Multi-layer Mosaics in the Presence of Motion and Depth Effects

Changki Min†, Qian Yu‡, and Gérard Medioni†
University of Southern California
†Integrated Media Systems Center, ‡Institute for Robotics and Intelligent Systems
Los Angeles, CA 90089, USA
{cmin, qianyu, medioni}@usc.edu

Abstract

In this paper, we present a new segmentation-based 2D mosaic framework. Most of current mosaic algorithms do not explicitly remove moving objects from images before registration, so that they often fail when the size of the moving objects is relatively large. To solve this problem, we first segment moving objects from the input images using the tensor voting framework, and then only the remaining backgrounds are processed for the background mosaic. The second mosaicking step is straightforward because the first motion segmentation step also produces very accurate dense matches. By providing comparative examples, we show that the quality of the background mosaics can be significantly improved by our framework.

1. Introduction

A mosaic is created by geometrically aligning a set of images and stitching them together. The mosaic is one of the oldest research topics in computer vision, and more attention has been given to it recently due to widely used digital cameras and camcorders. Although those popular image capturing devices provide us easy ways to create everyday images, it is still hard to obtain wide-angle pictures of surroundings due to limited optical capabilities. One solution for such a wide-angle image is to take several partial pictures of the surroundings, and stitch them together. In general, the image motion of the obtained partial pictures is more complicated than a simple translation. Therefore, some alignment techniques are required before stitching the pictures.

Many algorithms have been proposed for mosaics. Shum and Szeliski [7] applied both global and local alignment techniques to reduce accumulated registration errors and small motion parallax, Uyttendaele et al. [9] proposed methods to deal with blurry regions due to moving objects and exposure changes across images, and Davis [3] focused on the registration problem in the presence of moving object based on the Mellin transform. Brown and Lowe [2] used the popular SIFT features for robust image registration, and provided a complete framework which automatically generates panoramas without any user inputs. This system is called *Autostitch*, and we compare our proposed approach with it in the experimental section. Those approaches, however, do not explicitly remove both moving objects and strong parallax regions prior to the parameter estimation step, so that they often fail when sequences have some large-scale moving objects or a large portion of the images suffers from strong parallax.

In our proposed approach, we segment objects which move independently or have strong parallax *before* computing motion parameters of the background. This prior segmentation process provides us many advantages. The most important one is, of course, the robust motion parameter estimation. Since the objects which have different motion to the background are already excluded from images, we can compute very accurate motion parameters of the background resulting in the seamless mosaic of the images. Also the final mosaic does not include the moving objects meaning that it is a true background mosaic.

Our algorithm consists of two steps: motion segmentation and background stitching. The main contribution is the first step because it produces 1) background extraction and 2) dense trajectories of pixels. Thus, the following stitching step is straightforward because two outputs from the first step provide enough information for stitching. Note that the motion parameters can be easily computed from the given dense trajectories.

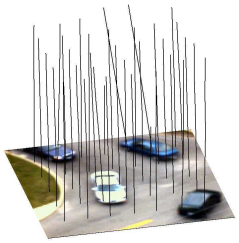
2. Motion segmentation by tensor voting

The tensor voting framework is one of the successful perceptual organization tools, and widely used in many computer vision applications. Its main function is to extract geometrical features (i.e., point, curve, surface, etc.) from a set of N-D points. Due to limited space, we refer readers to

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

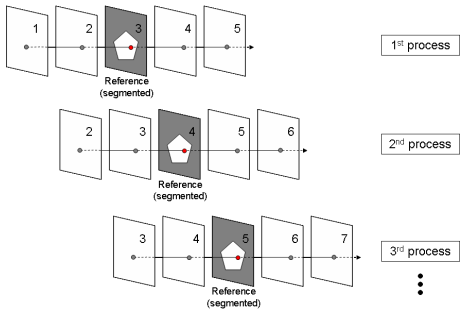
(a) Input sequence

(b) Segmented cars

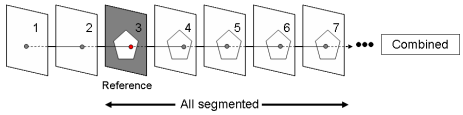


(c) Temporal trajectories

Figure 1. Multiple frame motion segmentation.



(a) Repeat the segmentation process shifting reference images



(b) Combine the trajectories with respect to the frame#3

Figure 2. Segmentation of all images for a mosaic.

[5][4] for more details about the framework.

We have proposed a new motion segmentation algorithm for an image sequence using tensor voting [6]. The approach is based on a spatiotemporal representation of an image sequence, where each object (including the background) forms a smooth layer. The first step is to find dense pixel-correspondences of the reference image in other images using a simple cross-correlation. Thus, it produces initial temporal trajectories of all pixels in the reference image in a 3D (x, y, t) volume – (x, y) are image coordinates and t is time. Due to occlusions and other several factors, those initial trajectories are not smooth. In order to enforce the spatiotemporal smoothness constraint which is our only motion model, we convert the 3D trajectory representation into a higher 5D space that has an additional velocity domain, (v_x, v_y). In this new 5D space, (x, y, t, v_x, v_y), each individual object is represented as a layer. Initially, the layers are not smooth yet because they are generated from the noisy trajectories. In order to make the layers smooth, we apply tensor voting to the layers, and separate them to identify each object. The smoothed dense temporal trajectories (i.e., refined matches of the pixels across images) are obtained by converting the 5D representation back to the 3D temporal trajectories.

Therefore, the smoothing process in the 5D space provides us both matches and segmentation simultaneously. Also note that the approach does not make restrictive assumptions on the observed scene or on the camera motion. For instance, even a non-rigid motion can be extracted as long as it satisfies the spatiotemporal smoothness constraint, the camera is allowed to move in any directions, and multiple objects can be simultaneously extracted. Figure 1 shows an example of the approach.

For the mosaic task, we need to remove moving objects from *all* images, while our approach produces the segmentation only for the reference image. Thus, we simply repeat the segmentation process by shifting the reference image across the sequence as can be seen in Figure 2(a). The trajectories obtained from each process are combined together in the next stage by simply averaging them, and this is illustrated in Figure 2(b). Since all the images are aligned to the first reference image in the following mosaic process, we need temporal trajectories (i.e., matches of pixels) of only the first reference image which is the frame#3 (#1 and #2 are ignored in the mosaic).

3. Background mosaics

3.1. Image registration

Given the assumption that the background is planar, the background images can be registered by a group of homographies, namely $I_i = H_{ij} I_j$. When only small changes exist in image position, the homography can be approximated as an affine transformation A_{ij}, which is the first order linearization of the homography [2].

$$A_{ij} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

A pair of correspondence points $(x, y), (x', y')$ computed from the previous segmentation step establishes two constraints on the affine transformation A_{ij} as shown in the

following equation:

$$\begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix}. \quad (2)$$

Instead of using all correspondences, we employ RANSAC to increase the accuracy of the affine parameter estimation. And linear least squares method is used to compute the parameters from the selected correspondences. Though global optimization methods such as bundle adjustment [8] might be used to reduce the registration error, the current registration is good enough to generate high-quality mosaics because the correspondences provided by tensor voting are very accurate.

3.2. Pixel blending

Given the affine transformations computed from the image registration, we can register each image to the selected reference image. Ideally, each pixel should have the same intensity as the matched pixels in other images. However, in practice the intensity consistency is often violated by illumination changes, shadows, and mis-registration. For each pixel in the background, we compute the mode from all corresponding pixels in other images, and use this mode to fill in the mosaic image.

Another issue is that due to the lack of the knowledge of foreground and background, the mixed blending between the moving objects and the background always ruins the mosaic as shown in Figure 3(a). In our case, the unwanted mixed blending does not happen since only the segmented background pixels are given. Our far better mosaic result which shows the complete background without the trail of the moving car is presented in Figure 3(b).

(a) With moving objects

(b) Without moving objects

Figure 3. Pixel blending results.

4. Experimental results

In order to evaluate the quality of mosaics generated by our proposed approach, we tested several UAV sequences and the 'Teddy bear' sequence [1]. The UAV sequences include one or more moving cars, while the 'Teddy bear' sequence is for a static scene. The latter, however, suffers from strong parallax so that it is a challenging example for mosaicking. For both cases, our results are qualitatively compared with the results generated by *Autostitch* [2] which is one of the most popular mosaicking frameworks.

UAV sequence

This is a typical UAV sequence where the flying camera looks the ground from the top, and the scene has a moving car. Figure 4(a) shows our background mosaic result, and (b) and (c) show the results of the Autostitch when it processes the original images and extracted background images only, respectively. Our result does not contain the moving car, and the lines on the street are straight. On the other hand, the Autostitch produces a somewhat distorted mosaic as can be seen in (b). This inaccurate alignment is caused by the moving car, and the improved mosaic is presented in (c) where the Autostitch processes our extracted background images. This example explains that processing only the background area is important for accurate image alignments in the presence of moving objects. The dark trail on the street in our result is caused by the shadows of the car on the ground, and it is not the problem of image alignment.

Teddy bear

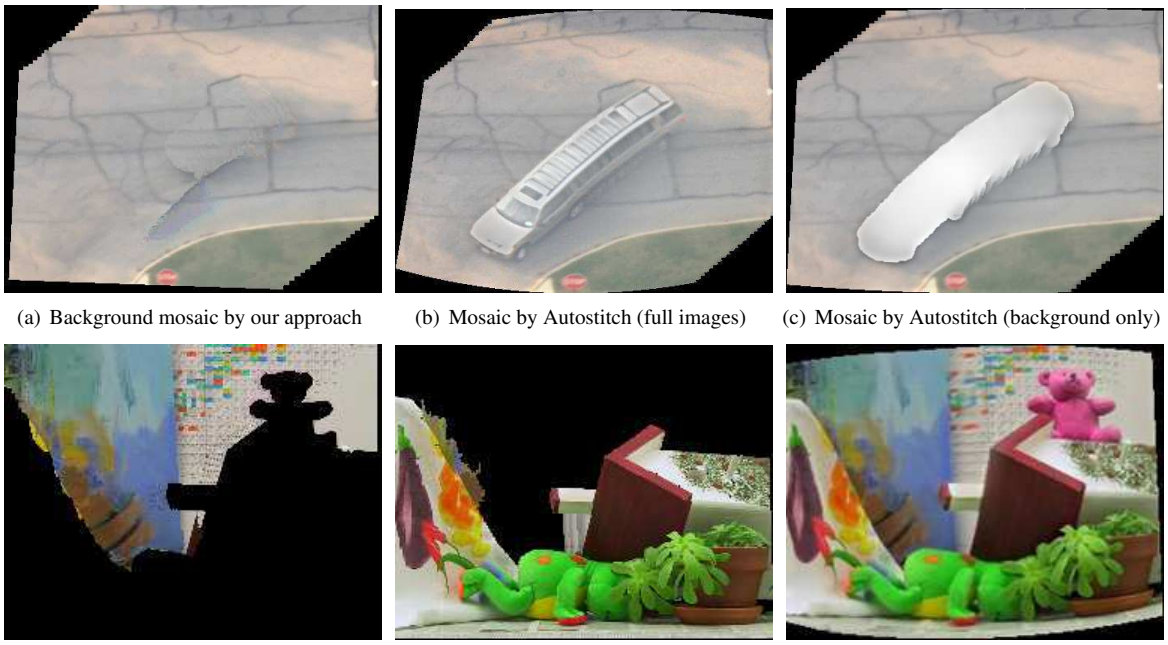
This sequence demonstrates the difficulties when sequences suffer from strong parallax. Figure 4(d) and (e) show our accurate image registration for the background and the foreground, respectively (Teddy bear is another independent object). If the background and the foreground are processed together in spite of strong parallax, a distorted mosaic is produced as shown in Figure 4(f). Also, this sequence is a good example of multi-layer mosaics: the foreground regions also generate their own mosaic in addition to the background mosaic forming two different layers of mosaics from a sequence.

Dynamic mosaics

Since we can generate a mosaic which consists of only the background regions, we are also able to incorporate moving objects into the *mosaic space* seamlessly as shown in Figure 5. This is a straightforward extension in that each original full image is simply overlayed on the resulting background mosaic using the same affine parameters. However, other methods such as the Autostitch cannot generate the dynamic mosaics because their mosaics include the trail of the moving objects as shown in Figure 4(b).

5. Conclusions

In this paper, we have proposed a new method for generating accurate background mosaics in the presence of moving objects or strong parallax. Although some other approaches also consider them as parts of their algorithms, it



(a) Background mosaic by our approach (b) Mosaic by Autostitch (full images) (c) Mosaic by Autostitch (background only)

(d) Background mosaic by our approach (e) Foreground mosaic by our approach (f) Mosaic by Autostitch (full images)

Figure 4. Comparison between our approach and the Autostitch. (a)-(c) is for a UAV sequence, and (d)-(f) is for the Teddy bear sequence.

Figure 5. Dynamic mosaic

is usually not enough to obtain accurate image alignments if the size of the moving objects is relatively large. By comparing our method with the well-known Autostitch, we could show that the mosaics can be significantly improved when each independent moving part is separated from the images before alignments. We believe that our background mosaics can be used in many applications (the dynamic mosaic is a good example), so that we are going to study those applications as the extension of our current framework.

Acknowledgment

The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152, and U.S. National Science Foundation grant IIS 03 29247. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

References

[1] http://cat.middlebury.edu/stereo/.
[2] M. Brown and D. Lowe. Recognising panoramas. In *ICCV*, pages 1218–1225, 2003.
[3] J. Davis. Mosaics of scenes with moving objects. In *CVPR*, pages 354–360, 1998.
[4] G. Medioni and S. Kang. *Emerging Topics in Computer Vision*. Prentice Hall, 1st edition, 2004.
[5] G. Medioni, M. Lee, and C. Tang. *A Computational Framework for Segmentation and Grouping*. Elsevier, 1st edition, 2000.
[6] C. Min and G. Medioni. Motion segmentation by spatiotemporal smoothness using 5d tensor voting. In *POCV*, 2006.
[7] H. Shum and R. Szeliski. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *IJCV*, 36(2):101–130, February 2000.
[8] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. *Bundle Adjustment: A Modern Synthesis*. Springer-Verlag, 1999.
[9] M. Uyttendaele, A. Eden, and R. Szeliski. Eliminating ghosting and exposure artifacts in image mosaics. In *CVPR*, pages II:509–516, 2001.