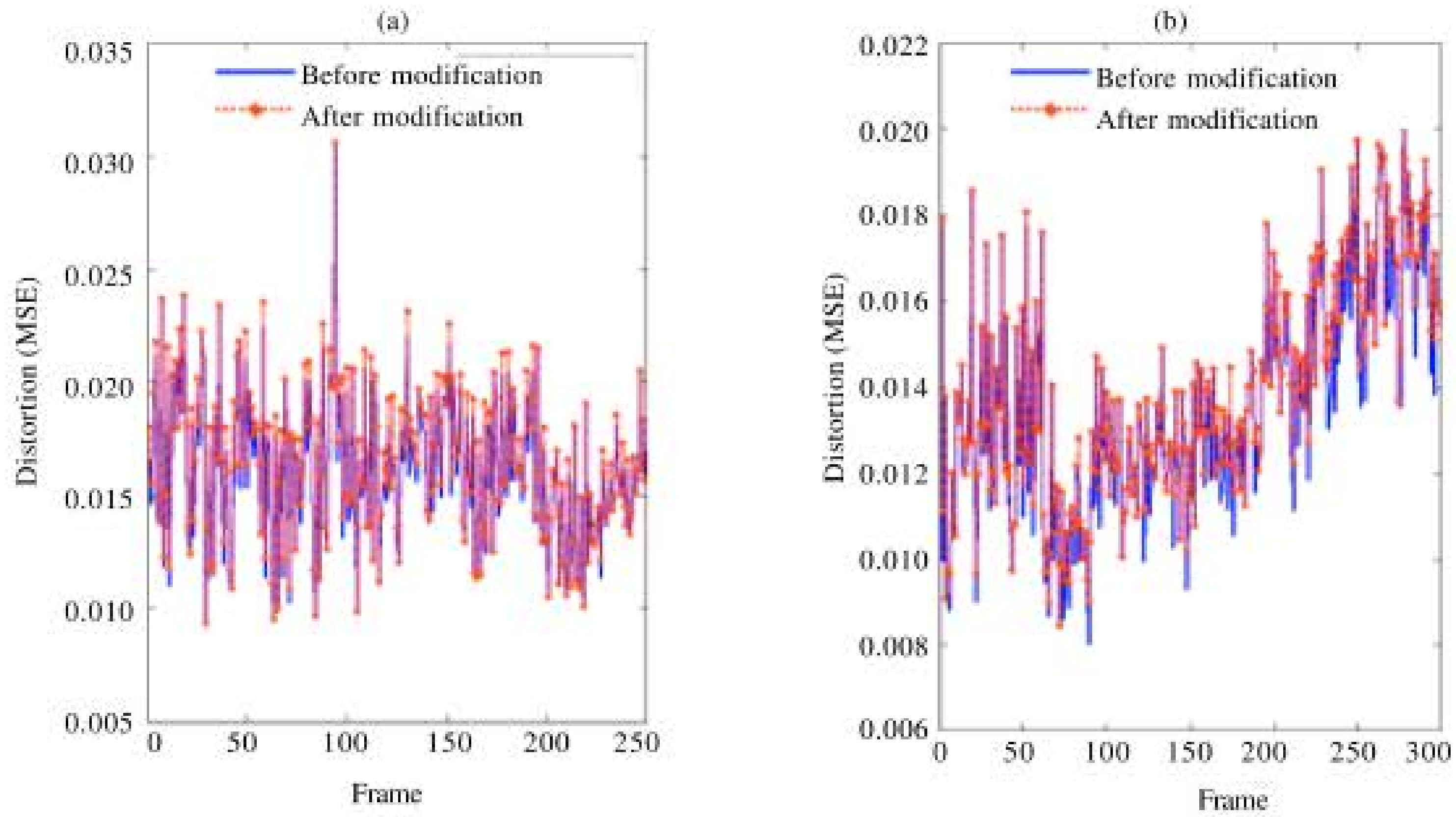

Fig. 5: The mean MSE between pre-and-post algorithm running (a) Flower sequence and (b) Coastguard sequence

CONCLUSION

In this study, we firstly classify the existing main R-D models into three categories: the analytic R-D model, the empirical model and the semi-analytic model. Their related performances are analyzed and compared. Some theory reasons are introduced to explain their some performance and property differences. And then the experimental building method of the piece linear R-D model is introduced and a method of property analysis and modification for the piece linear R-D model is proposed. The experiment results show the piece linear R-D model is accuracy and the method of property analysis and modification is valid. Present experiment results also confirm that the piece linear R-D model best-suits the P2P video streaming media system. The future study is to design and implement a P2P video streaming media system, based on the piece linear R-D model and the optimized algorithms employed it.

ACKNOWLEDGMENTS

This research is supported in part by National Natural Science Foundation of China under Grant No.60673179 and Natural Science Foundation of Zhejiang Province under Grant No. Z106727.

REFERENCES

Dai, M. and L. Dmitri, 2003. Analysis of rate-distortion functions and congestion control in scalable Internet video streaming. Proceedings of ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video, June 1-3, Monterey CA, pp: 60-69.

Dai, M., 2004. Rate-distortion analysis and traffic modeling of scalable video coders. Ph.D Thesis. Department of Electrical Engineering, Texas A and M University.

Dai, M., L. Dmitri and H.M. Radha, 2006. Rate-distortion analysis and quality control in scalable Internet streaming. IEEE Trans. Multimedia, 8: 1135-1146.

Hang, H. and J. Chen, 1997. Source model for transform video coder and its application part-I: Fundamental theory. IEEE Trans. Circuits Syst. Video Technol., 7: 287-298.

Hsu and M. Hefeeda, 2007. Optimal bit allocation for fine-grained scalable video sequences in distributed streaming environments. Proceedings of ACM/SPIE Multimedia Computing and Networking Conference (MMCN'07), Jan. 29, San Jose, CA, USA., pp: 1-12.

Park, H.J. and T.W. Lee, 2006. Capturing nonlinear dependencies in natural images using ICA and mixture of Laplacian distribution. Neurocomputing, 69: 1513-1528.

Sun, J., W. Gao, D. Zhao and Q. Huang, 2006. Statistical model, analysis and approximation of rate distortion function in MPEG-4 FGS videos. IEEE Trans. Circuits Syst. Video Technol., 16: 535-539.

Zhang, X., A. Verto, Y. Shi and H. Sun, 2003. Constant quality constrained rate allocation for FGS coded video. IEEE Trans. Circuits Syst. Video Technol., 13: 21-130.