SIMULTANEOUS DEBLUR AND SUPER-RESOLUTION TECHNIQUE FOR VIDEO SEQUENCE CAPTURED BY HAND-HELD VIDEO CAMERA

Yuki Matsushita, Hiroshi Kawasaki
Kagoshima University, Department of Engineering, Japan

Shintaro Ono, Katsushi Ikeuchi
The University of Tokyo, Institute of Industrial Science, Japan

ABSTRACT

Nowadays, video camera is commonly used everywhere and demand of retrieving a single shot from video sequence is increasing. Since resolution of video camera is usually lower than that of digital camera, simply cutting out a frame from a video sequence ends up with low quality. Further, because of the necessity of high fps on video camera, video data inevitably contains motion blur and it leads mis-registration between frames which is critical for multi-frame super-resolution. In this paper, we propose a method to restore high-resolution image from a video sequence considering motion blur. Since the frame-rate of a video camera is high, motion of the object in successive frames is small, and thus, stable feature tracking during short sequences is possible even if there is a blur. Thus, we adopt a division/integration approach to realize robust tracking for long sequence. We also propose a simultaneous deblur and super-resolution technique using multiple images based on MAP estimation. Experimental results are shown to prove the strength of our method.

Index Terms— motion-deblur, super-resolution, video analysis, feature tracking

1. INTRODUCTION

Demand for retrieving a high quality single image from video sequence is increasing, such as surveillance, personal purposes and so on. Since image quality of video camera is usually lower than that of digital camera, simple frame capturing is often insufficient for the actual purpose. Motion blur is another important reason for low quality.

In order to solve the problems, methods such as deblurring and super resolution have been researched [1, 2, 3]. Generally, these methods often conduct deblurring and super-resolution by estimating single blur kernel or registration parameter for the entire frame. However, this becomes problematic when the depth of the object is different within a scene, or when there are independent motions in the scene. When this happens, the process becomes challenging to apply the past methods which assume one specific motion in a scene. In addition, in regard to super-resolution, if motion blur exists, it is not only treated as noise, but also causes registration errors, resulting in severe degradation [4].

Several methods have been proposed in the past in order to solve the problem. The method includes dividing the image into small segments which represent planer parts of the scene, and then, applying the process for each segments [5]. However, since the method consists of two separate processes, such as motion deblur and super-resolution, there remains a severe problem on feature tracking during a long sequence. To solve the problem, we propose a new tracking method which integrates all the tracking results of short sequences with refinement algorithm. In addition, simultaneous deblur and super-resolution technique for multiple blur images is proposed. The contribution of the method is as follows.

1. Pixel based blur kernel estimation is introduced.
2. High precision and robust tracking method for a long sequence which contains motion blur is proposed.
3. Simultaneous method of deblurring and super-resolution for multiple images is presented.

By using the techniques, it is possible to create high-resolution still image from a video sequence which is low resolution and contains motion blur.

2. RELATED WORK

In terms of deblurring techniques for blurry image which is caused by camera motion, since the blur is a convolution process, restoration technique has been proposed as a deconvolution technique with known kernel [6, 7, 8]. Recently, blind deconvolution techniques are also intensively researched [2, 9]. In both techniques, since they assume single kernel, they cannot be applied if the scenes have multiple objects with different motion or depth. Recently, several methods have been proposed to solve the problem by estimating the kernel independently with scene segmentation [10, 3]. However, since those methods are for a single image, consistent deblurring for an image sequence is not considered; this is because motion blur severely occurs on camera, but not on video sequence. Although a few researches are conducted on video [11, 12], they only consider single blur kernel for each frame and cannot be applied to our case.

In respect to super-resolution, there are mainly two methods that utilize single or multiple images; multiple images approach is considered advantageous on image quality and stability. Besides, there is also another method, example-based method which uses image datasets of high-resolution patch and low resolution one [13]. This method requires large and precise datasets to restore complicated images and not suitable for common use. The core technique for multiple image approach is image registration with sub-pixel accuracy [1]. To realize accurate registration, robust and precise tracking of feature points is necessary. Among common feature extraction [14, 15] and tracking methods [16], most of them are

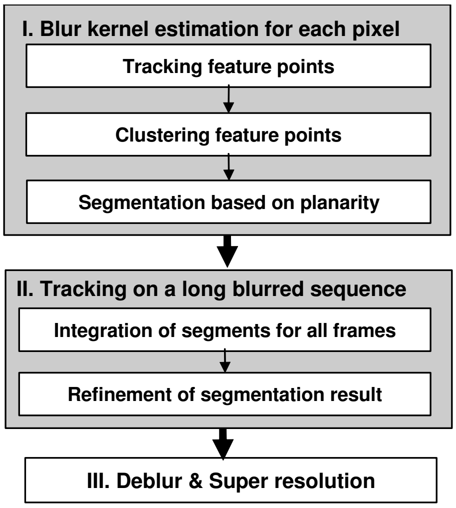


Fig. 1. Diagram of the entire algorithm

based on image gradient, and thus, they fail if high frequency components are lost; *e.g.*, by motion blur. Another problem on super-resolution on video sequence is the existence of multiple moving objects in the scene. Although a few methods are proposed [17, 5], since the methods do not consider motion blur, it is difficult to apply the methods to handheld video camera. Further, deblurring is not considered in their methods.

3. ALGORITHM OVERVIEW

For the scene that has multiple moving objects, we first divide the scenes into segments which represent independent objects before applying deblur and super-resolution. After segmentation is conducted, deblurring and super-resolution techniques are applied to each segment. Finally, all the segments are integrated to construct the final result. In our implementation, the process is divided into four steps as shown in Fig. 3.

In the first step I (Sec.4.1), feature points as well as planar segments in the scene which are calculated as clusters of feature points are extracted from image sequences. However, if there is a motion blur, it is difficult to apply the technique for long sequences. In out method, the entire frames are divided into short sequences to realize a stable process. Then, the extracted segments are integrated. Since simple integration accumulates errors, refinement technique is applied in step II (Sec.4.2). Finally in step III (Sec.4.3), deblurring and super-resolution are simultaneously applied.

4. DETAILED METHOD

4.1. Blur kernel estimation for each pixel

In order to apply motion deblur and super-resolution on video data with multiple motion and depths, it is necessary to estimate blur kernel for each region. In this paper, we first calculate the optical flow in the scene, and then, estimate the blur kernel for each pixel using the flow. To calculate the optical flow, we use the method proposed by Yamaguchi et. al. [5]. The method consists of three steps; such as **1)** tracking feature points, **2)** clustering the points whether they are expressed by the same homography, and **3)** segmentation by assigning the cluster to each pixel where the luminous difference is smallest. By using the segmentation result with corresponding homographies, motion vector of each pixel between the adjacent frames can be calculated. Here, we assume that the shape of blur kernel is linear, because shutter speed of camera is usually quite short. Note that the direction and the length of the kernel can be calculated here, however, scaling parameter for the kernel cannot be determined, and thus, it is estimated afterwards.

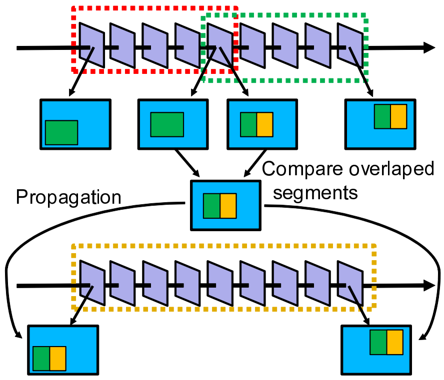


Fig. 2. Solving inconsistency at integration step.

4.2. Long sequence tracking for blurry images

4.2.1. division/integration approach

If the motion of the object is getting larger, tracking of feature points through long sequence becomes more difficult. This is due to the increased errors of tracking caused by the bigger motion blur and the changes of appearance. Our method solves this problem by first dividing the long sequence into a number of short sequences, and then, integrating them. Since the segmentation results in each short sequence are different even if they are successive, integration is not a simple task. For example, an object which splits into two parts makes different segmentation results in the sequence before and after the split. To solve the problem, when we divide the video into short sequences, we include the same frame in both successive sequences to share the same frame. Then, we merge segmentation results at the shared frame with OR operation so that each segment represents independent planer objects in the frame as shown in Fig. 2. Then, the merged segmentation results are propagated to the entire frames in the sequence. Further, the process is also recursively applied to the entire sequences.

There are segments that are not assigned to any planar object because of occlusion and noise. Those segments are regarded as undetermined and are re-estimated by closest pixel in the frame. Furthermore, since integration of multiple sequences accumulates errors, refinement algorithm is required.

4.2.2. Refinement of feature tracking and segmentation result

The precision of the segmentation, which represents a planar object in the scene, decreases due to the tracking error mainly caused by motion blur. To solve the problem, we refine the tracking result of feature points. However, the magnitude of blur and its direction differs from frame to frame, it is difficult to improve the precision with common methods. In our method, we use the blur invariant phase only correlation matching method (BIPOC) which provides highly precise registration even with an image with motion blur [18]. However BIPOC, by principle, can only be applied for parallel movements. Therefore, for free movements expressed by homography transformation, it is difficult to achieve high precision registration by simply applying BIPOC.

To overcome the problem, we implement an iterative approach with two steps; **1)** improvement of tracking precision by using BIPOC and **2)** updating homography transformation by using the results obtained from the first step. By doing the refinement steps repeatedly, high precision image registration is achieved.

4.3. Image restoration by MAP estimation

In super-resolution methods that utilize multiple images, MAP estimation is widely used. This is done by registering the input images and simulating image degradation process by pixelization. If there is a motion blur in the image, the result ends up up with low quality as the blur is considered as noise [4].

On the other hand, deblurring methods [7, 6] generally use a single image as the input, then deconvolution is applied with single blur kernel. Sine deconvolution is an unstable process, if the motion blur is large, quality of the process becomes degraded due to ringing effect, etc.

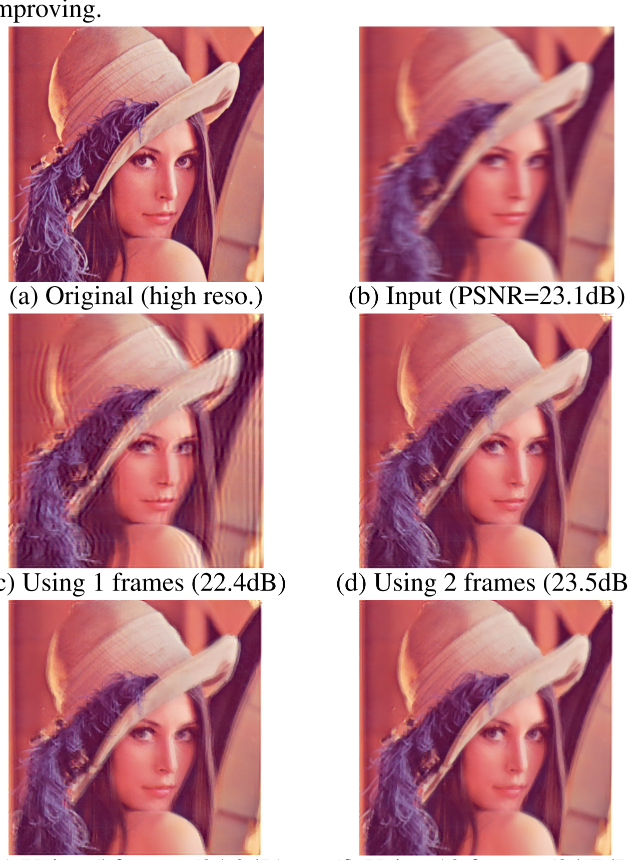
In this paper, to solve the both aforementioned problems, we propose the MAP estimation method which considers not only degradation by pixelization, but also motion blur. By simultaneously considering the pixelization and motion blur, it is expected to solve the both problems; the image degradation on super-resolution and the ringing effect by deblur. In our method, we estimate super-resolved image as well as scaling parameter of blur kernel. Let us consider x as the vector for high resolution image, y as the input image vector, s as the scaling parameter of blur kernel, N as the number of input frames, and P as the number of pixel of the high resolution image. Then, we can formulate the restoration as follows:

$$\underset{\boldsymbol{x},\boldsymbol{s}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{N} \sum_{i=1}^{P} ||\boldsymbol{y}_k i - \mathrm{D}\mathrm{B}_k(s,i)\mathrm{W}_k \boldsymbol{x}_i||^1 + \lambda ||\mathrm{C}\boldsymbol{x}||^2 \right\},$$

where D describes the sub-sampling kernel and $B_k(s,i)$ describes the blur kernel and W_k describes the registration parameter for pixel i, respectively. C is a prior information of the high resolution image for regularization purpose and λ is its weight. In this paper, we use Bilateral Total Variation as a prior and steepest descent method is applied to optimize the evaluation function. In the actual process, we first estimate s by fixing x, and then, estimate x by fixing s, and repeat the process until convergence.

5. EXPERIMENT

First, we use the simulation images to evaluate our method. We use a 400x400 pixel image for high resolution image as shown in Fig. 3(a). Then, we simulate parallel motions with 17-19 pixel length to random direction and apply motion blur along the directions with the fixed scaling 0.7. Then we downsampled the image into 200x200 pixel as shown in Fig. 3(b). Using the image set, we performed super-resolution with 2x2 times. With our method, the scaling parameter was estimated as 0.71, which has enough accuracy for MAP estimation. As shown in Fig. 3(c)-(f), by increasing the input number, we can confirm that the ringing effect is decreasing as well as PSNR is improving.



(a) Original (high reso.) (b) Input (PSNR=23.1dB)
(c) Using 1 frames (22.4dB) (d) Using 2 frames (23.5dB)
(e) Using 5 frames (24.2dB) (f) Using 10 frames (24.7dB)

Fig. 3. Simulation image results.

Next, we performed an experiment with real images taken by video camera in hand where a newspaper is moved in all directions with different velocity as shown in Fig. 4(a). In the experiment, 30 frames (350x265 pixel) are used to perform 2x2 magnification. Since the newspaper was moved with independent rotation and translation, there are different blurs occurred on each frame. Then, we performed three different algorithms to compare the result as shown in Fig. 4(b)-(d). By comparing our method to the deblurring-only and the super-resolution after deblurring method, we can confirm that the proposed method is the best among them.

Finally, we apply the technique to more complicated scenes where multiple humans are moving captured by hand-

(a) Input (b) Deblurring only

(c) SR only (d) SR + deblur (proposed)

Fig. 4. Real scene experimental results.

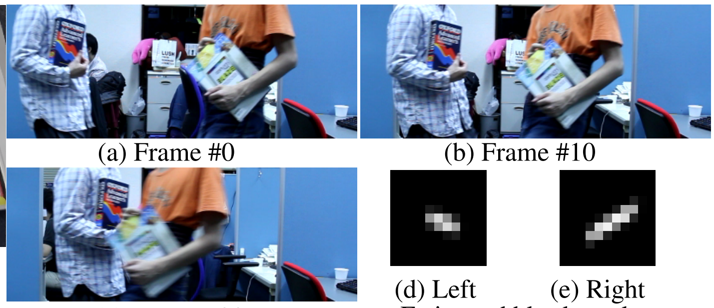
held camera as shown in Fig. 5(a)-(c). We use 20 frames (420x160 pixel) to perform 2x2 magnification. Since there are several motions in the sequence, segmentation results for short sequences are not the same, however, those are consistently integrated as shown in Fig. 6. Results with several methods are shown in Fig. 7. As shown in the results, we can confirm that our method is the best among them.

6. CONCLUSION

In this paper, we propose a method which realizes both deblurring as well as super-resolution simultaneously for an image sequence with multiple moving objects. In the method, we first divide the image into short sequences, and then, perform segmentation based on planar structure in the scene followed by integration of them. In addition, we successfully refine the precision of feature tracking result using BIPOC. Finally, we propose a method to simultaneously perform deblur and super-resolution for multiple images with scaling parameter for kernels. In the experiment, we successfully restore the high resolution image using the proposed method for both synthetic and real images. As our future work, simultaneous refinement of blur kernel is planned.

7. ACKNOWLEDGMENT

This work was supported in part by KAKENHI No.24500120 and SCOPE in Japan.



(a) Frame #0 (b) Frame #10

(c) Frame #20 (d) Left (e) Right

Estimated blur kernels

Fig. 5. (a)-(c) Sample frame of input sequence (d)-(e) estimated blur kernel for left and right person on frame #0.

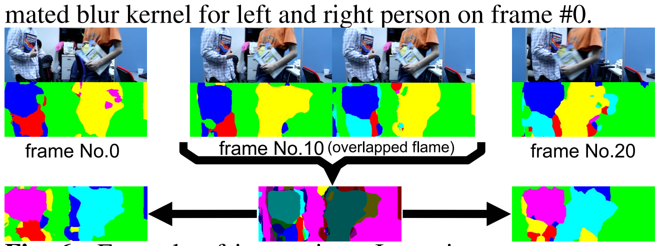
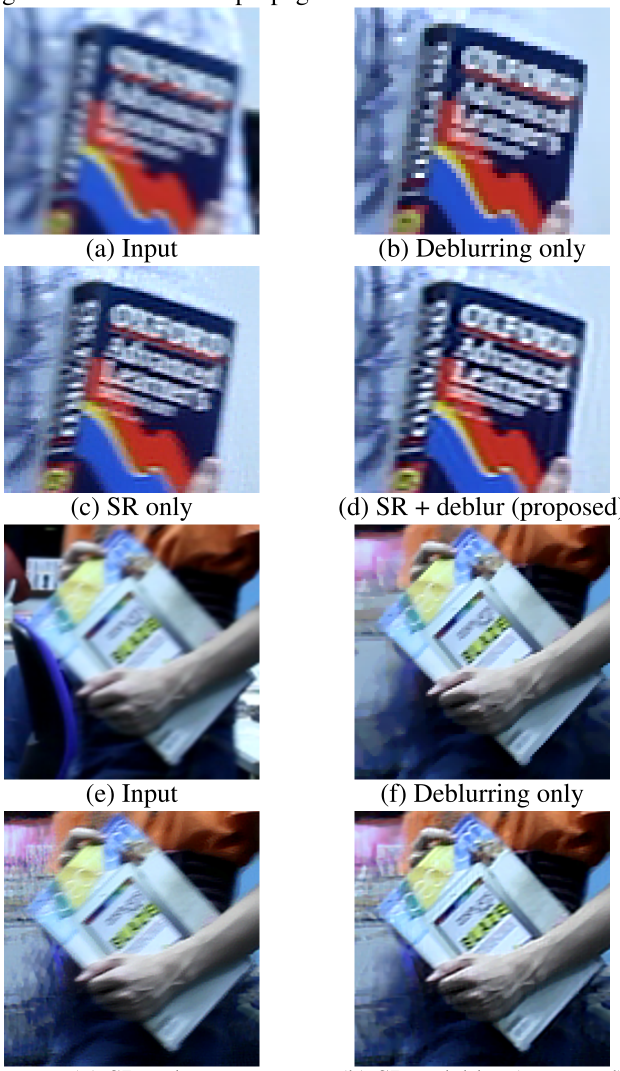


Fig. 6. Example of integration. Inconsistent segments are integrated and results are propagated to the entire frame.



(a) Input (b) Deblurring only

(c) SR only (d) SR + deblur (proposed)

(e) Input (f) Deblurring only

(g) SR only (h) SR + deblur (proposed)

Fig. 7. Result of multi objects scene: (a)-(d) input and result of left book and (e)-(h) input and result of right book.

8. REFERENCES

[1] Sina Farsiu, Michael Elad, and Peyman Milanfar, "A practical approach to superresolution," in *Visual Communications and Image Processing*, 2006, vol. 6077.

[2] Li Xu, Shicheng Zheng, and Jiaya Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2013, CVPR '13, pp. 1107–1114, IEEE Computer Society.

[3] Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee, "Dynamic scene deblurring," December 2013.

[4] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[5] Takuma Yamaguchi, Hisato Fukuda, Ryo Furukawa, Hiroshi Kawasaki, and Peter F. Sturm, "Video deblurring and super-resolution technique for multiple moving objects.," in *ACCV (4)*, Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, Eds. 2010, vol. 6495 of *Lecture Notes in Computer Science*, pp. 127–140, Springer.

[6] L B Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical Journal*, vol. 79, no. 6, pp. 745–754, 1974.

[7] Hadley William Richardson, "Bayesian-based iterative method of image restoration," *Journal of Optical Society of America*, pp. 55–59, 1972.

[8] Michal Irani Irani and Shmuel Peleg, "Improving resolution by image registration," *CVGIP*, vol. 53, pp. 231–239, 1991.

[9] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *SIGGRAPH*, 2006, pp. 787–794.

[10] Chandramouli Paramanand and Ambasamudram N. Rajagopalan, "Non-uniform motion deblurring for bilayer scenes," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 1115–1122, 2013.

[11] Sunghyun Cho, Jue Wang, and Seungyong Lee, "Vdeo deblurring for hand-held cameras using patch-based synthesis," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 64:1–64:9, 2012.

[12] Yunpeng Li, Sing Bing Kang, Neel Joshi, Steven M. Seitz, and Daniel P. Huttenlocher, "Generating sharp panoramas from motion-blurred videos," in *CVPR*, 2010, pp. 2424–2431.

[13] Weisheng Dong, D Zhang, Guangming Shi, and Xiaolin Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1838–1857, 2011.

[14] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *In ECCV*, 2006, pp. 404–417.

[15] Sei Nagashima, Takafumi Aoki, Tatsuo Higuchi, and Koji Kobayashi, "A subpixel image matching technique using phase-only correlation," in *Intelligent Signal Processing and Communications, 2006. ISPACS'06. International Symposium on*. IEEE, 2006, pp. 701–704.

[16] Carlo Tomasi and Takeo Kanade, "Detection and tracking of point features," Tech. Rep., International Journal of Computer Vision, 1991.

[17] Masayuki Tanaka, Yoichi Yaguchi, and Masatoshi Okutomi, "Robust and accurate estimation of multiple motions for whole-image super-resolution," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 649–652.

[18] Ville Ojansivu and Janne Heikkilä, "Image registration using blur-invariant phase correlation," *IEEE Signal Process. Lett.*, vol. 14, no. 7, pp. 449–452, 2007.