

MPEG-7 Transcoding Hints for Reduced Complexity and Improved Quality

Peter M. Kuhn¹, Teruhiko Suzuki¹, Anthony Vetro²

¹ SONY Corporation, Information & Network Technologies Laboratories, Tokyo, Japan {pkuhn, teruhiko}@av.crl.sony.co.jp

² MERL - Mitsubishi Electric Research Laboratories, Murray, Hill, NJ, USA avetro@merl.com

ABSTRACT

This paper introduces a new transcoding framework that is based on the use of meta-data specified by the MPEG-7 standard. Such transcoding related meta-data is referred to as a transcoding hint. The main objective of this paper is to provide an overview of these transcoding hints and elaborate on their extraction and usage. We focus on three major aspects of the transcoder that affect video transmission over packet networks: complexity reduction, mode decisions and adaptable object-based delivery.

Keywords: Transcoding, MPEG-7, Metadata, Motion Estimation, Mode Decision, CBR-to-VBR, Bit Allocation

1 Introduction

The amount of audiovisual (AV-) content that is transmitted over optical, wireless and wired networks is growing. Furthermore, within each of these networks, there are a variety of AV-terminals that support different formats, cf. Fig. 1. This scenario is often referred to as Universal Multimedia Access (UMA), cf. [Moh 99]. The involved networks are often characterized by different network bandwidth constraints, and the terminals themselves vary in display capabilities, processing power and memory capacity. Therefore, it is required to represent and deliver the content according to the current network and terminal characteristics. For example, audiovisual content stored in a compressed format such as MPEG-1, 2 or 4 [MPEG 4 VIS], has to be converted between different bit-rates and frame-rates, but must also account for different screen sizes, decoding complexities and memory constraints of the various AV-terminals.

To avoid storing the content in different compressed representations for different network bandwidths and different AV-terminals, low-delay and real-time transcoding from one format to another is required. Additionally, it is critical that subjective quality be preserved. To better meet these requirements, Transcoding Hints are defined by the emerging MPEG-7 standard. Within the MPEG-7 standardization committee, the term Transcoding Hints refers to meta-data that can be used to support transcoding applications. One requirement of the transcoding hints is to reduce storage size and facilitate distribution (e.g. broadcast to local AV-content servers). To do so, a highly compact representation is required. The second requirement of the transcoding hints is to preserve the subjective audiovisual quality as much as possible, while reducing the complexity of the transcoding process.

To facilitate transcoding and to fulfill both of the above requirements for various compressed target formats, transcoding hints are generated in advance and stored together or separately with the compressed AV-content.

This paper describes Transcoding Hints that have been proposed by the authors of this paper to MPEG and are now included as part of the MPEG-7 Standard, [MPEG 7 MDS]. These meta-data can be used for complexity reduction as well as for quality improvement in the transcoding process. Among the transcoding hints that we describe are the Difficulty Hint, the Shape Hint and the Motion Hints. The Difficulty Hint describes the bitrate coding complexity. One example use of this hint is for improved constant bitrate (CBR) to variable bitrate (VBR) conversion. The Shape Hint specifies the amount of change in a shape boundary over time and is proposed to overcome the composition problem when encoding or transcoding multiple video objects with different frame-rates. The Motion Hints describe i) the motion range, ii) the motion uncompensability and iii) the motion intensity. These meta-data can be used for a number of tasks, including anchor frame selection, encoding mode decisions, frame-rate and bit-rate control, as well as bitrate allocation among several video objects for object-based MPEG-4 transcoding. These proposed meta-data, especially the search range meta-data, aim also on computational complexity reduction of the transcoding process.

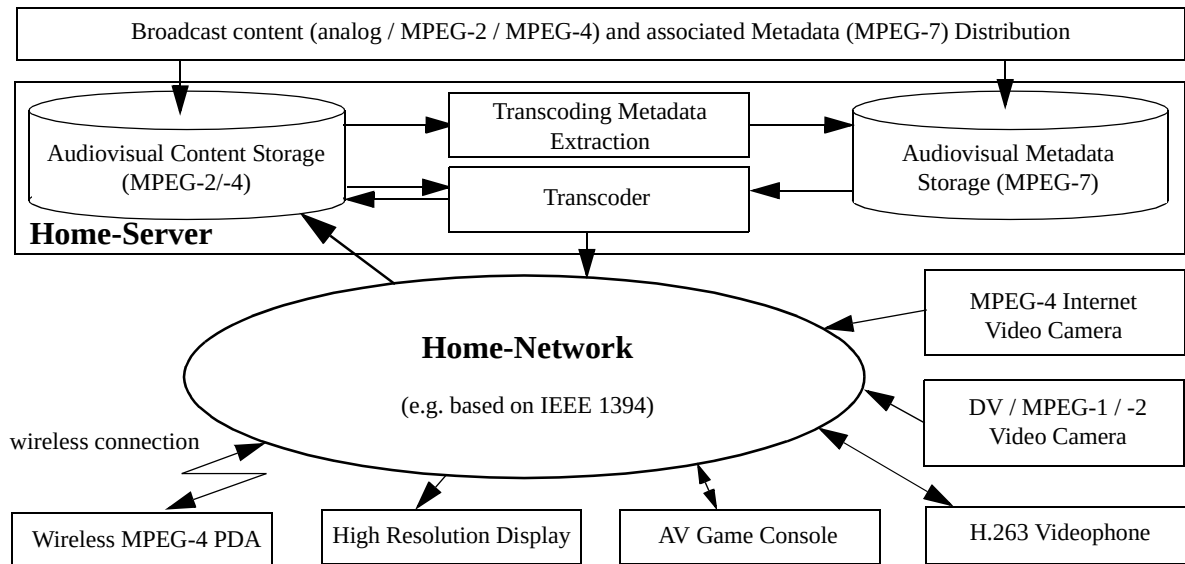


Fig. 1: Transcoding in the Home Network content Server

The rest of this paper is organized as follows. In the next section, improved mode decisions within the transcoder are considered, including the position of anchor frames, quantizer selections for CBR-to-VBR conversion and quantizer selections for transcoding of video objects. In section 3, adaptable object delivery is discussed. In section 4, reductions in complexity through constraints on the search range for motion estimation are described. In section 5, simulation results for the various techniques and applications are presented. Finally, the main contributions of this work are summarized in section 6.

2 Improved Mode Decisions

2.1 Anchor Frame Selection

For anchor frame selection and mode decision a motion uncompensability descriptor is used. This motion uncompensability descriptor is independent of the source content format and gives the sum of the amount of new content per frame within an AV segment and is defined as following. The motion uncompensability metric is based on feature point tracking. Feature points can be extracted e.g. by the Lucas/Kanade feature point tracker, [Luc 81], for every single frame of a video sequence. However, other methods are also possible. For example, a fast feature point extraction and tracking method based on [Luc 81] from MPEG compressed domain is described in [Kuhn 00]. For this algorithm a constant number of feature points is used per frame. It is tried to track each feature point from one frame to the next frame. In case a feature point can not be tracked from one frame to the next frame, a counter named "number_of_new_feature_points_per_frame" is incremented for every new feature point which could not be tracked from the previous frame. This counter is initialized to 0 at the beginning of an AV segment. At the end of the AV segment the number_of_new_feature_points_per_frame counter is normalized by the number of frames within the AV segment and by the maximum number of feature points per frame. With this method the source content independent metric of motion uncompensability is gained which states the amount of new content per frame.

As an example for transcoding from arbitrary content to an MPEG-1/-2/-4 video format, this descriptor can be used to assist the rate control of the transcoder and in determining the mode selection, GOP structure (IPB pattern), bitrate allocation of IPB frames, anchor frame selection, etc. resulting in improved visual quality. The search area of the motion estimation and therefore the computational complexity of the motion estimation of the transcoder can be reduced, when the motion uncompensability value is high. For this case, no significant gain in a Rate/Distortion sense is expected by a high motion search range. The computational complexity is reduced further, as no trial encoding has to be performed for this AV segment to find the best IPB structure. The motion uncompensability is normalized between 0.0 and 1.0, as only the relative value is important. In case the motion uncompensability is near 0.0, the next frame can be predicted by a P or B frame. In case the motion uncompensability is near 1.0, no significant relation exists between the successive frames and a transcoder may insert an I-frame here. Fig. 2 gives an example on how the motion uncompensability metric changes from the first frame t0 to the frame t1 (camera motion) to the frame t2 (new content appears in the scene). Fig. 3 depicts the motion uncompensability for the first 1000 frames of the MPEG-7 test sequence news1.mpg.

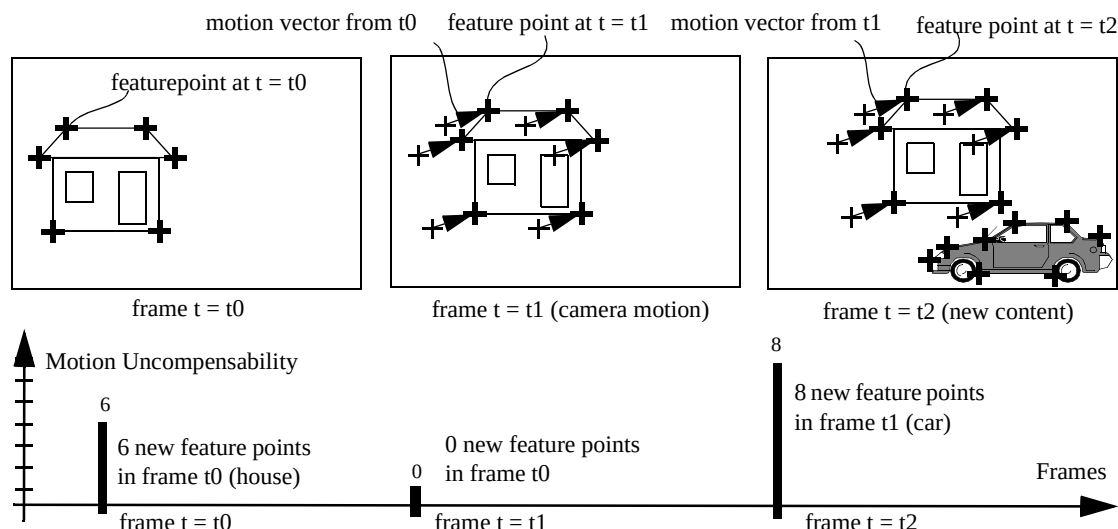


Fig. 2: Motion Uncompensability explanation

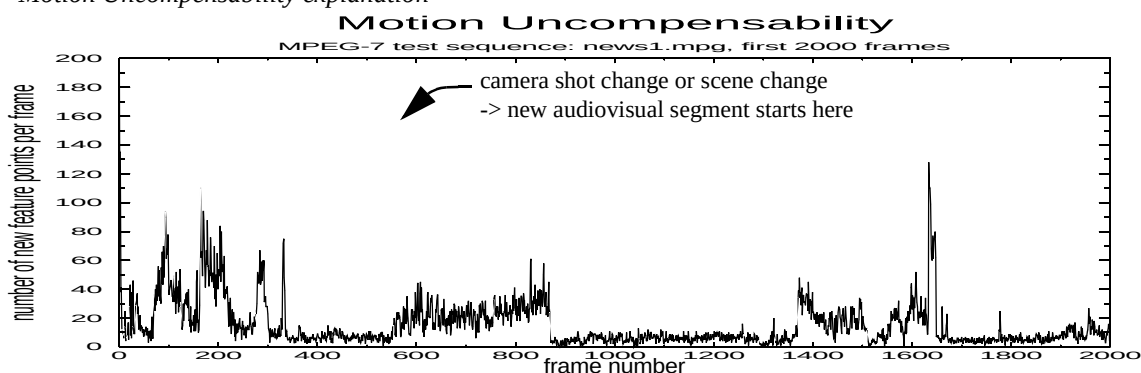


Fig. 3: Motion Uncompensability Descriptor

The transcoding metadata in this paper is defined for a video segment, which is a number of temporal subsequent frames. The video segment boundaries are determined by thresholding the peaks of the motion uncompensability metadata. A new AV segment is started when the motion uncompensability metric exceeds the predetermined threshold. This means, that a high number of feature points could not be tracked from the last frame to the next frame and thus a scene change can be inferred. In a number of cases the video-segment boundaries determined by this method were equal to semantic AV-segment boundaries, which can be beneficially used for human assisted semantic video segmentation for video browsing and summarization.

2.2 CBR-to-VBR Conversion

The Transcoding Difficulty Metadata depicts a measure of the difficulty to compress the source content. It targets on describing the required bitrate for an AV segment for compression by the transcoder. A low difficulty value depicts a video segment which can be easily encoded using a low bitrate. At a high difficulty value more bits have to be spent for this video segment to achieve a satisfactory subjective quality. The transcoder or encoder can assign an appropriate amount of bits to a certain AV segment using the difficulty hint. One of the main applications of the difficulty metadata is the transcoding from a constant bitrate (CBR) coded bitstream as used for network transmission to a variable bitrate (VBR) coded bitstream as used for storage devices (home server, DVD), resulting in increased coding efficiency.

In general, the coding efficiency of VBR is higher than CBR. The encoder can assign more bits to a difficult scene and remove bits from a relatively easy scene. Since mainly the pictures with low quality are improved, the overall subjective quality is significantly improved. In order to encode efficiently by VBR, the encoder has to know the encoding difficulty of the whole sequence in advance. Since the encoding difficulty changes over time, these meta-data are defined for each video segment. The absolute value of the Transcoding Difficulty metadata is of minor importance, while the relative value within the content is required in order to allocate the bitrate efficiently. The Transcoding Difficulty value is normalized within the content. The minimum value of the Transcoding Hints is 0.0 and the maximum is 1.0, which indicates the most difficult part of the content.

The Transcoding Difficulty can be extracted by the following steps:

- 1.) Encode de-/uncompressed content without motion compensation at quantizer $Q=1$
- 2.) Calculate generated bits/frame for each segment
- 3.) Normalize the value calculated in step 2.

In a simple case, the allocated bits of a segment for CBR-VBR conversion can be calculated by the following formula (LengthofSegment: number of frames of the segment, Diff: Difficulty meta-data value):

$$\text{Target bitrate} = (\text{Total_bits} * \text{Diff} * \text{LengthofSegment} / (\text{Sum of } (\text{Diff} * \text{Length of Segment}))) * \text{FrameRate} / \text{LengthofSegment} \quad (1)$$

Based on the Difficulty Hints the bitrate is calculated for each video segment, taking into account the length of the segment and the overall bitrate to be achieved.

2.3 Bit Allocation Among Video Objects

This section extends the difficulty hint defined in the previous section to multiple arbitrarily shaped video objects. The process of selecting QP's for each object relies on a model that defines the rate-quantizer (R-Q) relationship. See for example the quadratic R-Q model first proposed in [Chiang 97] and later used for multiple video objects in [Vetro 99]. Based on this model, the difficulty hint is used to compute the rate budget for each object.

The hint itself is a weight, w_i in the range [0,1], which indicates the relative encoding complexity of each segment. A segment may be defined as a single object or a group of objects. Defining the segment as a group of objects is advantageous to the transcoder in that future encoding complexity can be accounted for. Given that N objects are being considered, the rate for the i -th object at time t is simply given by: $R_i(t) = w_i \cdot R(t)$, where $R(t)$ is a portion of the total bit rate that has been allocated to the current set of objects at time t . Currently, $R(t)$ is equally distributed across all coding times. However, with the difficulty hint, unequal distribution may provide a means to improve the temporal allocation of bits. However, at this point, we are only interested in the spatial allocation among objects.

It should be noted that the difficulty hint is computed based on the actual number of bits spent during encoding. Although the incoming bits to the transcoder may be sufficient, quantization may hide the true encoding complexity. Therefore, when possible, the difficulty hint should be computed from encoding at fine QP's. The semantics of the MPEG-7 description scheme does not make this mandatory, but it should be noted that results may vary.

3 Adaptable Object Delivery

3.1 Key Object Identification

In recent years, key frame extraction has been a fundamental tool for research on video summarization, see for example [Han 99]. The primary objective of summarization is to present the user with a consolidated view of the content. The requirement is simple: give the user a general idea of the content that can be viewed in a short amount of time. In this paper, we would like to extend such concepts to the transcoder. Time constraints for summarization can be likened to bandwidth constraints for the transcoder. Given this, the transcoder should adapt its delivery so that the content is reduced to the most relevant objects. In this paper, we introduce a simple method to identify the key objects in the scene. Specifically, we consider a binary decision on each object, whether it is sent or not. To determine the importance of each object, a measure based on the intensity of motion activity, σ , and bit complexity, B , is proposed:

$$\eta = (\sigma + c) \cdot \frac{B}{\gamma} \quad (2)$$

where $c > 0$ is a constant to allow zero motion objects to be decided based on the amount of bits, and γ is a normalizing factor that accounts for object size. Larger values of η indicate objects of greater significance.

3.2 Variable Temporal Resolution

To motivate our reasons for considering a shape hint, the composition problem is discussed. Consider a video sequence with two objects, a foreground and background. In one case, both objects are encoded at 30Hz; in another case, the foreground is encoded at 30Hz and the background at 15Hz. When these two objects are encoded at the same temporal resolution, there is no problem with object composition during image reconstruction in the receiver, i.e., all pixels in the reconstructed scene are defined. However, when the objects are encoded at different temporal resolutions, holes will become present in the reconstructed scene. These holes are due to the movement of one object, without the updating of adjacent or overlapping objects.

The holes are uncovered areas of the scene that cannot be associated with either object and for which no pixels are defined. The holes disappear when the objects are resynchronized, i.e., background and foreground are coded at the same time instant.

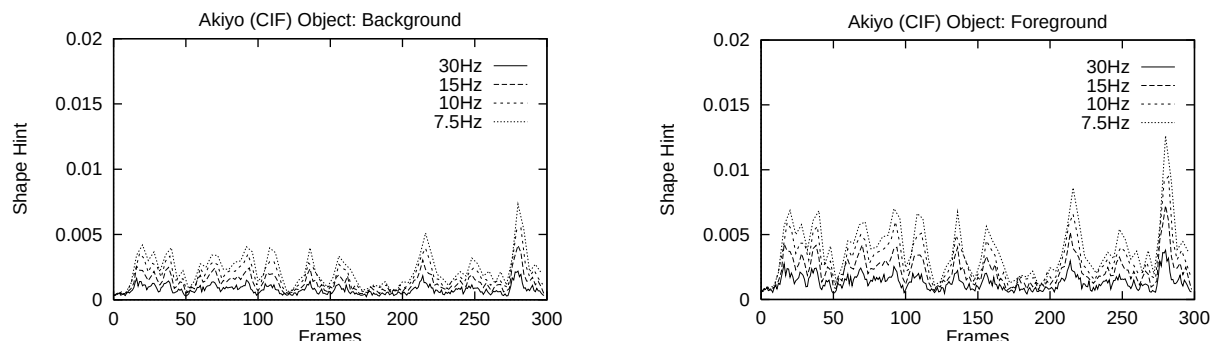


Fig. 4: Shape Hints for Akiyo. The shape hints for this sequence indicate very low movement of objects. This would be a very good candidate for variable temporal reduction. Depending on other content characteristics, an encoder may reduce the resolution of either object.

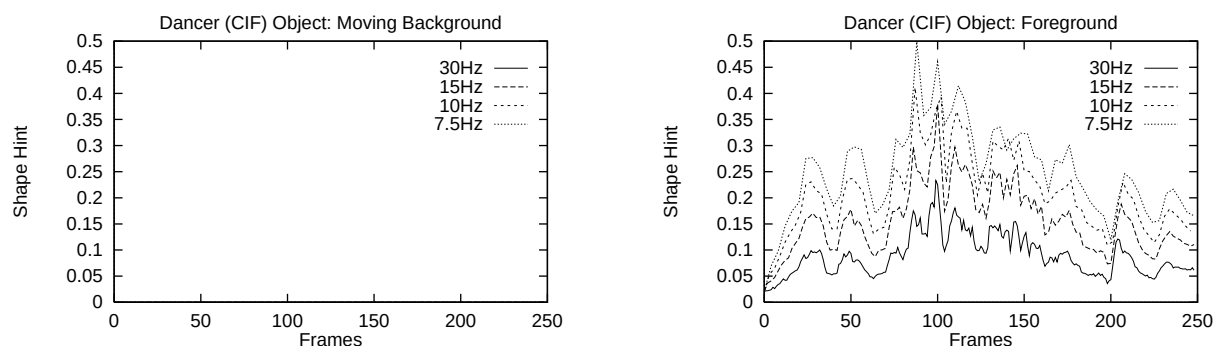


Fig. 5: Shape Hints for Dancer. The shape hint for the background indicates no change (it is a full rectangular frame), while the foreground object is experiencing large movement in shape. It is possible to have variable resolutions with this scene, since the full background is coded and it will never cause holes. However, due to high degree of shape movement in the foreground object, one may decide to encode this object at the full temporal rate.

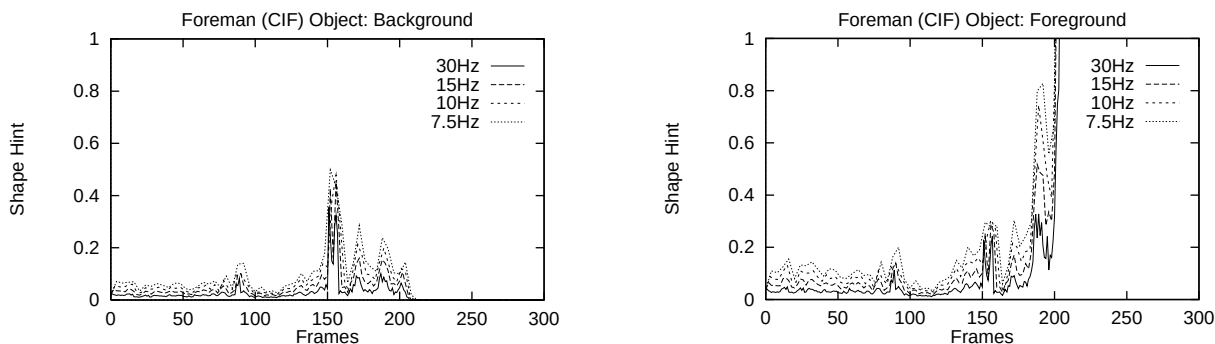


Fig. 6: Shape Hints for Foreman. The movement of the background and foreground agree with each other and show large movement for the first 200 frames. After that, the shape hint indicates that the foreground no longer exists and the background eventually fills the entire frame.

In [Erol 00], two shape distortion measures have been proposed for key frame extraction: the Hamming and Hausdorff distances. Both measure distortion from one frame to the next, however such shape metrics can also relate to the scene at various other levels, e.g., at the segment level over some defined period of time. As defined by the shape hint in MPEG-7, the above measures should be normalized in the range [0,1] over the video segment. We normalize the distortion by the number of pixels in the object. A value of zero would indicate no change, while a value of one would indicate that the object is moving very fast. Then to get the desired value for the segment, it is sufficient to average the sequence of shape hints to yield a single shape

hint value for the VOP over a given period of time. Actually, the shape hints themselves can be used to provide a temporal segmentation of the VOP's so that the averaged shape hint is more reliable for the specified duration of time.

To illustrate the above concepts, the shape hints are extracted for each object in a number of sequences at 30Hz, 15Hz, 10Hz and 7.5Hz. See Fig. 4, Fig. 5, and Fig. 6. For each of the sequences, we interpret how the shape hint may be used by a transcoder. In many cases, the final outcome and transcoder decision is based on additional information.

Given that the shape hints have been extracted, we now describe how it may be used for video object encoding or transcoding. In Fig. 6, we illustrate the steps that would be taken to compute variable temporal rates for each object. First, the shape hints for each VOP are extracted according to the methods discussed in the previous section. The collection of hints are passed into an analyzer whose function is to determine if composition problems will occur at the decoder if variable temporal rates are used. If variable temporal rates are allowed, then the temporal rates for each object can be computed with the assistance of additional transcoding hints or content characteristics. On the other hand, if variable temporal rates are not allowed, the encoder or transcoder may still adjust the outgoing temporal rate equally for all objects.

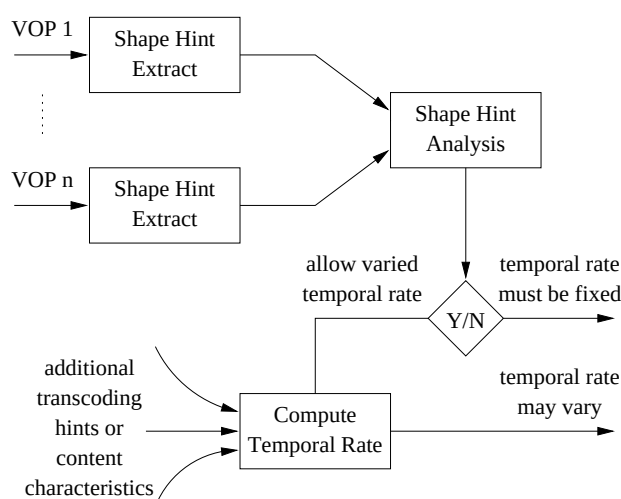


Fig. 7: Block diagram to illustrate use of shape hint and how variable temporal resolution could be computed

4 Complexity Reduction

The motion search range transcoding metadata targets on describing the motion within a video segment to reduce the computational complexity requirements for hardware and software implementations of the transcoder by describing the maximum motion vector range of the AV segment. This metadata is especially useful for transcoding from uncompressed or non-motion compensation based source format (e.g. DV, Intra only MPEG) to a motion-compensation (e.g. IPB-MPEG) based target format.

The feature points are extracted by a modified Lucas/Kanade/Tomasi [Luc 81] feature point tracker, as already used to determine the motion uncompensability metadata for video segmentation. By the feature point tracking method the motion vector is gained for each single feature point. After filtering, the maximal motion vector components for this video-segment in positive and negative x- and y-direction are saved in the motion search range metadata instantiations of this video-segment. This motion search range metadata assists the transcoder to select the appropriate motion estimation (ME) search range and the ME algorithm. This descriptor is different from "fcode" in MPEG-2 which does not specify the search range independently in each of their four directions.

5 Simulation results

In this section, the results of some experiments are described. The performance of a transcoder with Transcoding Hints is compared to the case without them. The evaluation criteria take the computational complexity as well as the subjective and quantitative visual quality into account.

5.1 Results on Reduced Complexity

The experimental conditions are:

- Source format: news1.mpg (CD 17 MPEG-7 test content), 352x288 pel, MPEG-1, 2.009600 Mbit/s, 40k VBV, 25 fps, Target format: MPEG-2 MP@ML, 352x288, N=15, m =3, 750 kbit/s,
- ISO MPEG-2 reference encoder: original motion estimation algorithm (fast spiral search), which was constrained in the search range by the transcoding hints in every positive or negative x- or y-direction,
- Max. search range for the P-frame (15 pel / 7 pel), max. search range for B-forward frames: 1/2 of the P-frame search range (truncated), max. search range for B-backward frames: 1/4 of the P-frame search range (truncated),
- Complexity analysis: x86 Linux, gcc 2.7.2.3, iprof041 [Kuhn 99].

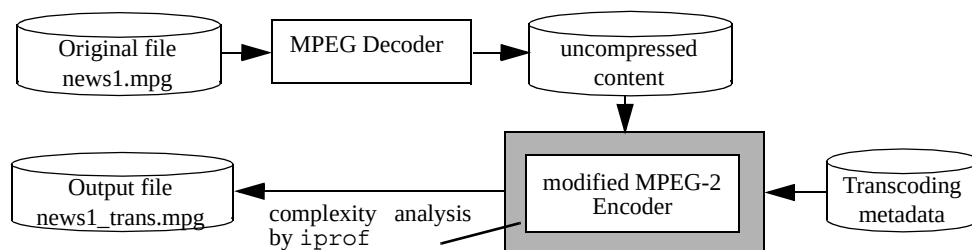


Fig. 8: Experiment Set-up

Experiments were conducted for the first 10000 frames of news1.mpg which depict a complexity reduction by up to 40 % towards a fast motion estimation algorithm (spiral search in the ISO MPEG-2 codec) while preserving the PSNR, Fig. 9, Fig. 10, and Fig. 11. For this experiment transcoding was performed in the baseband by using a modified MPEG-2 encoder, which was provided by the authors to the MPEG-7 Experimentation Model (MPEG-7 XM) software. The modified MPEG-2 encoder can take the XML-based transcoding hints for the encoding decisions into account. For complexity analysis the iprof methodology [Kuhn 99] is used, which provides an exact analysis of the CPU instructions used by the program, e.g. the number of multiplications, divisions, additions, memory read and write accesses, etc. for the instrumented transcoder.

Search Area	Transcoding Hints	PSNR Y (mean)	Bits (total)	Instructions total (1+e6)	Complexity reduction to 100%
+/- 31	yes	34.664	37500936	3523574	- 41.1 %
+/- 31	no	34.733	37500803	5986398	100.0 %
+/- 24	yes	34.650	37500870	2854603	- 35.0 %
+/- 24	no	34.742	37500810	4393058	100.0 %
+/- 15	yes	34.628	37500973	1906546	- 21.4 %
+/- 15	no	34.684	37500858	2425185	100.0 %
+/- 7	yes	34.299	37500722	1188563	- 7.0 %
+/- 7	no	34.522	37498348	1279229	100.0 %

Fig. 9: Simulative Results

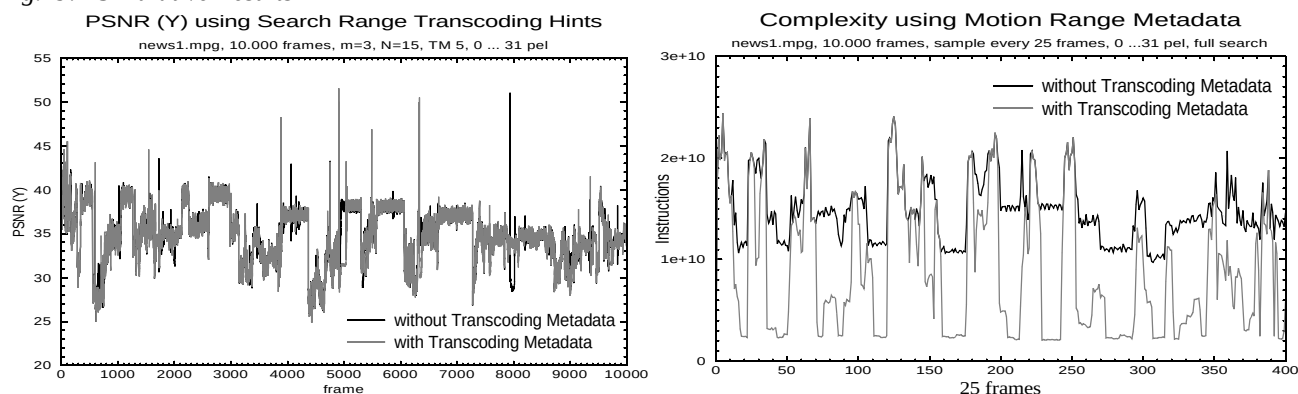


Fig. 10: PSNR (left) and Complexity Reduction (right) using the Motion Search Range Metadata

5.2 Results for Anchor frame selection

The motion uncompensability hint is employed to determine the video segment boundaries as stated before. In this experiment, variable video segment lengths between 12 and 400 frames are selected automatically by the motion uncompensability value. Every video segment marks a new scene and starts with an I-frame and consists of GOPs of variable size. Starting every video segment with an I-frame has the advantage of fast and simple video browsing, as the first frame of the segment decoded from the bitstream can be displayed immediately.

Variation of the GOP size dependent on the motion_uncompensability metric offers several advantages: For example, when the motion uncompensability is higher more new content appears in a video segment, and a smaller distance between the I-frames (i.e. smaller GOP size N) and a different distance between I and P frames (different M) may be chosen. While trying to keep the overall rate/distortion ratio constant, this results in increased subjective visual quality and increased transmission error robustness for fast changing scenes, as more I-frames are inserted in the scenes where the content of the video sequence is changing faster. However, besides optimization for transmission error robustness, the motion uncompensability metadata can be also employed to improve coding efficiency.

5.3 CBR-to-VBR Conversion results

Based on the Difficulty Hints an appropriate number of bits is assigned for each scene at a certain target bitrate in case of VBR. In this experiment, the target bitrate is assumed to be 750 kbps. Based on the Difficulty Hints the bitrate is calculated for each video segment, taking into account the length of the segment and the overall bitrate to be achieved. The results of the comparison between CBR and VBR is shown in Fig. 11. In case of VBR, the encoder assigns a number of bits based on the segment based Transcoding Hints including difficulty. The picture quality of a more difficult scene is improved in case of VBR. The overall subjective quality is improved for VBR in contrast to CBR. VBR improved the PSNR by about 2 dB for a very difficulty scene while keeping the picture quality for an easy scene. The advantage of VBR over CBR is quite obvious from these results.

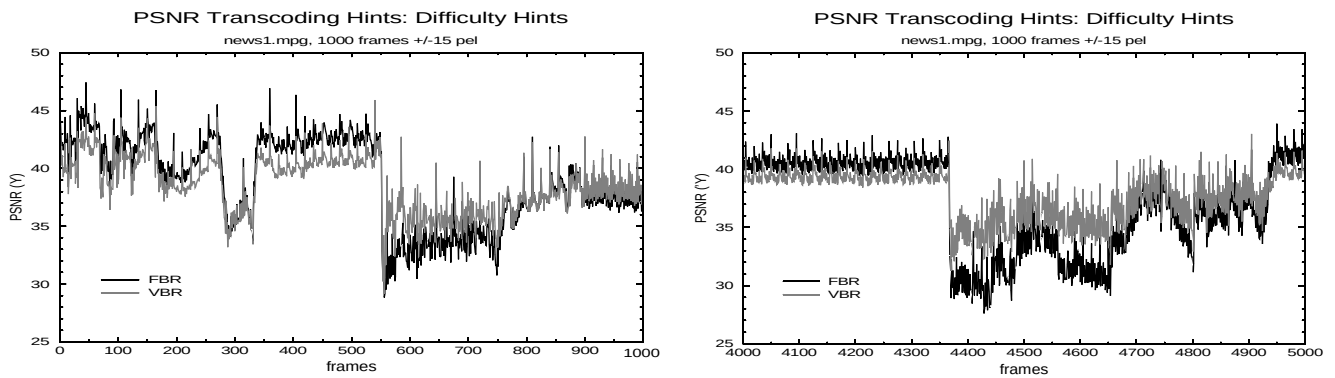


Fig. 11: Transcoding using the Difficulty Hint

5.4 Bit allocation among objects

To study bit allocation in the transcoder, several results are simulated with the Coastguard and News sequences. Both sequences are CIF resolution (352 x 288) and contain 4 objects each. The rate control algorithm used for encoding is the one reported in [Vetro 99] and implemented in [MPEG 4 REF]. To achieve the reduced rate, three methods are compared. The first method is a brute-force approach that simply decodes the original multiple video object bitstreams and re-encodes the decoded sequence at the outgoing rate. The next two methods have much lower complexity than the reference, where one method simulates the dynamic programming approach presented in [Vetro 00] and the other method simulates the proposed approach using difficulty hints. A plot of the PSNR curves comparing the reference method and two transcoding methods is shown in Fig. 12. The Coastguard sequence is simulated with $R_{in} = 512$ kbps, $R_{out} = 384$ kbps, a frame rate of 10 fps, and GOP parameters, $N=20$ and $M=1$. Similarly, the News sequences was simulated with $R_{in} = 256$ kbps, $R_{out} = 128$ kbps, a frame rate of 10 fps, and GOP parameters, $N=300$ and $M=1$. It is quite evident from the plots that the transcoding method that using difficulty hints is always better than the DP-based approach. This is expected since the difficulty hint can provide information about the scene that is not readily available. From the plots, we can also observe that for the Coastguard sequence, the transcoded results using the difficulty hint provide a very close match to the reference method. In fact, for many frames it is even slightly better. This indicates that with the use of non-causal information, it is possible to overcome some of the major drawbacks that conventional low-complexity open-loop transcoders suffer from - mainly drift.

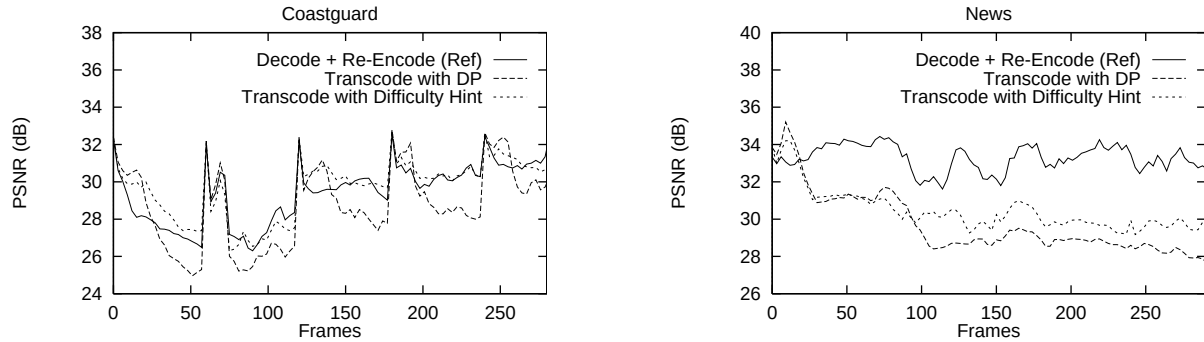


Fig. 12: PSNR comparing reference and transcoded frames using dynamic programming (DP) approach and meta-data based approach with difficulty hints. (Right) Coastguard, 512 -> 384kbps, 10fps, $N=20$, $M=1$ (Left) News, 256kbps -> 128kbps, 10fps, $N=300$, $M=1$.

5.5 Results with Key Object Identification

To demonstrate how key object identification can be used to adapt content delivery and balance quality, the News sequence is considered at CIF resolution, 10fps. The News sequence contains 4 objects of varying complexity and of distinctly varying interest to the user. The sequence was originally encoded at 384kbps, then transcoded down to 128kbps.

Fig. 13 illustrates several transcoded versions of frame 46 from the News sequence. The versions differ in the number of objects that are actually transmitted. In the left-most version, all objects are transmitted with relatively low quality, whereas in the right-most version, only one object is transmitted with relatively high quality. The center image lies somewhere in between the other two in that only the background is not sent. It can be considered a balance between the two in terms of quality and content. For all versions, the objects are identified using the computed measure of importance based on motion and difficulty hints, while the bit allocation for objects under consideration were done using the difficulty hint only.



Fig. 13: Illustration of adaptable transcoding of News sequence with key object identification, 384 -> 128kbps. The motion intensity and difficulty hints are used to order the importance of each object and the difficulty hint alone is used for bit allocation. Fewer objects can be delivered with noticeable higher quality.

5.6 Results with Varying Temporal Resolution

Given that the shape hint will identify scenes (or segments) that can take advantage of variable temporal resolution, we provide results to demonstrate the gain. The Akiyo sequence has been identified as a scene that can be encoded or transcoded with variable temporal resolution, without creating disturbing artifacts due to composition. To examine this further, we generate the R-D curves for two cases. The first case is our reference and simply encodes both foreground and background objects at 30Hz with a constant QP. This is referred to as the fixed case. As a comparison, we maintain the temporal resolution of the foreground at 30Hz, but reduce the temporal resolution of the background to 7.5Hz. This case is referred to as the variable case. A comparison of the fixed and variable R-D curves is shown in Fig. 14. The set of QP's that were used to generate the curves were $Q = 9, 15, 21$ and 27 . The finest QP was 9 since reducing the temporal resolution for high bandwidth simulations is not of much interest. It should be emphasized that the PSNR for all simulations are computed with respect to the original 30Hz sequence. The curves in this figure show that a 25 % reduction in bit rate can be achieved by reducing the resolution of the background only, with no loss of quality. As indicated by the shape hints for this sequence, the movement among shape boundaries is very small. Consequently, any holes in the reconstructed image are negligible and do not impact the PSNR values.

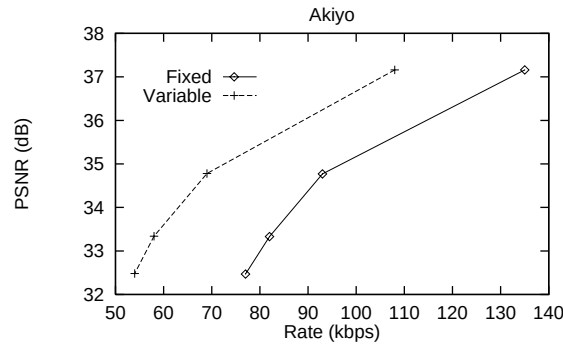


Fig. 14: R-D curves comparing Akiyo sequence coded for fixed and variable temporal resolution. With fixed temporal resolution, both foreground and background are coded at 30Hz. With variable temporal resolution the foreground is still coded at 30Hz, but the background is coded at 7.5Hz.

6 Conclusions

This paper presented extraction methodologies and usage of MPEG-7 meta-data for video transcoding. This meta-data gives "Hints" to the transcoder that describe the motion, difficulty and shape characteristics for audiovisual segments. These hints were proposed by the authors to MPEG and have been included as part of the MPEG-7 Standard. Difficulty Hints are meta-data to control the bitrate for texture coding, which can be used in a transcoder for improved constant bitrate (CBR) to variable bitrate (VBR) conversion and improved bit allocation among multiple video objects. The Motion Hints describe i) the motion range, ii) the motion uncompensability, and iii) the intensity of motion activity, which can be used for anchor frame selection, mode decision, adaptable object-based delivery and computational complexity reduction of the transcoding process. The Shape Hint specifies the amount of change in a shape boundary over time and is proposed to overcome the composition problem when encoding or transcoding multiple video objects with different frame-rates. Automatic extraction, representation and usage of the MPEG-7 transcoding hints are described in this paper and the simulation results depict the effectiveness of the described transcoding hints. The results show that the use of MPEG-7 transcoding metadata can preserve visual quality in terms of PSNR, provide new means of object-based delivery, and reduce the computational complexity significantly.

References

- [MPEG 4 VIS] ISO/IEC 14496-2:1999 Information technology - coding of audio/visual objects - Part 2: Visual.
- [MPEG 4 REF] ISO/IEC 14496-5:2000 Information technology - coding of audio/visual objects - Part 5: Reference Software.
- [MPEG 7 MDS] ISO/IEC CD 15938-5:2000 Information technology - Multimedia content description interface - Part 5: Multimedia Description Schemes
- [Chiang 97] T. Chiang, Y-Q. Zhang, "A new rate control scheme using quadratic rate-distortion modeling", IEEE Trans. Circuits Syst. Video Technol., Feb 1997
- [Erol 00] B. Erol, F. Kossentini, "Automatic key video object plane selection using shape information in the MPEG-4 compressed-domain", IEEE Trans. Multimedia, vol. 2, no. 2, June 2000
- [Han 99] A. Hanjalic, H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis", IEEE Trans. Circuits Syst. Video Technol., Dec 1999
- [Kuhn 99] Peter Kuhn: "Algorithms, Complexity Analysis and VLSI-Architectures for MPEG-4 Motion Estimation", Kluwer Academic Publishers, June 1999, ISBN 792385160, <http://www.wkap.nl/book.htm/0-7923-8516-0>
- [Kuhn 00] Peter Kuhn: "Camera Motion Estimation using feature points in MPEG compressed domain", IEEE International Conference on Image Processing (ICIP), Vancouver, CA, Sept. 2000
- [Luc 81] Bruce D. Lucas, Takeo Kanade: "An Iterative Image Registration Technique with an Application to Stereo Vision", International Joint Conference on Artificial Intelligence, 1981, p 674-679
- [Moh 99] R. Mohan, J. R. Smith, C.-S. Li: "Adapting Multimedia Content for Universal Access", IEEE Transactions on Multimedia, Vol. 1, No. 1, Mar 1999
- [Vetro 99] A. Vetro, H. Sun, Y. Wang, "MPEG-4 rate control for multiple video objects", IEEE Trans. Circuits and Syst. Video Technol., Feb. 1999
- [Vetro 00] A. Vetro, H. Sun, Y. Wang, "Object-based transcoding for scalable quality of service", IEEE Int'l Symp. Circuits and Syst., Geneva, CH, May 2000