Automatic Object Segmentation Algorithms for Sprite Coding using MPEG-4

Andreas Krutz[1], Alexander Glantz[1], Thomas Sikora[1], Paulo Nunes[2], and Fernando Pereira[3]

[1] TU Berlin, Communication Systems Group, Einsteinufer 17, EN-1, 10587 Berlin
[2] Instituto Superior de Ciências do Trabalho e da Empresa – Instituto de Telecomunicações, Lisboa, Portugal
[3] Instituto Superior Técnico – Instituto de Telecomunicações, Lisboa, Portugal
E-mail: *{krutz,glantz,sikora}@nue.tu-berlin.de, {paulo.nunes,fp}@lx.it.pt*

Abstract – *Object-based video coding, as standardized in MPEG-4 Part 2, can result in superior performance in comparison to common hybrid motion-compensated DCT-based approaches. We consider sprite coding which increases significantly the objective as well as the subjective quality of the coded video. The main challenge of this approach is the pre-segmentation of the video and the video content itself. To apply sprite coding, the input video has to be firstly segmented into foreground and background objects. We evaluate automatic object segmentation methods based on global motion estimation and background sprite generation. These algorithms are evaluated using the standardized MPEG-4 Visual Main Profile (sprite coding).*

Keywords – *Object-based video coding, MPEG-4 Visual*

1. INTRODUCTION

Sprite Coding [1] was developed and standardized in MPEG-4 some years ago. The main idea of this coding approach is to segment the video content of the input video into two parts, foreground object and background object, in a pre-processing step. Furthermore, a so-called background sprite image is generated which contains all the background information of a certain number of frames of the sequence. Global motion estimation (GME) techniques initiate this sprite generation process. The background sprite image and the foreground objects are then coded separately. At the decoder, original images from the video sequence are reconstructed from the background sprite image (containing only background information). These reconstructed frames are merged with the foreground objects and the whole sequence is decoded. Figure 1 illustrates this approach. The sprite coding within the MPEG-4 standard is described in [2] more in detail. It has been recently shown that the latest standardized motion compensated DCT-based video codec [3] can be improved by these sprite-based object representation of the video sequence and video content [4], [5].

A very important issue for this coding method is the pre-processing step (foreground/background segmentation, GME, and sprite generation). For the global motion estimation and background sprite generation, very popular methods have been proposed in [6]. The biggest challenge is the segmentation problem. It has been assumed in MPEG-4 that the separated foreground and background objects of the video content are known.

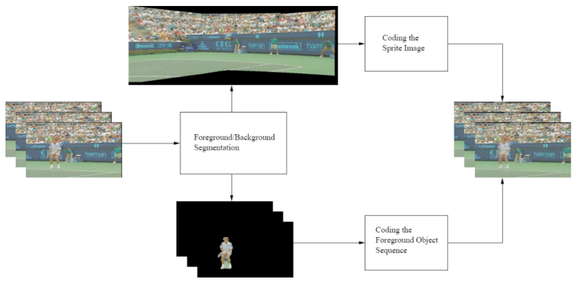

Fig. 1. Simplified Sprite Coding Scheme

There was also work reported which uses pre-processing techniques such as the global motion estimation for automatic object segmentation [9]. Another approach is the use of the already generated background sprite as a background model for a background subtraction method [10]. Inspired by this work, we developed new automatic object segmentaton algorithms for sequences with a moving camera. These algorithms are used in this work to separate the foreground and the background objects of the input video. Afterwards, the segmented video data is coded using the MPEG-4 Main Profile video codec.

The paper is organized as follows. The next Section introduces the pre-processing methods very shortly. Section 3 gives an overview of the applied segmentation methods. Results of the experimental evaluation are outlined in Section 4. The last Section concludes the paper.

2. PRE-PROCESSING TECHNIQUES

2.1 Global Motion Estimation

The performance of the entire sprite coding technique critically depends on the automatic and accurate estimation of the background object motion. Therefore, it is very important to apply an image registration technique with very accurate estimation of the higher-order motion parameters, in our case of the 8-parameter perspective motion model. For this, a gradient-based approach is applied using additional techniques, such as windowed phase correlation based initialization, image pyramid decomposition, and image up-sampling applying wavelet filter. A detailed description of the utilized image registration algorithm can be found in [7].

2.2 Multiple Sprite Generation

Concatenating the short-term perspective camera parameters (homographies $H_{n-1,n}$) in a recursive way yields non-exact long-term parameters representing the transformation $H_{0,n}$ between any frame and the reference frame. These homographies are the base for a robust but coarse camera calibration technique, published in [8]. Here we exploit the fact that for common camera setups the homographies can be decomposed in a product of intrinsic and extrinsic camera parameter matrices

$$\begin{aligned} \mathbf{H}_{0,n} &= \mathbf{F}_n \mathbf{R}_{0,n} \mathbf{F}_0^{-1} \\ &= \frac{1}{\alpha_{0,n}} \begin{pmatrix} r_{00} & r_{01} & f_0 r_{02} \\ r_{10} & r_{11} & f_0 r_{12} \\ r_{20}\alpha_{0,n}/f_0 & r_{21}\alpha_{0,n}/f_0 & r_{22}\alpha_{0,n} \end{pmatrix} \end{aligned} \quad (1)$$

where $R_{0,n}$ is the rotation matrix between frame 0 and n and F_n and F_0 contain focal length values of both views. After computing the focal length ratio $\alpha_{0,n} = f_0/f_n$ we calculate the focal length of the reference frame as the median of all solutions resulting from Eq. (1). This is done by exploiting orthogonality and constant vector norm constraints for the matrices $H_{0,n}$. Knowing all focal lengths the rotation angles can finally be computed using trigonometrically properties of the center points of every image [8]. Figure 2 shows an example of creating multiple sprites of a large camera pan.

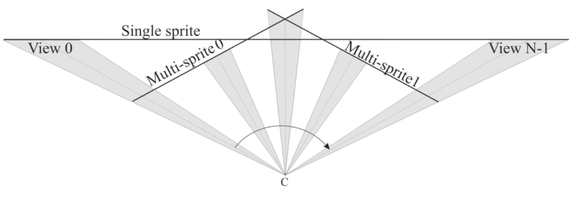

Fig. 2. Partition of a sequence into multiple background sprites for panning camera with constant focal length

3. AUTOMATIC OBJECT SEGMENTATION FOR SPRITE CODING IN MPEG-4 PART 2

In this section we introduce three automatic object segmentation algorithms based on global motion estimation and background sprite generation. All presented algorithms rely on background subtraction after background modeling. The resulted error frame is then processed with a segmentation algorithm proposed in [11]. In this method, an anisotropic diffusion filter is used as a pre-processing step. Afterwards, the filtered error image is converted to a binary image. Common morphological operators remove false detects and reshape the foreground objects. This algorithm is fixed for all considered methods.

3.1. Foreground/Background Segmentation using Global Motion Estimation

The first approach to apply automatic object segmentation is the use of global motion estimation. Here, global motion compensation means that the background motion of two consecutive frames of a video sequence is estimated. If an error image is calculated using the reference frame and the compensated frame, the background pixels are ideally removed completely. However, the foreground object regions appear from the reference and the compensated frame in the error image. To extract the foreground objects more precisely, two error images are taken into account: i) the error image calculated by the reference frame and the compensated previous frame, and ii) the error image calculated by the reference frame and the next frame. The segmentation algorithm described above is applied on these two error frames. Two binary masks are obtained and combined together by an AND-operator to achieve the final foreground object mask. Figure 3 illustrates the whole processing chain.

This approach is inspired by earlier work where only global motion estimation techniques where used for the pre-pocessing step [9].

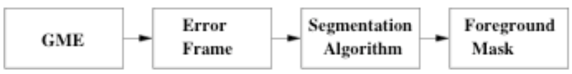

Fig. 3. GME-based object segmentation (Alg.1)

3.2. Global Motion Estimation and Multiple Background Sprites for Segmentation

The already generated background sprites can also be used as a background model for object segmentation. A very promising method has been proposed in [10]. Here, the efficiency of generating multiple sprites for segmentation has been outlined. In [8], a multiple sprite generation algorithm based on camera calibration has been published. This multiple background sprites are used together with short-term global motion compensated error images to calculate a more precise binary object mask. Figure 4 shows the block chart of this method. Advantages of both techniques, i.e., GME-based and background sprite-based, are combined. A detailed description for this segmentation approach can be found in [11].

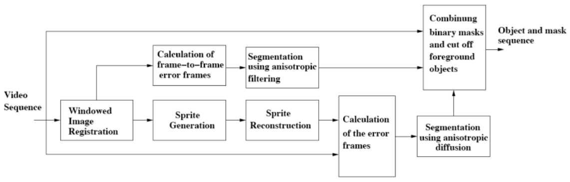

Fig. 4. Object Segmentation combining GME and Background Sprites (Alg.2)

3.3. Background Modeling using "Local" Background Sprites

The third object segmentation algorithm used includes a new background modeling. In practice, accumulated long-term parameters, which set up the sprite generation process, produce distortions in the background sprite image. Furthermore, the warping process has to be applied two times, first to align each considered image to the background sprite plane and second to reconstruct the image from the background sprite plane. For the new technique, we generate so-called "local" background sprites. That means, we produce background sprites for each frame of the video.

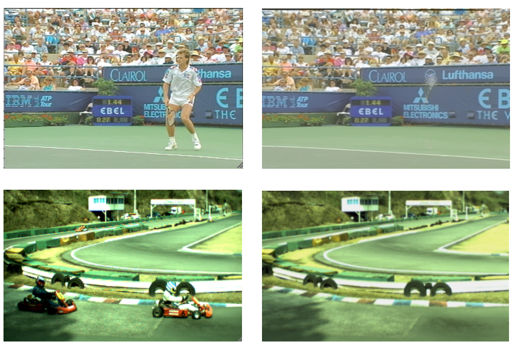

Fig. 5. "Local" background sprites and correspon-ding original frames ("Stefan" top, "Race1" bottom)

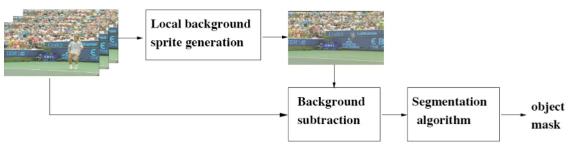

Fig. 6. Object Segmentation using "local" back-ground sprites (Alg.3)

During the generation process, we only consider pixels which fall in the current reference frame. Using this method, we are able to produce highly accurate background models for each frame of the sequence. Figure 5 shows two examples. The whole segmentation algorithm is then processed in a common background subtraction scheme (see Figure 6).

4. EXPERIMENTAL EVALUATION

For the experimental evaluation, we consider two MPEG test sequences. The first is the well-known "Stefan" sequence, 352x240 pixel, 300 frames. The second test video is the first view of the multi-view sequence "Race" (Race1), 544x336 pixel, 100 frames. For the "Stefan" sequence, multiple background sprites and single sprites are generated. For "Race", the camera pan is not high enough that multiple sprites are necessary. Therefore, only a single background sprite is taken into account. We applied the automatic object segmentation methods described for setting up the MPEG-4 sprite coding. We used the Main Profile of the MPEG-4 codec reference software. For this evaluation, we kept the quantization parameter constant for the background sprite images and varied the parameter for the foreground object sequence. We coded the sequences using MPEG-4 simple profile for comparison. For the prediction scheme, we used an *IPPP* – structure for both basis codecs to ensure comparability. Figure 7 and 8 show rate-distortion curves of the two test sequences considered. It can be seen that the latest segmentation algorithm (Alg.3) using "local" background sprites approaches the performance of one using the hand-segmented ground truth mask (see Figures 7 and 8, "GT"). For the "Race1" sequence, this algorithm achieves almost the same results as the ground truth mask. We achieve bit rate savings of up to 65% when applying the Alg.3 and the multiple sprites for the "Stefan" sequence and more than 65% for "Race1" sequence using Alg.3 and single sprites in comparison to the MPEG-4 Simple Profile coding. To this end, we have achieved the almost perfect object segmentation for sprite coding with the considered test sequences.

5. CONCLUSION

We have proposed automatic object segmentation algorithms for video sequences with a moving camera towards automatic object representation for sprite coding as standardized in MPEG-4 Part 2. We have achieved outstanding results with our latest segmentation method. Having the ability to segment the video content in several objects, sprite coding becomes more and more applicable.

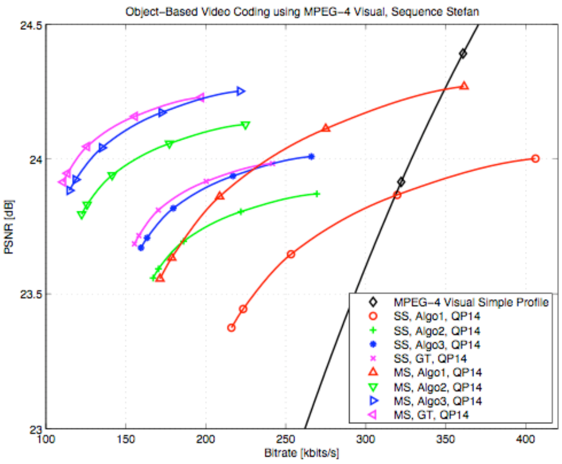

Fig. 7. Comparing MPEG-4 visual simple profile with Main Profile applying different types of background sprites and different types of segmentation algorithms ("Stefan")

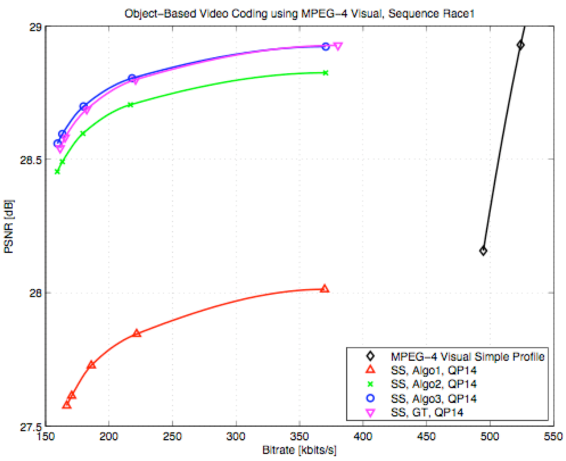

Fig. 8. Comparing MPEG-4 visual simple profile with Main Profile applying different types of segmentation algorithms ("Race1")

ACKNOWLEDGEMENT

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.

REFERENCES

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

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

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)*, Vol. 13, No. 7, July 2003.

[4] M. Kunter, A. Krutz, M. Droese, M. Frater, and T. Sikora, "Object-based multiple sprite coding of unsegmented videos using H.264/AVC", *IEEE Int. Conf. on Image Processing (ICIP),* San Antonio, USA, September 2007.

[5] A. Krutz, M. Kunter, M.Droese, M. Frater, and T. Sikora, "Content-adaptive video coding using object-based video coding and H.264/AVC", *Picture Coding Symposium (PCS),* Lisbon, Portugal, November 2007.

[6] A. Smolic, T. Sikora, and J.-R. Ohm, "Long-term global motion estimation and its application for sprite coding", *IEEE TCSVT*, Vol. 9, No. 8, July 1998, pp. 1227-1242.

[7] A. Krutz, M. Frater, M. Kunter and T. Sikora, "Windowed Image Registration for Robust Mosaicing of Scenes with Large Background Occlusions", *IEEE ICIP,* Atlanta, USA, October 2006

[8] M. Kunter, A. Krutz, M. Mandal, T. Sikora, "Optimal multiple sprite generation based on physical camera parameter estimation", *Visual Communication and Image Processing (VCIP'07),* San Jose, USA, January 2007.

[9] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera", *First Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS),* Louvain-la-Neuve, Belgium, June 1997.

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

[11] A. Krutz, M. Kunter, M. Mandal, M. Frater, and T. Sikora, "Motion-based object segmentation using sprites and anisotropic diffusion", 8^{th} *WIAMIS,* Santorini, Greece, June 2007.