Enhanced reconstruction of partially occluded objects with occlusion removal in synthetic aperture integral imaging

Zhiliang Zhou (周志良)[1*], Yan Yuan (袁 艳)[2], Xiangli Bin (相里斌)[1,3], and Qian Wang (王 潜)[2]

[1]*Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, China*

[2]*Key Laboratory of Precision Opto-Mechatronics Technology, Ministration of Education, Beihang University, Beijing 100191, China*

[3]*Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100190, China*

[*]*Corresponding author: zzlabc@mail.ustc.edu.cn*

Received September 26, 2010; accepted November 26, 2010; posted online March 28, 2011

Synthetic aperture integral imaging provides the ability to reconstruct partially occluded objects from multi-view images. However, the reconstructed images suffer from degraded contrast due to the superimposition of foreground defocus blur. We propose an algorithm to remove foreground occlusions before reconstructing backgrounds. Occlusions are identified by estimating the color variance on elemental images and then deleting it in the final synthetic image. We demonstrate the superiority of our method by presenting experimental results as well as comparing our method with other approaches.

OCIS codes: 100.0100, 100.3010, 100.4999.
doi: 10.3788/COL201109.041002.

Synthetic aperture integral imaging utilizes a translating camera[1] or a lenslet/camera array[2,3] to simulate a virtual lens with a large relative aperture. By focusing on the occluded surfaces of interest, foreground occlusions are severely blurred out due to defocus, whereas the objects behind them are sharpened to be seen on the reconstructed image. This ability of seeing beyond occlusions is potentially valuable in surveillance and tracking[2−6]. In synthetic aperture images, the defocus blur of foreground occlusions is superimposed on the image of in-focus objects, reducing the contrast and clarity of the synthetic images. Researchers have suggested several approaches to improving the reconstruction of occluded objects by eliminating foreground occlusions. Wilburn *et al.* assumed that occluded objects moved behind still occlusions[3], but their method required collecting continuous pictures over a period of time. Vaish *et al.* developed various cost functions for reconstructing occluded surfaces and compared their results with the reconstruction by voxel coloring method[7]. Before imaging the occluded objects, a totally blue or white screen was placed behind the occluders; hence, these colors were different from the background represented by the occluding objects. Yong *et al.* proposed a contrast-enhancing method for the reconstruction of heavily occluded objects, but their method was limited for very large occlusions with similar colors[8]. Shin *et al.* suggested occlusion removal techniques for computational integral imaging systems, which involved a lenslet array and a pickup camera[9,10]. Their approaches need two transforms between elemental images and sub-images before and after the removal of the occlusions. They also require either a reference image of the unoccluded objects[9] or an estimated depth map[10].

In this letter, we propose an approach to enhance the synthetic image quality by removing foreground occlusions based on variance estimation on the sequence of elemental images. Comparing our method with the two other related ideas in multi-perspective imaging is important. DaneshPanah *et al.* introduced the spectral radiation pattern and estimated its variance at different depths for the profilometry of three-dimensional (3D) objects[11]. Although their approach was effective in measuring convex objects, they did not consider the case of occlusions. Concurrent with the submission of this letter, Lee *et al.* published their work on occlusion removal techniques[12]. However, they did not remove the occlusion pixels but only set them to zero. This decreases the synthetic image contrast in the process of reconstruction because the zeros would be averaged on the pixels of the occluded objects. The effect is more significant in the case of occlusions that cover a range of depths. We will address these issues in our algorithm.

Figure 1(a) illustrates a typical schematic of synthetic aperture integral imaging using a camera array. The layout is depicted in two-dimensional (2D) for simplicity. Suppose there are N elemental images of size $m \times n$: I_1, I_2, $\cdots$, I_N. To reconstruct the occluded object, we project the elemental images to the background occluded plane. The synthetic aperture image is obtained by shifting and averaging the elemental images[6]:

$$I_{\rm b}(i,j) = \frac{1}{N_{\rm b}(i,j)} \sum_{k=1}^{N} I_k(i + \Delta i_{k{\rm b}}, j + \Delta j_{k{\rm b}}),$$
$$i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, n, \quad (1)$$

where $I_{\rm b}$ is the reconstructed image, $\Delta i_{k{\rm b}}$ and $\Delta j_{k{\rm b}}$ are the shift values determined by the projecting depth, and $N_{\rm b}$ is the number of overlapped pixels contributing to $I_{\rm b}$. Without generality, we only consider two points P and Q on the occluded and occluding surfaces, respectively. As described in Fig. 1(b), pixels (in dotted box) corresponding to P are aligned and sharpened, whereas pixels (in blue) corresponding to Q are dispersed to different

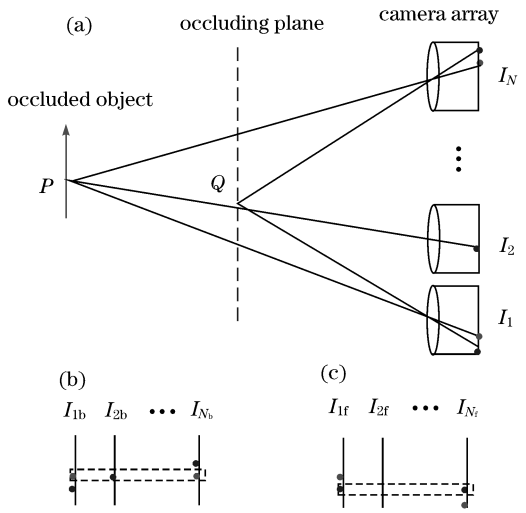

Fig. 1. (Color online) (a) Schematic of synthetic aperture integral imaging; (b) elemental images projected to the background occluded object; (c) elemental images projected to the foreground occluding plane.

locations and blurred out.

In Fig. 1(a), Q occludes P in the second camera. Reconstructing the occluded point involves redundant pixels representing the occluding point. To identify and remove the pixels corresponding to Q, we project elemental images to the occluding plane. In Fig. 1(c), these pixels (in dotted box) are aligned to the same location in each projected image. Different from P, Q is completely unoccluded and clearly visible in any camera. The irradiance of Q will be approximately equal in every camera if we assume the scene as Lambertian surfaces and ignore natural vignetting effects. Therefore, the overlapped pixels at the location corresponding to the occluding point should appear in similar colors. In contrast, if the pixels are overlapped at a location that does not correspond to an occluding point, they will be projected from different points and have diverse colors. This allows us to use the measure of color variance to distinguish the locations where the projected pixels are alike. The variance v of pixels at the location (i, j) is computed by the following equations:

$$I_{kf}(i,j) = I_k(i + \Delta i_{kf}, j + \Delta j_{kf}), \quad k = 1, 2, \cdots, N, \quad (2)$$

$$I_f(i,j) = \frac{1}{N_f(i,j)} \sum_{k=1}^{N} I_{kf}(i,j), \quad (3)$$

$$v(i,j) = \frac{1}{N_f(i,j)} \sum_{k=1}^{N} [I_{kf}(i,j) - I_f(i,j)]^2, \quad (4)$$

where I_{kf} is the elemental image projected to the foreground occluding plane, Δi_{kf} and Δj_{kf} are the shift values depending on the foreground depth, N_f is the number of overlapped pixels (typically identical to the number of elemental images), and v is the created variance map.

To detect the locations where pixels vary slightly, we segment the variance map using thresholding techniques. Instead of manually choosing a fixed value, we compute the threshold automatically using the mean of the variance map. The threshold is calculated by

$$t = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} v(i,j). \quad (5)$$

Variances under the threshold indicate the pixels with similar colors. These pixels are identified as foreground occlusions and should be deleted in elemental images. Therefore, the matted elemental images are acquired by

$$I'_{kf}(i,j) = \begin{cases} I_{kf}(i,j), & \text{if } v(i,j) > t; \\ \text{null}, & \text{otherwise.} \end{cases} \quad (6)$$

Equations (2)–(6) demonstrate the removal of a single occluding plane. If the occlusions spread over a range of depths, they can be eliminated layer by layer from the closest to the farthest with respect to the cameras. $N_f(i,j)$ in Eqs. (3) and (4) may be not identical to the number of elemental images. During the elimination of occlusion layers, $N_f(i,j)$ tends to decrease as the removed pixels (which are null in Eq. (6)) should not be taken into account. With a complete sweep of the foreground occlusions, we project the matted elemental images back to the background plane and compute their average. The final synthetic image is then obtained as

$$I'_{kb}(i,j) = I'_{kf}(i - \Delta i_{kf} + \Delta i_{kb}, j - \Delta j_{kf} + \Delta i_{kb}), \quad (7)$$

$$I'_b(i,j) = \frac{1}{N'_b(i,j)} \sum_{k=1}^{N} I'_{kb}(i,j), \quad (8)$$

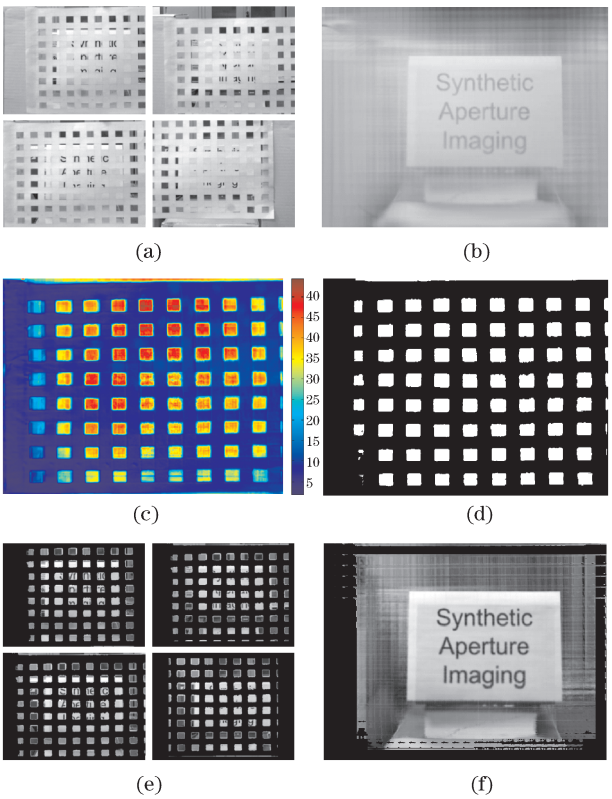

Fig. 2. (a) Elemental images; (b) synthetic aperture image without removal of occlusions; (c) variance map of elemental images projected to the foreground; (d) occlusion image created by thresholding the variance map; (e) matted elemental images; (f) synthetic aperture image with removal of occlusions.

Fig. 3. Reconstructed images of the occluded CD case using (a) entropy, (b) median, (c) focus, (d) stereo, (e) voxel coloring, and (f) our algorithm.

where $I'_{k\mathrm{b}}$ is the matted image projected to the background, N'_{b} is the number of overlapped pixels contributing to $I'_{k\mathrm{b}}$, and I'_{b} is the synthetic aperture image without occlusions. As mentioned above, $N'_{\mathrm{b}}(i,j)$ may also be smaller than the number of elemental images.

We implemented experiments to demonstrate the process of our algorithm. In our setup, a plane with the text Synthetic Aperture Imaging was placed behind a white planar mask composed of square holes. We captured the elemental images (Fig. 2(a)) using a single camera translating on an X-Y platform. As shown in Fig. 2(b), the occluded text is reconstructed by directly focusing on the occluded plane and averaging the shifted elemental images. The synthetic image is strongly degraded by the defocus blur of the foreground mask. In Fig. 2(c), the variance map is created from the elemental images projected to the occluding mask. Thresholding the variance map yields the occlusion image in Fig. 2(d) and the matted elemental images in Fig. 2(e). By focusing the matted images on the occluded text, the synthetic aperture image is shown after removing occlusions, as illustrated in Fig. 2(f). Compared with Fig. 2(b), Fig. 2(f) shows significant improvement on image contrast and clarity.

To evaluate the reconstruction performance, we compared our algorithm with the approaches in Ref. [7], where four cost functions and voxel coloring were explored for reconstruction of occluded objects. Figures 3(a)−(d) show the reconstructed compact disk (CD) case (occluded behind plants) using entropy, median, focus, and stereo, respectively. Without removal of occlusions, all the images suffer from quality degradation due to foreground blur. In Fig. 3(e), the image is reconstructed using voxel coloring, which also enables removing foregrounds but causes some pixels to be completely unrecoverable. The proposed approach produces the reconstruction in Fig. 3(f), which recovers almost the whole picture of the CD case.

We implemented the numerical evaluation of Fig. 3 using recognition algorithms based on the mean absolute difference (MAD)[13]. MAD finds the similarity between two images. We compared the reconstructed images with a reference image of an unoccluded target by calculating the MAD matrix:

$$s(x,y) = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} |I(i+x, j+y) - J(i,j)|, \quad (9)$$

where I is the reconstructed image of the scene, J is the original image (with size $m \times n$) of the unoccluded target, and x, y denote the tracking location in the reconstructed scene. The minimum $s(x, y)$ represents the best match between the original and the reconstructed targets. The recognition results are illustrated in Fig. 4. For illustrative purpose, we depict the reciprocal of the MAD values (s^{-1}) with respect to the tracking locations in x- and y-directions. The s^{-1} peaks, whose values are labeled in Fig. 4, indicate the locations of the occluded target (the CD case). By comparing the sharpness of the peaks, our reconstruction recognizes the target more precisely than do the others.

In conclusion, we have presented an approach to improve the reconstruction of partially occluded objects in synthetic aperture integral imaging. By removing the occlusions before background reconstruction, the proposed algorithm enables the enhancement of the synthetic image contrast and the improvement of target recognition. The performance of our reconstruction can be further improved with a prior color calibration, which decreases

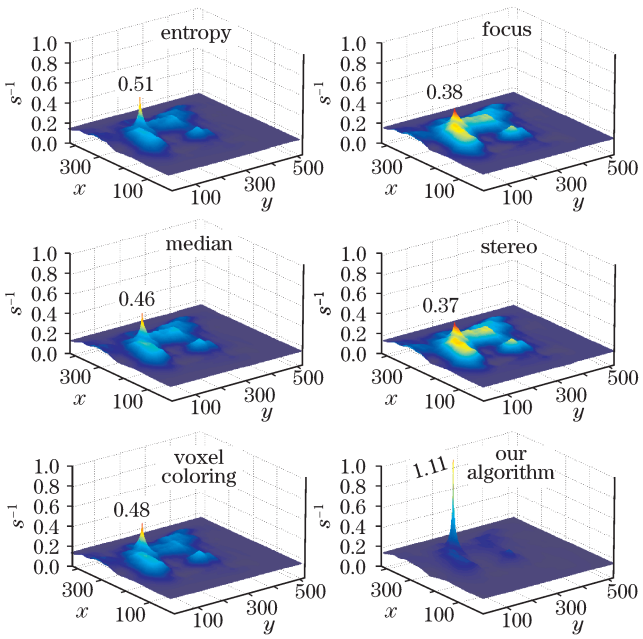

Fig. 4. Comparison of the recognition results. The labeled numbers denote the peak values.

color errors due to natural vignetting and sensor inaccuracy.

This work was supported by National Natural Science Foundation of China under Grant No. 60972088.

References

1. A. Isaksen, L. McMillan, and S. J. Gortler, in *Proceedings of ACM SIGGRAPH 2000* 297 (2000).
2. J.-S. Jang and B. Javidi, Opt. Lett. **27,** 1144 (2002).
3. B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, in *Proceedings of ACM SIGGRAPH 2005* 765 (2005).
4. B. Javidi, R. Ponce-Diaz, and S.-H. Hong, Opt. Lett. **31,** 1106 (2006).
5. N. Joshi, S. Avidan, W. Matusik, and D. J. Kriegman, in *Proceedings of IEEE Conference on Computer Vision* 1502 (2007).
6. M. Cho and B. Javidi, Opt Lett. **33,** 2737 (2008).
7. V. Vaish, R. Szeliski, C. L. Zitnick, and S. B. Kang, and M. Levoy, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 2331 (2006).
8. Y. S. Hwang and E.-S. Kim, Proc. SPIE **7237,** 723725 (2009).
9. D.-H. Shin, B.-G. Lee, and J.-J. Lee, Opt. Express **16,** 16294 (2008).
10. D.-H. Shin, H. Yoo, C.-W. Tan, B.-G. Lee, and J.-J. Lee, Opt. Commun. **281,** 4589 (2008).
11. M. DaneshPanah and B. Javidi, Opt. Lett. **34,** 1105 (2009).
12. B.-G. Lee, H.-H. Kang, and E.-S. Kim, 3D Res. **1(2),** 6 (2010).
13. H. Gharavi and M. Mills, IEEE Trans. Circuits Syst. **37,** 649 (1990).