ESTIMATION OF MPEG-7 TRANSCODING HINTS IN THE COMPRESSED DOMAIN

Marco Fumagalli and Rosa Lancini

CEFRIEL – Politecnico di Milano, via R.Fucini, 2 – 20135 Milano, Italy. E-mail: [fumagall,rosa]@cefriel.it

ABSTRACT

Video transcoding is a mechanism to convert video bit-streams from one coding format to other formats. In this operation, the computational complexity and picture quality are the two most important concerns.

MPEG-7 standard defines media Transcoding Hints (THs) in order to both reduce the complexity and improve the visual quality during transcoding process.

The basic assumption is that THs are generated in advance and they are stored together with the host video itself. Unfortunately there exist some applications for which it is unpractical, for time or complexity constraints, to compute the THs in advance.

This work proposes to estimate the THs exploiting the data coming from the compressed domain of a video in order to save computational complexity.

1. INTRODUCTION

Video transcoding is an efficient mechanism used to convert video bit-streams from one coding format to other formats, including syntax, bit-rate and resolution conversions. In realizing the transcoder, the computational complexity and picture quality are usually the two most important concerns.

The MPEG-7 standard [1] provides a rich set of standardized tools to describe multimedia contents. Both human users and automatic systems that process audiovisual information are within the scope of MPEG-7. It aims at offering a comprehensive set of audiovisual description tools to create descriptions, which form the basis for applications enabling the needed *quality access to content*.

MPEG-7 defines media Transcoding Hints (TH) aiming at reducing complexity and improving the quality of generic transcoding processes. The possible applications of THs include but not limited to constant bit-rate and variable bit-rate conversion, encoding mode decision, frame- and bit-rate control, complexity reduction for motion re-estimation and video skimming and browsing.

The most important and used THs [3] are *Difficulty Hint*, which describes the encoding complexity of segments within a video sequence, *Motion Uncompensability*, that describes the amount of new contents in a video segment and *Motion Range*, which helps a transcoder in setting an appropriate Motion Vector (MV) search range during motion re-estimation.

The basic assumption is that THs are generated in advance and they are stored together with the host video itself (or stored separately but with a proper linkage to the video). Unfortunately there exist some applications for which it is unpractical, for time or complexity constraints, to compute the THs in advance, e.g., real-time applications or access to a vast video archive for a 'one-shot' transcoding where it is useless to pre-compute and store a wide number of meta-data. In these situations an interesting way to obtain the THs is to estimate them using the coding information coming from the bit-stream of the video that need to be transcoded.

The aim of this work is to estimate the THs of a video sequence by exploiting its compressed domain in order to save computational complexity. In fact, parsing the input bit-stream, some coding information useful for the estimation of the THs (e.g., the MacroBlock (MB) bit-rate, the used quantizers (*QP*), and the MVs) can be accessed. The advantage of this approach is that we can obtain an approximation of the THs without any significant additional computational complexity that is instead typical of traditional THs calculation techniques.

The paper is organized as follows: Section 2 briefly explains the traditional computation of the THs. Section 3 proposes the estimation the THs values from the video bit-stream information and Section 4 shows some simulation results. Conclusions are given in Section 5.

2. TRADITIONAL THS CALCULATION TECHNIQUES

According to the traditional technique [3], the Difficulty Hints of a video sequence are extracted by encoding the original sequence using the best quantizer (*QP*=1) in Intra mode, then calculating and normalizing the achieved bit-per-frame for each segment. As regard of Motion Uncompensability and Motion Range, they can be computed, respectively, as the number of features points (FP) – extracted e.g., by using KLT algorithm - not tracked frame-by-frame and the greatest shifting of tracked FP within the considered segment of frames.

0-7803-9332-5/05/$20.00 ©2005 IEEE

All these approaches request the access to video data (and thus it requests the decoding process if only the bit-stream domain is available) and specific high-cost operations that cannot be justified e.g., in the case of 'one-shot' usage or in real-time applications.

3. THE PROPOSED ESTIMATION OF THE THS IN THE COMPRESSED DOMAIN

In this work we propose to compute the three above-mentioned THs using only the information coming from the coding process of the sequence or, in the case of pre-encoded video, the information available by parsing the bit-stream.

Difficulty Hints

In order to obtain the Difficulty Hints, we propose to use a rate-quantization model of each MB. According to [3], the rate-quantization model is well approximated by Equation (1)

$$R = \alpha + \frac{\beta}{QP^{\gamma}} \qquad (1)$$

where R is the bits associated to the considered MB, QP is the corresponding quantizer and α, β and γ are parameters. According to the Difficulty Hint definition, we need to estimate the number of bits R that should have been generated if we had used QP=1.

For each MB of the video we know from the coding process or from the parsing of the bit-stream the actual values of the pair $(\overline{QP}, \overline{R})$. The goal is to estimate, MB-by-MB, the values of the three parameters of the model in Equation (1), i.e., α, β, and γ, in order to compute R for QP=1. Given the number of equations is greater than that of constraints, we need two additional relationships among the parameters.

We found empirical relationships between α and γ and between β and γ - with a high degree of independency by the specific video sequence - that can be advantageously used for our aims. The empirical models are presented in Eq. (2) and (3) and their parameters in Table 1.

$$\alpha = \alpha(\gamma) = \frac{\gamma + p_1}{e^{(\gamma + p_2)}} + p_3 \qquad (2)$$

$$\beta = \beta(\gamma) = p_4 \cdot \gamma^2 + p_5 \cdot \gamma + p_6 \qquad (3)$$

Table 1. Empirical values of the parameters of Eq. (2) and (3)

p_1	-1551.4	p_4	680.5
p_2	-5.5	p_5	-1586.4
p_3	51.5	p_6	2432

Figure 1 shows the actual values of the parameters (α, β, γ) for each MB of 'Rugby' CCIR 601 sequence (MPEG-2, CBR=5Mbps) compared with the empirical models (similar results can be obtained with other sequences). In particular Figure 1(a) and (b) correspond to the Eq. (2) and (3) while (c) is a cross-comparison between the modeled α and β and the actual ones. The results show a good accuracy of the parameters relationships.

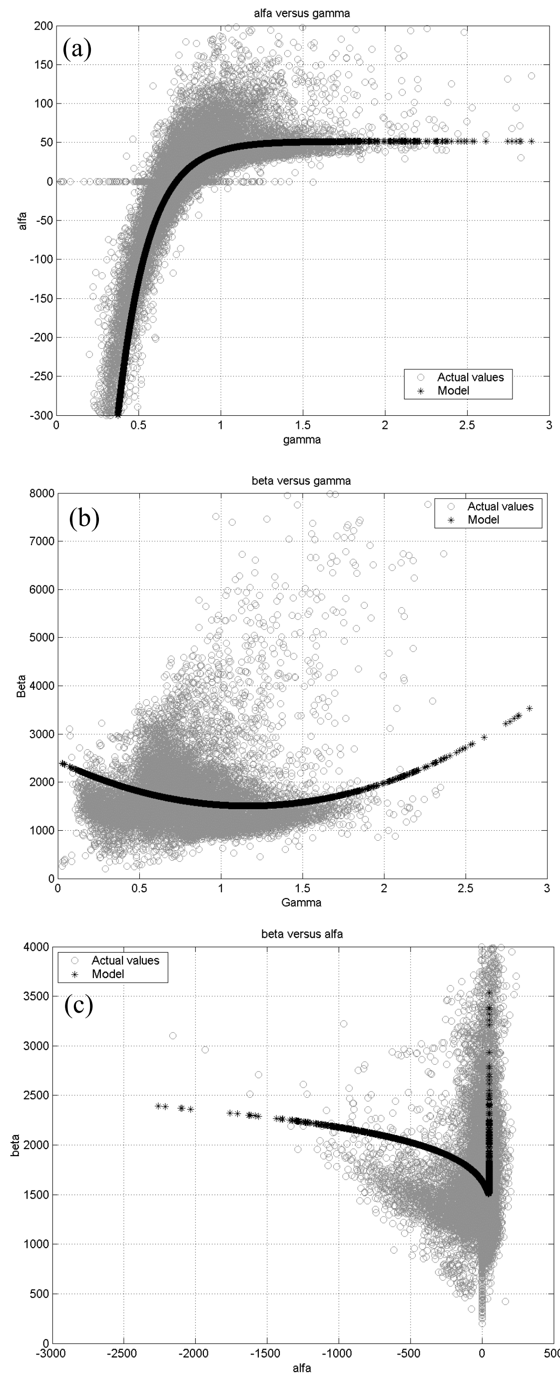

Figure 1. 'Rugby' CCIR 601 sequence. The actual value of the parameters for each MB versus the empirical model: (a) alfa vs. gamma, (b) beta vs. gamma, and (c) beta vs. alfa.

Eq (1) can now be rewrites as in Eq. (4)

$$R = f(QP,\gamma) = \alpha(\gamma) + \frac{\beta(\gamma)}{QP^{\gamma}} \qquad (4)$$

and thus, for each MB, it can be inverted to obtain $\gamma = g(\overline{QP}, \overline{R})$. Then, given the value of γ, α and β are calculated using Eq. (2) and (3). At this point we have all the information to estimate the number of bits R needed to code the considered MB for QP=1 using Eq. (1). Then the obtained R for each MB are summed along the considered video segment and normalized, as in the traditional technique.

To conclude, we highlight the inversion of Eq. (4) - introducing Eq. (2) and (3) - turns to be analytically complicated. To tackle this problem we use a table that approximates the inversion operation. Figure 2 shows the values of R of Eq. (4) for a set of values of QPs and γ parameters: for each MB, the actual $\overline{QP}$ value is used to select a specific curve of the figure, then the value of $\overline{R}$ is used to individuate γ over that specific curve. Given the curves are not monotonic, using this process we can find two different values of γ: in this case the greatest is chosen for its smaller gradient that guarantees better performance on approximation errors. We highlight the table depicted in Figure 2 is independent by the specific video sequences.

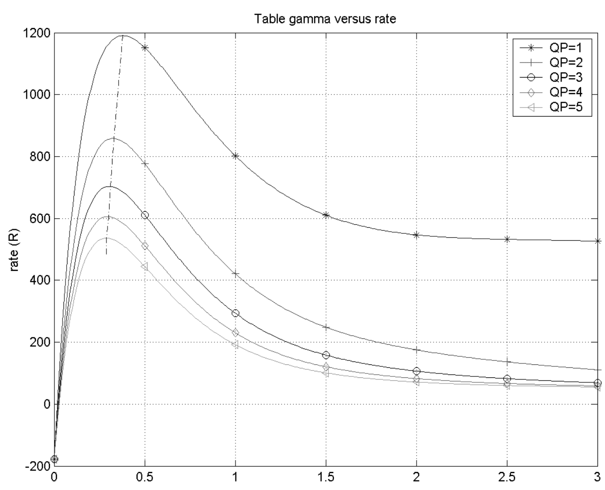

Figure 2. Relationship between gamma and rate R for a given QP according to Eq.(4).

Motion Range

In order to estimate the Motion Range using the information coming from the compression process or from the bit-stream parsing, we propose to approximate the greatest FP displacement (candidate to become the TH in the traditional approach) with the value of the N-th greatest MV in the same segment, where N is an empirical threshold. An intuitive explanation of this point is that, considering the displacement of each FP equal to the MV of the MB it refers to, there could be some MBs, not containing any FP, that have associated MVs greater than the maximum displacement of any FP. In other words, statistically speaking, there could be in each considered video segment, some MVs greater than the maximum FP displacement. This consideration justifies the introduction of the threshold N and its dependency by the video resolution. For a set of CCIR 601 sequences, the simulation results show an optimal empirical threshold between $N=50$ and $N=100$ (that corresponds to the minimum of MSE in Figure 3 for 'Rugby' and 'Fries' CCIR 601 sequences).

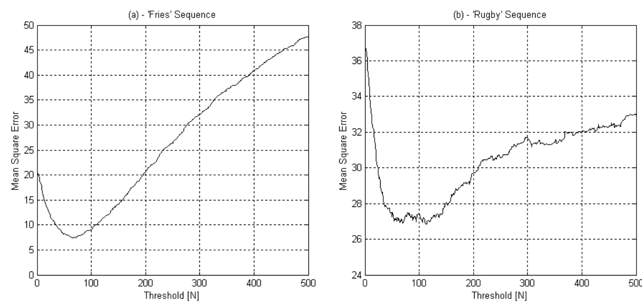

Figure 3. 'Rugby' and 'Fries, CCIR 601 sequences. MSE between the actual TH value and the estimated ones for different choices of the threshold N.

Motion Uncompensability

We propose to approximate the Motion Uncompensability with the percentage of Intra MBs in the considered video segment: i.e., we calculate the number of Intra MBs compared to the total number of MBs in that segment.

4. SIMULATION RESULTS

We test the proposed approaches on several sequences. Figure 4(a)-(c) depicts, respectively, the values of Difficulty Hints, Motion Uncompensability and Motion Range computed using the traditional and the proposed techniques for 'Fries' CCIR 601 sequence (MPEG-2, 20 Mbps). We can assert the proposed techniques properly approximate THs extracted using traditional methods [2].

To test the effectiveness of the proposed THs estimation method, we perform the transcoding of 'Fries' sequence from the MPEG-2 @ 20 Mbps to MPEG-1 @ 700 Kbps using the THs computed by both the traditional and proposed techniques [2]. Figure 4(d) shows the luminance PSNR tracks of the two experiments where the similar trend demonstrates the accuracy of the proposed techniques in preserving THs accuracy. Similar results are achieved for different sequences.

5. CONCLUSIONS

To conclude we can affirm that in real-time or 'one-shot' transcoding application, an accurate approximation of the THs of a video sequence can be obtained using the information coming from the compressed domain: it leads to save computational complexity without significant quality loss.

Further researches tend to improve the proposed techniques, e.g., a better extraction of Motion Hints, analyzing the FP selection mode, can reduce even more the computational complexity augmenting the estimation quality.

REFERENCES

[1] O. Avaro, P. Salembier, MPEG-7 Systems: Overview, IEEE Transaction on Circuit and System for Video Technology, 11, 6, June 2001.

[2] A. Yamada, M. Pickering, S. Jeanning, L. Cieplinski, J.R. Ohm, MPEG-7 Visual Part of Experimentation Model Version 8.0, ISO/IEC/JTC1/SC29/WG11/N3673, October 2000.

[3] P.M. Kuhn, T. Suzuki, A. Vetro, MPEG-7 Transcoding Hints for Reduced Complexity and Improved Quality, Proc. Int'l Packet Video Workshop, Kyongju, Korea, April 2001.

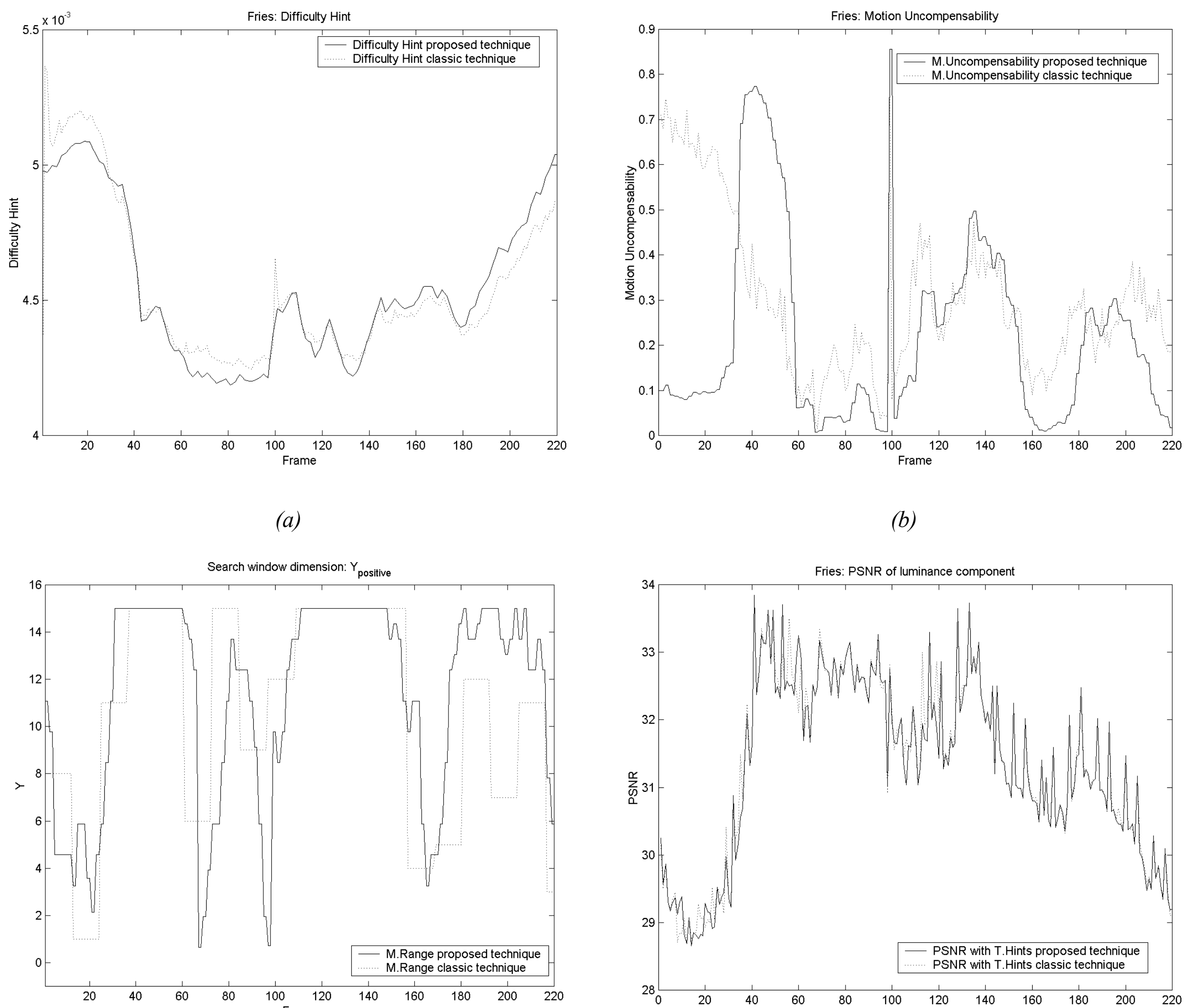

Figure 4: MPEG-7 Transcoding Hints for 'Fries' CCIR 601 sequence (MPEG-2 @ 20 Mbps): (a) Difficulty Hint, (b) Motion Uncompensability, and (c) Motion Range. (d) Y-PSNR track after transcoding (from MPEG-2 @ 20 Mbps to MPEG-1 @ 700 Kbps) using the proposed estimation techniques and the traditional ones.