MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Combined Rate Control and Mode Decision Optimization for MPEG-2 Transcoding with Spatial Resolution Reduction

TR2003-117   December 2003

Abstract

This paper presents a new algorithm for MPEG-2 transcoding with spatial resolution reduction. The proposed method combines rate control and mode decision to achieve optimal transcoding performance using Lagrange multiplier algorithm. Since the proposed method incorporates motion vector mapping and mode decision into a single Lagrange multiplier formula, by minimizing the Lagrangian cost function, the optimal solution for mode decision can be obtained. The proposed transcoding scheme has demonstrated better subjective and objective results compared with other methods.

Publication History:

1. First printing, TR-2003-117, September 2003

COMBINED RATE CONTROL AND MODE DECISION OPTIMIZATION FOR MPEG-2 TRANSCODING WITH SPATIAL RESOLUTION REDUCTION

Hao-Song Kong, Anthony Vetro, and Huifang Sun

Mitsubishi Electric Research Labs
201 Broadway, Cambridge, MA 02139.
hkong@merl.com, avetro@merl.com, hsun@merl.com

ABSTRACT

This paper presents a new algorithm for MPEG-2 transcoding with spatial resolution reduction. The proposed method combines rate control and mode decision to achieve optimal transcoding performance using Lagrange multiplier algorithm. Since the proposed method incorporates motion vector mapping and mode decision into a single Lagrange multiplier formula, by minimizing the Lagrangian cost function, the optimal solution for mode decision can be obtained. The proposed transcoding scheme has demonstrated better subjective and objective results compared with other methods.

1. INTRODUCTION

There are increasing demands for video transcoding with spatial resolution reduction. Such requirements come from the heterogeneous network constraints and characteristics of terminal devices. For example, people want to have ubiquitous access to the high quality video stored in the server, but their devices may have different processing and display capabilities, such as desktop workstations, lap top PCs, PDAs and mobile phones. Therefore, the compressed video contents need to be transcoded at lower bit rate and downscaled to lower spatial resolution.

Several papers have addressed the transcoding issues specifically emphasizing on the spatial resolution reduction [1]-[4]. There are two major issues in the spatial resolution reduction transcoding. One is motion vector mapping from high resolution to low resolution. Another is mode decision making for the downscaled macroblocks. There have been many discussions on the motion vector mapping methods [5]-[7], concentrating on weighting the input motion vectors. These methods are effective for small motion videos. However, with high motion video, they do not work properly. Several papers have presented mode decision methods [3], [6]-[8]. Most of them select the mode from the original input modes by using majority-voting mechanism. Some of them use other criteria for making the mode decision, but they are limited to intra and inter decision. Recently a new mode decision method [1] has been proposed, which considered frame and field pictures. However, the main idea of the method is still a simple "Go-With-the-Majority" strategy. Therefore the resulting modes may not be optimal.

In this paper, we propose a new algorithm to make the mode decisions for the downscaled macroblocks, which takes the input motion vectors into account. In the algorithm, we make the macroblock prediction in terms of the input motion vectors with different modes. For each candidate mode, the corresponding input motion vectors are averaged to form the candidate motion vector. Then, a Lagrangian cost function for the mode decision is minimized in a rate-distortion sense. The mode with the minimum cost value is selected as the transcoding mode. The proposed transcoder is compared to a cascaded transcoder and a MPEG-2 encoder (TM5 model). The experimental results show that the proposed scheme outperforms these two schemes. One of the goals of this paper is to obtain an upper bound for achievable video quality, which can be served as a metric measure to reflect subjective viewing judgment.

The rest of the paper is organized as follows. In section 2, the optimization problem is formulated. Then, the mode decision scheme is presented in section 3. The transcoder architecture is illustrated in section 4. Experimental results are shown in section 5. Conclusions are given in section 6.

2. PROBLEM FORMULATION

In order to achieve optimal transcoding performance, it is important to realize that coding modes should be determined jointly with rate control because the best coding mode depends on the operating point for the bit rate [9]-[11]. The optimization must choose the most efficient coding mode for each macroblock in the rate-distortion (R-D) sense. This task is complicated by the fact that the various coding modes show varying efficiency at different bit rates. Intuitively, an improved R-D performance is expected if the modes could be applied judiciously to different macroblocks. We use the Lagrange multiplier method to make the macroblock-based mode decision by minimizing the Lagrangian cost function:

$$J_i(\lambda, M_k, q_i) = \min_{M_k}\{D_i(M_k, q_i) + \lambda R_i(M_k, q_i)\} \quad (1)$$

where the M_k is varied over the coding mode set (7 modes for P picture, 11 modes for B picture), q_i is the quantizer step size $\in \{q_{1,} q_2, ... q_N\}$, $\forall i = 1,...N$, and N is the macroblock numbers of each frame. The mode decision and quantization parameter that are assigned to the macroblocks generate different R-D characteristics. Our goal is to determine a set of quantizer step sizes for all macroblocks of each frame, such that the total distortion D is minimized and the total number of bits R complies with the target budget imposed by the constraint, $R_{picture}$. The constrained problem is then formulated as:

$$\min D \quad \textit{subject to} \quad R < R_{picture} \quad (2)$$

with D and R given by

$$D = \sum_{i=1}^{N} d_i(q_i) \qquad R = \sum_{i=1}^{N} r_i(q_i) \quad (3)$$

For a particular value of the Lagrange multiplier, λ, if a set of $q_i^*(\lambda)$ minimizes the following expression:

$$\min_{q_i}\{d_i(q_i) + \lambda r_i(q_i)\} \;\; \forall i = 1,...N \quad (4)$$

then this set of $q_i^*(\lambda)$ corresponds to an optimal solution to equation (2).

To find the optimal operating point on the R-D curve, we searched for an optimal slope, λ^*, in equation (4), such that, $R(\lambda^*) < R_{picture}$. A fast convex search algorithm [12], [13] has been implemented in this paper and is outlined in the following steps.

Step-1) Initialize two values of λ, λ_1 and λ_2, with $\lambda_1 < \lambda_2$ which satisfies the relation:

$$\sum_{i=1}^{N} R_i(\lambda_1) < R_{picture} < \sum_{i=1}^{N} R_i(\lambda_2)$$

Step-2) $\lambda_{next} = \frac{\lambda_1 + \lambda_2}{2}$.

Step-3) Substitute λ_1 and λ_{next} into expression (4), minimize the expression and derive $q_i^*(\lambda_1)$ and $q_i^*(\lambda_{next})$, $\forall i = 1,...N$, respectively.

Step-4) If $[R(\lambda_1) - R_{picture}][R(\lambda_{next}) - R_{picture}] < 0$ substitute λ_2 by λ_{next}, otherwise substitute λ_1 by λ_{next}.

Step-5) If $\left|\frac{R(\lambda_{next}) - R_{picture}}{R_{picture}}\right| < \varepsilon$, where ε is a preset small positive number, the optimal slope λ^* is found and q_i^*, $\forall i = 1,...N$ is the optimal quantizer step size for each macroblock; else, go to Step-2.

3. MODE DECISION

Since the optimal q_i^*, $\forall i = 1,...N$ is derived for each macroblock subject to the constraints of $\sum_{i=1}^{N} r_i(q_i^*) < R_{picture}$, equation (1) becomes:

$$J_i(\lambda, M_k \mid q_i) = \min_{M_k}\{D_i(M_k \mid q_i) + \lambda R_i(M_k \mid q_i)\} \quad (5)$$

The minimum of the Lagrangian rate distortion function is now obtained by setting its derivative to zero, i.e.,

$$\frac{\partial J}{\partial R} = \frac{\partial D}{\partial R} + \lambda = 0$$

which yields

$$\lambda = -\frac{\partial D}{\partial R}$$

Since q_i is given to each macroblock, therefore λ can be solved by the following approximation:

$$\lambda = -\frac{\partial D}{\partial R} \approx -\frac{\Delta D}{\Delta R} = \frac{D(q) - D(q-1)}{R(q-1) - R(q)} \quad (6)$$

For each candidate mode, the cost function (5) is calculated and the one that has the minimum cost is selected as the coding mode for the macroblock.

4. PROPOSED ARCHITECTURE

For the convenience of discussion, only I- and P-frames are demonstrated in this paper. Intra-frames do not contain any motion information as they are coded independently. There are seven modes for P-frame coding.

Four modes {intra, no MC (motion compensation), MC frame, MC field} are considered as estimation mode for transcoding. Both I- and P-frames are rate-distortion optimized. For each macroblock, the coding mode as well as quantization parameter is determined such that the Lagrangian R-D functional *J* is minimized. The proposed transcoder architecture is shown in Figure 1.

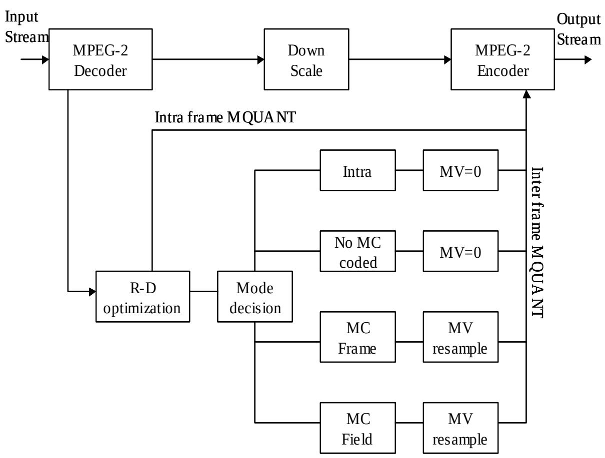

Figure 1. Rate control and mode decision optimized transcoding architecture for I and P frames only.

The block diagram in the figure illustrates the proposed transcoding scheme. The input video stream with higher resolution is first decoded using the standard MPEG-2 decoder. Motion vectors are extracted for the purpose of re-sampled motion vector and mode decision for the downscaled video. If the input frame is an I-frame, the decoded image is downscaled to the desired resolution and the rate-distortion optimization is performed to generate the quantization step size for each macroblock. The standard MPEG-2 encoder uses these optimal quantization parameters to encoder this frame as an I-frame in the downscaled video. If the input frame is a P-frame, except for the rate-distortion optimization operation, a mode decision procedure is performed based on the input motion vectors and different coding mode. The mode with the minimum cost in terms of rate-distortion is selected as the coding mode. The corresponding motion vector is re-sampled for the downscaled video.

5. EXPERIMENTAL RESULTS

The performance of the proposed rate control and mode decision optimized transcoding scheme is evaluated using the ‘Sprinkle’ test sequence for high level video and ‘Mobile & Calendar’ test sequence for main level video. In order to show the transcoding efficiency of the proposed method, we compared our transcoding results with those encoded by the cascade transcoder (with full decoding, motion estimation and re-encoding) and MPEG-2 TM5 encoder (direct encoding). The original ‘Sprinkle’ sequence is encoded at 30Mbps and its resolution is 1920x1080. The test sequence ‘Mobile & calendar’ is originally encoded at 6Mbps and has a resolution of 704x480. All experiments shown in this paper were downscaled by two.

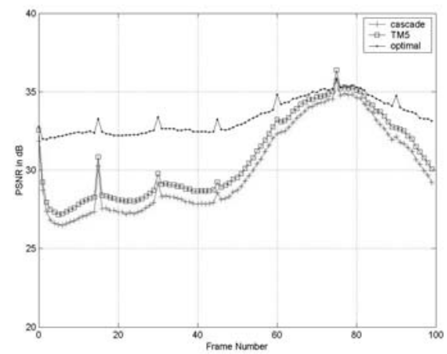

Figure 2. PSNR comparison among three coding schemes on ‘Sprinkle’ sequence.

Figure 2 shows the PSNR comparison among three coding schemes on ‘Sprinkle’ video sequence. The bit stream is transcoded from 30Mbps to 9Mbps using a cascaded transcoder (decoding, filtering and down-sampling, motion estimation and re-encoding) and our proposed transcoder with optimized rate control and mode decision algorithm. For reference, the video sequence is also encoded using MPEG-2 TM5 encoder at the same bit rate (the original YUV sequence is down-sampled by 2 and directly encoded).

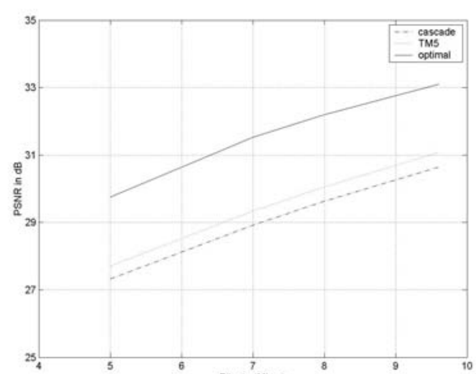

Figure 3. PSNR vs. bit rate of ‘Sprinkle’ sequence.

As shown in Figure 2, our transcoding scheme achieves higher signal to noise ratio gain compared to the other two schemes. Similar results have been obtained from different conversion ratios. Figure 3 shows the results of three peak signal-to-noise ratios in which our optimally transcoded video demonstrates the best quality among the three schemes.

Table 1 shows the experimental results of 'Mobile & Calendar' sequence. The proposed transcoding scheme not only has higher PSNR but also has lower rates than the other two schemes. These results are also visually shown in Figure 4.

Table 1. PSNR Comparison for Mobile sequence.

Mobile 100 frames	Cascaded transcoder	MPEG-2 TM5	Proposed Method
Target bit rate (Mbps)	PSNR (dB) Bytes	PSNR (dB) Bytes	PSNR (dB) Bytes
1	24.7816 416703	25.4464 416863	26.2404 401315
2	25.8779 833444	26.9276 833320	27.8646 823089
3	26.927 1249902	28.3058 1250007	29.337 1099985
4	27.6807 1666792	29.4685 1666732	30.6176 1367364
5	28.0028 2083485	30.1433 2083463	31.6768 1626213

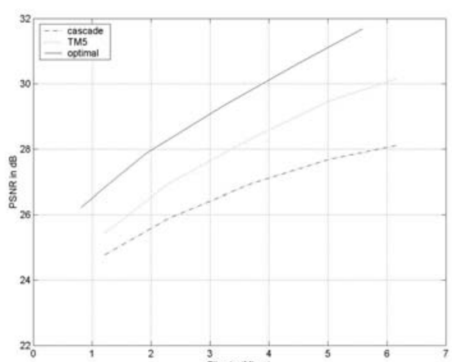

Figure 4. PSNR vs. bit rate of 'Mobile & Calendar' sequence.

6. CONCLUSIONS

In this paper, an optimal rate control and mode decision scheme has been presented for obtaining an upper bound to achievable performance for MPEG-2 spatial transcoding. The constrained optimization problem was first formulated in an operational rate-distortion sense and solved by the Lagrange multiplier method. Then, the transcoder architecture using the optimal coding scheme was presented and followed by experimental results. The better video quality obtained by the proposed transcoding method has proved that the compressed video can be used as a benchmark for evaluating the transcoding performance.

REFERENCES

[1] J. Xin, M. T. Sun, B. S. Choi, and K. W. Chun, "An HDTV-to-SDTV spatial transcoder," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 12, No. 11, pp. 998-1008, Nov. 2002.

[2] P. Yin, A. Vetro, B. Liu, and H. Sun, "Drift compensation for reduced spatial resolution transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 12, No. 11, pp. 1009-1020, Nov. 2002.

[3] T. Shanableh, and M. Ghanbari, "Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats," *IEEE Trans. Multimedia*, Vol. 2, No. 2, pp. 101-110, June 2000.

[4] G. Shen, B. Zeng, Y.-Q. Zhang, and M. L. Liou, "Transcoder with arbitrarily resizing capability," *IEEE ISCAS 2001*, pp. 25-28.

[5] B. Shen, I. K. Sethi, and B. Vasudev, "Adaptive motion-vector resampling for compressed video downscaling," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 9, No. 6, pp. 929-936, Sep 1999.

[6] M. Sugano, Y. Nakajima, H. Yanagihara, and A. Yoneyama, "An efficient transcoding from MPEG-2 to MPEG-1," *IEEE ICIP 2001*, pp. 417-420.

[7] M. R. Hashemi, L. Winger, and S. Panchanathan, "Compressed domain motion vector resampling for downscaling of MPEG video," *IEEE ICIP 1999*.

[8] N. Bjork, and C. Christopoulos, "Transcoder architectures for video coding," *IEEE Trans. Consumer Electronics*, Vol. 44, No. 1, pp. 88-98, Feb 1998.

[9] H. Sun, W. Kwok, M. Chien, and C. H. Ju, "MPEG coding performance improvement by jointly optimizing coding mode decisions and rate control," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 7, No. 3, pp. 449-458, June 1997.

[10] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 6, No. 2, pp. 182-190, April 1996.

[11] P.A.A. Assuncao, and M. Ghanbari, "Optimal transcoding of compressed video," *IEEE ICIP 1997*.

[12] Y. Shoham, and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustics Speech and Signal Processing*, Vol. 36, No. 9, pp. 1445-1453, September 1988.

[13] K. Ramchandran, and M. Vetterli, "Best wavelet packet bases in a rate-distortion sens," *IEEE Trans. Image Processing*, Vol. 2, No. 2, pp. 160-175, April 1993.