RATE-DISTORTION OPTIMIZATION FOR
AUTOMATIC SPRITE VIDEO CODING USING H.264/AVC

Andreas Krutz*, Alexander Glantz*, Michael Frater**, and Thomas Sikora*

*Communication Systems Group	**School of IT and EE
TU Berlin	University of New South Wales
Berlin, Germany	Canberra, Australia

ABSTRACT

Sprite-based video coding offers higher compression efficiency than conventional block-based hybrid video coders. In sprite coding a sequence is divided into a model of its background, i.e. a so-called background sprite image, and a foreground object sequence. These are then encoded and transmitted to the receiver. At the decoder the output video sequence is synthesized using content previously transmitted. The background information of the video sequence is, other than the foreground object sequence, encoded indirectly, meaning for a sequence with N frames one single image is encoded and not N background frames. The background for a single frame is reconstructed by coordinate transformation at the decoder. An important issue to address here is an appropriate rate-distortion technique for optimization of the sprite-based coding approach, since foreground and background are encoded independently. In this paper, the Lagrangian cost function is considered for rate-constrained encoder control. Experimental evaluation shows the correct use of the optimization method and the gain in terms of rate-distortion performance compared to H.264/AVC.

Index Terms— Video Coding, Rate-Distortion Theory, Sprite Coding, H.264/AVC

1. INTRODUCTION

Sprite-based or object-based video coding (OBVC) has been developed within the activities for standardization of MPEG-4 Part 2/Visual [1]. The motivation for this approach has been outlined in [2]. In object-based video coding a sequence is divided into a model of its background, i.e. a so-called background sprite image, and a foreground object sequence. These are then encoded and transmitted to the receiver. At the decoder the output video sequence is synthesized using content previously transmitted. Recent work proposed an extension of the state-of-the-art hybrid video coding approach H.264/AVC with sprite coding capabilities and embedded segmentation [3]. This work showed that for a certain kind of input video sequence bit rate savings of up to 50% compared to common H.264/AVC are possible. Background sprite images have been used to improve the coding efficiency especially in lower bit rate ranges.

The behavior of the object-based video coding approach in terms of rate-distortion performance is different compared to common hybrid video coders. This is mainly due to the fact that the background

information of the input video sequence is encoded indirectly. This means that for a sequence with N frames one single background sprite image is constructed and not N background frames or models. A representation of the original content is then synthesized at the decoder by combining the foreground object sequence with a background frame that has been reconstructed from the sprite image using coordinate transformation. Since in the approach used herein, background sprite image and foreground object sequence are encoded independently the question is how to optimize coding in a rate-distortion sense.

A lot of effort has been put into the rate-distortion optimization (RDO) of video coding. Previous work on rate control of object-based video coders deals with MPEG-4 Part 2/Visual [4], [5]. However, the use of sprite images as representation for the background of a video sequence has not been considered therein. The state-of-the-art optimization method for common block-based hybrid video coders is minimization of the Lagrangian cost function. The minimization problem is given by

$$\min\{J\}, \text{ where } J = D + \lambda R, \qquad (1)$$

where D is the distortion between original and reconstructed sequence, R is the bit rate necessary for transmission and $\lambda \geq 0$ is the Langrange parameter. A comprehensive description and evaluation has been published in [6] and [7]. Using the Lagrangian optimization approach, a critical point is the choice of the Lagrange parameter λ. In [6] a relationship between λ and the quantization parameter QP has been found experimentally. This has been further examined in [8]. A theoretical approximation which lead to this relationship has also been described in [6] and [8]. The same experiments have been performed using H.264/AVC leading to slightly different approximations for the Lagrange parameter [7].

Based on the work described above, the Lagrangian optimization approach is taken into account for rate-distortion optimization of the object-based video coding technique. Since the OBVC approach has different behavior to common hybrid video encoders, the optimization process has to be redesigned. This is done in Section 2. In Section 3, experimental results are shown and Section 4 concludes this paper.

2. RATE-DISTORTION OPTIMIZATION FOR AUTOMATIC SPRITE CODING

2.1. General Approach

A block diagram of the object-based video coding approach used for rate-distortion optimization is shown in Fig. 1. It is based on H.264/AVC, which here is treated as a black box. For a given in-
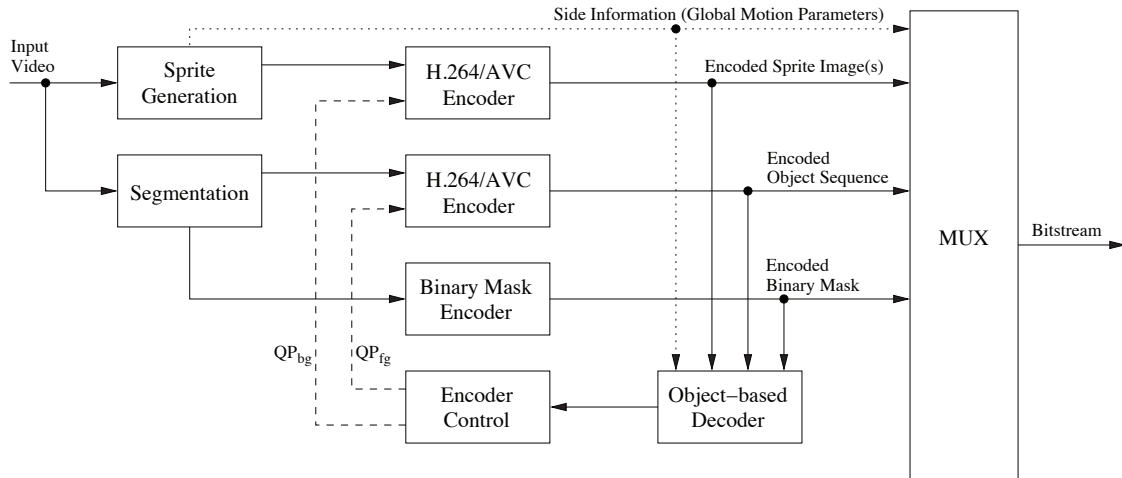
Fig. 1. Rate-distortion optimized object-based video encoder

put video sequence the OBVC encoder first generates a background sprite and a block-based foreground object sequence.

During the generation of the background sprite image, global motion parameters are computed that describe the global camera motion between every two consecutive frames of the video sequence. Here, the global motion parameters are based on the well-known 8-parameter perspective motion model. These so-called short-term parameters are combined to form long-term parameters for frame to sprite coordinate transformation [9]. The sprite image is coded as intra frame using H.264/AVC. The parameters have to be transmitted as side information to the receiver for reconstruction of the video content. For construction of the foreground object sequence, a segmentation step is performed which yields a binary foreground object mask [10]. The mask and the foreground object sequence are then extended to fit the 16×16 macroblock structure of H.264/AVC. The foreground object sequence is coded using common H.264/AVC and additionally the binary foreground object mask is coded using a binary mask encoder [11].

The main coding parameters in object-based video coding are the quantization step sizes for foreground object sequence and background sprite image. These are controlled by the values QP_{fg} and QP_{bg} respectively. The objective is, to set these values in a rate-distortion optimized way. The Lagrangian cost function to be minimized here is

$$J(I_N \,|\, \mathrm{QP}_{bg}, \mathrm{QP}_{fg}, \lambda_{opt}) =$$
$$D_{rec}(I_N \,|\, \mathrm{QP}_{bg}, \mathrm{QP}_{fg}) + \lambda_{opt} R_{rec}(I_N \,|\, \mathrm{QP}_{bg}, \mathrm{QP}_{fg}), \quad (2)$$

where I_N is the input video sequence that comprises of N frames, D_{rec} is the mean squared error between input video sequence and reconstructed video sequence, R_{rec} is the bit rate for the encoded sequence, and λ_{opt} is the optimal Langrange parameter that is yet to be defined. For the object-based video encoder this means that both background sprite image and foreground object sequence have to be locally decoded and a representation of the original content has to be synthesized for the encoder control unit to optimally set foreground and background quantization step sizes. This can be seen in Fig. 1.

The difference to previous approaches using the Lagrangian optimization technique for rate-distortion optimization is that it here
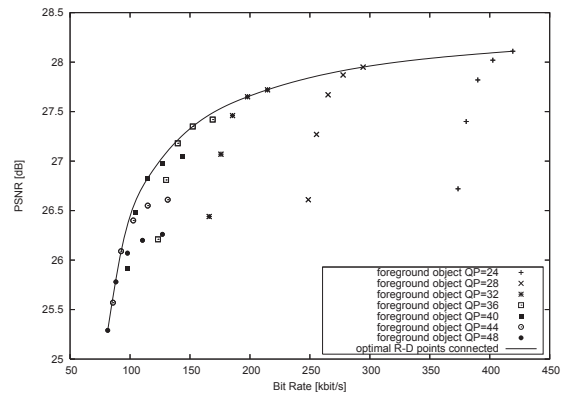

Fig. 2. Convex hull defining optimal R-D points for the product set of foreground object sequence $\mathrm{QP}_{fg} \in \{24, 28, 32, 36, 40, 44, 48\}$ and background sprite $\mathrm{QP}_{bg} \in \{20, 24, 28, 32, 36\}$; optimal points are connected by a line; sequence *Tempete*

is applied on the complete video sequence instead of on single macroblocks. The theoretical assumption leading to the Lagrange parameter is the relation between the quantization step size and the distortion of the signal. This is not completely applicable here, since for the background sprite image this relation does not directly hold: The background sprite image is coded as a single intra frame and the reconstructed signal is built from it using coordinate transformation. However, the functional form of the relationship between λ and QP is expected to be similar.

2.2. Lagrangian Multiplier Selection for Optimal Sprite Coding

In [6], [7], and [8], a set of λ values has been assigned for finding the optimal Lagrange parameter. The normalized distribution of quantization parameters QP has been measured which lead to an average QP for each λ. This approach could then be used to find a relationship between λ and QP and thus derive an appropriate value for the constant c in the equations
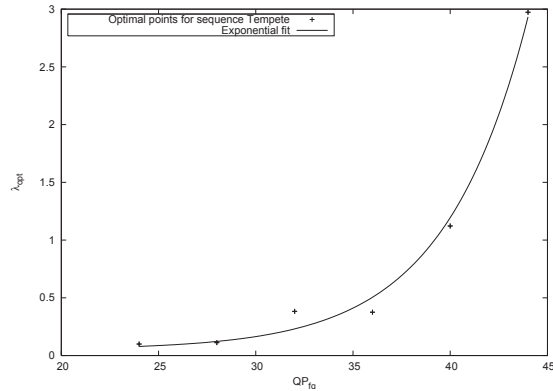
Fig. 3. Experimentally found optimal points for sequence *Tempete* in λ_{opt}-QP$_{fg}$-plane and exponential fit for data (the point for QP$_{fg} =$ 48 has been omitted due to scaling reasons)

$$\lambda_{opt} = c \cdot \mathrm{QP}^2 \qquad (3)$$

$$\lambda_{opt} = c \cdot 2^{\frac{\mathrm{QP}-12}{3}} \qquad (4)$$

For the object-based video coding approach Equ. 4 is used, since it is based upon H.264/AVC and takes into account its logarithmic increase in quantization step size.

For estimation of an optimal Lagrange parameter λ_{opt} a video sequence (*Tempete*) is coded using the OBVC coder shown in Fig. 1 using all possible combinations, i.e. the product space, of QP$_{fg} \in \{24, 28, 32, 36, 40, 44, 48\}$ and QP$_{bg} \in \{20, 24, 28, 32, 36\}$. Analogous to Lagrangian cost function minimization in block-based hybrid video coding, for a given foreground quantization parameter QP$_{fg}$ all possible background quantization parameters QP$_{bg}$ define the "modes". The outcome of this experiment can be seen marked with points in Fig. 2. The convex hull around all points in the rate-distortion plane in Fig. 2 yields an optimal coding point for each QP$_{fg}$.

Finally, each of these optimal coding points is associated with a specific Lagrange parameter λ_{opt}, when minimizing Equ. 2 for the given QP$_{fg}$ and all "modes" QP$_{bg}$. Fig. 3 shows the points in the λ_{opt}-QP$_{fg}$-plane that have been experimentally found (the point for QP$_{fg} = 48$ has been omitted due to scaling reasons). Using a curve fitting approach on the given data yields the following equation for the Lagrange parameter:

$$\lambda_{opt}(\mathrm{QP}_{fg}) = 0.001773 \cdot 2^{\frac{\mathrm{QP}_{fg}-12}{3}} + 0.0508 \qquad (5)$$

To confirm this relationship, an experimental evaluation is presented in the next section.

3. EXPERIMENTAL RESULTS OF OPTIMIZED FOREGROUND/BACKGROUND ENCODING IN OBVC

For experimental evaluation, the rate-distortion optimization technique presented above has been applied using a set of test sequences that are described in Tab. 1. For the H.264/AVC black box the reference software JSVM 9.1 has been used with hierarchical B-frames and CABAC entropy coding enabled. For the calculation of the resulting bit rate 25 frames/s are supposed for the test video sequences and 4.8 kbit/s are used for the transmission of the global motion

Sequence	Description
Allstars	CIF resolution, 250 frames, showing a soccer game
Entertainment	CIF resolution, 250 frames, singer performing on stage
Stefan	SIF resolution, 300 frames, camera following a tennis player
Tempete	CIF resolution, 260 frames, zoom-out,lots of foreground objects

Table 1. Test video sequences

parameters. To set up a fair comparison between the rate-distortion optimized object-based video encoder and H.264/AVC these settings are fixed coding the complete sequence with H.264/AVC and for coding only the foreground object sequence.

All sequences have been coded using combinations from the product space of QP$_{fg} \in \{28, 32, 36, 40, 44, 48\}$ and QP$_{bg} \in \{20, 24, 28, 32, 36\}$. The resulting rate-distortion points can be seen marked in Fig. 4(a) to 4(d). In the same figures, the results using the rate-distortion optimized OBVC can be seen using solid lines. Since nearly all the points lie below these lines, the experimentally found relationship between λ_{opt} and QP$_{fg}$ can be considered correct. This means the rate-distortion optimized object-based video coding environment always leads to the best result in terms of rate-distortion performance.

When compared to coding the test video sequences using H.264/AVC, i.e. the dashed lines in Fig. 4(a) to 4(d), it can be seen that the rate-distortion optimized OBVC outperforms H.264/AVC in lower bit rate ranges. Bit rate savings of up to 50% and an average of 30% can be reached. These improvements are obtained especially in lower bit rate ranges, which is due to the fact that the use of background sprite images for background representation imposes upper PSNR limits: When the background of a specific frame is reconstructed from a single sprite image, distortion is always introduced due to non-ideal interpolation and global motion estimation error.

4. CONCLUSION

In this paper, a rate-distortion optimization technique for sprite-based or object-based video coding has been proposed. The well established Lagrangian optimization method has been applied for a given object-based video encoder. This OBVC uses background sprite images for background representation and synthesizes a representation of the sequence at the receiver using the data previously transmitted. Due to the indirect coding of the background content, a new relationship between the Lagrange parameter λ and the quantization step size for the foreground object sequence, controlled by the quantization parameter QP$_{fg}$, had to be found. For all test video sequences, the rate-distortion optimized object-based video encoder chooses the optimal coding points automatically. Compared to H.264/AVC, bit rate savings of up to 50% can be reached. These improvements are obtained especially in lower bit rate ranges, which is due to the fact that the use of background sprite images for background representation imposes upper PSNR limits. The next step is to apply an enhanced background sprite generation method to increase this limit. In that way, object-based video coding should also result in improved performance in higher bit rate ranges.
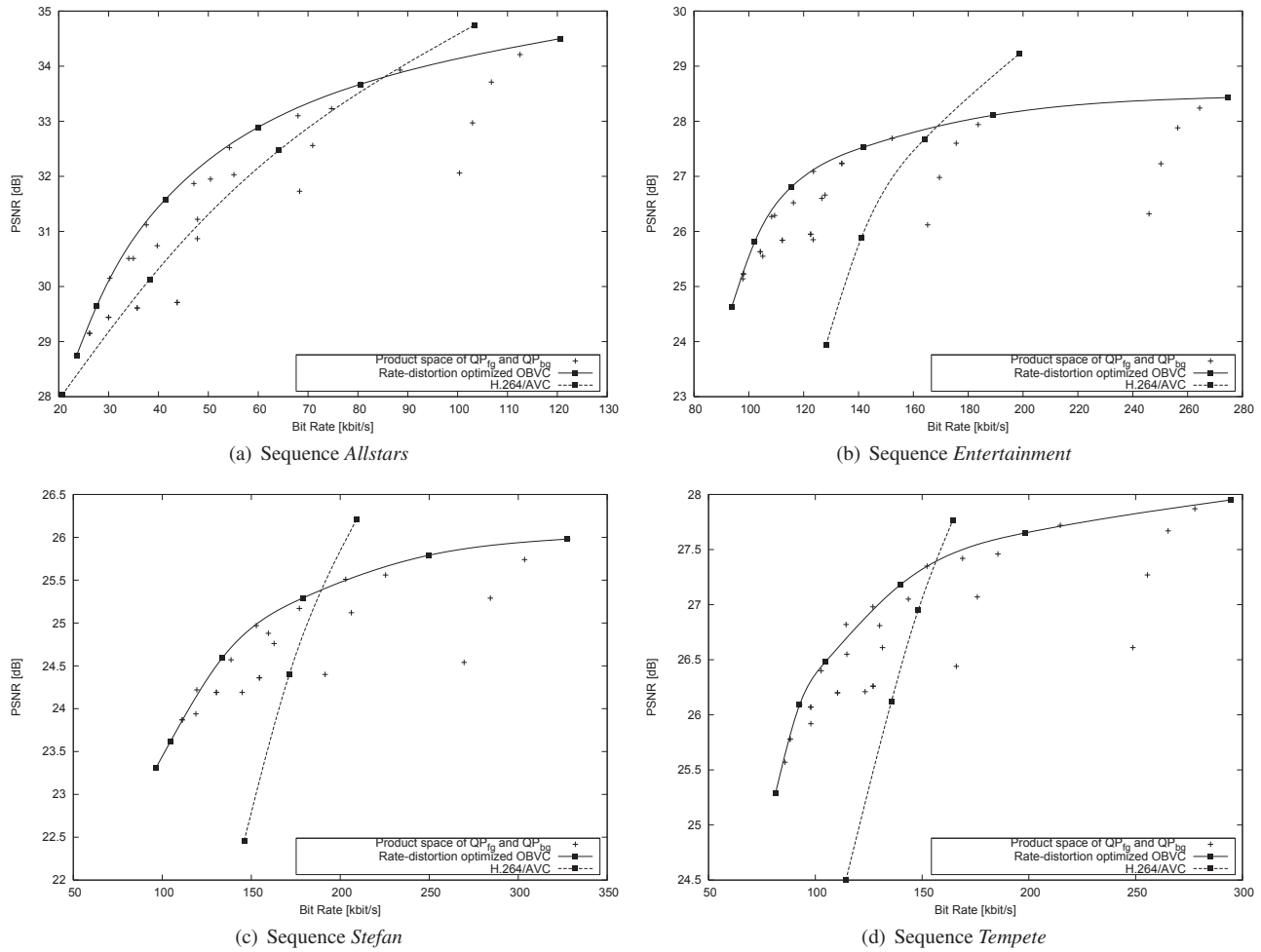
(a) Sequence *Allstars*

(b) Sequence *Entertainment*

(c) Sequence *Stefan*

(d) Sequence *Tempete*

Fig. 4. Results for rate-distortion optimized OBVC in comparison with H.264/AVC using all test sequences

5. REFERENCES

[1] F. Pereira and T. Ebrahimi, *The MPEG-4 Book*, Springer, 2002.

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

[3] 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.

[4] P. Nunes and F. Pereira, "Object-based rate control for the MPEG-4 visual simple profile," in *Proc Workshop on Image Analysis for Multimedia Interactive Services WIAMIS*, Berlin, Germany, May 1999.

[5] J. I. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate control and bit allocation for MPEG-4," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1243–1258, December 1999.

[6] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, November 1998.

[7] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688–703, July 2003.

[8] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *IEEE International Conference on Image Processing (ICIP2001)*, Thessaloniki, Greece, Oct. 2001.

[9] 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.

[10] A. Krutz, A. Glantz, T. Borgmann, M. Frater, and T. Sikora, "Motion-based object segmentation using local background sprites," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, Taipei, Taiwan, Apr. 2009.

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