CONTENT-ADAPTIVE VIDEO CODING COMBINING OBJECT-BASED CODING AND H.264/AVC

Andreas Krutz, Matthias Kunter*, Michael Dröse*, Michael Frater**, and Thomas Sikora**

*Communication Systems Group
TU Berlin
Berlin, Germany

**School of IT and EE
University of New South Wales
Canberra, Australia

ABSTRACT

In recent years advanced video codecs have been developed, such as standardized in MPEG-4. The latest video codec standardized, the H.264/AVC, provides compression performance superior to previous standards, but is based on the same basic motion-compensated-DCT architecture. However, for certain kinds of videos, it has also been shown that it is possible to outperform the H.264/AVC using an object-based video codec. The challange now is to develop a general-purpose object-based video coding system. In this paper, we present an automated approach to separate a video scene into shots that are coded either with an object-based codec or the common H.264/AVC. Using this idea of applying different video codecs for different kinds of content, we achieve a higher coding gain for the whole video scene considered. For the first experimental evaluation, we consider a football sequence.

1. INTRODUCTION

During the last two decades, efficient video codecs have been developed and standardized (e.g. MPEG1,2 and 4). Recent video coding research has led to the latest video coding standard H.264/AVC [1], whose compression performance significantly exceeds previous standards. All these codecs rely on the well-known hybrid video coding scheme standardized first in MPEG1.

Object-based video coding using video mosaics provides an alternative approach [2], [3], [4]. In such systems, the video content is segmented into foreground and background objects. A background mosaic is built over a group of frames that have similar background. The mosaic and the segmented foreground objects are coded separately. At the decoder, both are then merged together to the original video. For a certain kind of video content this approach can significantely outperform the common hybrid video coding scheme. There are two critical points applying this kind of coding, the foreground/background segmentation and the generation of the mosaic. A lot of work has been done in this area (e.g. [5], [6]). For coding the separated video data, several coding approaches have been also developed and standardized in MPEG-4 Video.

The biggest drawback of the object-based video coding using background mosaics is the computational cost of the sophisticated pre-processing steps in segmentation and generating a video mosaic at the encoder. Furthermore, the use of different coding algorithms for the mosaic and segmented foreground objects also increases the complexity at the decoder. To make this object-based codec more applicable, a new object-based video codec has been proposed recently for single- and multi-view video [7],[8]. Here, the video mosaic is generated first. An object segmentation algorithm is then applied using the recontructed frames from the video mosaic, integrating the foreground/background segmentation into the mosaic generation. We emphasize that foreground/background segmentation is performed fully automatically. Thus, no user-assisted segmentation is required. For coding the separated video data, the H.264/AVC is used. We show that our proposed approach ouptperforms the H.264/AVC.

The task is now to detect automatically material where the object-based codec can be applied. To achieve this, the content of the considered video has to be analyzed first. We then apply a frame-to-frame image registration algorithm to detect the motion of the camera. The first approach is to use the achieved shot-term motion parameters for further analysis of the video, including shot-boundary detection. These parameters are also used to build a criterion for choosing the most appropriate video codec (i.e. either the object-based codec or H.264/AVC). Every shot is then coded using the video codec decided. At the decoder, all shots are decoded and merged to the original scene. An overview of this content-adaptive video coding system is given in Fig.1.

The remainder of this paper is structured as follows. In Section 2, the motion-based video content analysis is described. Section 3 describes the object-based video codec in more detail. Experimental results are presented in Section 4.

2. VIDEO CONTENT ANALYSIS

The video content analysis used here relies completely on the estimated short-term higher-order motion parameters. In the next two subsections, the shot boundary detection and the codec-decision criterion is described.

2.1. Motion-based Shot Boundary Detection

Much previous work has been done in shot boundary detection. Many recent approaches allow fades and wipes and other effects to be integrated into the shots on either side of the boundary. For our purpose, we need a technique that detects the shot boundary frame- accurately. We define a shot boundary as a point in the video sequence where the mosaic generation needs to be restarted. Based on this definition, it is often desirable to classify inter-shot effects such as fades as separate shots.

The simple shot-boundary-detection algorithm presented here relies only on the estimation of the camera motion. The first step for

The work presented was developed within VISNET2, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

M. R. Frater was supported in this work by the Australian Research Council under project DP0667074

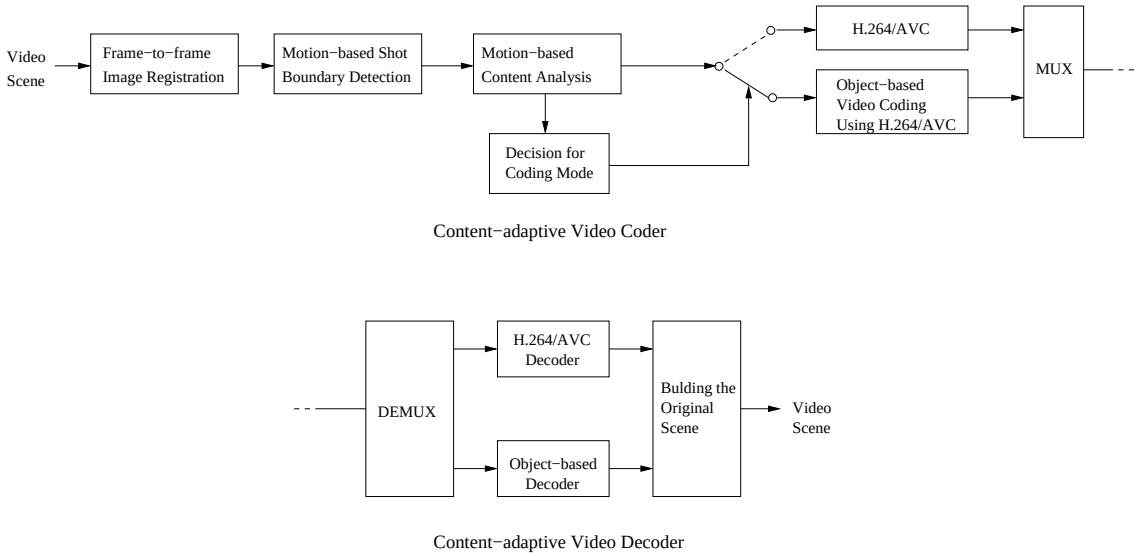

Fig. 1. Content-adaptive Video Codec

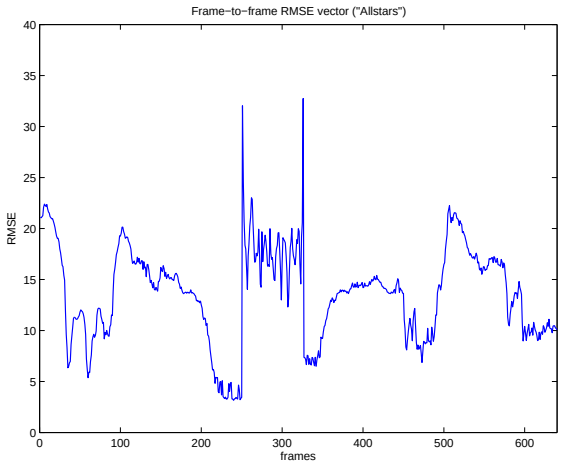

Fig. 2. Frame-to-Frame RMSE evaluation

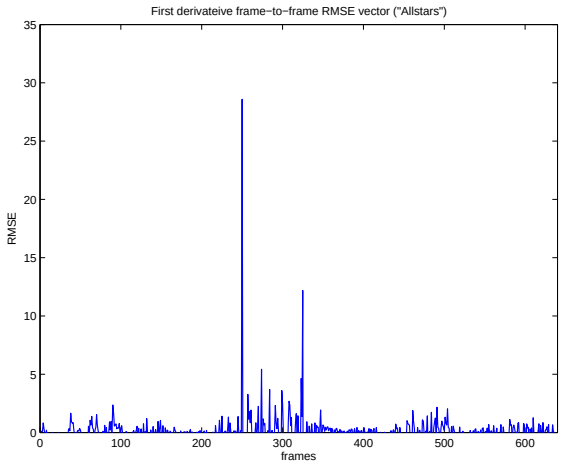

Fig. 3. First numerical derivative of the RMSE in Fig. 2

analyzing the scene is the estimation of the camera motion. We use the perspective motion model, thus 8 motion parameters per frame of the this motion model are calculated using a frame-to-frame image registration method [9]. This motion data is used to generate a prediction of each frame with respect to the previous frame. The $RMSE$ between the frame and its prediction is then calculated. This is done for each frame of the considered scene. The $RMSE$-curve of the whole scene from a football match is depicted in Fig.2. For a better interpretation of the curve, the absolute value of the first numerical derivative is calculated, which is shown in Fig.3. Large differences between two frames can be more easily detected in this derivative. There are two peaks in this curve which indicate that the two frames at these points can not be compensated by the adjacent frame. This means that a shot boundary is very likely. The two peak values are recognized using thresholding. For the threshold, the variance of the differential $RMSE$-vector is calculated. In this case, a tuning factor has to be defined to adjust the threshold. The factor is set to 2. For the considered scene (641 frames), two shot boundaries are detected (frame 319/320 and frames 403/404. Figure 4 shows a keyframe of each shot.

2.2. Coding Criterion

To find a criterion for the decision of the video codec, the differential $RMSE$-curve is considered, as for shot boundary detection. In the case presented above, the sequence is segmented into three shots. We calculate the variance of each curve segment. It can be seen in the figure that the variance of shot 1 and 3 is less than for shot 2. We know from earlier examination that the object-based video codec achieves a higher coding gain for sequences like shot 1 and 3. For shot 2, it is not possible to build a video mosaic because of this very close camera shot with a large foreground object. In that case, it is very hard to segment. Finally, we know from our recent experiments that the background object has to be much larger than the foreground objects to gain more coding efficiency. So we need a criterion which distiguishes shot 1, 3 and shot 2. The variance

(a) Keyframe (shot1)

(b) Keyframe (shot2)

(c) Keyframe (shot3)

Fig. 4. Keyframes of the three shots detected

value of the differential $RMSE$ for shot 1 and 3 are 0.4 and 2.4, respectively. The variance for shot 2 is 15.6. That means that the short-term frame-to-frame image registration is very unstable for the second shot. These motion paramters set up the mosaic generation algorithm and if the accuracy of the background estimation varies in that way an accurate mosaic cannot be generated. Considering these variance values a threshold has to be defined, we calculate the mean of the three variance values in a pre-processing step.

3. CONTENT-BASED VIDEO CODING

This section introduces the object-based coding scheme used and summarizes the complete content-based video coding system.

3.1. Object-based Coding Approach

The object-based video codec (OBVC), which has been presented in [7],[8], combines the advantages of the object-based coding idea using background mosaics and the excellent coding performance of the H.264/AVC. As pre-processing, a video mosaic is generated which contains all the background information of the sequence. By applying a blending technique, nearly all foreground objects can be removed from the background mosaic image. Figure 5 shows the background mosaics for shot 1 and 3 of our considered test scene. The video sequence is then reconstructed from the mosaic and all the frames contain only background information. This background video sequence is used for an in-built foreground/background segmentation algorithm which relies on a background subtraction technique and some further algorithms. The segmentation algorithm is described more in detail in [10]. Having the segmented video data, the background mosaic image, the foreground objects sequence, the foreground/background binary mask (which is needed at the decoder) and the motion parameters (which are not coded) the H.264/AVC is used to code the video segments. At the decoder, the segments are merged together to the reconstructed video sequence.

3.2. The Content-based Video Coding System

All the techniques described are combined in the coding system shown in Fig.1. The coding-mode decision relies only on the camera motion estimation. Two video codecs, object-based video coding (OBVC) and H.264/AVC, are considered for coding shots of a scene separately. The video data is then transmitted and decoded with the related video decoder. Afterwards, the scene is set together from the separated shots. In the next section, the first experimental results are presented and these results show the suitability of content-adaptive coding.

(a) Mosaic (shot1)

(b) Mosaic (shot3)

Fig. 5. Mosaics of shot 1 and 3

4. EXPERIMENTAL RESULTS

The experiments are examined with the football test sequence "Allstars" (704x576 pixels, 641 frames, 30 frames/s). The sequence is coded using the proposed content-adaptive OBVC and only with H.264/AVC. For the AVC, we use the latest examined prediction scheme, hierarchical B-frames, with a GOP of 15 frames. These settings are fixed for the content-adaptive codec and the use of H.264/AVC. Shot 1 (250 frames) and 3 (316 frames) can be coded using the OBVC. Figure 6 and 7 show rate-distortion curves for these two sub-sequences. It can be seen that especially for shot 1 the OBVC achieves a much higher coding performance in comparison to the H.264/AVC. The difference of the $PSNR$-values is up to 3 dB and higher. For shot 3, there is also an imrovement of the coding performance (up to 2 dB), however, here the coding limit is reached earlier because of the presence of more foreground objects in the scene. The shot 2 is coded with the H.264/AVC for both cases, so there is no benefit of our approach over H.264/AVC. Figure 8 shows the rate-distortion curve for the whole scene. Due to the coding gain of shot 1 and 3 of the OBVC, the content-based video coding system outperforms the H.264/AVC over a bit-rate range of up to 250 kbits/s. We achieve gains of up to 2 dB in quality for the same bit rates, or save more than 30% of the bit rate for the same quality. This can be stretched by providing more bits for shot 2 (last point of the curve). It can be seen that despite the limit of shot 3 a coding gain can be held in that range for the whole scene. Figure 9 shows parts of frames taken from the decoded videos from shot 1. It can be seen that the subjective quality as well as the objective quality of the OBVC-coded video is higher than for that coded with the H.264/AVC.

5. CONCLUSION

We have presented an approach for combining two different video coding approaches to outperform the use of only one of them. For a certain kinds of sequences, it is possible to achieve higher coding gain using object-based coding in comparison to the H.264/AVC. In other cases, object-based coding is not possible or brings less coding efficiency and here the H.264/AVC is used. We have shown that this content-adaptive video coding system outperforms the H.264/AVC with the considered test scene. We expect that for many video scenes significant gain in coding efficiency in comparison to only use AVC. A comprehensive experimental evaluation is the prime goal of further work.

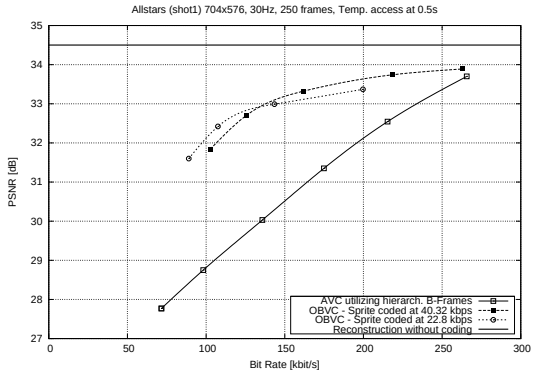

Fig. 6. "Allstars" (shot1)

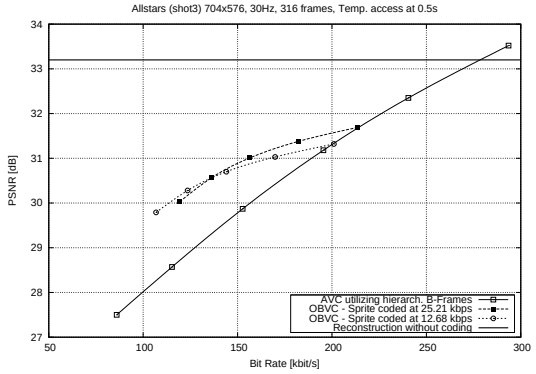

Fig. 7. "Allstars" (shot3)

6. REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 302–311, July 2003.

[2] T. Sikora, "Trends and perspectives in image and video coding," *Proceedings of the IEEE*, vol. 93, pp. 6–17, January 2005.

[3] T. Sikora, "The mpeg-4 video standard verification model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 19–31, 1997.

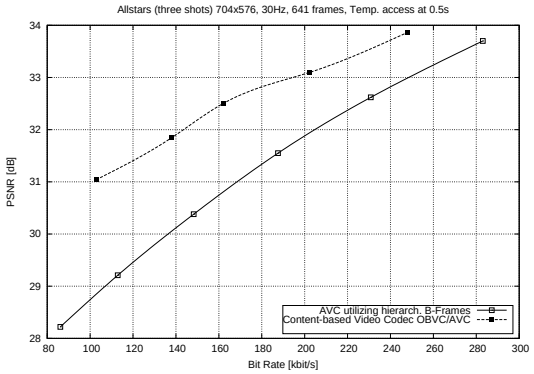

Fig. 8. "Allstars" (complete scene)

(a) PSNR=31.35 dB, R=174.88 kbits/s (H.264)

(b) PSNR=33.32 dB, R=161.64 kbits/s (OBVC)

Fig. 9. Comparison of decoded frames (parts)

[4] A. Smolic, T. Sikora, and J.-R. Ohm, "Long-term global motion estimation and its application for sprite coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1227–1242, December 1998.

[5] D. Farin, P. H. N. de With, and W. Effelsberg, "Video object segmentation using multi-sprite background subtraction," in *Int. Conf. on Multimedia and Expo (ICME)*, Taipei, Taiwan, June 2004.

[6] M. Kunter, J. Kim, and T. Sikora, "Super-resolution mosaicing using embedded hybrid recursive flow-based segmentation," in *IEEE Int. Conf. on Information, Communication and Signal Processing (ICICS'05)*, Bangkok, Thailand, Dec. 2005.

[7] A. Krutz, M. Droese, M. Kunter, M. Mandal, M. Frater, and T. Sikora, "Low bit-rate object-based multi-view video coding using mvc," in *First International 3DTV-Conference*, Kos, Greece, May 2007.

[8] M. Kunter, A. Krutz, M. Droese, M. Frater, and T. Sikora, "Object-based multiple sprite coding of unsegmented videos using h.264/avc," in *IEEE International Conference on Image Processing (ICIP2007)*, San Antonio, USA, Sept. 2007.

[9] A. Krutz, M. Frater, M. Kunter, and T. Sikora, "Windowed image registration for robust mosaicing of scenes with large background occlusions," in *Int. Conf. on Image Processing (ICIP06)*, Atlanta, USA, Oct. 2006.

[10] A. Krutz, M. Kunter, M. Mandal, M. Frater, and T. Sikora, "Motion-based object segmentation using sprites and anisotropic diffusion," in *8th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Santorini, Greece, June 2007.