ROYALTY COST BASED OPTIMIZATION FOR VIDEO COMPRESSION

Emrah Akyol, Onur G. Guleryuz, and M. Reha Civanlar
{eakyol, guleryuz, rcivanlar}@docomolabs-usa.com
DoCoMo USA Labs, 3240 Hillview Avenue, Palo Alto, CA 94304, USA

ABSTRACT

A video compression standard incorporates many tools and technologies which must be licensed by systems that deploy the standard. The licensing determines the royalty costs that must be paid to the holders of intellectual property on the respective tools. With current abundance of well understood and effective video compression tools, one can imagine the formation of cross cutting tool libraries with tools drawn from different video compression standards. This allows dynamic selection from a large pool of tools, having potentially overlapping functionality, when encoding individual video sequences. In this paper we examine the royalty cost aspect of the scenario where video is encoded using a library of royalty bearing tools by considering encoding that jointly optimizes rate, distortion, and royalty cost. We provide a system that optimizes video delivery under various licensing conditions imposed on tool intellectual property. We present an example of royalty based encoding (using assumed royalty costs) to show the merit of the proposed framework.

1. INTRODUCTION

In recent video coding standards (such as the H.264/AVC standard [1]) significant improvements in rate-distortion performance have been made possible by the incorporation of a substantial number of new tools such as variable block sized motion estimation, intra prediction, quarter pixel motion compensation, multi-frame and multi-hypothesis motion estimation (ME), adaptive de-blocking filter, and context based arithmetic coding, just to name a few. Using just the toolsets collected from currently deployed standards, it is clear that today one can transport video through a variety of networks using a vast range of tools that correspond to a vast range of efficiencies in the end to end delivery. The currently developed MPEG-RVC (Reconfigurable Video Coding) standard [1] allows the construction of tool libraries and provides a description language where an encoder can signal to a decoder which subset of the tools within a library are utilized on a particular video sequence. When combined with the proposed work, such cross cutting libraries can facilitate many applications that do not readily fit into the target application domain of a single standard or profile.

If one is concerned with the highest efficiency media delivery, one is often confined to more recent, state-of-the-art tools that tend to have high royalty costs. If on the other hand, one allows some inefficiency, it may be possible to accomplish delivery with reduced royalty costs or even free of any royalty costs. Given that the encoded media itself may also have content licensing costs, the system we consider finds the optimal trade-off by encoding media in a way that minimizes the combined royalty cost for the desired media quality level and effective bandwidth of the transport medium. In essence, the encoding we propose optimizes performance over a function that describes optimal encoding points for the allowed range of triplet values formed by distortion, rate, and royalty cost.

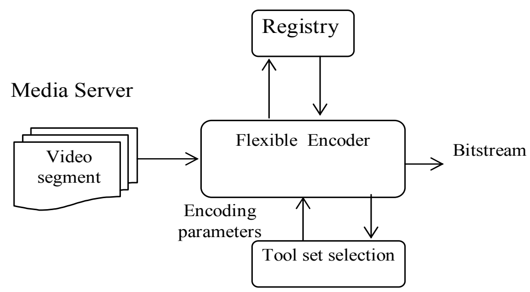

Figure 1: The general scheme of the proposed framework

In this paper we provide a system, composed of a server, a registry, and users, which optimizes video delivery under various licensing conditions imposed on tool intellectual property. The optimization is such that at the same decoded video quality, video encoded for high bandwidth environments can be decoded using lower efficiency but reduced royalty cost tools, whereas video encoded for low bandwidth environments can be decoded using higher efficiency but increased royalty cost tools. A general scheme for the proposed royalty based encoding is shown in Figure-1. In this framework each video segment is encoded with possibly different set of tools depending on the total rate, distortion, and royalty cost constraints in a content adaptive fashion. We assume a flexible encoder can use any subset of the toolset, Ω, which is known to both at the encoder and decoder.

We assume that each tool in Ω has a royalty cost that determines the cost of licensing that tool for use in coding the media. The royalty cost can be flexible, for example, the cost can be fixed no matter how many times the tool is used during encoding, it can be based on the number of times the tool is used, it can depend on other tools used (for example a royalty bearing tool can become royalty free if combined with other royalty free tools), or it can be based on the expected quality of the decoded media (for example a tool may have different royalty costs when video is delivered to a device with a low resolution display as opposed to a high resolution display).

The royalty costs and conditions of each tool are stored in a registry, which takes the media coding conditions and tools used by the encoder and determines a certificate Φ_i for the i^{th} media segment. For the rest of this work, for notational convenience, we will parameterize these conditions with a vector Γ_i for media segment *i*. Γ_i contains the relevant conditions such as targeted media quality Q_i, effective bandwidth B_i, utilized tools $\Omega_i \subset \Omega$, media specific data, etc.

For segment *i* of media data, the encoder determines the desired conditions for media coding and chooses a toolset $\Omega_i \subset \Omega$ that minimizes the royalty cost for the segment under the desired conditions. The conditions and the toolset are then used to form Γ_i. The encoder obtains the certificate Φ_i for the i^{th} segment by contacting the registry with Γ_i. The media data is then encoded and the coded media is delivered to the decoder together with the certificate Φ_i, the list of utilized tools, and coding parameters.

The decoder decodes the coded media and optionally determines the utilized toolset and conditions to construct its version of Γ_i, which we term $\hat{\Gamma}_i$ can be used to ascertain the legitimacy of the coded media and the toolset used in decoding it. For example the decoder can transmit $\hat{\Gamma}_i$ to the registry which can validate that the media and toolset are legitimate.
Once the i^{th} segment is transported, the encoder proceeds with the encoding of the remaining segments. However the granting of certificates associated with the remaining segments can be conditional on the validation of one or more of the previous segments by the registry.

In this work, in addition to introducing the royalty based encoding framework, we attempt to answer this question: What is the optimal set of tools for each segment (such as MB, a slice, a frame or group of frames (GOP) depending on the target granularity) for a total royalty cost target?

2. PROBLEM FORMULATION

In the rest of the paper we utilize the following notation.
Ω : The set of all available tools.
Ω_i : The set of tools used in GOP *i.*
$R(\Omega_i)$: The rate obtained by using the tool set Ω_i.
$D(\Omega_i, R(\Omega_i))$: The distortion at rate $R(\Omega_i)$ using tool set Ω_i
C_j : The royalty cost of the tool j
λ_R : Lagrange parameter for rate
λ_C : Lagrange parameter for royalty cost

Problem Definition:

For each of the *M* segments, find the set of tools to use such that distortion is minimized under total rate and total royalty cost constraints, i.e.,

$$\Omega^{opt} = \arg\min_{\Omega_i \subset \Omega} \sum_{i=1}^{M} D(\Omega_i, R(\Omega_i)) \quad \text{st.} \sum_{i=1}^{M} \sum_{j \in \Omega_i} C_j \le C_T, \sum_{i=1}^{M} R_i \le R_T \qquad (1)$$

This constrained problem can be converted to an unconstrained one by introducing Lagrange multipliers (λ_R, λ_C) to the cost function. Then the problem becomes finding the optimal set of tools which minimizes the Lagrangian cost:

$$\Omega_i^{opt} = \arg\min_{\Omega_i \subset \Omega} \left\{ D(\Omega_i, R(\Omega_i)) + \lambda_R R(\Omega_i) + \lambda_C \sum_{j \in \Omega_i} C_j \right\} \qquad (2)$$

where (λ_R, λ_C) are determined to satisfy total cost and rate constraints. The values of λ_R and λ_C can be found, for example, by a two dimensional convex search on the *R-C* plane [3][4].

It is clear that depending on the sophistication of the license terms, the optimization in (2) can become involved. In order to expose the main outlines of the impact of the proposed work, in this paper we do not delve into detailed optimization issues but simply evaluate all tool set options for each segment and choose the one with minimum total Lagrangian cost. Depending on the toolset granularity and the royalty cost assignment scenario, this optimization can be performed for each macroblock, slice, frame or GOP. For example, if the cost of using a tool is assigned based on the number of times it is used, then the optimization should be performed on a macroblock basis. If the cost is assigned at GOP level, then optimizing the tool set for each GOP is sufficient. Note also that in cases where the optimization is computationally too complex (such as real time encoding), some rate-distortion models [5][6] for different tools can be used to decrease the complexity.

3. EXPERIMENTAL RESULTS

In our experiments we used four tools within h.264/MPEG-AVC standard and JM10.2 software [7] with hypothetical royalty cost assignments. We considered two different types of cost assignment: i) All royalty costs are identical ii) The royalty cost of a tool is assigned proportionally to the rate gains it provides on average. For case ii) values are deduced from Table-I, which shows the rate reduction obtained by using the tools multiple reference frames, subpixel accurate motion estimation, in-loop deblocking filter and advanced entropy coding for different video sequences. Specifically, we encoded each sequence in I-P-P… coding structure, with motion estimation with five reference frames or one, with advanced entropy coding (CABAC) or VLC tables, with or without the loop filter, and with quarter pixel accurate or integer motion estimation. Hence, to measure the rate change obtained by multi-reference frames, we set the number of references to one and measured the rate change between this encoding and the previous one at the same quantization parameter. Similarly, the rate difference between CABAC entropy coding and VLC, using integer or quarter pixel accurate MVs and using or disabling the loop filter is measured. The rate differences along with associated PSNR values are given in Table-I. Table II illustrates the determined costs. It is clear from Table-I that compression performance depends on the rate and video content differently for each tool.

Table I: Rate and distortion changes with different tools. The utilized tools are subpixel accurate ME, loop filter, advanced entropy coding, and multiple reference frames respectively. The rate reduction is shown as percentage with respect to the baseline along with PSNR gain at QP=25.

	Video-1 foreman	Video-2 akiyo	Video-3 coast.	Video-4 mobile	Video-5 bus
Tool-1 Rate	56.51%	55.93%	14.93%	51.56%	50.05%
ΔPSNR	0.40	0.48	0.72	0.64	0.80
Tool-2 Rate	1.05%	0.37%	0.76%	0.02%	0.45%
ΔPSNR	0.05	0.25	0.01	0.06	0.06
Tool-3 Rate	7.21%	4.69%	10.85%	7.81%	4.85%
ΔPSNR	0.06	0.02	0.06	0.06	0.05
Tool-4 Rate	15.28%	2.45%	0.65%	19.30%	5.87%
ΔPSNR	0.18	0.06	0.02	0.25	0.12

Table II: Assigned cost of using each tool for two different cases

	Case-1	Case-2
Tool-1 (Subpixel MV)	10	10
Tool-2 (Loop filter)	10	0.13
Tool-3 (CABAC)	10	1.55
Tool-4 (Multiple Ref.)	10	1.82

In our experiments, we compared segment based tool optimization to sequence based tool optimization under the two types of royalty cost assignment shown in Table-II. Specifically, we first generated a test video sequence by concatenating 10 MPEG test videos (mobile, akiyo, foreman, bus, coastguard, news, container, silent, carphone, mother) with different content at 30fps CIF resolution. To show the merit of content adaptive selection of the tools, we performed two sets of encodings: i) we used the same set of tools over the entire sequence ii) we adaptively change the tool set for each GOP. For the latter case, optimization is performed for each GOP and for both cases a total royalty cost is shown as a fraction (such as 12%, 25%, 50%, 75%) of the total maximum royalty cost obtained by using all of the tools.

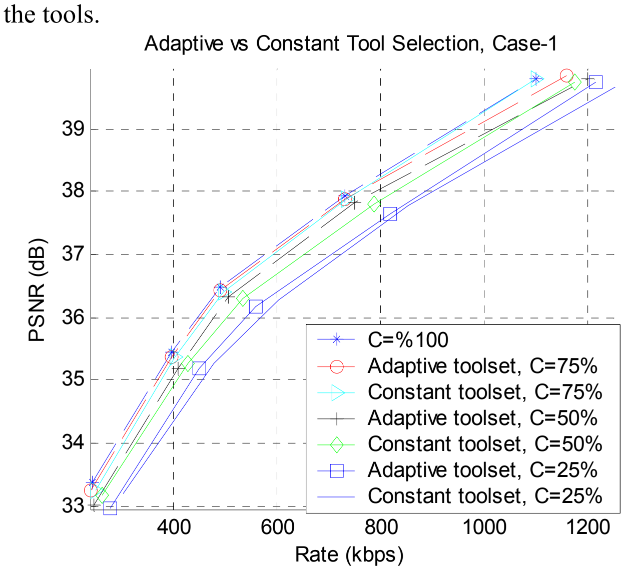

Figure 2: Comparison of using identical tools over the entire sequence and changing the tools with respect to content, case -1: identical costs for every tool QP= 23, 26, 28, 30, 33

Figure-2 and Figure-3 show the rate distortion curves for each tool selection method along with that of using all the tools. Figure-2 illustrates that when the royalty cost assignment is uniform, one can accomplish significant reduction in royalty costs with small loss in compression efficiency using both types of optimization. As seen from Figure-3 however, this is not the case when the royalty cost assignment is based on average rate reductions. Figure-3 also shows that segment adaptive selection of the tools provides larger gains over using fixed tools, when the royalty costs are fairly chosen, i.e., if the compression efficiency is considered in assigning the royalty costs of tools. Another important observation is that, segment adaptive tool selection can provide 25% royalty cost reduction with very small loss in compression efficiency, at least for this example.

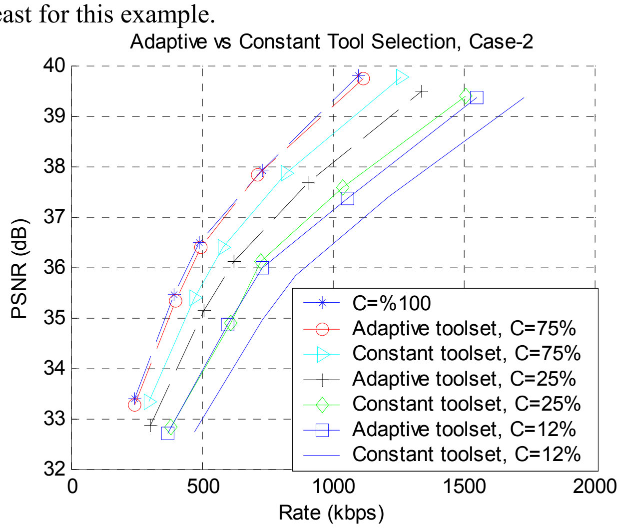

Figure 3: Comparison of using identical tools over the entire sequence and changing the tools with respect to content, case-2: different costs for each tool, QP= 23, 26, 28, 30, 33.

In Table-III, we present the tool usage for different total royalty costs, different royalty cost assignment cases, and different toolset selection optimizations (constant and adaptive toolset). It is interesting to note that, in case-2 where the royalty cost of each tool is different, the toolset changes significantly with the total cost change (which is not observed for case-1). This is because some tools have very low costs and are preferred when the total cost is small.

Table III: Tool usage for different royalty costs, adaptive and fixed tools for different cost assignments

	Tool-1	Tool-2	Tool-3	Tool-4
Case-1 C=75% Adaptive	97%	48%	98%	57%
Fixed	100%	0%	100%	0%
Case-1 C=25% Adaptive	69%	0%	20%	10%
Fixed	100%	0%	0%	0%
Case-2 C=75% Adaptive	76%	72%	98%	67%
Fixed	100%	100%	0%	0%
Case-2 C=25% Adaptive	29%	41%	2%	18%
Fixed	0%	0%	100%	100%

For the considered example, we have also noticed that under uniform royalty cost assignments, tools that do not contribute to rate-distortion heavily are automatically cut off when doing fixed optimization. Adaptive optimization allows such tools to keep contributing as needed by the varying content. Incorporation of more tools in the optimization is expected to generate finer granularity in the optimal encoding decisions.

4. CONCLUSION

In this paper we introduced a new framework for video compression by considering the royalty costs of the tools used in the encoding. We formulated the problem and provided an optimization approach for the royalty cost constrained rate-distortion problem. We showed the merit of the proposed optimization for different royalty cost assignment scenarios, and we showed the superiority of the content based toolset adaptation over using a fixed tool set for the entire sequence. Our work can be generalized to include content costs as well as more elaborate licensing conditions. Future work will consider results on the case where costs are assigned based on the number of times the associated tool is used on a per-macroblock basis (the optimization should be performed for each macroblock since every macroblock will have different tool usage), results with more tools that account for the interdependency among tools (as noted by one of the reviewers, the cost assignment method for case ii- ignores effects due to interdependencies among tools), as well as the nature of optimal decisions and tool usage based on content, bitrate, and buffer constraints among others.

5. REFERENCES

[1] Draft ITU-T Rec. and Final Draft International Standard of Joint Video Specification, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, Joint Video Team, May 2003.

[2] Study of Reconfigurable Video Coding V2.0, MPEG2006/N8040 Montreux, April 2006.

[3] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, pp. 399–417, 1963.

[4] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15 no. 6, pp. 23–50, November 1998.

[5] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding", IEEE Transactions on Image Processing, vol. 9 , no. 2, February 2000.

[6] P. Yin, A. Vetro M. Xia, B. Liu,, "Rate-Distortion Models for Video Transcoding", SPIE Conference on Image and Video Communications and Processing, Vol. 5022, pp. 467-488, January 2003

[7] JVT Reference Software Version 10.2 [Online]. Available:http://iphome.hhi.de/suehring/tml/index.htm