Space-Time Coded Imaging for Robust Depth and Motion Deblurring

Takeyuki Kobayashi†, Ryosuke Naito†, Fumihiko Sakaue†, Jun Sato (member)†

Abstract Recent progress in coded imaging has enabled us to recover sharp images from undesirable image blurs caused by depth and motion. However, the existing methods are limited to single types of blurs, such as depth blur or motion blur. The coded imaging must also tackle the problem of image noise caused by decreasing of input light. In this paper, we propose a method, which enables us to deblur both depth blur and motion blur simultaneously by coding image capture both in space and time. In particular, we show that by changing the aperture pattern efficiently, accurate image deblurring can be achieved under the existence of image noise.

Key words: Coded imaging, PSF, Deblur, Motion Deblur

1. Introduction

Recent progress in coded imaging has enabled us to recover sharp deblurred images from captured images with undesirable image blurs caused by the out of focus (i.e. depth blurs) and the relative motion between cameras and objects (i.e. motion blurs). Veeraraghavan et al. derived a coded aperture pattern which maximizes the minimum value of the Fourier spectrum of aperture, and showed that it enables us to deblur images obtained from varying depth scenes. Raskar et al. proposed a method which codes exposure time of the camera for obtaining broadband frequency characteristics of the PSF under motion blur, and showed that it enables us to deblur motion blur efficiently. However, these existing deblurring methods can only be applied to single types of blurs, such as a depth blur or a motion blur. Thus, we in this paper propose a method which enables us to recover from depth blur and motion blur simultaneously.

Although we can simply combine a coded aperture and a coded exposure for deblurring both the depth blur and the motion blur, it degrades the image quality of deblurred images much. This is because the simple combination of a coded aperture and a coded exposure drastically diminishes the amount of input light to the sensor, and causes a severe degradation in the signal to noise ratio.

To cope with this problem, we in this paper propose a robust space-time coding of image capture. In our method, the image capture is coded both in space and in time. That is, the coded aperture pattern changes dynamically during the exposure time keeping the same amount of input light with that of the existing coded aperture or the coded exposure. The proposed method enables us to deblur both depth blur and motion blur simultaneously without degrading image quality.

2. Images under Depth and Motion Blur

Suppose a 3D object is not at in-focus depth, and the captured image of the object has depth blur. In this case, the blurred image \( s(x,y) \) can be described by using the point spread function (PSF) of the depth blur \( h_d(x,y) \) and the original unblurred signal \( f(x,y) \) as follows:

\[
s(x,y) = h_d(x,y) * f(x,y)
\]

where, \( * \) denotes the convolution of two signals.

If the 3D object is not at the in-focus depth and moves in the 3D space, the captured image of the object also has motion blur. Assuming that the depth blur is unchanged during object motion, the blurred image \( g(x,y) \) is described by using the PSF of motion blur \( h_m(x,y) \) and depth blur image \( s(x,y) \) as follows:
From (1) and (2), the relationship between image $g(x, y)$ blurred by depth and motion and original unblurred signal $f(x, y)$ can be described as follows:

$$g(x, y) = h(x, y) * f(x, y)$$  \hspace{1cm} (3)

where, $h(x, y)$ is the PSF of combined blur, and is described as follows:

$$h(x, y) = h_m(x, y) * h_d(x, y)$$  \hspace{1cm} (4)

Thus, if we know $h(x, y)$, the original signal $f(x, y)$ can be recovered by performing the deconvolution on $g(x, y)$ based on $h(x, y)$. However, if we perform the deconvolution on images captured by a standard camera, the recovered images suffer from an outbreak of ringings and imperfect deblurring. This is because the PSF of the standard circular aperture and flat exposure has many zero crossings and drastic changes in its power spectrum, and they cause loss of original signal information and instability in the deconvolution of image signals.

3. Space-Time Coding of Image Capture

To cope with the problem of standard apertures and exposures, we in this paper consider coding of image capture both in space and in time. That is, we change the coded aperture patterns sequentially during the exposure time as shown in Fig. 1, so that it enables us to deblur the depth blur and the motion blur simultaneously. This can be achieved by extending the idea of Raskar et al., who extracted the optimal exposure pattern by maximizing the minimum in the frequency characteristics of exposure pattern. Unfortunately, their method cannot be applied directly in the case of combined blur of depth and motion, since the aperture pattern derived by maximizing its frequency characteristics does not maximize the frequency characteristics of an image captured under combined blurs. Thus, we in this paper derive a sequential aperture pattern, not from its own frequency characteristics, but from the frequency characteristics of the PSF observed under a sequential aperture pattern.

Let us consider the Fourier transform $H(u, v)$ of PSF, $h(x, y)$, observed under combined blur, and let $\Phi(H(u, v))$ be its log power spectrum. The PSF of combined blur $h(x, y)$ can be computed by projecting a moving 3D point into the image through the sequential coded aperture during the exposure time. For deblurring images stably, it is desirable for the PSF to have flat and high frequency responses through all frequencies. Thus, we consider the following cost function $E$, and find a sequence of coded aperture patterns, which maximizes $E$:

$$E = \min \left[ \Phi(H(u, v)) \right] - \alpha \cdot \text{var}[\Phi(H(u, v))]$$  \hspace{1cm} (5)

where $\min[\cdot]$ denotes the minimum value and $\text{var}[\cdot]$ denotes the variance. The $\alpha$ is an appropriate constant. The sequence of coded aperture patterns which maximizes $E$ has flat and high frequency responses through all frequencies, and thus the deconvolution of images with the PSF generated by the sequence of apertures becomes much more stable than that of standard circular apertures and flat exposures. The derived space-time coding is also superior to the existing coded apertures and coded exposures, since the derived space-time coding is optimized for the combination of depth blur and motion blur, while the existing methods are optimized just for a single blur.

If there are no restrictions on space-time coding, the sequential aperture pattern which maximizes $E$ will be a pinhole aperture at a single time, since it has the best frequency characteristics. However, we cannot obtain a sufficient amount of light under a pinhole aperture at a single instance, and as a result the signal to noise ratio of the obtained image will become very poor. Thus, we in this paper maximize $E$ subject to opening half the area of the aperture at each time instant. As a
result, the amount of input light reduces only up to \(1/2\) in the proposed space-time coding, while the simple combination of coded aperture and coded exposure reduces input light up to \(1/4\), since the coded aperture reduces the input light with \(1/2\), and the coded exposure further reduces the light with \(1/2\). Thus, the proposed space-time coding enables us to deblur depth and motion blurs simultaneously without degrading the signal to noise ratio of deblurred images.

4. Optimal Space-Time Coding

For deriving an optimal sequential aperture pattern, we first consider the number of aperture alteration during the exposure. For this objective, we evaluated the relationship between the number of apertures in a sequence of apertures and the cost function \(E\). In this evaluation, we derived a globally optimal sequential pattern of \(3 \times 3\) apertures varying the number of apertures in the sequence from 1 to 7, and evaluated the cost function \(E\) of these sequential apertures with \(\alpha = 6\). The exhaustive search was used in this derivation. We chose these aperture size and number, since it is difficult to derive globally optimal sequential patterns under higher resolution apertures with longer sequences because of combinatorial explosion. We derived the optimum regularization parameter \(\alpha = 6\) from various experimental data.

The optimal space-time coding was derived under the following conditions. The image sensor was situated 600 mm from the lens center, and since the focal length was 300 mm, the object situated 600 mm from the lens center was focused without any depth blur. However, since the object was situated at 700 mm in this experiment, the image of the object has depth blur. The diameter of aperture was 50 mm, and thus the diameter of depth blur was 16.7 pixels. Furthermore, since the object moved horizontally with a speed of 20 pixels/frame, the image also has motion blur.

Fig. 2 shows the globally optimal sequence of apertures derived at each number of apertures, and Fig. 3 shows the change in cost function \(E\) with respect to the change in the number of apertures. The horizontal axis is the number of apertures in the sequence, and the vertical axis is the cost function \(E\).

From Fig. 3, we find that the value of cost function increases and the frequency characteristics improve as the number of apertures increases. However, the improvement declines when the number of apertures increases. Also, it becomes more difficult to find globally optimal aperture patterns as the number of apertures increases. Therefore, we in this paper derived the optimal space-time coding fixing the number of apertures to 4 and the resolution of apertures to \(6 \times 6\). By maximizing \(E\) under these conditions using the hill climbing method with ten thousand random initial patterns, we find that the optimal sequence of apertures is as shown in Fig. 4.

To see the efficiency of the derived space-time coding, we compare the frequency characteristics of the derived space-time coding and the conventional apertures and exposures. Fig. 5 shows the frequency characteristics of the proposed space-time coding and the conventional methods, i.e. a simple circular aperture with a flat exposure, Veeraraghavan’s coded aperture\(^1\) with a flat exposure and Raskar’s coded exposure\(^11\) with a circular aperture. These frequency characteristics are observed along the direction of motion. The vertical axis is the log power spectrum and the horizontal axis is the normalized frequency. As shown in this figure, the proposed space-time coding has much flatter and higher frequency responses over all frequencies than the standard circular aperture with a flat exposure and the existing coded imaging. This means the
proposed space-time coding enables us to deblur images more accurately and reliably than the existing methods.

5. Real Image Experiments

We next show the results from real image experiments. In this experiment, a flat object with a "Lena" image was mounted on a moving stage and moved horizontally 0.1 mm/sec during an exposure time of 10 sec, so that the captured images have motion blur. Since the object is not at the in-focus position, the captured images also have depth blur. **Fig. 6** shows our experimental system. The distance from the lens to the center of the object is 250 mm. The diameter of aperture is 20 mm. The lens is 0.10× with optical magnification, F=2.8, W.D=205 mm and the distance to the focal plane from the camera is 454 mm. The camera is 1024 × 768 pixels, interline color CCD camera. The size of the image sensor is 4.81 mm × 3.62 mm.

**Fig. 7** (a), (b), (c) and (d) show captured images under the standard circular aperture with a flat exposure, Raskar’s coded exposure\(^{11}\) with a circular aperture, Veeraraghavan’s coded aperture\(^{11}\) with a flat exposure and the proposed space-time coding. In the proposed method, we took 4 sequential images with an exposure time of 2.5 sec by changing the coded apertures (Fig. 4), and integrated them into a single image with an exposure time of 10 sec. Then, we deblurred these observed images by using the Winner filter. **Fig. 8** (a), (b), (c) and (d) show images deblurred from Fig. 7 (a), (b), (c) and (d). As shown in these images, the proposed method provides us the best results. In particular, the image recovered from the proposed method is much sharper and has fewer artifacts than those recovered from the standard imaging and the existing coded imaging.

6. Quantitative Evaluation

As shown in the real image experiments, our method can efficiently deblur both the depth blur and motion blur simultaneously. However, it is difficult to evaluate the accuracy of deblurring quantitatively in the real image experiments. Thus, we in this section evaluate the proposed method quantitatively by using synthetic images generated by a lens simulator. The lens simulator can generate observed images accurately by tracing all the rays which come into the lens system. It can also control the spatial pattern of aperture and the temporal pattern of exposure arbitrarily. Thus, the image of moving objects with depth variations captured with a specific aperture and a specific exposure pattern, can be generated accurately. However, it assumes an ideal lens, and thus it neglects the aberration and other distortions.
By using the lens simulator, we generated an image of an object whose surface texture is "Pepper". The image size is 480 × 480 pixels, and the image intensity varies from 0 to 255. The Gaussian noise with a standard deviation of 0.5 was added to the image intensity. The amount of motion blur and depth blur was the same as that in Section 4.

We compared the accuracy of deblurring by generating input images using the lens simulator and deblurring them by using the proposed method and the existing methods. Fig. 9 (a) shows the deblurred image from the standard circular aperture with a flat exposure. As shown in this figure, the image blur is not eliminated properly from the standard aperture and exposure. Fig. 9 (b) shows the result from Raskar’s coded exposure with a circular aperture, and Fig. 9 (c) shows the result from Veeraraghavan’s coded aperture with flat exposure. As shown in these results the coded imaging provides us better quality of images than the standard aperture and exposure. However, they still have image blur and some artifacts in the images. This is because the existing coded imaging methods only consider single type of blur in images, while the actual images have both depth blur and motion blur. However, the image recovered from the proposed method is very sharp and has much fewer artifacts as shown in Fig. 9 (d).

Table 1 shows the PSNR of the deblurred images relative to the original image. In this comparison, we deblurred three types of images, i.e. an image with just motion blur, an image with just depth blur, and an image with both motion and depth blurs. As shown in Table 1, Raskar’s coded exposure provides us good results under motion blur, and Veeraraghavan’s coded aperture provides us good results under depth blur. However, the PSNR of these methods degrade in the case of motion and depth blurs, while the proposed method provides us good results even under motion and depth blurs. As shown in Table 1, the proposed method provides us the same PSNR of 30.581 dB under pure motion blur in Table 1. Similarly, since the coded exposure does not work without depth blur, the two methods have the same PSNR of 30.069 dB under pure depth blur.

Fig. 10 shows the relationship between the object motion and the PSNR of deblurred images in each method.
method. The vertical orange line at motion of 20 pixels shows the values used for deriving the optimal patterns. We find that the proposed method has higher PSNR than the existing methods in motion blur from 7 pixels/frame to 44 pixels/frame. Since the space-time coding is optimized for the combinations of depth blur and motion blur simultaneously, the maximum value of PSNR for the proposed method shifted from the orange line in Fig. 10.

Fig. 11 shows the relationship between the object depth and the PSNR. The horizontal axis is depth from the focal point. We find that the proposed method has higher PSNR than the existing methods at depths from 60 mm to 430 mm. Thus, our method is efficient not only for a specific combination of depth and motion blur but also for a wide range of their combinations.

Fig. 12 shows the relationship between the magnitude of image noise (σ) and the PSNR. The amount of motion blur and depth blur is the same as those used in the experiment in Fig. 10 and Fig. 11. Thus, the values at σ = 0.5 are the same as those plotted at the vertical orange line in Fig. 10 and Fig. 11. We find that the proposed method has much higher PSNR than the existing methods, when σ is smaller than 1.5.

Fig. 13 shows other example images deblurred by the proposed method and the existing methods, and Table 2 shows their PSNR. The amount of motion blur and depth blur, and other conditions are the same as those used in the experiment in Fig. 9. As shown in these results, the proposed method outperforms all the other methods.

7. Conclusion

In this paper, we proposed the robust space-time coding of image capture for deblurring depth blur and motion blur simultaneously. In particular, we derived an optimal sequential aperture pattern for deblurring both depth and motion blurs simultaneously under relatively low image noise. We showed that the derived space-time coding have better frequency characteristics than the existing coded aperture and coded exposure methods, hence it provides us better deblurring results. The efficiency of the proposed method was shown by comparing it with the existing coded imaging methods in real image experiments as well as in synthetic image experiments. From these experiments, we found that the proposed method is very powerful for deblurring depth blur and motion blur simultaneously.

References


Takeyuki Kobayashi received the B.E. degree from Nagoya Institute of Technology, Japan in 1998, and the M.E. degree from Nagoya University, Japan in 2000. He is currently a Ph.D. candidate at the Department of Computer Science and Engineering, Nagoya Institute of Technology and a Director of ATA Co., LTD. His research interests include Coded Imaging.

Ryosuke Naito received the B.E. and M.E. degrees in Nagoya Institute of Technology, Japan in 2011 and 2013, respectively. His research interests include Coded Imaging.

Fumihiko Sakaue received the B.E., M.E. and Ph.D. degrees in Information technology from Okayama University, Japan in 2001, 2003 and 2006, respectively. He is currently a research assistant professor at the Department of Computer Science and Engineering, Nagoya Institute of Technology. During 2011–2012, he was a visiting researcher at the Robotics Institute, Carnegie Mellon University. His research interests include computer vision, photometric image analysis, pattern recognition and mixed reality.

Jun Sato received the B.Eng. degree from Nagoya Institute of Technology, Japan in 1984, and the Ph.D. degree in information engineering from the University of Cambridge, England in 1997. From 1996 to 1998 he was a research Associate at the Department of Engineering, University of Cambridge. During 1997–1998, he was an invited researcher at the ATR Human Information Processing Research Laboratories, Kyoto. He joined the Department of Electrical and Computer Engineering, Nagoya Institute of Technology in 1998 as an associate professor. He is currently a professor at the Department of Computer Science and Engineering, Nagoya Institute of Technology. His research interests include computer vision, visual geometry, coded imaging, coded projection and augmented reality. He was awarded the Best Science Paper Prize of BMVC twice in 1994 and 1997, the Best Paper Prize of SSII1999 and the MIRU Frontier Prize in 2012, etc. He is the editor of the International Journal of Computer Vision, and a member of the IEEE, IEICE, IPSJ, VRSJ and ITE.