MPEG-4 and Beyond - Trends and Perspectives for Image and Video Coding

Thomas Sikora

Heinrich-Hertz-Institute (HHI) for Communication Technology Berlin GmbH, Germany

Einsteinufer 37, D-10587 Berlin, Germany, Email: sikora@hhi.de

Abstract

The MPEG video group is currently developing the so-called MPEG-4 video coding standard, targeted for future interactive multimedia video communications calling for content-based functionalities, universal access in error prone environments and high coding efficiency. Besides the provisions for content-based functionalities the MPEG-4 video standard will assist the efficient storage and transmission of images and video in error prone environments over a range of bit rates between 5 kbit/s and 4 Mb/s. This paper outlines the techniques that are currently being investigated by MPEG-4 and discusses the scope of some of the promising techniques under investigation.

I. Introduction

The rapid development in the field of video and audio compression within the last ten years - and the associated research and development momentum generated - took many experts in the field by surprise. While the MPEG-2 standard [1] makes its way into the consumer market, this momentum is being retained with intensive research and development efforts worldwide dedicated towards the specification of even more efficient audio and video compression technology. Much effort is concentrated around the new MPEG-4 standardization phase which has the mandate to develop and standardized audio and video compression algorithms for multimedia applications.

The purpose of this paper is to discuss trends and perspectives of the techniques investigated in the context of the MPEG-4 standardization process - and to outline techniques that promise to be of interest for future applications.

II. MPEG-4 Functional Coding of Video

Anticipating the rapid convergence of telecommunications industries, computer and TV/film industries, the MPEG group officially initiated a new MPEG-4 standardization phase in 1994 - with the mandate to standardize algorithms and tools for coding and flexible representation of audio-visual data to meet the challenges of future Multimedia applications and applications environments [2]. The MPEG-4 development is already at an advanced stage with decisions on the major technical details of the algorithms being defined by October 1997 when issuing the MPEG-4 standard Committee Draft. In particular MPEG-4 addresses the need for

- *Universal accessibility and robustness in error prone environments* - Although the MPEG-4 standards will be network (physical-layer) independent in nature, the algorithms and tools for coding audio-visual data are designed with awareness of network peculiarities.

- *High interactive functionality* - It is envisioned that - in addition to conventional playback of audio and video sequences - the user need to access „content“ of audio-visual data to present and manipulate/store the data in a highly flexible way.

- *Coding of natural and synthetic data* - MPEG-4 will assist the efficient and flexible coding and representation of both natural (pixel based) as well as synthetic data.

- *Compression efficiency* - For the storage and transmission of audio-visual data a high coding efficiency, meaning a good quality of the reconstructed data, is required.

Bit rates targeted for the MPEG-4 video standard are between 5-64 kbits/s for mobile or PSTN video applications [3] and up to 4 Mb/s for TV/film applications. The release of the MPEG-4 International Standard is targeted for July 1998.

III. The MPEG-4 Video Standard Development

Similar to the MPEG-2 TM5 Test Model, the MPEG-4 standard process developed a Video

Verification Model (VM) which defines a "Common Core" video coding algorithm for the collaborative work within the MPEG-4 Video Group. Based on this core algorithm a number of "Core Experiments" are defined with the aim to collaboratively improve the efficiency and functionality of the VM - and to iteratively converge through several versions of the model towards the final MPEG-4 video coding standard algorithm by the end of 1997. To this reason the MPEG-4 Video Verification Model provides an important platform for collaborative experimentation within MPEG-4 and should already give some indication about the structure of the emerging MPEG-4 Video coding standard [2]. More importantly the Verification Model process, based on Core Experiments, is an efficient way to investigate the potential of various diverse algorithms proposed to MPEG. In other words, MPEG has aclearly defined mechanism to explore whether techniques work and improve or not - and what the implication of the algorithms is in terms of hardware and software complexity. In contrast to just reading articles in research journals and conference proceedings, within MPEG the algorithms are tested and benchmarked by various companies under controlled conditions.

Various techniques have been investigated and considered for the next generation MPEG video coding standard, including DCT-based technology, wavelets, matching pursuits and segmentation-based coding algorithms. Based on an evaluation of these proposals in formal subjective tests in October 1995, MPEG-4 settled on a hybrid block-based DCT/MC-based algorithm which has substantial similarities with existing standards. As additional element the provision for coding arbitrarily shaped regions in sequences is supported along with further functionalities, such as „Sprite" prediction.

The MPEG-4 video coding algorithms will eventually support all functionalities already provided by MPEG-1 and MPEG-2, including the provision to efficiently compress standard rectangular sized image sequences at varying levels of input formats, frame rates and bit rates. In addition content-based functionalities will be assisted.

The MPEG-4 video standard introduces the concept of Video Object Planes (VOP's) to support the so-called content-based functionalities. This concept is illustrated in Figure1. It is assumed that each frame of an input video sequence is segmented into a number of arbitrarily shaped image regions (video object planes) - each of the regions may possibly cover particular image or video content of interest, i.e. describing physical objects or content within scenes. The shape, motion and texture information of the VOP's belonging to the same VO is encoded and transmitted or coded into a separate VOL (Video Object Layer). In addition, relevant information needed to identify each of the VOL's - and how the various VOL's are composed at the receiver to reconstruct the entire original sequence is also included in the bitstream. This allows the separate decoding of each VOP and if required a flexible manipulation of the video sequence. Notice that the video source input assumed for the VOL structure either already exists in terms of separate entities (i.e. is generated with chroma-key technology) or is generated by means of on-line or off-line segmentation algorithms.

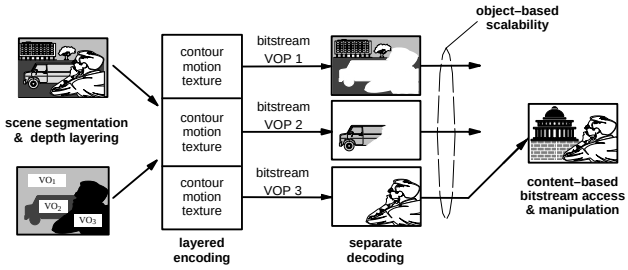


Fig. 1: The „object-layered" coding approach taken by the MPEG-4 video coding standard.

As illustrated in Figure 2, the MPEG-4 video standard will support the coding of rectangular size image sequences which is similar to conventional MPEG-1/2 coding approaches and involves motion prediction/compensation followed by DCT-based texture coding. For the content-based functionalities the image sequences are in general considered to be arbitrarily shaped - in contrast to the standard MPEG-1 and MPEG-2 definitions which encode rectangular size image sequences. The MPEG-4 content-based approach can be seen as a logical extension of the conventional MPEG-2 coding approach towards image input sequences of arbitrary shape. However, if the original input image sequences are not of arbitrary shape, the coding structure simply degenerates into a MPEG-1/2-like single layer representation which supports coding of conventional image sequences of rectangular shape.

Figure 3 provides a more detailed block-diagram of the MPEG-4 coding algorithm. The different motion prediction modes that are currently supported by the MPEG-4 Video Verification model include:

- Conventional block-based motion vector prediction (for blocks of 8x8 or 16x16 pels)
- Global motion compensation using affine motion parameters (rotation, zoom, translation), calculated for each frame and applied if required on a block basis.

- Static and dynamic sprite prediction using affine motion parameters (rotation, zoom, translation).

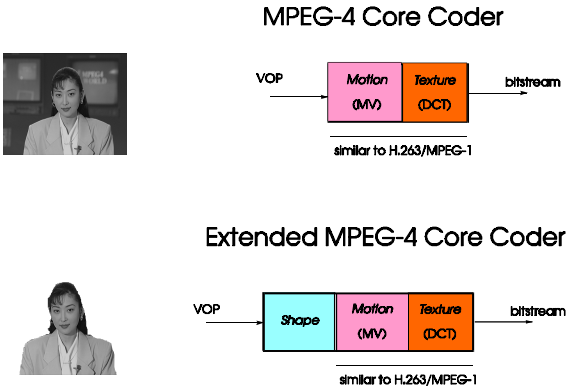


Fig. 2: The VLBV Core and the Extended Core

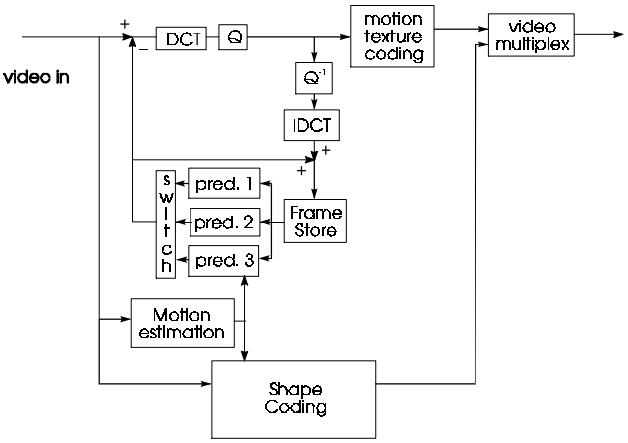


Fig. 3: Block Diagram of the MPEG-4 Video Coder

IV. Coding Efficiency - MPEG-4

The basic MPEG-4 core video coding algorithm is DCT-based and has strong similarities with existing standards. As mentioned earlier, improvements in terms of pure coding efficiency have not been the prime concern of the MPEG-4 standardization phase. However, every effort was made to improve image quality in the course of this process.

Coding of Sequences with Generic Content:

For coding conventional video sequences (rectangular) with generic content the MPEG-4 video standard algorithm operates on bit rates between 5 kbit/s and 4 Mb/s very efficiently. At very low bit rates between 5 - 100 kbit/s the MPEG-4 algorithm will achieve a coding efficiency which will be superior to that of the ITU H.263 standard and comparable with the new ITU H.263+ specification. At medium bit rates between 100 kbit/s and 1 Mb/s the MPEG-4 standard will provide better quality than MPEG-1. Around 4 Mb/s on interlaced sources a comparable (if not better) quality than the MPEG-2 standard is achieved. Random access functionality will be provided over this range of bit rates to allow Pause, FastForward and FastRevers within scenes.

The improvements in terms of coding efficiency are mainly due to an improved Slice Layer and Macroblock Layer syntax and improved motion prediction modes, followed by postprocessing of the blocking artifacts:

- Switched 8x8 and 16x16 pel motion compensation allows a more precise motion prediction and compensation,
- Block-overlapping motion compensation conceals blocking artifacts at lower bit rates,
- Global motion compensation mode improves for scenes with global camera motion content,
- Postprocessing filters reduce block and ringing artifacts at lower bit rates.

Coding of Sequences with Specific Content - the MPEG-4 Sprite Prediction Approach:

A number of tools are being investigated within the MPEG-4 video development which attempt to provide higher quality as well as additional content-based functionalities for sequences with restricted content. An interesting example is the MPEG-4 „Sprite“ prediction [3][4]. The „Sprite“ coding allows the efficient transmission of background scenes where the changes within the background content is mainly caused by camera motion. Thus a static sprite is a possibly large still image (i.e. static and flat background panorama) which is transmitted to the receiver first - and then stored in a frame store at both encoder and decoder. The camera parameters are transmitted to the receiver for each frame so that the appropriate part of the scene can be mapped (or warped - including zoom, rotation and translation within the Sprite image) at the receiver for display.

Consider the case that for a given video sequence the content in a scene can be separated into foreground object(s) and a (static) background Sprite. This may be done off-line by analysis of the content of a scene prior to coding. Figure 4 illustrates the Sprite (background) generation for a video sequence which contains a tennis match with high camera motion and texture. One tennis player is moving in front of a background scene. Starting from frame 1, through successive image analysis and with the help of the camera motion, the final Sprite background image is derived in frame 200.

Notice that the Sprite generation is not standardized, since it can be seen as a postprocessing tool.

Using the MPEG-4 Sprite coding technology the foreground content can be coded and transmitted separately from the receiver. If the background is static, only one frame needs to be transmitted at the beginning of a scene (i.e. frame 200 in Figure 4) - plus the camera parameters. The receiver composes the separately transmitted foreground and background to reconstruct the original scene. Figure 5 illustrates this concept using the example above. The foreground object tennis player is coded separately from the background as an object of arbitrary shape. The background (Figure 5, right) is reconstructed from the Sprite background image in Figure 4 stored at the decoder. Only 8 motion parameters were transmitted to the receiver to indicate which part of the Sprite is being used under what kind of perspective transformation. Only few bits are being spend for the background information.

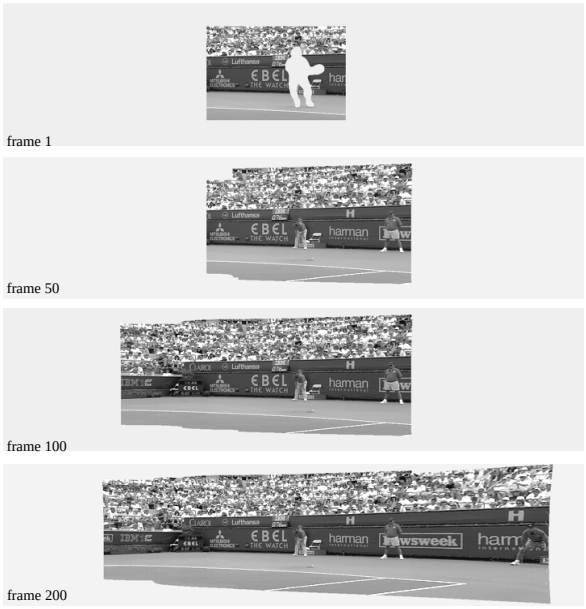


Fig. 4: Sprite Background Generation

The coding gain using the MPEG-4 Sprite technology over existing compression technology appears to be substantial in the example given above (Figure 6). Notice, however, that the technique described can not be seen as a tool which is easily applied to generic scene content. The gain described above can only be achieved if substantial parts of a scene contain regions where motion is described by simple motion models - and if these regions can be extracted from the remaining parts of the scene by means of image analysis and postprocessing. This certainly is an assumption that can be considered feasible to improve video quality for multimedia database applications - but most certainly not for Broadcast applications where on-line processing and coding is a necessity.

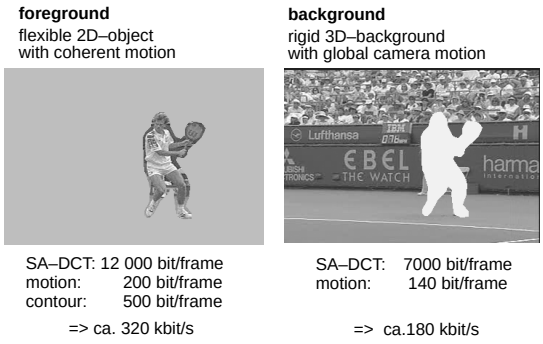
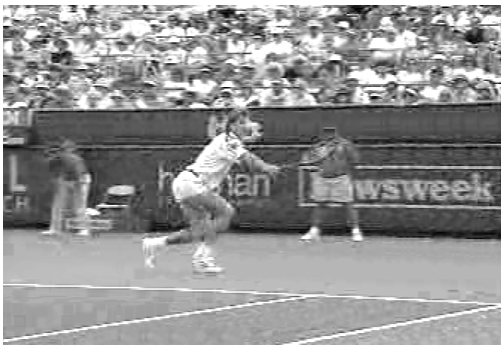
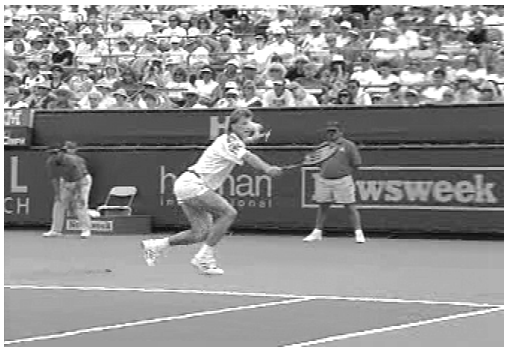


Fig. 5: Foreground Tennis Player and Background Sprite Coded Separately.



(A)



(B)

Fig. 6: Sequence Coded Using MPEG-1 (A) and MPEG-4 with Sprite Technology (B) at Appr. 1 Mb/s.

V. Coding Efficiency - Beyond MPEG-4

It appears that there is currently technology under development worldwide which promises much potential for the years to come. Most of this technology departs from the well investigated and successful block-based MC/DCT approaches used for the MPEG algorithms (including those of the MPEG-4 core algorithm) and has also been investigated within the MPEG-4 development

process - and has proven to perform very promising when benchmarked against the MPEG-4 Verification Model in the Core Experiment process. The fact that some of these techniques may not be considered for the MPEG-4 standard should not be taken as a criteria to judge the potential of these algorithms. MPEG evaluates and adopts technology based on the state-of-the-art performance in Core Experiments taking into account various criteria other than coding efficiency. In the following two of the promising Core Experiment algorithms - which take a substantial departure from standard MPEG technology - are outlined.

Quadtree-Based Motion Compensation using Polynomial Motion Models:

A number of segmentation-based video coding algorithms were developed over the last years with the primary aim to improve coding efficiency [3]. A very interesting algorithm within this particular class of techniques was introduced by Nokia, Finland [5]. The primary intent of the algorithm is to improve the motion compensation based on a quadtree segmentation of the motion vector field. Figure 7 illustrates this concept. A motion compensation is employed based on the previously coded frame N-1. In contrast to standard MPEG technology this technique does not employ block-based motion compensation - but rather identifies possibly large segments within images with same or similar motion. For each segments a flexible number of motion parameters (between 2 and 12 parameters to track complex motion) are coded and used for motion compensation. Next to the motion parameters also the quadtree structure needs to be transmitted to the receiver. The residual error is coded using a variable block-size DCT, but in general the technique is not restricted to a DCT approach.

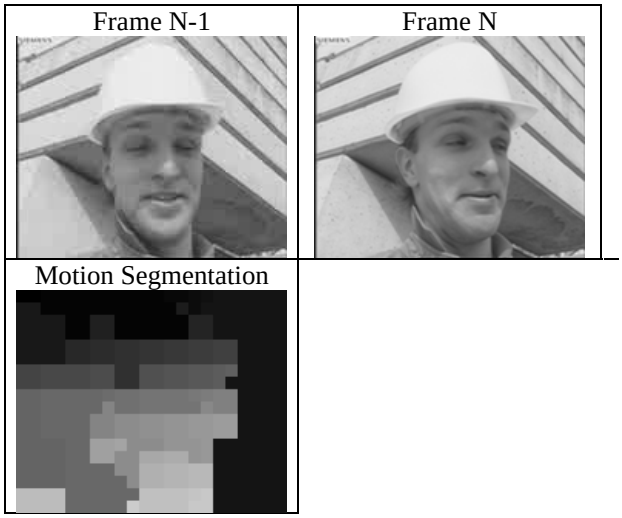


Fig. 7: Motion Segmentation Using a Quadtree Approach.

The quadtree segmentation method allows an excellent prediction of motion between frames and achieves a good trade-off between the higher degree of the motion model and the cost for transmitting the motion and segmentation parameters. The technique is computationally very demanding at the encoder since an iterative motion estimation technique is employed (and complexity increases with increased picture size). The decoder has a complexity comparable to standard MPEG decoders.

Texture Coding Using Matching Pursuits:

Many disturbing artifacts visible when coding video at lower bit rates using standard MPEG coders are so-called blocking or ringing artifacts. These artifacts are caused by the insufficient number of DCT-coefficients transmitted due to a constraint bit budget. A number of alternative techniques for coding textures or residual errors after motion compensation have been proposed in the last few years with the attempt to obtain more visually pleasant images compared to those reconstructed using a DCT approach [3].

An algorithm specifically tailored for coding residual errors at lower bit rates is based on a „Matching Pursuit“ approach [6]. The main novelty of the algorithm is an inner-product search to decompose motion residual signals on an overcomplete dictionary of separable Gabor functions. The method is not block-based at all and enables to code signals in a highly flexible way where they appear with highest energy - and to allocate the bits accordingly. A standard block-based motion compensation technique is used to remove motion redundancies, but the method is not restricted to this technique.

This texture coding strategy has much similarity with a vector quantization approach - where a look-up table with basis functions is provided rather than with picture elements. First the location of the most dominant signals in the residual images are identified and the locations are transmitted to the receiver (Figure 8A). Next, for each location, the most suitable basis functions are searched and matched from a large look-up table. Basis functions can have varying lengths and amplitudes (Figure 8B).

The matching-pursuit algorithm avoids the typical artifacts often apparent with MPEG technology resulting in smoother images which are usually more pleasant to the viewer. A particular disadvantage of the method is the computational burden at the encoder. Depending on the bit rate (or on the number of energy maxima to be coded for a residual image) the computational load at the encoder is increased compared to a DCT approach

between a factor of 2 up to factors of 100 or more. There appears scope for reducing the computational burden.

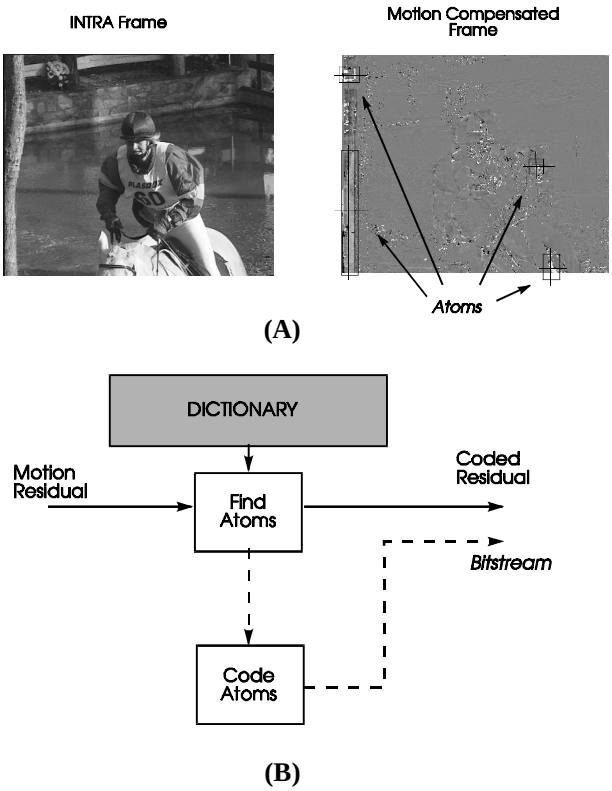


Fig. 8: Matching Pursuit INTER frame texture coding.

VI. Summary and Conclusion

The primary effort for the MPEG-4 standardization process is targeted towards future interactive multimedia video communications calling for content-based functionalities, universal access in error prone environments and high coding efficiency. Besides the provisions for content-based functionalities the MPEG-4 video standard will assist the efficient storage and transmission of video in error prone environments over a range of bit rates between 5 kbit/s and 4 Mb/s.

The MPEG-4 core technology is based on a hybrid MC/DCT approach similar to conventional MPEG-1/2 algorithms with additional provisions for efficient coding of arbitrarily shape content in video sequences. These provisions include techniques for coding shape and transparency information for arbitrarily shaped video objects as well as algorithms for coding Sprites. A number of motion prediction modes are defined which improve coding efficiency for particular scene content over a large range of bit rates, both for generic scene content as well as for scenes with restricted content, i.e. static background.

During the MPEG-4 development process a number of techniques were proposed and investigated in Core Experiments which depart from the conventional algorithms standardized by MPEG-1 and MPEG-2 - and which hold much promise for potential improvements in terms of coding efficiency for video signals for the coming years.

References

[1] R.Schäfer and T.Sikora, „Digital Video Coding Standards and Their Role in Video Communications“, Proceedings of the IEEE, Vol.83, No.6, June 1995.

[2] T.Sikora, „The MPEG-4 Video Standard Verification Model“, IEEE Trans. CSVT, Vol.7, No.1, Feb.1997.

[3] IEEE Trans. Circuits and Systems for Video Technology, "Special Issue on MPEG-4", Vol.7, No.1, Feb.1997.

[4] M.-C.Lee et al., „A Layered Video Object Coding System Using Sprite and Affine Motion Model“, IEEE Trans. CSVT, Vol.7, No.1, Feb.1997.

[5] Nokia, Finl. - M.Karczewicz, „P9: Core Experiment on Motion Compensated Prediction Using Quadtree Segmentation and Polynomial Motion Fields“, MPEG96/M1189, October 1996.

[6] R.Neff and A.Zakhor, „Very Low Bit-Rate Video Coding Based on Matching Pursuits“, IEEE Trans. CSVT, Vol.7, No.1, Feb.1997.