A Method for Parameter Translation in Direct MPEG-2 to H.264 Video Transcoding with Downscaling

Stanislav Očovaj, Branislav Pjetar, Željko Lukač

Abstract **— In this paper, a method for translation of coding parameters such as macroblock type, quantizer scale and motion vectors from MPEG-2 to H.264 video format is presented. The presented method can be used for direct MPEG-2 to H.264 video format transcoding, where no additional refinement is used. The paper is mainly focused on transcoding of spatially downscaled video sequences, but it can also be used in situations where no downscaling is performed.**

Keywords **— Digital Television, H.264, MPEG-2, Video Transcoding**

I. Introduction

MPEG-2 is currently one of the most used standards for video compression and it is used in many standards for digital television and digital video storage such as DVB (Digital Video Broadcast) and DVD (Digital Video Disc).

In recent years, many new standards for video compression were published, including ISO MPEG-4 and ITU H.264 (also included in the MPEG-4 standard as part 10). Compression tools defined in new standards provide much higher compression ratios than MPEG-2, while preserving the same picture quality.

In order to provide interoperability between systems based on MPEG-2, and systems based on new standards, video bit stream transcoding is required. One way to do the video transcoding is to implement the complete MPEG-2 decoder and H.264 encoder. This approach is the most efficient in terms of picture quality and compression gain, but is also the most computationally demanding. The other possibility is to re-use some of the coding parameters from the MPEG-2 video bit stream in order to decrease the complexity of the encoder [3]. However, if the coded video sequence must be spatially scaled because the target system uses different resolution, coding parameters can not be directly used. This paper presents a method for parameter translation that can be used in situations where picture scaling is required. Section II of the paper describes the translation of parameters such as macroblock type and quantizer scale. Section III describes translation of motion vectors in detail. In section IV, the testing results are presented.

This work is supported in part by the Serbian Ministry of Science and Environmental Protection under Grant TR-6136B.

S. Očovaj, Fakultet tehničkih nauka, Novi Sad, Srbija (telefon: 381-21-4801-128; e-mail: Stanislav.Ocovaj@Micronas.com).

B. Pjetar, MicronasNIT, Novi Sad, Srbija; (e-mail: Branislav.Pjetar@Micronas.com).

Ž. Lukač, Fakultet tehničkih nauka, Novi Sad, Srbija (e-mail: Zeljko.Lukac@Micronas.com).

II. Macroblock Parameter Translation

When translating macroblock parameters from the original picture to the resulting picture scaled horizontally and vertically by a factor 2, four original macroblocks are merged into one resulting macroblock. That means that the parameters from four original macroblocks are used to calculate the parameters for each resulting macroblock.

However, when video sequence in HD resolution (1280x720, for example) is transcoded to the sequence in SD resolution (720x576), horizontal and vertical scaling ratios have non-integer values. This means that non-integer number of original macroblocks must be translated into one resulting macroblock. To overcome this problem, the parameters from each macroblock that covered (completely or partially) by resulting macroblock are used for translation and a weight factor corresponding to the percentage of the covered macroblock area is assigned to each original macroblock. Weight factors associated with original macroblocks are calculated as follows.

First, the horizontal and vertical indices of macroblocks that are covered by resulting macroblock are calculated. Horizontal index X_{min} of the left-most original macroblock is calculated as:

$$X_{\min} = trunc(X \cdot SR_x) \quad (1)$$

Where X is horizontal index of the resulting macroblock, SR_x is horizontal scaling ratio and trunc() is a function that removes the fractional part of the number. Horizontal index X_{max} of the right-most original macroblock is calculated as:

$$X_{\max} = trunc((X+1) \cdot SR_x) \quad (2)$$

If the fractional part of $(X+1)SR_x$ is 0, X_{max} is subtracted by 1. Weights are calculated separately for horizontal and vertical direction. Horizontal weight associated with the left-most original macroblock is calculated as:

$$W_0 = X_{\min} - X \cdot SR_x + 1 \quad (3)$$

Horizontal weight associated with the right-most original macroblock is calculated as:

$$W_{X_{\max}-X_{\min}} = (X+1) \cdot SR_x - X_{\max} \quad (4)$$

Obviously, all other horizontal weights must have the value 1.

Vertical indices and weights are calculated similarly. The resulting weight factors for original macroblocks are calculated by multiplying the appropriate horizontal and vertical weight factors.

Macroblock mode is determined as follows. If any of the MPEG-2 macroblocks are intra coded, the resulting macroblock will be intra coded. If all MPEG-2 macroblocks are skipped, the resulting macroblock will be skipped. In all other cases, the resulting macroblock will be inter coded. For inter coded macroblocks, prediction direction is determined based on the prediction direction of the MPEG-2 macroblocks and weight factors. If most of the MPEG-2 macroblocks use forward prediction, the resulting macroblock will use forward prediction and if most of the MPEG2 macroblocks use backward prediction, the resulting macroblock will use backward prediction.

Quantization parameter of the resulting macroblock is calculated as the weighted average of quantization parameters of the MPEG-2 macroblocks.

III. Motion Vector Translation

Motion vector translation is the key element in direct video transcoding, because of its impact on the bit rate of the resulting video bit stream. Its purpose is to map multiple original motion vectors into one resulting motion vector that represents motion of the resulting macroblock as good as possible. In order to translate the motion vectors, the following 3 operations must be performed:

- Motion correction
- Mapping
- Scaling

All motion vector mapping methods assume that all the original motion vectors are of the same kind, i.e. that all of them are in the same format (field or frame) and that all of them point to the same reference picture.

Since only field-to-field transcoding is considered in this paper, all motion vectors will be in the field format. However, MPEG-2 assigns additional field selection bit to each field motion vector, which tells if the motion vector points to the top or bottom reference field. Motion vectors from different original macroblocks may point to the different reference fields. Reference field for the resulting macroblock is chosen based on the reference fields of the original macroblocks and weight factors. If most of the MPEG-2 macroblocks use top field as a reference, the resulting macroblock will also use top field as a reference. Otherwise, bottom field will be used.

When the reference field is chosen, motion vectors that point to the field of the opposite parity must be corrected prior to mapping. If a field of the same parity as the coded field is chosen as a reference, motion vectors that point to the reference field of the opposite parity must be multiplied by the factor C_1/C_2. Similarly, if a field of the opposite parity than the coded field is chosen as a reference, motion vectors that point to the reference field of the same parity must be multiplied by the factor C_2/C_1. Motion vector scaling factors C_1 and C_2 are calculated under the assumption that motion vectors represent motion of the objects in the picture, and that this motion is linear with constant speed. For example, if the top field (Field 0 in Fig. 1.) is chosen as a reference for the top of a P frame (Field 6 in Fig. 1.), motion vectors that point to the bottom reference field (Field 1 in Fig. 1.) must be multiplied by a factor 6/5.

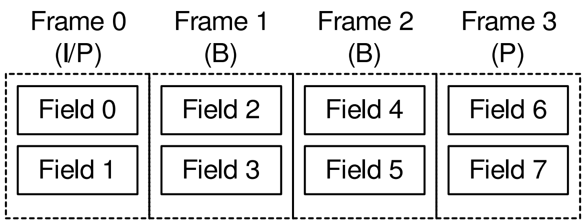

Fig. 1. A sequence of 4 frames with 2 consecutive B frames

Tables 1, 2 and 3 summarize the values of C_1 and C_2 for P, first B and second B frames, respectively (it is assumed that two consecutive B frames are coded between the two reference frames).

TABLE 1: CORRECTION FACTORS FOR P FRAMES.

Field	C_1	C_2
First	6	5
Second	6	1

TABLE 2: CORRECTION FACTORS FOR FIRST B FRAMES.

Field	Forward reference		Backward reference	
	C_1	C_2	C_1	C_2
First	2	1	4	5
Second	2	3	4	3

TABLE 3: CORRECTION FACTORS FOR SECOND B FRAMES.

Field	Forward reference		Backward reference	
	C_1	C_2	C_1	C_2
First	4	3	2	3
Second	4	5	2	1

For mapping of motion vectors, four different methods were used:

Average: 16x16 macroblock partition with motion vector derived as a weighted average of the MPEG-2 motion vectors

Dominant: 16x16 macroblock partition with motion vector from the dominant MPEG-2 macroblock (macroblock with the largest weight factor)

16x8 Average: 16x8 macroblock partition with two motion vectors derived as a weighted average of the MPEG-2 motion vectors

Median: 16x16 macroblock partition with motion vector derived as a weighted median of MPEG-2 motion vectors (the vector with the lowest distance to other MPEG2 motion vectors).

Finally, the resulting motion vector (or a pair of motion vectors in the case of 16x8 average) is scaled by the picture size scaling ratio. Also, motion vectors must be multiplied by two in order to convert them from half-pixel precision (used in MPEG-2) to quarter-pixel precision (used in H.264).

IV. Results

Motion vector translation methods presented in this paper were tested using two test sequences, "Pedestrian Zone" and "Park Run" in 1920x1080 resolution. Both sequences were coded in MPEG2 format using the reference encoder. The GOP (Group of Pictures) size used was 24 for both sequences, and two consecutive B frames were coded between two I/P frames. Each sequence was encoded at bit rate that gives the same PSNR as the sequence in 720x576 resolution coded at 6Mbit/s. The sequences were then decoded using the reference MPEG2 decoder and scaled to 720x576 resolution using AviSynth. Coding parameters were extracted from the MPEG2 bit stream and stored into a textual file. The downscaled sequences were encoded with modified H.264 encoder at different bit rates. Fig. 2. and Fig. 3. show the resulting rate-distortion curves for the two test sequences. PSNR of the MPEG-2 coded sequence is also included for reference.

V. Conclusion

In this paper, a method for translation of coding parameters such as macroblock type, quantizer scale and motion vectors from MPEG-2 to H.264 video format was presented. The presented method can be used in direct MPEG-2 to H.264 video transcoding when picture downscaling by an arbitrary ratio is required. The results show that the motion vector translation method named "median" gives the best rate-distortion performance. The quality of the sequences coded using "median" translation method without any motion vector refinement is comparable to the original MPEG-2 sequence.

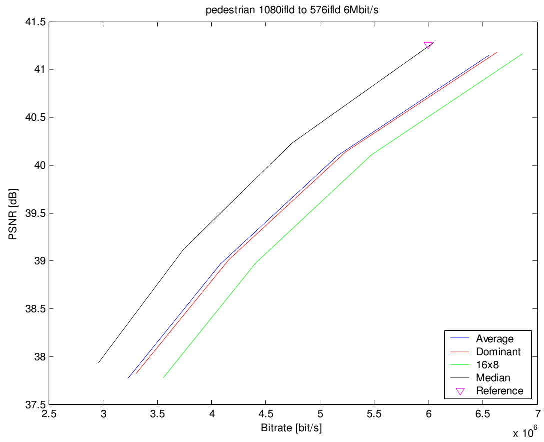

Fig. 2. Rate-distortion curve for "Pedestrian Zone" sequence

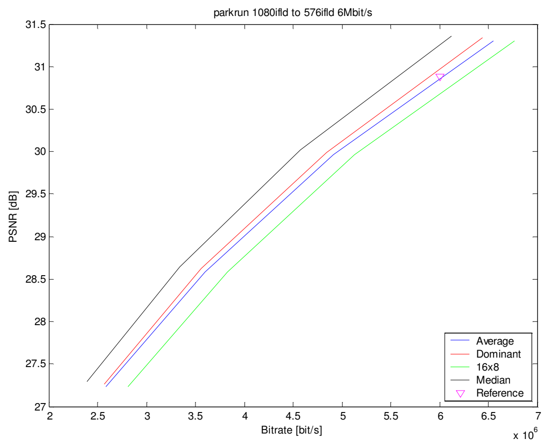

Fig. 3. Rate-distortion curve for "Park Run" sequence

References

[1] ITU-T Recommendation H.262 "Information technology – Generic coding of moving pictures and associated audio information: Video", International Telecommunication Union, 2000

[2] ITU-T Recommendation H.264 "Advanced video coding for generic audiovisual services", International Telecommunication Union, 2005

[3] Ishfaq Ahmad, Xiaohui Wei, Yu Sun and Ya-Qin Zhang "Video Transcoding: An Overview of Various Techniques and Research Issues", IEEE Transactions On Multimedia, Vol. 7, No. 5, October 2005

[4] AviSynth, http://avisynth.org/