MODE MAPPING METHOD FOR H.264/AVC SPATIAL DOWNSCALING TRANSCODING

Peng Zhang[1], Yan Lu[2], Qingming Huang[3], Wen Gao[1]
[1] Institute of Computing Technology, Chinese Academy of Science, Beijing, 100080, China
[2] Microsoft Research Asia, Beijing, 100080, China
[3] Graduate School, Chinese Academy of Sciences, Beijing, 100080, China
{peng.zhang, qmhuang, wgao}@jdl.ac.cn; t-yanlu@microsoft.com

ABSTRACT

Transcoding is a very effective method for universal media adaptation. Spatial resolution downscaling is one of the most popular transcoding applications. Although a number of transcoding methods have been proposed in the past years, they are not very suitable for the most up-to-date H.264/AVC standard due to its many new features. In this paper, a coding mode mapping method is proposed for H.264/AVC spatial downscaling transcoding. Different from the traditional schemes, the proposed method is focused on the mode decision part, which plays a key role in H.264/AVC video encoding. The mapping scheme is exploited based on the statistics and analysis that reveals the inherent relationship between the frames having the same content but with different spatial resolutions.

1. INTRODUCTION

Recently, Universal Multimedia Access (UMA) has become a very promising research topic with the main objective of enabling clients with limited communication, processing, storage and display capabilities to access rich multimedia content. Due to the expansion and diversity of multimedia applications and the current communication infrastructure comprised of different underlying networks and protocols, there has been a growing need to perform video transcoding to accommodate network limitations and terminal constraints [1]. Spatial resolution downscaling transcoding is revealed to be one of the most popular transcoding applications, because it solves the common problem of mobile terminals with small screen to, for example, access DVD or other large spatial resolution multimedia streams.

In the past years, several papers have addressed the problem of spatial downscaling transcoding [2]~[7]. In the literature, there are two major transcoding architectures: cascaded pixel-domain transcoder (CPDT) [1, 8] and DCT-domain transcoder (DDT) [9]. Due to its flexibility and drift free, CPDT is mostly adopted in spatial downscaling transcoding. The previous researches mainly dealt with motion vectors mapping [4] or motion vectors refinement [5], DCT domain down conversion [2], drift error [6], and error resilience [7]. All these methods are based on the standards of MPEG and H.26x series. Motion compensation in the standard such as MPEG-2 and H.263 is based on blocks with constant block size. Therefore, motion vector manipulation and DCT-domain down-sampling for transcoding based on these standards are very easy to be implemented.

However, the traditional transcoding schemes may be unsuitable for the most up-to-date H.264/AVC standard [10], because it employs variable block-size motion compensation and directional spatial prediction for intra coding so as to achieve high coding efficiency. To select the best coding mode, the H.264/AVC encoder usually try every possible mode for each macroblock. This procedure is computational complex, and therefore against the general purpose of transcoding. As a consequence, how to fast and efficiently estimate the mode of each macroblock is the new problem that other standards never encountered. To tackle this problem, we propose a mode-mapping method for spatial down-scaling transcoding, which explores and utilizes the pre-encoded mode information in the H.264/AVC bitstreams.

The rest of this paper is organized as follows. In Section 2, firstly we analyze the problem of H.264/AVC transcoding, and therefore we propose a new transcoding architecture. In section 3, we present the detailed algorithms of mode mapping, including both the simple mode mapping and the improved method by taken into account motion vectors. Experimental results showing the effectiveness of the proposed mode-mapping method is presented in section 4. Finally, Section 5 concludes this paper.

2. TRANSCODING ARCHITECTURE

Compared to prior video coding methods, H.264/AVC employs many new features, such as variable block-size motion compensation with small block sizes, quarter sample accurate motion compensation, motion vectors over picture boundaries, multiple reference picture motion compensation, weighted prediction, improved skipped and direct motion reference, directional spatial prediction for intra coding [11]. These entire enable H.264/AVC achieve high coding efficiency at the cost of computational

0-7803-8554-3/04/$20.00 ©2004 IEEE.

complexity. For example, H.264/AVC employs directional spatial prediction for intra coding and variable block-size motion compensation with small block sizes for inter coding. This means that each macroblock of Intra (I) frames can either select 16x16 or 4x4 mode. If 4x4 mode is selected, there are 9 direction options to be chosen further. For Predictive-coded (P) frames, in addition to the previous intra modes, more inter modes have been defined. Concretely, the macroblock can be partitioned into 16x8, 8x16 and 8x8 blocks for motion compensation, and 8x8 block can be further partitioned into 8x4, 4x8 and 4x4 sub-blocks. And also, H.264/AVC supports skip mode for P frames. To select the best mode for each block, H.264/AVC employs Rate-Distortion Optimization (RDO) scheme expressed as follows [13]:

$$J = D + \lambda R , \qquad (1)$$

where D and R denote the distortion and rate, respectively, and λ is the Lagarange parameter. The target of RDO is to select the mode that results in the minimum coding cost J.

Therefore, the encoder must try every candidate mode in mode selection. The quarter-sample resolution motion estimation makes it much more computational complex. Intuitively, if we can achieve the mode information of each macroblock without looping every candidate mode, the complexity can be greatly reduced. Fig.1 illustrates the comparison of time cost of encoding foreman sequence with and without the mode information of each macroblock. Although the prediction direction for Intra 4x4 (I4x4) mode and the sub-mode for Predictive 8x8 (P8x8) mode still have to be further searched, the encoding process with coding mode pre-selected can still save about 50% time cost.

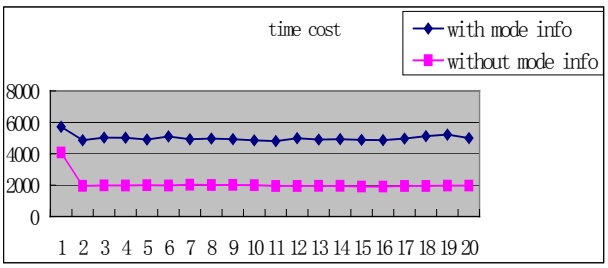

Fig.1 Time cost comparison of encoding foreman sequence with and without the mode information of each macroblock.

The general problem of spatial downscaling is shown in Fig. 2, where MB1, MB2, MB3 and MB4 are the four original macroblocks, and MB is the corresponding downsized macroblock. There is some correlation between the four modes of the original macroblocks and the mode of the corresponding downsized macroblock. Fig. 3 shows an example of the mode distribution of I frame and P frame in foreman CIF and QCIF sequences, respectively. The number in the figure represents the mode of its located macroblock, with the white and colored grids representing the inter mode and intra mode, respectively. Based on the statistics of the relationship, we propose a mode mapping method to estimate the mode of the downsized macroblock in terms of the information of the pre-encoded four macroblocks.

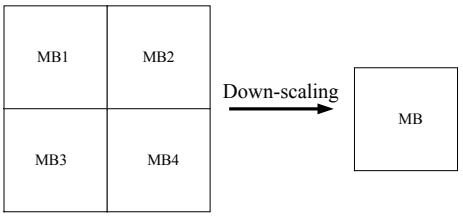

Fig. 2 Spatial resolution down-conversion

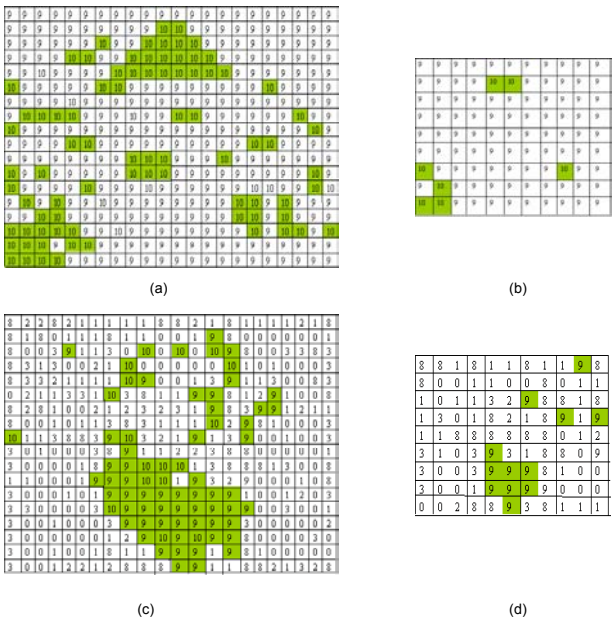

Fig. 3. Coding model distribution for I frame and P frame of foreman sequence. (a) I frame of the 1st picture, CIF; (b) I frame of the 1st picture, QCIF; (c) P frame of the 2nd picture, CIF; and (d) P frame of the 2nd picture, QCIF.

The architecture of the proposed transcoder is shown in Fig. 4. Basically, the proposed scheme is a kind of CPDT transcoder. After Variable Length Decoding (VLD) is performed, the coding mode information is extracted as the input of mode mapping so that the mode of the down conversed macroblock can be estimated. The estimated mode will instruct the encoding process. We denote the first scheme as simple mode mapping. In this scheme, only the coding mode of the input MB is used. In other words, we only estimate the MB mode while the sub-mode of P8x8 mode and the predictive direction for I4x4 are not concerned.

Then, in the second scheme, we take into account motion vectors to further reduce the complexity of the transcoder. In this scheme, motion vectors are considered

as one major factor for mode decision, in which the sub-mode of P8x8 mode can be estimated as well.

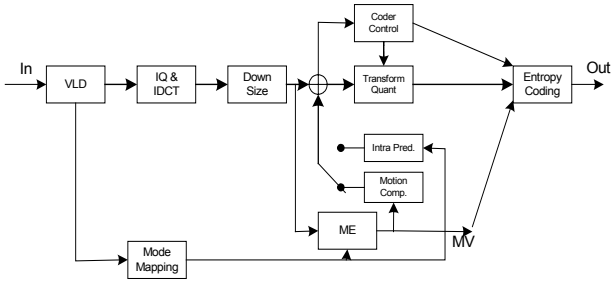

Fig. 4. Transcoding architecture with mode mapping.

We take a cascaded decoder-encoder scheme as comparison to the two proposed mode-mapping methods. For CIF-to-QCIF downscaling, 16x16 macroblock can be directly mapped into 8x8 blocks, as indicated in [12], and in addition, 8x8 mode can be mapped into 4x4 sub-mode. Fig. 5 shows the direct mode-mapping scheme, in which a 16x16 macroblock coding mode is mapped into an 8x8 sub-mode. The decision of the final coding mode of a macroblock will be further explained in the next section.

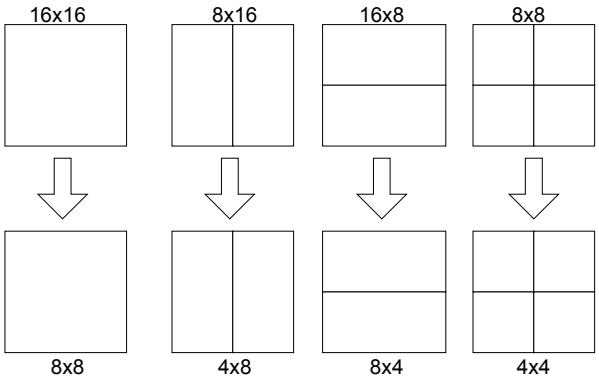

Fig. 5. Illustration of the direct coding mode mapping.

3. MODE-MAPPING ALGORITHMS

3.1 Simple Mode Mapping Method

As mentioned above, the simple mode-mapping scheme only uses the mode information of the four pre-encoded macroblocks to estimate the mode of the corresponding macroblock in the transcoded bitstream. Concretely, this mode decision algorithm is composed of two parts for both I and P frames. The following paragraphs first present the mode decision of I frames transcoding, and then present P frames in detail.

For the I frame coding, I4x4 mode is usually used for coding the details in the frame. Therefore, if there are more than one macroblocks of the four pre-encoded macroblocks coded with I4x4 mode, the mode of the corresponding macroblock in the downsized frame is selected as I4x4; it is selected as Intra 16x16 (I16x16).

For the P frame coding, there are much more coding modes that must be taken into account. The detailed implementation of the proposed algorithm is presented as bellow:

- If the coding modes of the four pre-encoded macroblocks are all Intra-16x16, the mode of the corresponding macroblock in the downsized frame is also selected as I16x16;
- If more than one MBs are intra mode, the mode of the corresponding MB in the downsized frame is decided as I4x4;
- If more than three of the four pre-encoded macroblocks correspond to Predictive 16x16 (P16x16) or skip mode, the mode of the corresponding macroblock in the downsized frame is selected as P16x16; otherwise, it is selected as P8x8. If P8x8 mode is selected, the four sub-modes are decided according to the four modes of their corresponding MBs the same as direct-8x8 mode decision algorithm.

3.2 Mode Mapping with Motion Vectors

In the above simple mode-mapping scheme, Predictive 16x8 (P16x8) and Predictive 8x16 (P8x16) are not utilized. Some macroblocks with P8x8 modes might also be processed in RDO. To improve the previous simple mode-mapping method, motion vectors are taken into account, and the sub-modes for the P8x8 marcoblock are estimated as well.

The described algorithm performs the same processes as the previous simple mode mappoing, except that P8x8 mode is selected. If P8x8 mode is selected, then the distances of motion vectors of the input four macroblocks are calculated. To make the mode-mapping algorithm more efficient, a bottom-up block-merging algorithm similar to the method proposed in [14] is applied as follows:

If all of the distances of motion vectors between the neighboring macroblocks among {MB1, MB2, MB3, MB4} are less than a threshold *TH*, Inter 16x16 mode is selected; else if both the distance between MB1 and MB3 and the distance between MB2 and MB4 are less than *TH*, Inter 16x8 mode is selected; else Inter 8x8 is selected; else Inter 8x16 mode is selected;

4. EXPERIMENTAL RESULTS

In this section, we present some experimental results to compare the different mode decision methods in H.264/AVC video transcoding. To remove the effects of different motion vector manipulation techniques and the DCT-domain downsampling, we employ the cascaded

decoder-encoder architecture. The experiments are performed with the H.264/AVC reference software JM 7.3. The hardware platform is a PC with the processor of Intel 2.0 GHz, 256 MB RAM. In the experiments, the input is the bitstream with CIF format at 1024kbps, and the output is the bitstream with QCIF format at various bit rates from 64kbps to 384kbps. Sequences of Foreman, Mobile and Mother&Daughter are tested, respectively. For simplicity, B frames are not considered in the experiments.

Fig.6 shows the RD curves of the four transcoders. The quality of the cascaded decoder-encoder with RDO (EncDec) is the best, followed by the mode mapping with motion vectors (MapMV) and the simple mode mapping (SimMap). The direct downscaling scheme (DirMap) has the worst performance. Furthermore, at lower bit rate, our mode mapping schemes perform much better than the direct 8x8 downscaling. Although the cascaded decoder-encoder with RDO outperforms the proposed mode mapping schemes, it has much higher computational complexity than ours.

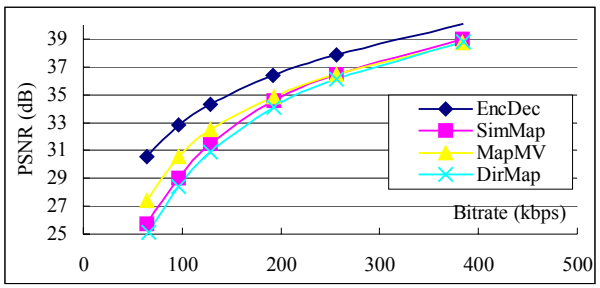

Fig. 6. RD curves of the different mode decision schemes.

5. CONCLUSION

In this paper, a mode mapping method has been presented for H.264/AVC spatial downscaling transcoding. Different from the traditional transcoding methods applied on the previous standards, the proposed method has been focused on mode decision due to its important role in H.264/AVC video encoding and transcoding. The mapping scheme is exploited based on the statistics that reveals the inherent relation between the frames with different spatial resolutions but having the same content. How to further improve the performance at low bitrate still remains a future work.

6. ACKNOWLEDGEMENT

This research has been partially supported by NSFC under contract No. 60333020, 863 Program of China under contract No. 2002AA118010, 973 Program of China under contract No. 2001cca03300 and Bairen Project of CAS.

REFERENCES

[1] H. Sun, W. Kwok, and J. W. Zdepski, "Architectures for MPEG compressed bitstream scaling," *IEEE Trans. Circuits Syst. for Video Technol.*, Vol. 6, No. 2, April 1996.

[2] B. Shen, I. K. Sethi and V. Bhaskaran, "Adaptive motion vector resampling for compressed video down-scaling," in *Proc. IEEE int'l Conf. Image Processing*, Oct. 1997.

[3] W. Zhu, K. H. Yang, and M. J. Beacken, "CIF-to-QCIF video bitstream down-conversion in the DCT domain," Bell labs Tech. J., vol. 3, no. 3, pp. 21-29, July-Sept. 1998.

[4] P. Yin, M.Wu, and B. Lui, "Video transcoding by reducing spatial resolution," in *Proc. IEEE Int. Conf. Image Processing*, Vol. 1, pp. 972-975, Oct. 2000.

[5] J. Youn, M. T. Sun, and C. W. Lin, "Motion vector refinement for high performance transcoding," *IEEE Trans. Multimedia*, Vol. 1, pp. 30-40, Mar. 1999.

[6] P. Yin, A. Vetro, B. Liu and H. Sun, "Drift compensation for reduced spatial resolution transcoding," *Transactions on Circuits and Systems for Video Technology*, Vol. 12, Issue 11, pp.1009-1020, November 2002

[7] A. Vetro, P. Yin, B. Liu and H. Sun, "Reduced spatial-temporal transcoding using an intra refresh technique," in Proc. IEEE Int. Symp. Circuits and Systems, Scottsdale, AZ, vol. 4, pp. 723-726, May 2002

[8] J. Youn, M.T. Sun, and J. Xin, "Video transcoding architectures for bit rate scaling of H.263 Bit Streams," *ACM Multimedia Conference*, 1999.

[9] P.A.A. Assuncao and M. Ghanbari, "A frequency-domain video transcoder for dynamic bitrate reduction of MPEG-2 bit streams," *IEEE Trans. Circuits Syst. for Video Technol.*, Vol. 8, pp. 953-967, Dec. 1998.

[10] "Draft ITU-T recommendation and final draft standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14496-10 AVC" in Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT G050, 2003

[11] T. Wiengand, G. J. Sullivan, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 13, pp. 560-576, July 2003.

[12] A. Vetro, C. Christopulos and H. Sun, "Video transcoding architectures and techniques: an overview," *IEEE Signal Processing Magazine*, Mar. 2003.

[13] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 13, pp. 560-576, July 2003.

[14] Y. K. Tu, J. F. Yang, Y. N. Shen, and M. T. Sun, "Fast Variable-Size Block motion estimation using merging procedure with an Adaptive threshold," IEEE ICME 2003.