MOTION-BASED OBJECT SEGMENTATION USING LOCAL BACKGROUND SPRITES

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

*Communication Systems Group
Technische Universität Berlin
Berlin, Germany

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

ABSTRACT

It is well known that video material with a static background allows easier segmentation than that with a moving background. One approach to segmentation of sequences with a moving background is to use preprocessing to create a static background, after which conventional background subtraction techniques can be used for segmenting foreground objects. It has been recently shown that global motion estimation and/or background sprite generation techniques are reliable. We propose a new background modeling technique for object segmentation using local background sprite generation. Experimental results show the excellent performance of this new method compared to recent algorithms proposed.

Index Terms— Object segmentation, background modeling, mosaicing, global motion estimation

1. INTRODUCTION

Segmentation of moving objects in video sequences is an important research field. Application scenarios in which object segmentation algorithms are used range from surveillance systems to low-level preprocessing for extraction of semantic information from video sequences. For some applications the cameras used to capture the video content are static. In many other applications, image processing techniques must cope with moving cameras. The potential for using background sprites in object-based video coding has been summarized in [1]. However, the separation of foreground objects from the background remains an open issue. During the MPEG-4 standardization, several methods were proposed using e.g. motion, color or texture [2]. For segmentation of moving foreground objects with a moving camera, a global motion estimation algorithm with a higher-order motion model has been used [3]. A background sprite generated over a certain number of frames can be used as a background model for the segmentation step. An approach using multiple background sprites has demonstrated that this kind of background model is very promising [4]. An enhanced object-based video codec using background sprites and H.264/AVC video coding has been introduced [5].

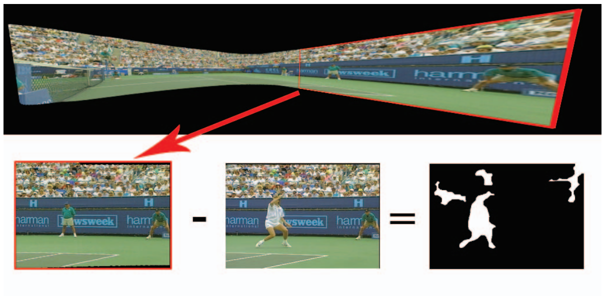


Fig. 1. Problems using common background sprites for object segmentation

However, the mapping of pixel content from various frames in a scene into a single sprite or a collection of multiple sprites may cause severe geometrical distortion of the background especially in border regions. For reconstruction of the background of a single frame a second mapping needs to be performed which causes additional distortion which is due to non-ideal interpolation or registration errors. Object segmentation using background subtraction is one application that is degraded by such distiortion. These problems are exemplified in Figure 1. In the proposed algorithm a mapping of content from many frames in a scene is performed for each individual frame for background construction. In other words, global motion estimation is performed from many adjacent frames into the frame where the background needs to be reconstructed. No backward mapping is required. Thus, our background sprites are local and there are as many individual sprites generated as frames exist in a sequence. This will result in a more precise background reconstruction compared to conventional global sprites.

This paper is organized as follows. Section 2 describes the new background modeling approach. Section 3 describes the segmentation of error frames used. Section 4 provides experimental evaluation of the new approach and Section 5

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

We would like to thank Dr. M. Kunter and Dipl.-Ing. M. Karaman for the technical support and discussions.

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

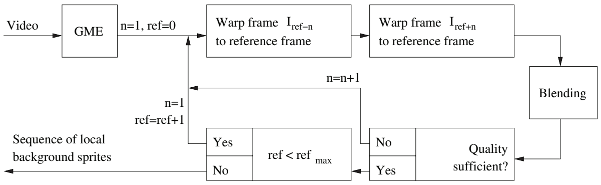

Video
GME
n=1, ref=0
Warp frame I_{ref-n} to reference frame
Warp frame I_{ref+n} to reference frame
Blending
n=n+1
n=1
ref=ref+1
Sequence of local background sprites
Yes
No
ref < ref $_{max}$
No
Yes
Quality sufficient?

Fig. 2. Algorithm for generation of local background sprites

summarizes the paper.

2. LOCAL BACKGROUND SPRITES

A local background sprite specifies a model of the background of a video sequence. Other than common background sprites one model is built for every frame and not one model for the whole sequence. This means there exist as many background models as frames in the sequence. For background model generation only the local temporal neighborhood is taken into account and the size of a local background sprite matches the size of its corresponding reference frame. The idea is to minimize geometrical distortion and to avoid distortion resulting from a second mapping into the reference frame's coordinate system which is needed when using common background sprites. A block diagram for local background sprite generation is shown in Fig. 2. Its different parts are explained in this section.

2.1. Global Motion Estimation

For global motion estimation a hierarchical gradient descent approach based on the Gauss-Newton method is applied as used in [6]. The algorithm estimates the displacement between every two consecutive frames in a sequence using an 8-parametric higher-order motion model describing the motion by means of translation, scaling, rotation, sheering and perspective transformation. The background modeling algorithm uses these short-term parameters to compute long-term parameters related to the current reference frame. Since the short-term parameters are used several times while creating all local background sprites, they are computed in a preprocessing step.

2.2. Warping and Blending

For every reference frame a local background sprite is built. The algorithm iteratively transforms temporally neighboring frames into the coordinate system of the reference which produces a dynamically growing image stack. This approach can be seen in Figure 3.

In every step the images in the stack are merged together to build a preliminary local background sprite. For this purpose a so-called blending filter is used which is implemented here as a median filter.

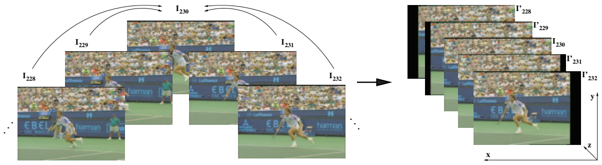
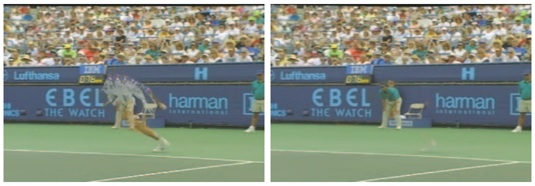


Fig. 3. Warping and Blending



(a) Step $t = 2$ (b) Step $t = 8$

Fig. 4. Preliminary local background sprites, sequence "Stefan", reference frame 230

By successively adding temporally neighboring frames the foreground objects in the preliminary local background sprites are removed step by step. In Figure 4 one can see two preliminary local background sprites for the "Stefan" sequence. The foreground object has nearly completely vanished after eight blending steps. An approach for an adaptive break-up criterion is presented next.

2.3. Quality Evaluation of Local Background Sprites

A possible measure for the difference between two images is the root mean square error (RMSE). The RMSE between a reference frame $I_{ref}(x, y)$ and its preliminary local background sprite $I_{bs,t}(x, y)$ in step t is defined by

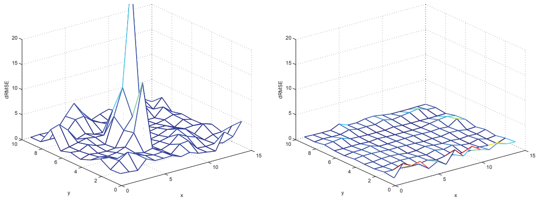
$$RMSE_t = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I_{ref}(i,j) - I_{bs,t}(i,j))^2} \quad (1)$$

where M and N are the dimensions of the reference frame and the preliminary local background sprite respectively. Since the foreground objects vanish step by step the RMSE value increases successively. Therefore, the difference of the RMSE values in two consecutive steps

$$\Delta RMSE_t = RMSE_t - RMSE_{t-1} \quad (2)$$

decreases. When the foreground objects are completely eliminated, the values $RMSE_t$ and $\Delta RMSE_t$ change only marginally.

Additionally, we define matrices containing the blockwise calculated value $\Delta RMSE_t$, which we call dRMSE-matrices. Reference frame and preliminary local background sprite are divided into blocks of fixed size. The value $\Delta RMSE_t$ is then calculated for every block independently. Therefore, no averaging over the whole frame takes place. Distinct areas



(a) Step $t = 2$

(b) Step $t = 8$

Fig. 5. dRMSE-matrices using blocks of size 25×25, sequence “Stefan”, reference frame 230

in the preliminary local background sprite can be evaluated independently. Figure 5 shows the corresponding matrices for the example in Figure 4. After eight blending steps the matrix is nearly flat in all regions. This corresponds with the results in Figure 4.

The generation of the local background sprite is completed when there is no more information added in any region, meaning the dRMSE-matrix is flat in every region. We use the maximum value of these matrices as break-up criterion. Generation of a local background sprite is stopped when its maximum value lies below a predefined threshold.

3. OBJECT SEGMENTATION

The foreground object segmentation is based on a background subtraction approach. Subtracting the background model from the original frame produces an error frame. This error frame ideally has high values in foreground regions and near-zero values in background regions. To produce a binary foreground object mask the algorithm first uses an anisotropic diffusion filter to smooth the error frame while preserving sharp edges. After rescaling of all values the error frame is binarized using an adaptive threshold. Several morphological operators eliminate problems like holes in foreground objects and remove very small objects that have been erroneously segmented.

This segmentation algorithm is identical for all background modeling techniques compared in the next section. For further information refer to [7].

4. EXPERIMENTAL EVALUATION

We have evaluated the new approach using three test sequences. The first sequence is called “Mountain” (352×192, 100 frames) and is from a BBC documentary showing a leopard chasing an animal. The second sequence is called “Race1 (View 0)” (544×336, 100 frames) and is part of an MPEG multiview test sequence showing a kart race. The third sequence is the well-known “Stefan” sequence (352×240, 300 frames).

Firstly, we evaluated the quality of our background models compared to common background models. We consider three types of common background models – single, multiple [5] and super-resolution background sprites [8]. We therefore compute the background PSNR between the original frames and the background model using manually segmented groundtruth masks to only take background pixels into account. Table 1 shows the mean background PSNR for the three test sequences used. One can clearly see that the proposed method highly outperforms common background models by up to 6 dB.

Sequence	Background Model	PSNR [dB]
Mountain	Single sprite	28.8877
	Local background sprites	**35.1592**
Race1	Single sprite	30.1350
	Local background sprites	**34.6666**
Stefan	Multiple sprites	24.7312
	Super-resolution sprite	27.4404
	Local background sprites	**29.5146**

Table 1. Mean background PSNR comparing reconstructed single/multiple/super-resolution sprites with local background sprites

For evaluation of automatic foreground object segmentation quality we compare the proposed algorithm to two other approaches. The first algorithm (Algo1) is inspired by [3]. For a given reference frame global motion estimation and compensation are performed using its predecessor and successor frames. This produces two error frames which are segmented using the algorithm from Section 3. The two resulting binary masks are then combined using a pixel-wise logical AND-operation. The second algorithm (Algo2) combines an object mask provided by global motion compensated frames similar to the first algorithm with a background subtraction method using reconstructed background sequences from common background sprites [7]. For all three test sequences we produced manually segmented foreground object masks i.e. groundtruth masks. The used evaluation metrics are the well-known precision, recall and F_1-measure.

Table 2 shows the mean precision, recall and F_1-measure for all three test sequences. It can be clearly seen that for all test sequences the proposed automatic segmentation algorithm performs best by means of F_1-measure. The gain lies between 4%-25% compared to the first algorithm and 3%-6% compared to the second algorithm. Furthermore, the mean recall value increased significantly for all test sequences. Considering the object-based video coding application, it is of at most importance to have a very accurate segmentation of moving foreground objects. Achieving such high recall values, we expect to improve the overall quality of our object-based video coding system. Figure 6 shows examples of automatically segmented foreground objects using the proposed algorithm. Especially the high-frequency background of the

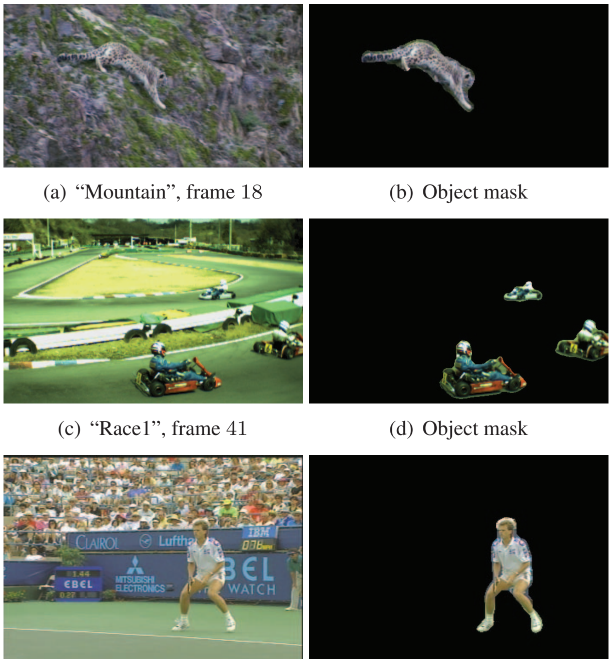
Sequence	Algorithm	P	R	F
Mountain	Algo1	0.83	0.85	0.84
	Algo2	0.85	0.86	0.85
	Proposed	0.86	0.90	**0.88**
Race1	Algo1	0.81	0.70	0.75
	Algo2	0.91	0.88	0.89
	Proposed	0.88	0.94	**0.91**
Stefan	Algo1	0.60	0.78	0.63
	Algo2	0.88	0.77	0.82
	Proposed	0.85	0.92	**0.88**

Table 2. Mean precision (P), recall (R) and F_1-Measure (F) for the test sequences used

"Stefan" sequence could be removed almost completely for the whole sequence.

5. SUMMARY

We have presented a novel background modeling approach for video sequences with moving camera. Common background sprites model the background of a sequence in one image which usually is of large size. Additionally a second mapping from the model into the coordinate system of the reference frame is needed. These drawbacks produce distortion that are avoided using the background modeling approach presented. It has been shown that the quality of the background model using this new technique clearly leads to improved object segmentation results compared to common background models when used in a background subtraction method.



(a) "Mountain", frame 18

(b) Object mask

(c) "Race1", frame 41

(d) Object mask

(e) "Stefan", frame 149

(f) Object mask

Fig. 6. Examples for automatically segmented foreground objects using the proposed algorithm

6. REFERENCES

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

[2] A.A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services-the european cost 211 framework," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, no. 7, pp. 802–813, Nov 1998.

[3] R. Mech and M. Wollborn, "A noise robust method for segmentation of moving objects in video sequences," *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 4, pp. 2657–2660 vol.4, Apr 1997.

[4] 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.

[5] 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 (ICIP'07)*, San Antonio, USA, Sept. 2007.

[6] A. Krutz, M. Frater, and T. Sikora, "Improved image registration using the up-sampled domain," in *Int. Conf. on Multimedia Signal Processing (MMSP'06)*, Victoria, Canada, Oct. 2006.

[7] 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.

[8] 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.