HIGH-RESOLUTION VIDEO MOSAICING

Aljoscha Smolic and Thomas Wiegand
Heinrich-Hertz-Institute (HHI)
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany
{smolic/wiegand}@hhi.de

ABSTRACT

The generation of high-resolution mosaics from multiple images is presented. The estimation of the high-resolution signal utilizes a motion-compensated filtering approach for video frames that are affected by spatial aliasing. For that, global motion models and the corresponding estimation algorithms are employed to provide a very accurate and continuous description of sub-pixel motion for the background of the video signal. The motion information is exploited to generate mosaics with four or sixteen times the resolution of the constructing video sequence or known mosaicing techniques. The mosaicing results indicate that the presented multi-frame methods provide superior visual quality. The approach is further extended to the generation of high-resolution video.

1   INTRODUCTION

Mosaicing is an important task in a variety of video processing applications. Video mosaics can be employed for segmentation, coding, video editing (e.g., stable replacement of parts of the background), content description, and for generating panoramic views of the whole background of a scene. Another application that is additionally provided with high-resolution mosaics is up-sampling of coded or non-coded video signals and format conversion.

An increasing demand for the latter is driven for instance by the convergence of traditional TV-broadcast, computer systems and (mobile) communications. This implies that a great variety of different video capturing devices, transmission channels, databases, storage media, terminals, and displays have to work together. Such a scenario has been addressed in [1], where images taken from a hand-held low-resolution video camera are resolution-adaptively embedded into an image mosaic. For that, the various images are categorized in different resolution classes, which are used to always patch the highest resolution image into the corresponding part of the mosaic.

The method presented here also relies on the assumption that global motion relates large portions of the observed video frames to each other and that the global motion is caused by camera operation. But in contrast to [1], we don not rely on various resolution classes that provide higher resolution patches of the video mosaic. Our high-resolution approach is more in spirit of the super-resolution approaches presented in [2], [3], [4]. The idea is to use image information from several frames for the interpolation of sub-pixel intensity values, exploiting sub-pixel motion information. Such an approach is also used in [3], [4] for instance for interlaced to progressive conversion. Effects on prediction and coding have been studied in [5]. More

precisely, the approach in [2], [3], [4] is to apply a dense motion vector field of translational displacements for estimation of a high-resolution image, with the advantage that it can be used with any kind of video material without a priori knowledge, since the model assumptions are not restrictive.

The performance of motion-compensated interpolation algorithms highly depends on very reliable and accurate motion description and parameter estimation. But a reliable and accurate estimation of a dense motion vector field is very difficult in many cases and typically contains many wrong estimates that do not coincide with the true motion. Moreover, the probability for such errors increases with the temporal distance between the two involved frames, which can be quite large in a multi-frame interpolation scenario. Further most of the approaches have a restriction of the accuracy of the sub-pixel motion description to typically half-pixel or quarter-pixel resolution.

The idea presented in this paper is to use the sub-pixel motion information of a global motion model to interpolate mosaics with a resolution that is higher than the one of the video sequences they are composed of, providing a combination of video format conversion and mosaicing. For regions that comply with the assumed global model, the motion description is very accurate and continuous with no restriction to a particular discrete sub-pixel accuracy. Estimation algorithms that provide very accurate and robust results have already been developed [6], [7], [8] and are employed in this paper. The next section introduces the estimation of high-resolution mosaics followed by a section describing experiments that verify the efficiency of the proposed techniques.

2   HIGH-RESOLUTION MOSAICING

The process of mosaicing denotes the transformation and blending of all images of a considered video sequence into a common reference coordinate system as delineated in Fig.1.
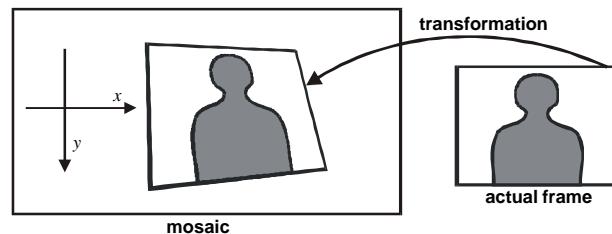

Fig.1 Process of mosaicing: warping and blending all frames of a video sequence towards a common refer-

ence system, controlled by estimated global motion parameters.

The common coordinate system can be either the local pixel coordinate system of one of the images or a further transformed coordinate system. In any case, accurate long-term motion parameters with respect to the common coordinate system have to be determined.

Hence, one prerequisite for the estimation of mosaics from multiple frames is an accurate description of motion [6]. If the camera operation is restricted to zoom and rotation (i.e. no translation), the global motion can be exactly described by the perspective motion model:

$$x_1 = \frac{a_1 + a_2 x_0 + a_3 y_0}{1 + c_1 x_0 + c_2 y_0}$$
$$y_1 = \frac{b_1 + b_2 x_0 + b_3 y_0}{1 + c_1 x_0 + c_2 y_0}. \qquad (1)$$

These well known equations describe the transformation of a position in the reference image (x_0, y_0) to the position in the actual image (x_1, y_1). The transformation is controlled by a set of motion parameters with elements a_i (influencing x-coordinates), b_i (influencing y-coordinates), and c_i (influencing both coordinates).

For reliable and very accurate estimation of the motion parameters, we utilize the iterative minimization of a postulated error measure with respect to the motion parameters. For details see [6], [7]. The stability and reliability of this gradient-based algorithm is provided by incorporation of robust M-estimators to cope with foreground objects as well as initial estimation of translation or processing in a resolution pyramid to cope with large displacements. Further, a recursive approach provides increased long-term stability.

Since the motion equations (1) represent continuous functions, mosaicing requires interpolation of sub-pixel intensity values. Despite usage of sophisticated interpolation filters, mosaics tend to look quite blurry [6]. This can be avoided by high-resolution video mosaicing. Figure 2 illustrates the approach. The pixels of the video frames are transformed into a mosaic of double resolution in both directions, controlled by the estimated motion parameters. First, the frame at time instant t_0 is written into the mosaic of double resolution leaving empty pixel positions in the mosaic at time instant t_0. Second, a half-pixel diagonal shift from frame t_0 to frame t_1 is assumed. In this case, the corresponding pixels in the mosaic are filled since they fall directly onto an integer-pixel position in the mosaic. In practice we use a tolerance range of e.g. ± 0.2 pixel units in the mosaic. This rule is employed to preserve the original sharpness of the video sequence by avoiding spatial interpolation. For a theoretical motivation of our approach see [9]. The remaining empty pixel positions are filled when processing more images.
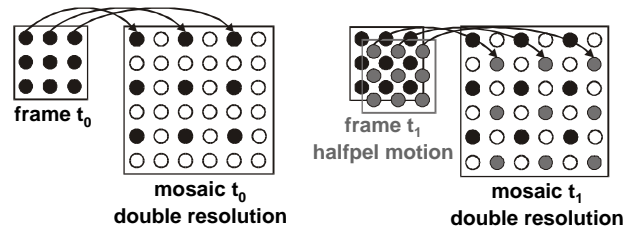

Fig.2 High-resolution mosaicing: transformation of video pixels into mosaic of double resolution in both directions, without interpolation of intensity values.
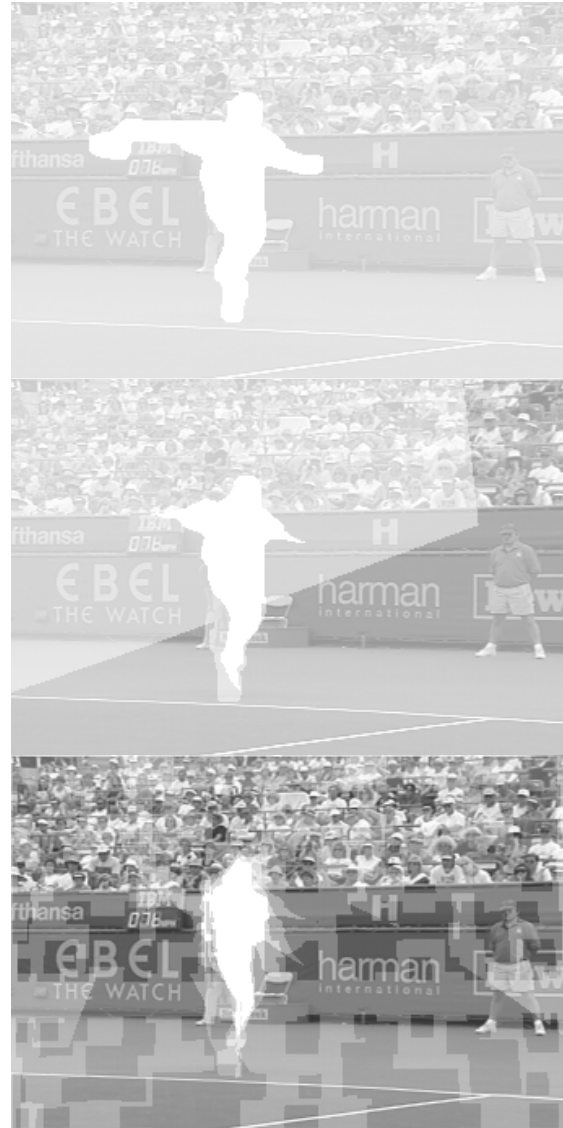
Fig.3 Successive filling of a mosaic of double resolution in both directions (704x480 pixel), for sequence Stefan (SIF, 352x240 pixel). Results when processing 1, 5, 15, 49 frames are shown.

Figure 3 illustrates the high-resolution mosaicing process for a realistic scenario. In the beginning (top-left image in Fig.3), the mosaic is quite sparse since only one frame is written into the mosaic leaving the sub-pixel positions of the mosaic empty. Typically, the multiple frames that are used to generate the mosaic are either temporally backwards or centered around the first frame having the same coordinate system as the mosaic. The number of images for the generation of the high-resolution mosaic depends on the global motion of the video sequence and the up-sampling factor. Thus our approach would fail, if only full pixel translational motion is contained in the video sequence. The remaining three images of Fig.3 are then estimated via applying the high-resolution algorithm as given above.

Remaining holes in the high-resolution mosaic are filled by standard bilinear interpolation. Finally foreground objects that have to be up-sampled using other methods, are pasted over the high-resolution mosaic. It should be noted that a high-resolution mosaic can be generated for every frame of a video sequence, finally resulting in a high-resolution video sequence.

3   EXPERIMENTAL RESULTS

We have tested the presented algorithm with a variety of sequences. The results obtained for the sequence Mobile & Calendar (SIF, 360x240 pixel, 30 Hz, 300 frames) are shown in Fig. 4.[1] It shows a mosaic with 16 times the resolution of the video sequence of 300 frames. The coordinate system of the mosaic corresponds to the coordinate system of the middle frame with number 150. For mosaic generation, global motion is estimated over a long period of time as described above. The accuracy of our estimation technique is validated by the fact that the high-resolution mosaic contains no artifacts. The high-resolution mosaic summarizes the whole background information that becomes visible. Holes remain especially near the borders, since these regions were only sparsely covered by video images. The characteristic patterns of holes result from global motion characteristics (pure pan and perspective model) interfering with the up-sampling grid.

Figure 5 compares a detail from the high-resolution mosaic of Mobile & Calendar with the corresponding high-resolution detail generated with bilinear interpolation of the original image. The high-resolution video mosaicing method clearly shows superior visual quality. It is much sharper, contains much more details, appears much less blocky, and most of all aliasing is highly reduced.

The difference is also visible when the corresponding high-resolution video sequences are compared. For the high-resolution mosaic, disturbing artifacts due to aliasing are drastically reduced. In fact the gain of the presented approach relies on the presence of aliasing in the source material. In experiments with low-passed video we have found, that the gain decreases with the cut-off frequency of the filter.

4   CONCLUSIONS

We have presented an algorithm for generation of high-resolution mosaics that exploits the very accurate and continuous motion description of a perspective global motion model. The approach taken extends the process of robust global motion estimation and video mosaicing. For video sequences with camera operation, the quality of the up-sampled background can be significantly improved.

REFERENCES

[1] S. Takeuchi, D. Shibuichi, N. Terashima and H. Tominaga, "Adaptive Resolution Image Acquisition Using Image Mosaicing Technique from Video Sequence", ICIP'2000, IEEE International Conference on Image Processing, couver, Canada, September 2000.

[2] R.R. Schultz and R.L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences", IEEE Transactions on Image Processing, Vol. 5 No. 6, June 1996.

[3] J.-R. Ohm and K. Rümmler, "Variable-Raster Multiresolution Video Processing with Motion Compensation Techniques", Proc. ICIP'97, Santa Barbara, California, USA, October 1997.

[4] A. Smolic, K. Rümmler, J.-R. Ohm, R. Schäfer and S. Bauer, „MPEG-4 Video Transmission over DAB/DMB: Joined Optimization of Encoding and Format Conversion", Proc. MOMUC'98, International Workshop on Mobile Multimedia Communicatuion, Berlin, Germany, October 12-14, 1998.

[5] T. Wedi, "A Time-Recursive Interpolation Filter for Motion Compensated Prediction Considering Aliasing", ICIP'2000, IEEE International Conference on Image Processing, Vancouver, Canada, September 2000.

[6] A. Smolic, T. Sikora and J.-R. Ohm, "Long-Term Global Motion Estimation and its Application for Sprite Coding, Content Description and Segmentation", IEEE Trans. on CSVT, Vol. 9, No.8, pp. 1227-1242, December 1999.

[7] A. Smolic and J.-R. Ohm, „Robust Global Motion Estimation Using a Simplified M-Estimator Approach", ICIP'2000, IEEE International Conference on Image Processing, Vancouver, Canada, September 2000.

[8] F. Dufaux and J. Konrad, „Efficient, Robust and Fast Global Motion Estimation for Video Coding", IEEE Transactions on Image Processing, Vol. 9, No. 3, March

---

[1] For more results see: http://bs.hhi.de/~smolic/hr-mosaics.html

2000.

[9]   A. Smolic, U.Gölz, P. Kauff, and T.Wiegand, "A theory for motion-compensated interpolation for high-resolution mosaics", in preparation.

Fig.4 Mosaic (2040x1020 pixel) with 16 times the resolution of sequence Mobile & Calendar (SIF, 360x240 pixel), generation temporally centered around middle frame no. 150 for all 300 frames.

Fig.5 Details from images with 16 times the resolution of the original, for frame no. 150 of sequence Mobile & Calendar, left: bilinear interpolation from original image, right: high-resolution mosaicing.