Super-resolution Mosaicing for Object based Video Format Conversion

Matthias Kunter and Thomas Sikora

Communication Sytems Group
Technische Universitaet Berlin, 10587 Berlin, Germany
{*kunter, sikora*}*@nue.tu-berlin.de*

Abstract This paper presents a new approach to spatial upsampling of digital video based on super-resolution mosaics. First, we robustly generate a background mosaic of higher resolution than the original video. In order to achieve that goal, we apply hierarchical global image registration estimating an optimal parabolic parameter set for each view of a scene shot. The final mosaic is generated using statistical and projection grid distance measures to avoid the impact of foreground objects and to accomplish super-resolution respectively. Second, arbitrarily moving foreground objects are segmented using MRF-based change detection methods based on the calculated mosaic. For the foreground objects an optical flow field between adjacent frames is computed. Third, we create new views with higher spatial resolution fusing re-projected background content from the mosaic together with super-resolution foreground objects obtained using optical flow field calculation. Results show that this method is able to convert videos into higher spatial resolution with very high objective and subjective quality.

1 Introduction

With the propagation of many different video recording, storage and display devices, video format conversion applications are attracting a great deal of attention. Since consumer devices allow everyone to create visual content in resolution of low level and medium level quality and due to limited transmission bandwith and storage capacities, spatial-temporal video upsampling is becoming more and more important.

Many methods for video format conversion have been proposed and applied during the last years, especially motion-compensation based methods and algorithms applied in the coding domain [5]. We present a super-resolution based approach to convert videos into higher spatial resolution. Video mosaicing hereby plays an important role summarizing the static background of a scene, recorded with camera motion like zoom, pan, tilt, and rotation. Because of the automatic elimination of freely moving foreground objects, super-resolution mosaicing is a very effective tool to raise the background resolution of such a scene [3]. Due to its signal enhancement characteristic while using all available subsets of image samples it has a high performance compared to simple block motion or region motion based video format conversion methods [5]. A second advantage of video mosaicing is the use for very efficient segmentation of freely moving foreground objects. In order to segment foreground objects, MRF-based change detection algorithms comparing original and mosaic based re-projected video scenes are applied [4]. Thus, a good side effect is the creation of an object based scenario as it is proposed in the MPEG-4 video coding standard [6].

For resolution enhancement of the arbitrarily moving foreground objects which are not covered by the mosaicing process an optical flow based super-resolution technique is applied. Since we are only interested in foreground flow no dense flow calculation between adjacent frames is necessary. This part is inspired by the work of Baker and Kanade who researched super-resolution optical flow very intensively [7].

Figure 1 shows the flowchart of our proposed method. The next three sections describe the techniques for super-resolution background construction, mosaic based object segmentation, and foreground super-resolution computation respectively. In section 5 we show first results and compare it to standard upsampling methods in an objective and subjective assessment. Section 6 concludes this paper.

2 Super-resolution Background Construction

We apply hierarchical global motion estimation for the registration of all frames of a video shot into the reference coordinate system. As underlying trans-formation

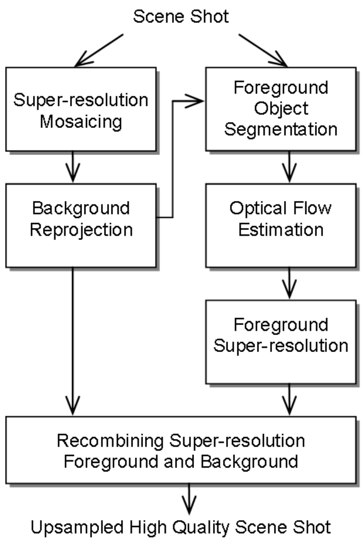

Fig. 1 Flowchart of the proposed mosaic based video conversion algorithm

model the parabolic transformation [2] is used. It is given by

$$\begin{aligned}(x',y')^T &= T\left(\mathbf{k};(x,y)^T\right) \quad (1)\\ \mathbf{k} &= (a_0,\dots,a_5,b_0,\dots,b_5)^T\\ x' &= a_0+a_1x+a_2y+a_3x^2+a_4y^2+a_5xy\\ y' &= b_0+b_1x+b_2y+b_3x^2+b_4y^2+b_5xy\,.\end{aligned}$$

It can be shown that this 12-parameter model yields much better registration results than the widely used perspective model (homography) [2], especially in the case of small translational camera motion. Since the model is nonlinear, concatenation and inversion have to be estimated numerically.

The short term motion parameters are calculated applying Levenberg-Marquardt error minimization, whereas the error function can be described by

$$E(t)=\frac{1}{2}\cdot\frac{1}{N_\Omega}\cdot\sum_{(x,y)\in\Omega}\left(I_{t-1}(x,y)-I_t(x',y')\right)^2. \quad (2)$$

I_{t-1} is the reference frame at time $t-1$ and I_t is the trans-formed and warped frame at time t. The energy term is normalized by the number of overlapping pixels N_Ω in the reference frame domain Ω . For initialization of the gradient based error minimization we calculate a robust feature-based affine transformation parameter set using a RANSAC - based technique. This monte-carlo-like method prevents the registration algorithm from getting stuck into local minima. Note that the affine transformation parameter set is the linear part of Eq. (1).

For global alignment of every frame we finally apply direct parabolic parameter estimation between a preliminary constructed mosaic using the first frame as reference coordinate system and the actual frame. In order

Fig. 2 Preliminary (left) and final (right) super-resolution mosaic of sequence "Charlottenburg"

to achieve this registration we apply again Levenberg-Marquardt energy minimization. As initial starting value the concatenation of the parabolic transformation parameter is used. This can be formulated in a recursive manner:

$$T_{0\to t}=T_{0\to t-1}\otimes T_{t-1\to t}. \quad (3)$$

The preliminary mosaic is a simple summarization of newly discovered image content (see Fig. 2).

The robustness of this method makes correct registration possible even if huge parts of the background are covered by foreground objects.

2.1 Blending Technique

To achieve super-resolution we search the frame I_n with the biggest zoom into the scene. Without loss of generality a scaled version of the first frame I0 can be used as reference coordinate system. The scaling factor s should be at least

$$s\geq 2\cdot\left|a_1^n b_2^n - a_2^n b_1^n\right|. \quad (4)$$

The term describes the determinant of the Jacobian of the affine part of Transformation $T_{0\to n}$ which is maximal for I_n.

For the construction of the final mosaic a statistical analysis of all candidate pixels for a mosaic pixel over time is performed. Here we assume that foreground objects have enough motion that simple median filtering would discover the static background. Farin [4] proposes to conduct further correlation based analysis which makes the process very complex. For our blending approach we follow the rule: A pixel candidate $I_t(x',y')$ belongs to the background only if

$$\begin{aligned}\left|I_t\left(T_{0\to t}(x,y)-m\right)\right| &\leq A\cdot\underset{\forall\tau}{\mathrm{median}}\left|I_\tau\left(T_{0\to\tau}(x,y)\right)-m\right|\\ \text{with } m &= \underset{\forall\nu}{\mathrm{median}}\left(I_\nu\left(T_{0\to\nu}(x,y)\right)\right). \quad (5)\end{aligned}$$

Fig. 3 shows all possible pixel candidates for a line of the super-resolution mosaic of sequence "Stefan". For super-resolution blending, pixel interpolation which is

Fig. 3 Image of mosaic pixel candidates over time t for the line $y = 90$ of sequence "Stefan" (see the foreground object)

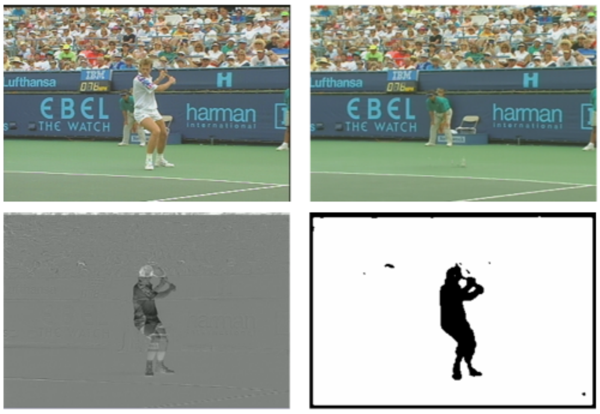

Fig. 4 Original (upper left), mosaic re-projection (upper right), luminance difference image (lower left), and generated change detection mask (lower right) of frame 12 - sequence "Stefan"

equivalent to low-pass filtering has to be avoided. Therefore, out of the background pixel candidates the on is chosen with the smallest distance dt to integer position [2].

3 Mosaic based Foreground Segmentation

Re-projecting the content of the finally calculated mosaic into the original frame coordinate systems generates very accurate background models for every single image I_t of a video shot. Thus, the problem of geometric adjustment [1] for object segmentation based on change detection is non-existent.

We consider significance based change detection for a block of pixels centered at every image pixel of I_t. Additionally, for exploitation of spatial and temporal consistency of the computed change mask a simplified Markov-Gibbs random field approach is applied. A pixel is marked as unchanged if its conditional probability of absolute block pixel difference exceeds a certain threshold [5].

$$p\left(D(x)|\mathrm{H}_0\right) > \tau \qquad (6)$$

$D(x)$ is the mean pixel difference for a block Ω_x and H_0 is the null hypothesis indicating the fact that a pixel belongs to the image background. Since $p\left(D(x)|\mathrm{H}_0\right)$ and the alternative conditional pdf $p\left(D(x)|\mathrm{H}_1\right)$ can be modeled by Gaussian random variables we end up with simple thresholding the block pixel differences.

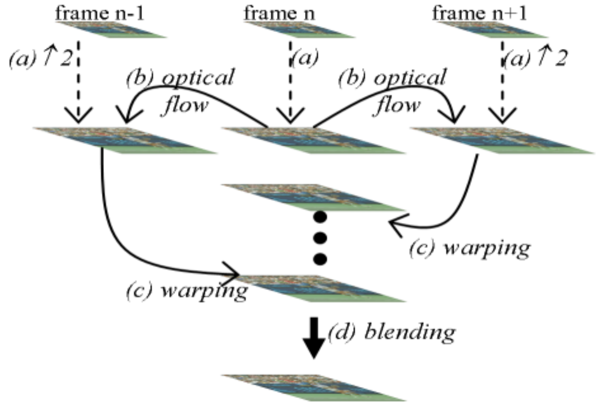

Fig. 5 Principle of flow based super-resolution foreground construction using two adjacent frames: (a) image interpolation, (b) optical flow calculation, (c) image warping, (d) robust blending

For incorporating spatial and temporal consistency we use the previously generated change mask as state map. Since the probability of a pixel to be marked as changed raises with the number of marked neighboring pixels in space and time, threshold is decreased linearly depending on the number of neighboring change pixels in the temporal previous change mask.

Figure 4 shows exemplarily the result of the segmentation method for frame 12 of sequence "stefan". Note, that also background regions are detected because of independently moving background objects.

4 Object Super-Resolution using Optical Flow

Foreground objects are not covered by our robust mosaic super-resolution process. Since it is very desirable to enhance the resolution of the foreground objects we apply an optical flow based super-resolution method proposed in [7]. The authors suggest a five step algorithm which's principle is shown in Fig. 5. Using the four neighboring frames of the actual frame In we compute the optical flow between the upscaled frames and apply a robust blending process to the warped pictures.

As optical flow estimation a hierarchical Lucas-Kanade algorithm is used [7] which is less robust for whole pictures but good enough for foreground objects where only small occlusion occurs. Any other flow algorithm could be used instead [8]. Results show that especially for foreground objects this process yields in more detailed pictures than with any common interpolation method (see Fig. 6).

5 Experimental Results

For the demonstration of the accuracy of our proposed method we compared its results applied on lowpass filtered and downsampled scenes with the original sequences.

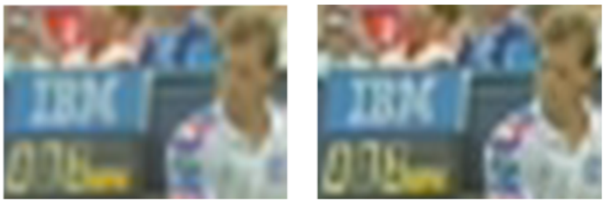

Fig. 6 Frame 3 of sequence "Stefan" after spline interpolation (left) and super-resolution optical flow (right); see slight differences for script region and face region

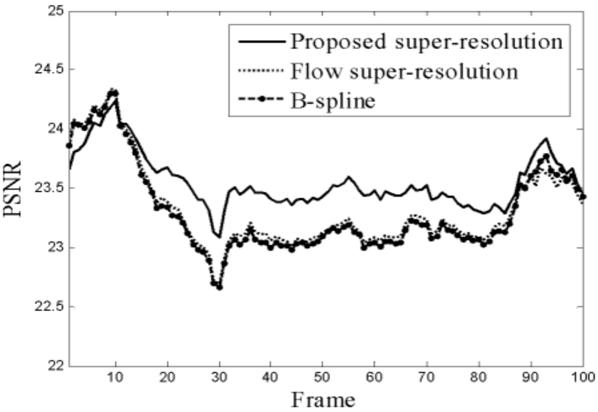

Fig. 7 Results after super-resolution format conversion for first 100 frames of sequence "Stefan", downsampled (176x120)

Figure 7 shows the calculated PSNR values for sequence "Stefan" for three different spatial upscaling approaches over 100 frames. The first is a simple interpolation approach using B-splines, known to be a precise type of image interpolation. Second we show the results for super-resolution based on optical flow super-resolution only (see section 4). Since the applied method for optical flow calculation is not reliable for occlusion regions the PSNR curve is only slightly improved in comparison to the B-spline approach. The third result is obtained applying our proposed method. The gain in terms of PSNR is up to 0.6dB. The mean values over 100 frames for the B-spline, optical flow based, and the proposed method average to 23.22dB, 23.33dB, and 23.56dB respectively.

To strengthen our results we applied the method for incorrectly subsampled scenes having spatial aliasing. Here, the improvement can be mainly assessed subjectively because the flickering caused by aliasing cannot be measured in terms of PSNR values. See Figure 8 for correct restoration of horizontal lines. The mosaic based multi sample reconstruction is very helpful to overcome the aliasing effect.

6 Summary and Conclusion

We presented a new approach for spatial up-conversion of video scenes based on super-resolution mosaicing. Since the mosaic process can only be applied for rigid background objects a optical flow super-resolution method for independently moving foreground objects is used. Main advantage of this approach is the simultaneous generation of meaningful objects due to the mosaic based object segmentation. Results show that for main parts of a video sequence we exceed commonly used upscaling methods, such as complex interpolation approaches. Additionally subjective assessment shows that we can mainly overcome the effects of spatial aliasing. Further work includes the incorporation of more reliable flow calculation algorithms.

Fig. 8 Results: frame 79 of sequence "stefan"- B-spline interpolation (left), proposed method (right) - see aliasing effects for the ground line in the left image

Acknowledgement: This work was developed within 3DTV (FP6-PLT-511568-3DTV), a European Network of Excellence funded under the European Commission IST FP6 programme.

References

1. R. J. Radke, et al., "Image Change Detection Algorithms: A Systematic Survey," *IEEE Trans. Image Processing*, vol. 14 , no. 3, pp. 294-307, March 2005.
2. M. Kunter, J. Kim, T. Sikora, "Super-resolution Mosaicing using Embedded Hybrid Recursive Flow-based Segmentation," *Int. Conf. on Information, Communications and Signal Processing (ICICS '05)*, Bangkok, Thailand, Dec. 2005.
3. G. Ye, et all. "A Robust Approach to Super-resolution Sprite Generation," *Int. Conf. on Image Processing (ICIP'05)*, Genova, Italy, Sept. 2005.
4. D. Farin, P. H.N. de With, and W. Effelsberg, "Video-Object Segmentation using Multi-Sprite Background Subtraction", *IEEE Int. Conf. on Multimedia and Expo (ICME)*, Taipei, Taiwan, 2004.
5. A. Smolic, et al., "MPEG-4 Video Transmission over DAB/DMB: Joined Optimization of Encoding and Format Conversion", *Int. Workshop on Multimedia Commun. (MoMuc'98)*, Berlin, Germany, Oct. 1998.
6. T. Sikora "Mpeg digital video coding standards," *Signal Processing Magazine*, IEEE, vol. 14, no. 5, pp. 82-100, Sep. 1997.
7. S. Baker, T. Kanade, "Super-resolution optical flow," Technical Report CMU-RI-TR-99-36, Carnegie Mellon University, 1999.
8. R. Fransens, C. Strecha, L. V. Gool, "A Probablistic Approach to Optical Flow based Super-resoltuion," *Conf. on Computer Vision and Pattern Recognition (CVPRW'04)*,Washington, DC, USA, June 2004.