TUAM 4.2

Building MPEG-7 Transcoding Hints from Intrinsic Characteristics of MPEG Videos

Jin-Hau Kuo, Chia-Chiang Ho, *Student Member, IEEE*, and Ja-Ling Wu, *Senior Member, IEEE*
Communications and Multimedia Lab., National Taiwan University, Taipei, Taiwan

Abstract--MPEG-7 transcoding hints are defined to provide better quality or reduced complexity of transcoding applications. This paper describes application scenarios, hierarchies, and generation methods for transcoding hints that are based on intrinsic characteristics of MPEG videos.

I. INTRODUCTION

Transcoding hints, formally as one description scheme *MediaTranscodingHint*, are defined in MPEG-7, for reducing complexity and improving quality of general transcoding processes [1]. Transcoding hints defined in the MPEG-7 standard include *motion hints, shape hints, coding hints, spatial resolution hint, difficulty* and *importance hints*. The possible usage of transcoding hints includes but not limited to constant bitrate (CBR) to variable bitrate (VBR) conversion, encoding mode decision, frame rate and bitrate control, complexity reduction for motion re-estimation, and also video skimming or browsing.

Two central goals of MPEG-7 are generality and flexibility, and both generation and consumption of transcoding hints are not formally defined [2]. The definition of video segment is also left for real applications. Focusing on MPEG videos, in this paper we 1) define practical transcoding hint hierarchies form both structural and semantic perspectives; 2) address three application scenarios from which transcoding hints can be generated; and 3) propose architectures of transcoding hint generators for different application scenarios, based on retrieved low-level MPEG characteristics. Our work provides roadmaps of implementations embedded on consumer electronics or multimedia software.

II. APPLICATION SCENARIOS AND TRANSCODING HINT HIERARCHIES

As far as MPEG videos are concerned, we classify three application scenarios within which transcoding hints can be generated.

Scenario 1-- raw input, compressed output: This scenario refers to those applications performing MPEG encoding. Examples are professional studio encoders, personal video recorders (PVR) and software MPEG encoders, etc.

Scenario 2-- compressed input, compressed output: This scenario refers to those applications performing MPEG video processing. Examples include video post-processing, video editing and video content analysis for indexing, browsing, or searching usages. These applications take MPEG videos as

input, and generate metadata and/or modified MPEG videos.

Scenario 3-- compressed input, raw output: This scenario refers to those applications performing MPEG video decoding. Both hardware and software decoders, which decode MPEG videos come from network or local storage, belongs to this scenario.

Fig. 1 shows the proposed transcoding hint hierarchies, in an entity-relationship fashion. Importance hint and spatial resolution hint are semantically determined through user interaction and thus are optional. Transcoding hint generators are responsible for other hints. Two styles of video representation are proposed. The *sequence-scene-frame* style coincides with human understanding of videos, and is semantic. The *sequence-GOP-frame* style coincides with internal structure of MPEG bitstreams, and is thus structural. These two styles of video representation may help for *improving efficiency and accuracy of transcoding in both* systematic and user-oriented ways.

III. ARCHITECTURES OF TRANSCODING HINT GENERATORS

A. Transcoding Hint Generator for Application Scenarios 1

In this scenario, an MPEG encoder is responsible for the generation of MPEG bitstreams. Most of transcoding hints can be generated by using coding information generated in the encoding process, as shown in the Fig. 2. This way, the extra *computation requirement is as low as possible*. Typical MPEG bitstreams contain three types of frames, that is, I-, P- and B-frames. For I-frames, additional motion estimation must be performed to yield corresponding motion hint. For B-frames, some manipulation of motion vectors should be performed to yield motion vector field as in the P-frame case. Motion hints are generated thereafter. Coding hint is directly derived from coding mode decision made by the encoder, specifically, the decision made by frame type selection and rate control modules. The generation of difficulty hint requires a trial encoding process. The scene segmentation module adopts the similar idea proposed in [3] to generate semantic video representations. After frame-level hints are generated, a normalization work is performed to form hints of higher levels.

B. Transcoding Hint Generator for Application Scenarios 2 and 3

Applications belonging to these scenarios process the MPEG compressed videos for various usages, such as playback, browsing, editing and transcoding. The transcoding hint generator proposed here is based on an MPEG decoder, as shown in Fig. 3. Most of transcoding hints, including

motion hint and coding hint, are easily contributed during the decoding process (and following normalization work, of course). Scene segmentation and trial encoding modules are used to generate semantic video representation and difficulty hint, respectively. It should be noted that generating transcoding hint from an MPEG video might be the side-product of a video processing task, instead of generating them from an independent generator. For example, an MPEG compressed movie title stored in a personal video archive must have once been playbacked, and the corresponding transcoding hint can be generated at that time.

IV. CONCLUSION REMARKS

The advantages of transcoding hints are confirmed by experiments reported in MPEG meetings. It is interesting that all transcoding hints are defined as optional, and this shows that the applicability of transcoding hints are very different for different media formats and target applications. In this paper, targeting on MPEG videos, we tackle the applicability and generation of transcoding hints by defining different application scenarios and proposing different transcoding hint generators for them. In our architectures, transcoding hints are generated together with typical video processing tasks, including encoding, editing and decoding.

Defining transcoding hints is not to solve obstacles of transcoding, but to increase quality and/or to reduce complexity. An intelligent transcoder should have two alternative configurations: one is without the transcoding hint and the other uses transcoding hints wisely. Generating metadata to aid the task of transcoding is of course a positive idea. It is our belief that generating media-format-dependent metadata may help more for the task of transcoding than those general transcoding hints may do. One previous work in our laboratory had transferred this idea into a practical system [4].
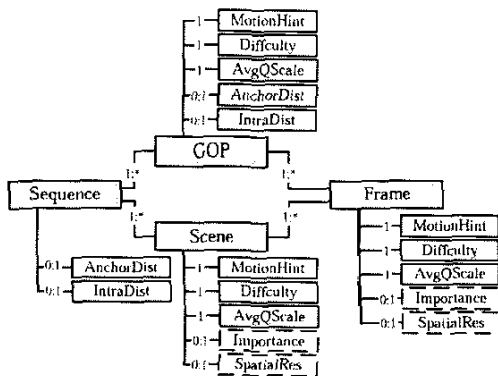

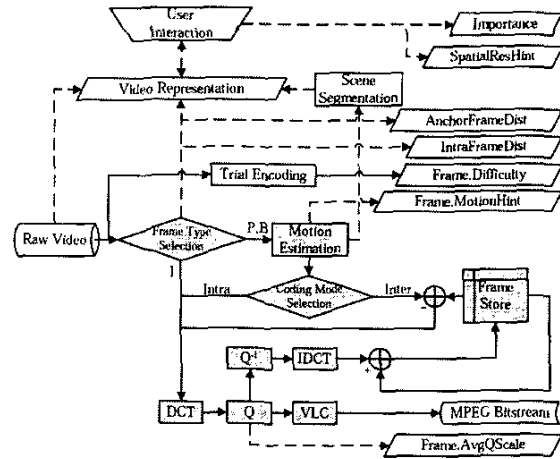Fig. 1. Structural and semantic transcoding hint hierarchies.

Fig. 2. Here is the proposed transcoding hint generator for application scenario 1. Grayed processes represent a typical MPEG encoder. The quantization module implicitly represents the rate control module.
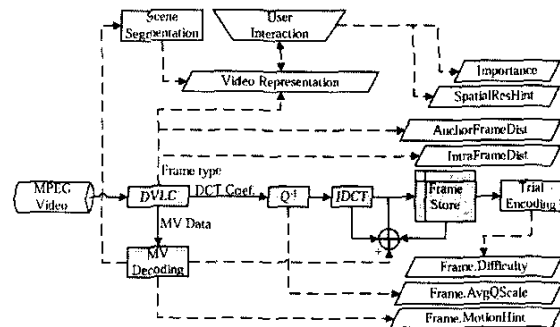

Fig. 3. Here is the proposed transcoding hint generator for application scenarios 2 and 3. Grayed processes represent a typical MPEG decoder.

REFERENCES

[1] *Information Technology-- Multimedia Content Description Interface, Part 5 Multimedia Description Scheme*, ISO/IEC 15938-5, 2001.

[2] S.-F. Chang, T. Sikora, and A. Purl, "Overview of the MPEG-7 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 688-695, June 2001.

[3] J.-H. Kuo and J.-L. Wu, "An efficient algorithm for scene change detection and camera motion characterization using the approach of heterogeneous video transcoding on MPEG Compressed Videos," *IEEE 3rd Int. Conf. On Information, Communications and Signal Processing*, Oct 2001.

[4] P.-K. Hsiao, Y.-S. Tung, Ja-Ling Wu, K.-L. Huang and H.-S. Chen, "A frame-based MPEG characteristics extraction tool and its application in video transcoding," *IEEE Trans. Consumer Electronics*, vol. 48, no. 3, pp. 522-532, Aug. 2002.