Mosaicing and Restoration from Blurred Image Sequence Taken with Moving Camera

Midori Onogi and Hideo Saito

Keio University, Department of Information and Computer Science,
Yokohama, Japan
{midori, saito}@ozawa.ics.keio.ac.jp
http://www.ozawa.ics.keio.ac.jp/Saito/

Abstract. A wide-area image can be synthesized from an image sequence taken with a moving camera by using image mosaicing techniques. However, motion blur caused by the motion of the camera may significantly degrade the quality of the synthesized image. In this paper, we propose a new method for generating a deblurred mosaic from an image sequence that is degraded by motion blur under the condition that we do not have any information about the intrinsic and extrinsic parameters of the moving camera during input acquisition. In this method, we assume the objects in the scene can be classified into two regions in order to handle depth. In this paper, the displacement vectors of the features, which are computed using the KLT feature tracker on the consecutive frames, are classified into two regions. Here, the classified vectors provide a Point Spread Function (PSF) of the blurred image, and a homography between two consecutive frames for segmentation and mosaicing. Experimental results show that the Signal to Noise Ratio of the generated images can be significantly improved by our proposed method.

1 Introduction

In general, the resolution and the field of view of a digital image are limited by the camera. Image mosaicing techniques [3, 9] have been used to synthesize a wide-area image from a number of images, which are taken from different camera pose and/or positions. Mosaicing in more general cases of camera motion can be performed by projecting thin strips from the images onto manifolds which are adapted to the motion[11]. These 2D image alignment methods for image mosaicing can be applied successfully when the scene can be approximated by one plane such as in aerial photography, where the scene can be considered flat because the camera is far from the scene. However, these methods fail when the scene is 3D which includes different depths. In order to handle 3D parallax, a depth invariant mosaicing method by computing the camera motion using space-time volume has been developed[14]. Zhigang Zhu et. al. proposed a method for generating stereoscopic mosaics from images captured by a video camera

S. Singh et al. (Eds.): ICAPR 2005, LNCS 3687, pp. 598–607, 2005.

mounted on an airborne platform with GPS/INS measurements using a parallel-perspective representation [13].

Since such image sequences are sometimes captured using a moving camera, the motion makes the captured images blurred. As motion blur due to camera motion may significantly degrade the image quality, a considerable amount of research has been dedicated to restore these images. Blurred images can be deblurred by using image deconvolution [5]. A general motion blur PSF can be recovered from various devices [1]. Motion blur correction from multiple images has recently been tried as well. Rav-Acha et.al. proposed a method for image deblurring from two images having motion blur in different direction[7]. Synthesizing a super-resolved image from multiple images is also an active research topic [3, 8, 10]. Motion deblurring has also been addressed in the context of temporal super-resolution [2].

In most of the research on image restoration, motion blur is considered shift invariant. However, in practice, the motion blur is shift variant because a 3D scene has multiple depths, so the applicability is limited when the scene can not be considered flat. In addition, as many image mosaicing approaches do not consider motion blur, the quality of the mosaic synthesized from a blurred sequence is degraded.

In this paper, we propose a new method for generating a deblurred mosaic from a blurred image sequence captured by a moving camera, in which the intrinsic and motion parameters are unknown. In our proposed method, we combine methods for motion image deblurring and image mosaicing so that we can synthesize mosaic images without motion blur from image sequences taken with a moving camera.

The proposed method is achieved by deblurring each frame of the input sequence and generating a mosaic image from the deblurred frames. The proposed method also takes into account multiple regions with different depth and blur by segmenting each frame based on displacement differences of tracked points, and the estimation of the homography[1] and PSF for each region.

The proposed method first tracks a number of feature points over the input image sequences using the KLT feature tracker. By assuming that the object scene can be represented by two layers of planar regions, the displacement vectors of the features on the consecutive frames are classified into foreground points and background points. For each region, the displacement vectors are averaged for estimating the PSF of the motion blur. By applying the Wiener filter [5] with the estimated PSF for each region, the input image sequence can be deblurred. Our method then merges all the images in the input image sequence by image mosaicing techniques. For the image mosaicing, the homography of each region between the consecutive frames is estimated from the displacement vectors within the region. After the image mosaicing, we can finally synthesize a deblurred wide-area image from the input image sequence. We also have conducted experiments with various scenes consisting of a foreground object and

[1] The homography maps the projected point from one plane in a 3D scene to another plane.

a background object. Experimental results demonstrate that our method can generate higher quality mosaics than images without motion blur restoration.

2 Proposed Method

2.1 Corresponding Points

The moving camera provides a sequence of images. By computing corresponding points between consecutive frames, we can obtain the homographies between the consecutive frames for each region and the parameters of the PSF representing motion blur caused by camera motion. We compute corresponding points using the KLT feature tracker [6].

2.2 Classifying Corresponding Points

In order to make a mosaic of a scene that can be approximated by two layers, we need to classify the corresponding points into foreground points and background points.

Let $\boldsymbol{p}_k^i = (x_k^i, y_k^i)$ denote the position of the ith feature point in the kth frame ($0 < k < m$, $0 < i < n$. m indicates the number of the images and n denotes the number of tracked feature points). First, we compute the displacement d_k^i between the consecutive frames of every corresponding point as

$$d_k^i = \sqrt{(x_{k+1}^i - x_k^i)^2 + (y_{k+1}^i - y_k^i)^2}. \tag{1}$$

We assume that the displacements of the points in the foreground region are sufficiently larger than those in the background region for the classification as shown in Fig. 1. Then, we classify the points into foreground points and background points using the Discriminant Analysis Method in which the threshold is determined by maximizing $F(t)$, the ratio between inner-class and inter-class variance. $F(t)$ is represented as follows:

$$F(t) = \frac{\sigma_B^2}{\sigma_I^2}, \tag{2}$$

where t is the threshold, σ_B^2 is inter-class variance and σ_I^2 is inner-class variance.

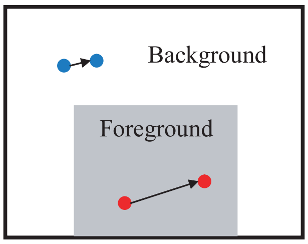


Fig. 1. Displacements of the corresponding points

2.3 Deblurring Images

As the distance from the camera to the foreground is different from the distance to the background region, the motion blur should be restored independently for each region. Since we assume that the translational component of motion of the camera is almost dominant, we can approximate the PSF as follows:

$$h(x,y) = \begin{cases} \frac{1}{w}, x\cos\theta + y\sin\theta \le \frac{2}{w} \\ 0, x\cos\theta + y\sin\theta > \frac{2}{w} \end{cases}, \quad (3)$$

where w and θ are the width and the angle of motion blur, respectively. Since we also assume that the motion vector between two consecutive frames is almost constant in each region, we can compute w and θ by an average vector of the displacement vectors of the corresponding points in each region. We can deblur each region of each frame using the PSF for the foreground region, $h_k^f(x,y)$, or the PSF for the background region, $h_k^b(x,y)$. Given the estimated PSF, we can deblur each frame using existing deconvolution algorithms. We deblur the images by a Wiener filter [5] in the frequency domain.

As the background regions of the frames are rarely blurred, these do not need to be deblurred in most of the cases. Artifacts from excessively enhancing the edges degrade the quality of images if the backgrounds are strongly deblurred.

2.4 Segmentation and Mosaicing

Given the deblurred frames, we generate a wide-area mosaic. When the scene can be considered flat such as in aerial photography, image mosaicing generally needs homographies between a base frame, which is selected as the standard image plane for merging, and every other frame. However, we consider our scene in 3D therefore a single plane is not enough. We generate mosaics for the foreground region and the background region separately by using a separate homography for each region rather than a single homography for the whole frame. It is necessary to segment the scene to handle this case.

Segmentation of the scene is done by splitting the scene into two layers of planer regions. First, we select a base frame, which is not necessarily the start frame, from the input frames. From the base frame, we manually select the vertices $\boldsymbol{v}_{base}^j$ ($0 < j < l$, l indicates the number of the vertices) on the border of the foreground region which approximates a polygon. The corresponding vertices in the kth frame $\boldsymbol{v}_k^j$ can be computed with

$$\boldsymbol{v}_k^j = \mathbf{H}_{base,k}^f \boldsymbol{v}_{base}^j, \quad (4)$$

where $\mathbf{H}_{base,k}^f$ is the homography of the foreground region. This is shown in Fig. 2. However, to obtain $\boldsymbol{v}_k^j$, we must compute the foreground homography $\mathbf{H}_{base,k}^f$. We can obtain $\mathbf{H}_{base,k}^f$ by the product of the homographies of consecutive frames up to the kth frame.

$$\mathbf{H}_{base,k}^f = \mathbf{H}_{base,base-1}^f \mathbf{H}_{base-1,base-2}^f \cdots \mathbf{H}_{k+1,k}^f. \quad (5)$$

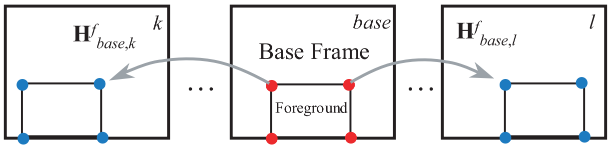


Fig. 2. Segmentation

The consecutive frame homographies can be obtained from the SVD of the foreground points that were obtained from corresponding point classification in section 2.2.

Each pixel in the foreground regions of the deblurred frames is transformed into a base frame by $\mathbf{H}^f_{base,k}$. The color values of the pixels in the same place are averaged, and then the foreground regions are merged into a foreground mosaic $I^f(x, y)$, while the background regions are also merged into a background mosaic $I^b(x, y)$ by the background homography $\mathbf{H}^b_{base,k}$ as shown in Fig. 3. $\mathbf{H}^b_{base,k}$ can also be computed in the same way as the foreground homography $\mathbf{H}^f_{base,k}$. Although each frame is deblurred as described in section 2.3, the artifacts from enhancing the edges remain in each image. By the averaging of the image mosaicing process, the effect of the artifacts can be decreased.

Given both foreground and background mosaics $I^f(x, y)$ and $I^b(x, y)$, we can finally generate the output image $O(x, y)$ as:

$$O(x, y) = I^f(x, y) \cdot M^f + I^b(x, y) \cdot \bar{M}^f, \tag{6}$$

where M^f is a segmentation mask for the shape of the foreground region obtained from the vertices of the polygon in the base frame.

2.5 Removing mistracked corresponding points

The tracked corresponding feature points sometimes may include some wrongly tracked feature points. When there are such mistracked points, the accuracy in estimating the PSF and the homography between consecutive frames is reduced. Because the accuracy of the homography is very important for segmentation and image mosaicing, it is important to improve the accuracy. We remove the as many as possible of the mistracked feature points during the segmentation process with the following technique.

For All Frames
1. Compute Homography $\mathbf{H}^f_{base,k}$, and vertices $\boldsymbol{v}^j_k$ as in section 2.4.
2. ***For All corresponding points***
If Foreground point
If Inside polygon Keep point
Else Delete point
If Background point
If Inside polygon Delete point
Else Keep point

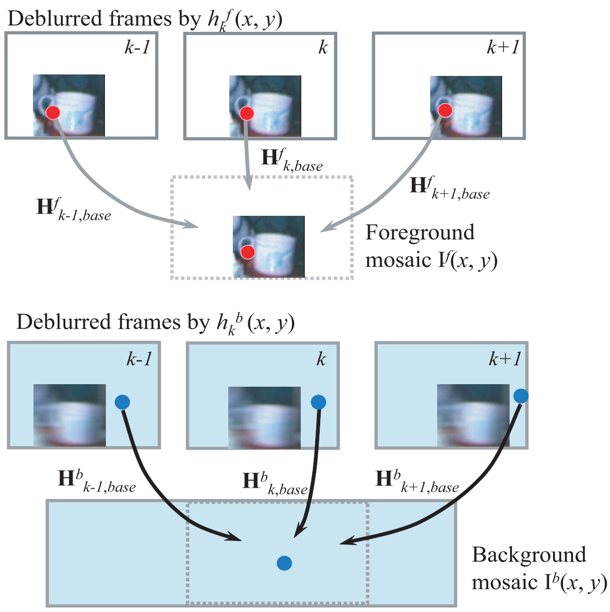


Fig. 3. Image mosaicing

3. **If** No deleted points Stop processing
Else Go to step4
4. Reclassify remaining points as in section 2.2. Go to step1.

With the new corresponding points, the input frames are deblurred as described in section 2.3. After deblurring images, we then segment the captured frames again, and merge the foreground and background images using the new points of each region.

3 Experimental results

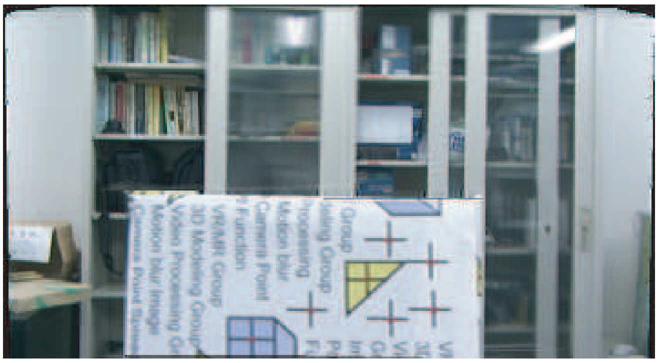
We recorded a video sequence of a number of scenes, whereby a planar foreground object is in front of the background scene.

3.1 Blur removal

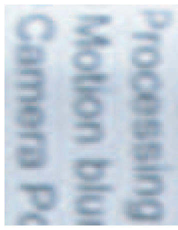
In Figs. 4a and 5a, we show examples taken from input image sequence that is blurred due to camera motion. Figs. 4b and 5b show individual frames deblurred by $h_k^f(x, y)$ using a Wiener filter. However, artifacts still remain in the deblurred frames. Figs. 4c and 5c show the output images generated by the proposed method. Figs 4d, e, and f show a close-up of the raw input image, a deblurred only output image, and a deblurred+mosaiced output image respectively. As can be seen from the figures, our method produces the clearest text.

(a) Raw input image sequence

(b) Deblurred foreground region input image sequence



(c) Deblurred mosaic



(d)Input image zoom-in (e)Deblurred only zoom-in (f)Deblurred mosaicing zoom-in

Fig. 4. A planar document is captured as the foreground region in the scene

3.2 Mistracked corresponding points removal

In Fig. 6, we show the input sequence in Fig. 4 including the mistracked corresponding points in the background region. These mistracked corresponding points are removed by image segmentation. We show a result when the mistracked corresponding points are not removed in Fig. 7. Since the accuracy of the homography computed from the points which include all the mistracked corresponding points is degraded, the input images are wrongly aligned. Consequently, the text in the foreground region can hardly be read. The width of the result image is also narrower than the image shown in Fig. 4c because of the poor accuracy of the homography.

3.3 S/N ratio

In order to validate the accuracy of the output image, we compute the Signal to Noise Ratio (S/N ratio) of the foreground regions of the images. S/N ratio SNR is expressed as:

(a) Raw input image sequence

(b) Deblurred foreground region input image sequence

(c) Deblurred mosaic

(d)Input image zoom-in

(e)Deblurred only zoom-in

(f)Deblurred mosaicing zoom-in

Fig. 5. A planar photo is captured as the foreground region in the scene

(a)Images before removing mistracked corresponding points

(b)Images after removing mistracked corresponding points

Red dots represent corresponding points between two consecutive frames.

Fig. 6. Examples taken from image sequence including mistracked corresponding points

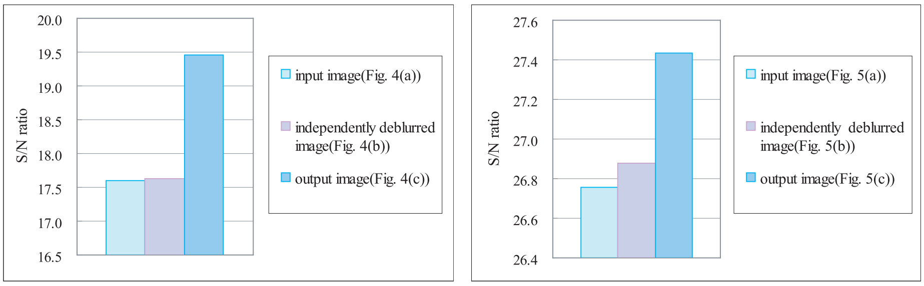
$$RSME = \sqrt{\frac{\sum [f(i,j) - F(i,j)]^2}{N}} \tag{7}$$

$$SNR = 20 \log_{10} \left(\frac{255}{RSME} \right), \tag{8}$$

where $f(i,j)$ represents evaluated image, and $F(i,j)$ represents the ground truth image captured without motion blur by using a tripod. N is the number of the

(a) Deblurred mosaic (b) Zoom-in of (a)

Fig. 7. A result without removing the mistracked corresponding points



(a) S/N ratio of the images shown in Fig4 (b) S/N ratio of the images shown in Fig5

Fig. 8. S/N ratio

pixels in the foreground region. $RSME$ is the root square mean error. In Fig. 8, we see that S/N ratio of the images generated by the proposed method exceeds that of the images deblurred at each frame independently. The results show that our method is effective in restoration of motion blur, and in decreasing artifacts by deconvolution of the PSF.

4 Conclusion

In this paper, we have presented a method for generating a deblurred mosaic image from motion images captured by a handy moving camera, under the condition that the scene consists of the layers with different depth. By deblurring each frame and mosaicing them, we generate the deblurred image with a high S/N ratio. The experimental results demonstrate the validity of the proposed method.

Our proposed method has several possible applications. It could be applied to image mosaicing from blurred sequences captured from moving cars or trains out of the city in which a building stands in front of the background. Our approach could also be useful for mosaicing a group photo when the people stand in front of a planar background.

References

1. Moshe Ben-Ezra and Shree K. Nayar: Motion-Based Motion Deblurring. IEEE Trans. on PAMI, June 2004, Volume 26, Issue 6.
2. E. Shechtman, Y. Caspi, and M. Irani: Increasing Space-Time Resolution in Video. In Proc.ECCV, pp. 753-768, 2002.
3. D. Capel: Image Mosaicing and Super-Resolution (Book). Springer, 2004.
4. B. Bascle, A. Blake, and A. Zisserman: Motion deblurring and super-resolution from an image sequence. In Proc.ECCV, pp. 312-320, 1996.
5. A.K. Katsaggelos: Digital Image Restoration (Book). SpringerBerlin, 1991.
6. Bruce D. Lucas and Takeo Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision. International Joint Conference on Artificial Intelligence, pp. 674-679, 1981.
7. A. Rav-Acha, S. Peleg: Two motion-blurred images are better than one. Pattern Recognition Letters, Vol 26, pp. 311-317, 2005.
8. S. Sei, H. Saito: Super-Resolved Image Synthesis from Uncalibrated Camera with Unknown Motion. IAPR Workshop on Machine Vision Applications (MVA02), pp. 420-423, Dec.2002
9. R. Szeliski: Video Mosaics for Virtual Environments. IEEE Computer Graphics and Applications, Vol. 16, No. 2, pp. 22-30, 1996.
10. I. Zakharov, D. Dovnar, Y. Lebedinsky: Super-resolution image restoration from several blurred images formed in various conditions. Prpc. International Conference on Image Processing 2003, pp. II-315-318,2003.
11. S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet: Mosaicing on Adaptive Manifolds. IEEE Trans. on PAMI, Oct. 2000, pp. 1144-1154
12. A. Zomet, S. Peleg, and C. Arora: Rectified Mosaicing: Mosaics without the Curl. CVPR'00, June 2000, Vol. II, pp. 459-465
13. Zhigang Zhu, Allen R. Hanson, and Edward M. Riseman: Generalized Parallel-Perspective Stereo Mosaics from Airborne Video. IEEE Trans on PAMI, February 2004, Volume 26, Issue 2
14. Alex Rav-Acha, Yael Shor, and Shmuel Peleg: Mosaicing with Parallax using Time Warping. Second IEEE Workshop on Image and Video Registration (IVR'04), Washington, DC, July 2004