MPEG-2 To MPEG-4 Transcoding

Wang Xing Guo[1, 2], Zheng Wei Guo[1], and Ishfaq Ahmad[1]
[1]Multimedia Technology Research Center, Hong Kong University of Science and Technology
[2]Zhe Jiang University, Hang Zhou, China, 310027

Abstract[1]

The MPEG-2 video compression standard is successfully adapted in the DVD and digital TV broadcast industries. The new MPEG-4 standard has much broader and ambitious perspective to support high and low bit-rate multimedia applications on existing and future networks. MPEG-4 aims to complement MPEG-2, not override it. Therefore, one can envisage that MPEG-2 and MPEG-4 will co-operate and coexist in a synergistic manner for a wide variety of application environments. This paper addresses the problem of transcoding MPEG-2 video content to MPEG-4 compliant video bitstream. The goal is to develop effective architectures and algorithms for efficient transcoding. We discuss two transcoding approaches: pixel domain method and compression domain method. We discuss their details to facilitate their applications in different situations. We report simulations authenticate the performance of the proposed transcoding solutions.

1. Introduction

A number of international standards have been established based on different applications, technology and era. While these standards can operate for a spectrum of applications, each is optimized for a certain class of application. Sometimes, it is not economical and feasible to make any single media sever or terminal to support all kinds of encoding and decoding. Transcoding techniques that convert one format to another, preferably in the compressed domain, solve the problem of inter-standard operability [3,4].

The ISO/IEC JTC1/SC29/WG11 working group has created a very successful international standard, ISO/IEC 13818, commonly known as MPEG-2 [1]. A broad and diverse group of industries, including computer, communications, consumer electronics, and television industries, contributed to the development of the MPEG-2 standard. Now a large variety of products based on the MPEG-2 standard are available on the market. MPEG-4, whose formal ISO/IEC designation is ISO/IEC14496 [2] was developed to fulfill the new needs of emerging multimedia applications, such as object based coding, interaction and natural and synthetic data integration.

[1] This work was supported by Research Grants Council of Hong Hong under contract # CRC98/01.EG05 and HKUST6228/99E.

The most important goal of MPEG-2 was to make the storage and transmission of digital AV material more efficient, by compressing the data. Therefore, it deals with 'frame-based video' and audio. Interaction with the content is limited to the video frame level and with its associated audio. MPEG-4 goes beyond these goals by specifying a description of digital AV scenes in the form of 'AV objects'. The new properties of MPEG-4 can be summarized as: 1) allow interactivity with each AV object; 2) integrate objects of different natures; 3) allow 'universal access' to multimedia information on a wide variety of networks; 4) provide object based high compression efficiency.

To realize interoperation between different systems and networks, some effective transcoders are desired for different multimedia applications. We develop two transcoding architectures. The first is a motion vector refining architecture that yields drift-free target video quality. The second is a compression domain transcoding architecture that aims for efficiency. The rest of this paper is organized as follows. Section 2 gives a brief overview of the conversion problems. Section 3 addresses two transcoding architectures. Section 4 discusses the transcoding distortion source and Section 5 presents the proposed mismatch MB retrieval techniques. Results and concluding remarks are presented in Section 6 and Section 7, respectively.

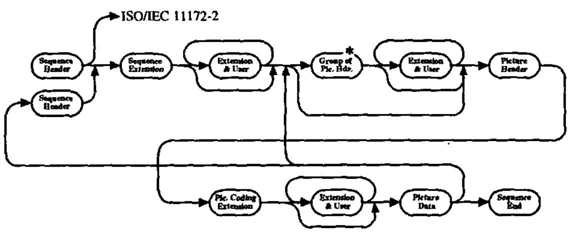
2. Transcoding Considerations for MPEG-4

In MPEG-4, rectangular video involves no shape coding process. The texture data is encoded just like MPEG-2 frame sequence, but the following important differences exist between the two standards.

- MPEG-4 encoding includes extra spatial prediction module of quantized DCT coefficient, e.g., spatial AC coefficients prediction for INTRA macroblock, for improved compression efficiency.
- The motion vectors prediction and represent formats are different, such as Mid-filtering in MPEG-4 encoding.
- The quantization process is different in the two standards.
- MPEG-4 supports new macroblock coding modes, such as direct mode, 8x8 motion compensation mode etc, which affects the encoding performance considerably. In addition it is very important that some MB encoding modes of MPEG-2 are not supported in MPEG-4. For example, Intra MB encoding mode of B-picture is only allowed in MPEG-

2. Another example is that skipped MB can only be allowed in direct mode of MPEG-4.

The data structure of MPEG-2 bitstream is depicted in Fig.1. The main part of MPEG-4 data structure is similar as Fig.1 when the visual object mode is selected as rectangular fame natural video, and the additional part is shown in Fig. 2. Obviously, MPEG-4 visual bitstream bears more elaborate header information as opposed to MPEG-2, implying that many extra parameters and marks should be reset during the transcoding process.



* After a GOP the first picture shall be an I-picture

Fig.1: High level bitstream data structure of MPEG-2.

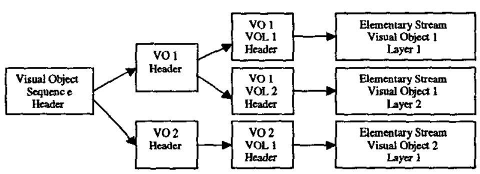


Fig. 2: The additional parts of MPEG-4 visual bitstream structure based on rectangular frame.

The following criteria are desired in transcoding:

1) The quality of target video should be high as direct MPEG-4 encoding.
2) The information of the original MPEG-2 stream should be re-used as much as possible to overcome multigenerational deterioration.
3) The operation of transcoding should be minimized to reduce the computation cost.
4) The transcoding delay and memory requirement should be minmized to meet real-time constraints.

3. Transcoding Approaches

An intuitive way for transcoding is to decode the MPEG-2 bitstream and re-encode it with MEPG-4 encoder. Obviously, this solution has very high computational, memory, and storage requirements. An alternative solution is that macroblock data, side information and other parameters are extracted during the decoding process of MPEG-2. Then these data can be processed and organized based on MPEG-4 standard. Many approaches based on the tradeoff between operation complexity and quality could be considered. We consider two efficient transcoding architectures for standard transcoding and propose techniques suitable to MPEG-4.

Fig.3 shows the Pixel domain transcoding structure, it is also named motion vector refining approach [5,6]. Better efficiency of MPEG-4 encoding always results from its rich motion compensation modes, such as the 8*8 motion mode, direct motion mode etc. Compared with MPEG-2, based on the motion vectors from original MPEG-2 bitstream, it is easy to estimate other motion vectors of new prediction mode without seriously operation increasing. The reason is that the search candidates could be reduced dramatically in a small range around the original motion vectors. This process could be called motion compensation refining function, which will virtually improve the target video quality. Because motion compensation module is involved in this architecture, no picture-drift occurs. It means that no special deterioration from error accumulation, which is often an important question produced in transcoding.

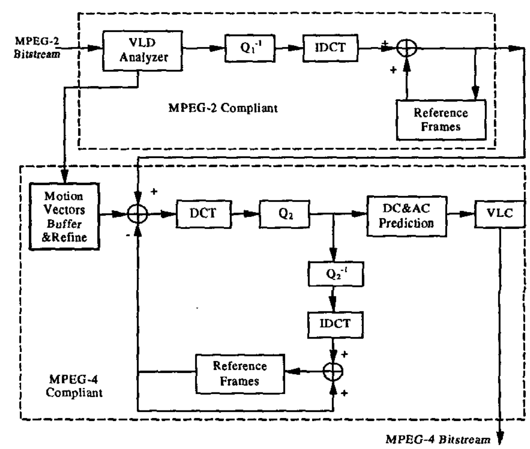


Fig. 3: Drift Free Pixel Domain Transcoding Diagram (Motion Vector Refining Transcoding)

In Fig.3 Motion Estimation is replaced by a motion vector-refining module. However, two decoding feedback loops and one encoding loop are involved that limit the reduction of complexity. In addition, since picture ordering is inevitable, the transcoding delay may be several frames, and buffer requirement is somewhat high.

Fig. 4 is the MC-DCT domain transcoding flow chart [7], indicating that no motion estimation, motion compensation, DCT and IDCT are involved. So not only the operations are mostly reduced but also the buffer need is the minimum, resulting in reduction of delay and hardware need.

No compensation exists in the transcoder, causing the picture-drift effect, the most important flaw of MC-DCT domain method. Fortunately, it is more obvious in the case of high bitrate stream, while most of MPEG-4 applications use low bitrate. Above all, the MC-DCT domain transcoding can easily be achieved in real-time and with low latency access and complexity.

In MC-DCT domain transcoding, it is important that some MB modes of MPEG-2 and MPEG-4 are different, such as no intra MB is allowed in BVOP, and skipped MB can only be allowed in the direct mode. In the absence of proper mechanism, the mismatch MBs will be not properly retrieved and may have to be reconstructed from the Inter-coded data. But that can cause serious blocking artifacts. To deal with this situation, we develop algorithms that can properly retrieve these mismatch MBs, which will be presented subsequently.

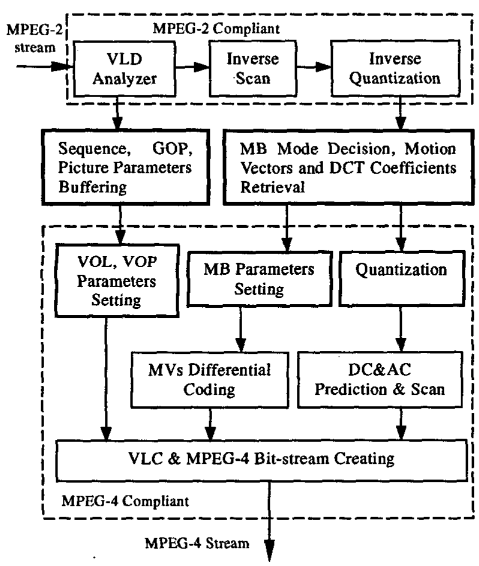


Fig. 4: MC-DCT domain transcoding.

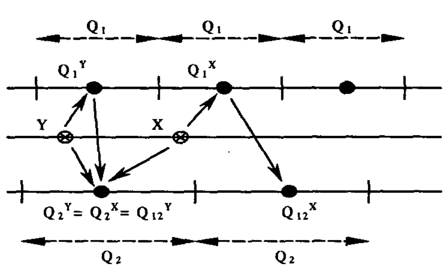


Fig. 5: Requantization error.

4. Transcoding Distortion

The transcoding distortion of IVOP is composed of three components: 1) Some nonzero DCT coefficients of the input frame that become zero after re-quantization; 2) The quantization error; 3) The re-quantization error introduced by transcoding, as illustrated in Fig. 5.

Re-quantization leads to a higher distortion than that produced by directly quantizing the original DCT coefficients using the same quantizer step size. In Fig.5 $Q_2^Y = Q_{12}^Y$, $Q_2^X \neq Q_{12}^X$

pictures of MPEG-2 bit-stream. How to process them will significantly affect the quality of the target MPEG-4 bit-stream. We present two algorithms to retrieve such MBs.

a) **Special motion estimation solution**:

- If the current MB is Intra of B-picture, search Intra MB in its temporal neighbor I-picture or P-picture.
- Force these MBs as prediction reference, set the motion vector for current MB to point it.
- Execute motion compensation in DCT domain, select the optimal prediction as the motion mode of current MB.

This method can yield good mismatch MB retrieval at the cost of a higher operation complexity.

b) **MB concealment solution**:

If the current MB is Intra MB of B-picture, select the best (see below, for optimal criteria) data of spatio-temporal neighbor MBs as its MB data.

It only conceals the mode mismatch MB rather than really retrieve the current MB data. The operation has a low complexity, but the quality is not very high. The following three criteria could be used to select the reference MB from the candidates.

- Short distance priority criterion;
- Consistent motion criterion;
- Small DCT coefficient criterion.

Small DCT coefficients mean that the corresponding prediction is more precise during the motion compensation.

6. Results

Fig. 6 is the simulation result of Stefan sequence, which was pre-encoded at 5Mbps with M=3, N=15, using MPEG-2 encoder. Then the MPEG-2 bitstream was converted into MPEG-4 bitstream using: 1) Pixel-domain architecture, 2) MC-DCT domain architecture. Concatenated MPEG-2 decoder and MPEG-4 encoder was also used to facilitate the transcoding performance comparison. As Fig. 6, pixel-domain transcoding had the same perceptual quality as the decoder-encoder cascade architecture without complex motion estimation. MC-DCT domain is an alternative architecture, although the target video quality is not very high, it can dramatically reduce the operation complexity since no motion estimation/compensation and DCT/IDCT are involved. From the PSNR profile, we can find the MC-DCT transcoding performance can be improved obviously by the proposed MB retrieval methods.

7. Conclusion

In this paper the problem of MPEG-2 to MPEG-4 bitstream transcoding was analyzed. Two efficient transcoding solutions were proposed, both of which reduce the operation complexity considerably. Pixel domain methods based on a drift-free architecture provide the ideal perceptual quality by motion parameters reusing and refining. MC-DCT domain transcoding incurs the least operation complexity, since no motion estimation/compensation, DCT/IDCT is involved. Furthermore, it also facilitates low latency and easy rate control function, making it quite suitable for some real-time applications. Two MB retrieval methods were presented to address the problem of MB mode mismatch between the two standards.

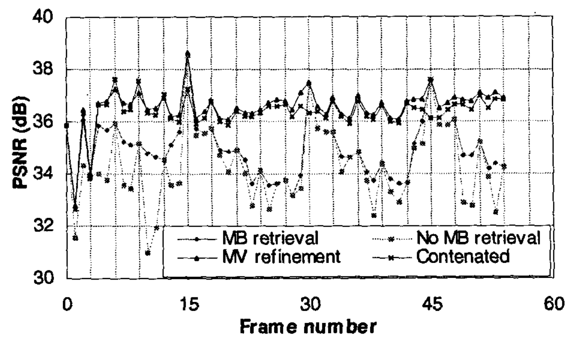


Fig. 6: Quantitative PSNR of different transcoding.

References

[1] ISO/IEC 13818-2 Amendment 2, "Generic Coding of Moving Pictures and Associated Audio", *International Organization for Standardization*, Jan. 1996.

[2] Information technology — Coding of audio-visual object: Visual," *ISO/IEC JTC1 CD 14496-2(MPEG-4)*, 1998.

[3] P.A.A. Assuncao and M.Ghanbari, "A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit streams", *IEEE Transaction on Circuits and Systems for video technology*, vol.88, pp.953~967, 1998.

[4] S.Acharya and B.Smith, "Compressed domain transcoding of MPEG", *IEEE International Conference on Multimedia Computing and Systems*, pp.295~304, 1998.

[5] Y. Jegngnam, M.T.Sun and C.W.Lin, "Motion vector refinement for high-performance transcoding", *IEEE Transactions on Multimedia*, vol.11, pp.30~40, March 1999.

[6] N.Bjork and C.Christopoulos, "Transcoder architectures for video coding", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.5, pp.2813-2816, 1998.

[7] Dogan S, Sadka AH, Kondoz AM, "Efficient MPEG-4/H/263 video transcoder for interoperability of heterogeneous multimedia networks," *Electronics Letters*, vol.35, no.11, 27 May 1999, pp.863-4, IEE, UK.