Analyzed Rate Distortion Model in Standard Video Codecs for Rate Control

Mehdi Rezaei, Moncef Gabbouj, Stephan Wenger
[1,2]Institute of Signal processing, Tampere University of Technology,
[3] Nokia Research Center, Tampere, Finland
Mehdi.Rezaei@tut.fi, Moncef.Gabbouj@tut.fi, Stephan.Wenger@nokia.com

Abstract— **In this paper we propose a general and precise Rate-Distortion (RD) model for standard video codecs, including MPEG-4, H.263 and AVC H.264. The model can be used in almost all video rate control algorithms for different applications, from one-pass to multi-pass, and from constant rate to constant quality. It is designed taking into account previous theoretical results, and based on new assumptions which are confirmed by comprehensive practical experiments. We conceptually divide the RD model into two separate parts. The first part is codec specific and reflects the codec's properties. The second part is video content specific and describes the interaction between the video content and the encoder. This division provides several advantages. The model performs more precisely but needs less updating efforts. Furthermore, in this paper we introduce a simple complexity measure for the video content. As an example, we parameterized the RD model by the complexity measure for the estimation of Intra frame bits. This example can be used for complexity estimation in a start frame and in the scene cuts. The experimental results show that both the RD model and complexity measure perform well, with high precision as estimation tools.**

I. Introduction

Digital video applications are increasingly important, and different applications have very different constraints. One of the main challenges in designing a digital video system is to provide encoded video with high visual quality and compression performance, while all constraints are obeyed. Several video compression standards have been specified, with an ever increasing number of coding parameters that allow tailoring the bit stream. Controlling these parameters to obey the (sometimes very different) constraints in various applications presents a major challenge. This task is performed by a rate control algorithm, which is not part of a standard. Many rate control algorithms (RCA) have been designed for different applications and different encoders. Almost all of them utilize a rate distortion (RD) model as a basic tool that describes the relationship between the rate and the quality of the encoded video. There are close relationships between the distortion and the quantization scale (QS) in all standard video encoders; hence, an RD model can be described as a relationship between the rate and the QS. Most of the RD models used in previous research works have a common base theory with different simplifications and assumptions. The RD function in [1], [2] is modeled as a second-order function of the inverse of a distortion measure. The distortion measure is assumed to be the average QS of a frame. The same model is used in [3] with some modifications. They scale the model according to the complexity of the video content. They also add another parameter to the model, which describes the header bits, motion vectors and shape information. These concepts have been used many times in different RCAs. Furthermore different first-order RD models, which are simplified versions of the second-order RD model, have been employed.

In this paper we propose a new RD model which is a generalized form of the RD models previously used. It seems to provide more precision than the previous models tested. We divide the RD model into two separate parts. The first part of the model is codec specific and hence reflects the codec's properties independently from the video content. The second part is video content specific and describes the interaction between the video content and the encoder.

Splitting up the RD model, and analyzing both parts independently has several advantages. First, the model provides more accuracy because each part can be parameterized more precisely compared to a joint model. Secondly, the codec specific part is be parameterized only once for each codec and can be used henceforth with no need for updates during encoding. This provides good accuracy with less computational complexity. The content specific part of the model can be parameterized and updated for different video content properties, similar to previous RD models. Conceptually, our new RD model can be "plugged" into all reviewed RCAs with minimal changes, and is expected to increase the performance of those RCAs. Our model can be used in macroblock level and frame level (or higher levels) of control. It can be parameterized for all types of frames we are aware of, including Intra frames, Inter frames and the frame types introduced in H.264/AVC.

The model is highly flexible. It can be parameterized to utilize the mean absolute difference of motion compensated block (MAD) in block or macroblock level RCAs for P or B frames. It can also be parameterized for Intra frames, so to

Mehdi Rezaei is researcher with the Institute of Signal Processing, Tampere University of Technology, Finland. Moncef Gabbouj is Professor and head of the Institute of Signal Processing. Stephan Wenger is a Principal Scientist with Nokia Research Center, Tampere.

estimate the number of encoded bits (in start frame and in the scene cuts) where no previous results are available. In this application the model can be updated according to the video content itself and not according to previous encoded results.

Finally, we introduce a comparatively accurate complexity measure, which can be used for parameterization of the RD model as discussed above. As an example, we parameterized the RD model with the complexity measure for the estimation of Intra frame bits. The experimental results show that the RD model and the complexity measure perform with good precision.

This paper is organized as follow: Section II provides an overview of the basic RD model. Section III presents the details of proposed RD model and complexity measure. Simulation results are available in Section IV. The paper closes with conclusions in Section V.

II. BASIC RD MODEL OVERVIEW

To illustrate the basic RD model we summarize the results derived in [1], [2]. Assuming that the video source signal has Laplacian distribution statistics

$$P(x) = \frac{\alpha}{2} e^{-\alpha|x|}, \text{ where } -\infty < x < \infty . \quad (1)$$

The distortion measure is defined as

$$D(x, \tilde{x}) = |x - \tilde{x}|, \quad (2)$$

then there is a closed form solution for RD function in [4]

$$R(D) = \ln\left(\frac{1}{\alpha D}\right), \quad (3)$$

where $D_{\min} = 0$, $D_{\max} = \frac{1}{\alpha}$, $0 < D < \frac{1}{\alpha}$.

The RD function can be expanded into a Taylor series

$$R(D) = \left(\frac{1}{\alpha D} - 1\right) - \frac{1}{2}\left(\frac{1}{\alpha D} - 1\right)^2 + R_3(D)$$
$$= -\frac{3}{2} + \frac{2}{\alpha} D^{-1} - \frac{1}{2\alpha^2} D^{-2} + R_3(D) \cdot \quad (4)$$

Based on the above observation, they presented a model to evaluate the target bit rate before performing the actual encoding in [1], [2] as follow:

$$R_i = a_1 \times Q_i^{-1} + a_2 \times Q_i^{-2}, \quad (5)$$

where R_i is the total number of bits used for encoding the frame *i,* Q_i denotes quantization scale used for encoding of the frame *i,* a_1, a_2 are the first and the second order coefficients.

In [3] they enhanced the RD model, by introducing two new parameters. They modified the RD model as

$$\frac{R_i - H_i}{M_i} = a_1 \times Q_i^{-1} + a_2 \times Q_i^{-2}, \quad (6)$$

where H_i denotes the bits used for header, motion vectors and shape information. M_i is a scaling factor that scales the model according to complexity of the video content. They used the MAD for motion compensated residual as the complexity measure.

Many other RD models have exactly the same structure and utilize similar concepts, or they are a simplified model with a first order RD function. Many RCAs have used the same model with different parameterization in different levels of control. While we use the results of the base theory, we introduce a new concept, which provides a precise RD model. We assume that the rate is not only is a function of the video content properties, but also of many building parts in the encoder structure. The header bits introduced in formula (6) is only one example of many parameters that can affect on the RD model at different operating points. Another important example is the entropy encoders as important parts in all video codecs. Each codec has several entropy encoders and each entropy encoder has a different performance in different operating points.

III. DETAILS OF PROPOSED RD MODEL

The proposed RD model and complexity measure are discussed in this section. We discuss and parameterize the model for the Intra frames, as an example, in standard video codecs, including H.263, MPEG-4 and H.264/ AVC. Generally it can be used for all type of frames including I, P, B, IDR, SI, and SP frames, and also it can be used for all video units including macroblock, slice, frame and GOP. We introduce the analysis concept which increases the accuracy of the model and simplifies the use of the model.

A. General RD Model

We assume that the number of bits employed for a coded video unit (macroblock, slice, frame …) and also the quality of encoded video unit (after reproduction) are the result of an interaction between the video encoder and the video content. In other words, the bitrate is a function of the video compression standard (denoted CS henceforth) and the video content.

$$R = F(CS, Content), \quad (7)$$

where R denotes rate and F expresses the functionality. The main idea of this paper is to conceptually divide the RD function into two separate, independent parts

$$R = F_1(CS) + F_2(CS, Content). \quad (8)$$

Each video compression standard has it own syntax and set of coding tools, including, for example, entropy code, coding tools, etc. These affect the bit rate for a given quality level and for a given content. Simply put, some video codecs are more efficient than others. We assume in our model that at least a part of this efficiency can be expressed independently from the content, and that is what is expressed in F_1.

A simple example: H.263 employs header bits different from the MPEG-4, and this can be reflected in the codec specific function F_1.

The second part of the model is the content specific function. It describes the closed interaction between the encoder and the video content. While a video encoder with fixed encoding parameters provides different results for different video contents, it means some properties or parameters are different in the video contents. Functionality between these parameters and rate and distortion can be expressed as F_2 in the proposed model. For example MAD which is used in [3] as a representation of complexity of video content can be placed as a parameter in the content specific function F_2.

B. Model Parameter Estimation

For simplicity, henceforth we assume the use of one given coding standard. We further disregard the use of different coding tools that coding standard may allow, and consider only a single tunable parameter, the Quantizer (Q). In this case the model is simplified to

$$R = F_1(Q) + F_2(Q, Content). \tag{9}$$

In other words, the encoder specific function appears as a function of QS and interaction of encoder with video content appears as a function of QS and content properties. In all standard video encoders the distortion is a monotonic increasing function of the quantization step. The quantization step in H.263 and MPEG-4 encoders is linearly proportional to the quantization scale (QS) or quantization parameter

$$Q_{Step} = 2Q, \quad 10)$$

Furthermore, in H.264 AVC encoder

$$Q_{Step} = 2^{\left(Q/6\right)}. \tag{11}$$

So the function between rate and QS can be considered as a RD model

$$R(D) = F_1(Q(D)) + F_2(Q(D), Content). \tag{12}$$

We parameterize the RD model for Intra frames, as examples, in three standard video codecs. The same approach can be used for other type of frames or for macroblocks. Before parameterization, we identify a complexity measure that specifies the relationship between rate and the video content properties. Different measures have been used in older work. Generally, the number of coded bits generated in a block is proportional to the block variance. In H.263 TMN8 [5], they model the rate of an intra or motion-compensated macroblock as a function of variance of luminance and chrominance values in the macroblock. In MPEG-4 Appendix-L [6] they classify the blocks according to their variance and then they use a look up table for estimation of encoded bits. In many RCA such as [3] they use the MAD (Mean Absolute Difference) for motion-compensated macroblocks as a measure of complexity. From simplicity point of view, the MAD is a good criterion because it calculated for motion estimation process. Also SAD (Sum Absolute Difference) for motion-compensated macroblocks as a measure of complexity is used in several RCAs such as [7]. MAD and SAD are appropriate for B and P frames, but cannot be used for Intra frames. On the other hand each complexity measure which is used for complexity estimation of the I-frames is applicable for residual information in P and B frames. The MAD and SAD are used because they have been calculated for motion estimation process already. Our study shows that from the accuracy point of view these complexity measures are not very accurate.

We tested several complexity measures. In most cases, variance provides a good accuracy in the estimation of complexity. However, experimental results show that it is insufficient when some kinds of textural structures are present in the video. A video frame with special textures may need a lot of bits for encoding while it has a low variance. On the other hand we can find some textural structures which have a high variance but they need only a few bits for encoding.

To reflect the shortcomings discussed above, we introduce the following complexity measure for complexity estimation of the Intra frame.

$$X = \left(\overline{V} + \overline{T_V} + \overline{T_H}\right), \tag{13}$$

where
$$V = \frac{1}{256}\sum_{i=1}^{16}\sum_{j=1}^{16}\left(P(i,j) - \overline{P(i,j)} \right)^2, \tag{14}$$

$$T_V = \sum_{i=1}^{16}\sum_{j=1}^{16}\left| P(i,j) - P(i,j-1) \right|, \tag{15}$$

$$T_H = \sum_{i=1}^{16}\sum_{j=1}^{16}\left| P(i,j) - P(i-1,j) \right|, \tag{16}$$

where X denotes the complexity measure. V is the variance of luminance pixels $P(i,j)$ in one Macroblock, T_V and T_H denote the vertical and the horizontal texture measures on the luminance pixels. $\overline{V}$, $\overline{T_V}$ and $\overline{T_H}$ are average values of V, T_V and T_H respectively on all macroblocks in the frame. The experimental results indicate that these parameters can be independent, and all of them play an important role in coding complexity. Since they have an equal degree of relevance, they get equal weighting factors in the final closed form formula (13). We selected randomly a large number of frames (1200 frame) from 25 well-known video clips, which are used in standardization processes. The selected frames were coded as intra frames with constant QS by the three standard video encoders. We define the coding complexity function as multiplication of QS and required bit budget for encoding

$$S = Q \times R, \tag{17}$$

Figure 1 shows the obtained coding complexity of encoded frames as a function of the complexity measure (X). Keep in mind that the above results are produced using a constant QS. For the tested operation point, the results indicate that the rate fits well into a first order function of the complexity measure. In other words, assuming a constant QS, the curve fitting results can be expressed as

$$R_{Q=Constant} = \alpha + \beta.X \,, \qquad (18)$$

Where R denotes encoded bits of I-frame. α and β are zero-order and first-order constant coefficients extracted from curve fitting. In general case in which QS is not constant we can assume

$$S = Q \times R(Q) = A(Q) + B(Q) \times X \,, \qquad (19)$$

Where $A(Q)$ and $B(Q)$ are replacement of α and β in general case. Now we repeat the above experiments for different values of QS. For each value of QS we find a value for $A(Q)$ and also another value for $B(Q)$ which are zero-order and first-order coefficients of the fitted curves. Results of experiments are shown in fig. 2. As it shown there are regular dependencies between complexity function coefficients (A, B) and Q in different standard video encoders. The function (19) can be manipulated as

$$R(Q) = \frac{A(Q)}{Q} + \frac{B(Q)}{Q} \times X = C(Q) + D(Q) \times X \,. \qquad (20)$$

Figure 3 shows the rate coefficients ($C(Q)$, $D(Q)$) for different video encoders. These two functions are corresponding to F_1 and F_2 in (9). The functionalities of $C(Q)$ and $D(Q)$ can be extracted by a simple curve fitting between values of these two function and QS, but we perform the curve fitting process between these two function and inverse values of QS, so to find a closed form model similar to the basic RD model. After the curve fitting we obtain a closed form function between rate and the complexity measure as

$$R(Q) = E(Q^{-1}) + F(Q^{-1}) \times X \,. \qquad (21)$$

The rate model coefficients ($E(Q^{-1})$, $F(Q^{-1})$) are defined by curve fitting. Results of curve fitting are shown in fig. 4. If we use, for example, second order polynomials for curve fitting the rate function can be expressed as

$$R(Q) = \left(\frac{e_2}{Q^2} + \frac{e_1}{Q} + e_0\right) + \left(\frac{f_2}{Q^2} + \frac{f_1}{Q} + f_0\right) \times X \,. \qquad (22)$$

where e_i, f_i are polynomial coefficients. While there is a closed relationship between Q and distortion this function can be considered as RD model. This is a typical solution for RD model in (9) and (21). If we manipulate the RD model as

$$R(Q) = \left(\frac{e_2 + f_2.X}{Q^2} + \frac{e_1 + f_1.X}{Q} + e_0 + f_0.X\right) . \qquad (23)$$

Comparing this model with previous models, in fact the previous models such as (5) and (6) are simplified version of this model.

Coming back to the new RD model (22) we observe two separate parts. The first part is the codec specific function, which is independent of video content. It can be parameterized once, with a high accuracy for each coding standard and henceforth used without updates. The second part is content specific, and it should be parameterized for each encoder and each complexity measure. From an estimation error point of view, the analyzed RD model provides less estimation error because at least the first part can be modeled with high accuracy, independently of video content, while the both part have equal degree of importance in the total value.

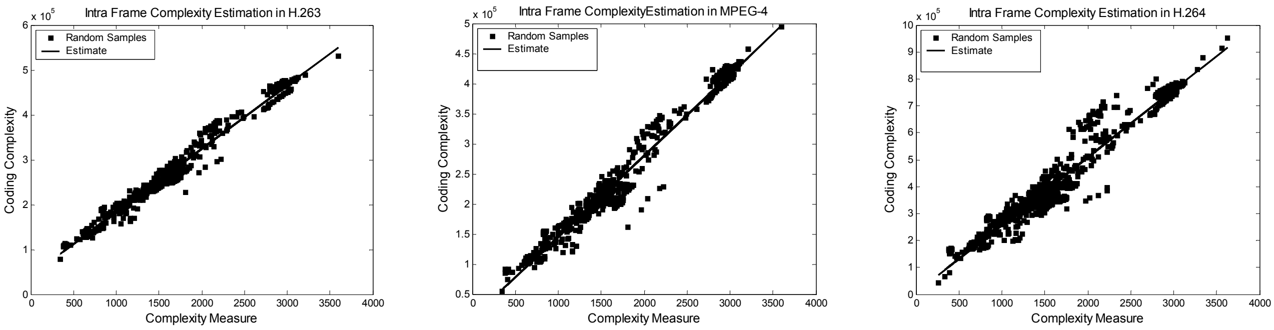
IV. RESULTS

To evaluate the new model, we select a large number of frames (1000 frames) from different video sequences. We encoded the test frames by the three video encoders complying with H.263, MPEG-4 and H.264. To evaluate the RD model and the complexity measure separately, we use different combinations of RD models and complexity measures for the estimation of Intra frame encoded bits. First, we parameterized our RD model by our complexity measure. Second, we used the variance as a generally accepted complexity measure, for parameterization of our RD model. Finally, we parameterized a general second-order RD model, including header bits such as formula (6), by our complexity measure. After parameterization we estimated the encoded bits of the Intra frames by the three models mentioned above, and compare the estimated values with experimentally obtained values for the encoded bits. The table 1 illustrates the percentage of the average estimation error on all encoded frames. We use the average estimation error per cent as

$$A.E.E\% = mean\left(\frac{100 \times abs(RealBits - EstimateBits)}{RealBits}\right) . \qquad (24)$$

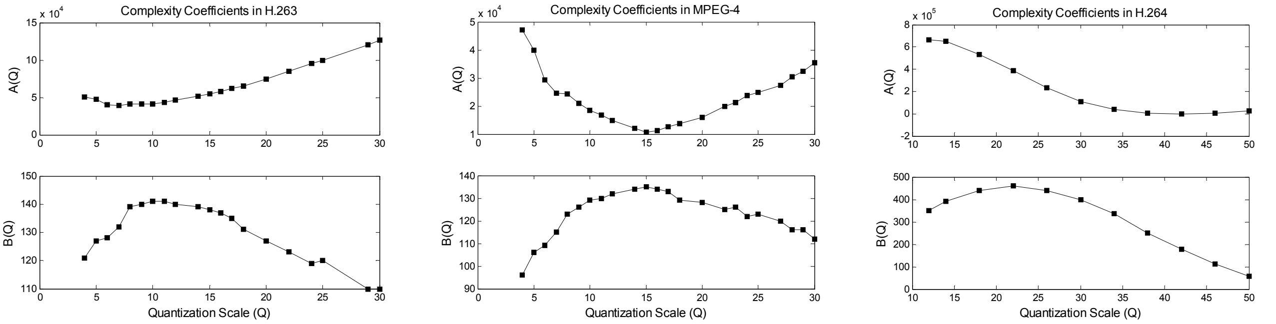
Two experiments by our RD model and two complexity measures indicate that the proposed complexity measure is more accurate than the variance. Also two experiments by our complexity measure and two different RD models prove that the proposed RD model performs much better than the general second-order RD model.

In average, on three video encoders, the bit estimation error of our RD model is only 49% of error obtained by the general second-order RD model. The complexity estimation error of our complexity measure is only 59% of it's the error obtained by using the variance.



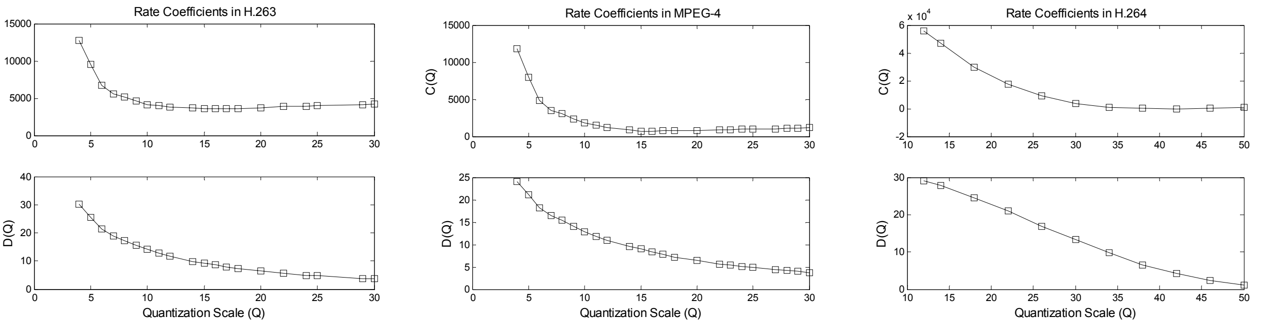
(a) (b) (c)

Fig. 1. Real coding complexity as function of the complexity measure in standard video encoders: a) H.263 b) MPEG-4 c) H.264



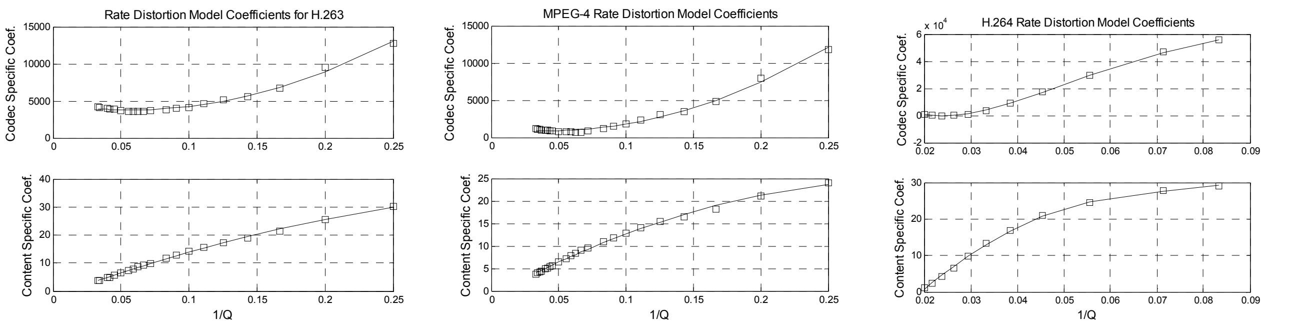
(a) (b) (c)

Fig. 2. Complexity coefficients as functions of quantization scale in different standard video encoders: a) H.263 b) MPEG-4 c) H.264



(a) (b) (c)

Fig. 3. Rate coefficients as functions of quantization scale in different standard video encoders: a) H.263 b) MPEG-4 c) H.264



(a) (b) (c)

Fig. 4. Rate Distortion model coefficients as functions of inverse of quantization scale in standard video encoders: a) H.263 b) MPEG-4 c) H.264

TABLE I. COMPARISON OF THE PROPOSED RD MODEL AND COMPLEXITY MEASURE WITH THE GENERAL SECOND-ORDER RD MODEL AND VARIANCE AS GENERAL COMPLEXITY MEASURE IN ESTIMATION OF INTRA FRAME BITS IN DIFFERENT STANDARD VIDEO ENCODERS.

RD Model and Complexity Measure	**Average Estimation Error %**		
	H.263	MPEG-4	H.264
New RD Model and New Coplexity Measure	5.09	7.44	10.81
New RD Model and Variance	10.87	12.97	15.32
2th-order Model+Header and New Coplexity Measure	10.81	11.58	25.45

The experiments generated different values for the average estimation errors for the three compression standards. This difference is not a result of the RD model. Considering the results of complexity measure in fig. 1, the difference in the estimation error is a result of the different accuracy of the complexity measure, dependent on the three codecs. The high concentration of samples near the fitted curve indicates a high accuracy of the complexity measure for H.263. The widely distributed samples around the fitted curve represent less accuracy of complexity measure for H.264.

Figure 5 shows the obtained bits and bits estimated by the proposed RD model and complexity measure, for a number of Intra frames. It also depicts the estimation error. The inaccuracy of the model is rarely larger than 15%, and well below 10% in most cases. Considering the very different nature of the content, that was randomly selected from a number of commonly used video sequences; we consider these results as excellent.

We believe that most RD models used in previous work are simplified versions of ours, in that the compression standard related term is weighted zero. We trust to have shown by our numerical results that all coefficients employed in our RD model have a considerable degree of importance. Hence, any kind of simplification increases the estimation error, perhaps considerably. We suggest that all reviewed older RD models produce an estimation error mostly because of the simplifications in the codec specific parts of our RD model.

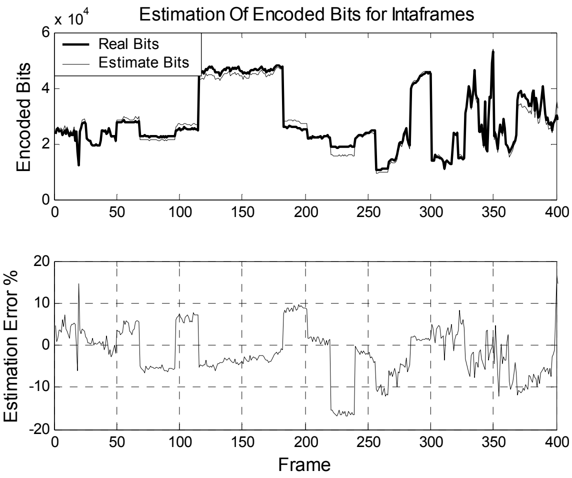


Fig. 5. Real value of encoded bits, estimated value and the estimation error by the proposed RD model and complexity measure in MPEG-4.

V. CONCLUSION

We introduced a new RD model, which comprises the base theory used in previous RD models but also considers the compression standard. We showed that the RD model should be conceptually divided into two separate parts: the codec specific part and the content specific part. Doing so has several advantages, mainly more accuracy with less updating effort.

The compression standard specific part is identified only once for a given standard and application. The content specific part can be parameterized, for each complexity measure, once or frequently depending on the application.

Finally, we proposed a new complexity measure for the estimation of the bit rate before encoding. As an example, we parameterized our RD model by this new complexity measure and applied to Intra frames encoding. The experimental results show that our RD model and complexity measure provide accurate estimation for the rate and the complexity. While we documented only the use on Intra frame complexity, the RD model is flexible, in that it can be parameterized for all types of frames, macroblock, block and GOP. It can also be used in conjunction with different complexity measure, such as SAD, MAD or even variance. Finally, we believe our RD model would be suitable also for other video compression standards once appropriately parameterized. .

REFERENCES

[1] T. Chiang, Y.-Q. Zhang, J. Lee, S. Martucci, H. Peterson, I. Sodagar, R. Suryadevara, and C. Wine, "A rate control scheme using a new ratedistortion model," Dallas, TX: JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 95/0436, Nov. 1995.

[2] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using a new rate-distortion model," *IEEE Trans. Circuits and Systems for. Video Technology,* pp. 246–250, Feb. 1997.

[3] Hung-Ju Lee, T.Chaing, Ya-Qin Zhang, "Scalable rate control for MPEG-4 Video," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 10, pp. 878–894, Sep. 2000.

[4] A. Viteribi and J. Omura, "A new rate control scheme using a new rate-distortion model," in *Principle of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[5] Thomas R. Gardos, "Video Codec Test Model, Near-Term, Version 8 (TMN8)," *ITU - Telecommunications Standardization Sector*, Document Q15-A-59, Portland, 24-27 June 1997

[6] "MPEG-4 Video Verification Model version 18.0," Appendix I: Rate Control , *ISO/IEC JTC1/SC29/WG11 N3908* ,January 2001/Pisa

[7] King N. Ngan, T. Meier, Zhenzhong Chen, "Improved single-video-object rate control for MPEG-4," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 13, pp. 385–393, May 2003.