Super-resolution Imaging With Occlusion Removal Using a Camera Array

Tingtian Li and Daniel P. K. Lun
Centre for Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Kowloon, Hong Kong
tingtianpolyu.li@connect.polyu.hk, enpklun@polyu.edu.hk

***Abstract*—In this paper, a novel algorithm which combines the super-resolution imaging and occlusion removal into a single and automatic procedure is proposed. By utilizing the visual parallax of objects at different depths and the sub-pixel information of the images captured by a camera array, we can estimate the shape of the occlusion and reconstruct the background at a higher resolution iteratively. The occlusion shape estimation is achieved by a new method called "seed growth", which treats the detected feature points of the occlusion as "seeds". These "seeds" will gradually grow until they reach the occlusion boundary. Experimental results show that the proposed algorithm can well remove the occlusion while super-resolving the background. It performs equally well when there are multiple occlusion objects or the object has irregular shape.**

Keywords—camera array; super-resolution imaging; occlusion removal; seed growth

I. INTRODUCTION

Occlusion is often an annoying problem in imaging. It is particularly the case in some image based monitoring or tracking applications where the operation may fail when encountering occlusions [1], [2]. To solve the problem, the synthetic aperture technology [3]-[6] were developed such that they can "see through" occlusions with a camera array. Such technology has been adopted in some applications like video tracking for relieving the occlusion problem [7]-[10]. However, these synthetic aperture techniques actually just blur the occlusion object in the image but not totally remove it. So shadows of the occlusion object often remain on the recovered background. While other methods try to totally remove occlusions using, for instance, the inpainting method [11]-[13]. However, those inpainting methods cannot detect the occlusion automatically. They require human to manually select the occlusion which limits their practical values. Furthermore, they work well for repeated texture regions but tend to fail for complicated textures and large unknown holes.

The huge resource requirement is also a major limitation of the synthetic aperture methods. While a camera array is needed to recover the occluded background, the amount of memory storage (or the data bandwidth for networking applications) for the images can be extremely high. To lower the resource requirement, one of the solutions is to reduce the resolution of the cameras in the array which however affects the image quality. As the image taken by each camera in the array can be considered as a sub-pixel lower resolution version of the target scene, super-resolution techniques can be applied to combine all images taken by the camera array into a single high resolution image. The study on super-resolution imaging has been conducted for long [14]-[16]. It is also suggested in [17] to use camera array to implement super-resolution imaging. However, as to our knowledge, there is no work that combines super-resolution imaging and occlusion removal into a single and automatic procedure.

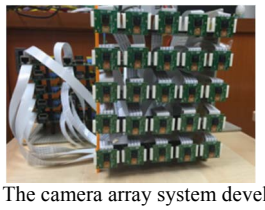


Fig.1. The camera array system developed by our team.

In this paper, a novel super-resolution imaging algorithm with automatic occlusion removal is proposed. The new algorithm firstly identifies the feature points of the target and occlusion objects using a scale-invariant feature transform (SIFT) [18] detector. And due to the different parallax, we can separate the feature points into the target and occlusion ones by using the Random Sample Consensus (RANSAC) algorithm [19]. Then we use a Lucy–Richardson (LR) deconvolution [17] method which can blur the occlusion to implement the super-resolution imaging. In particular, we add the Huber prior [16] to the iterative process to enhance the quality of the image edges. At every iteration, we estimate the shape of the occlusion object by using a novel thought called "seed growth" method. It is then removed gradually from the image during the iteration. In our experiments, a camera array system as shown in Fig.1 is used to capture the images. This system is composed of 25 Raspberry Pi modules each connected to a camera generating low resolution pictures. The experimental results show that, comparing with the traditional synthetic aperture methods, the proposed algorithm can well remove the occlusion object, while reconstruct the target at high resolution.

II. BASE MODEL

To simplify the discussion, let us assume that there is only one occlusion object in front of the target background. Then each image captured by the camera array with N cameras can be considered as the superimposition of the occlusion layer o and the background layer b, which can be expressed as follows:

$$y_i = K_i \cdot o_i + (\mathbf{1} - K_i) \cdot b_i \ , \quad (1)$$

where $i \in \{1,2,\dots,N\}$ denotes the camera index; K is a mask such that it is equal to 1 for occlusion appearing pixels and 0 otherwise. They are all vectors for mathematical convenience. Here, '$\cdot$' means element-wise multiplication and $\mathbf{1}$ is an all one vector. Although the above model has only one occlusion layer, it can be easily extended to multiple occlusion layers.

Let x_b denote the high resolution background vector we want to get. Then x_b is related to the low resolution background b_i captured by camera i by

$$b_i = M_i x_b \ , \quad (2)$$

where

$$M_i = DRW_{b,i} \ . \quad (3)$$

In (3), $W_{b,i}$ denotes the warping matrix for the background with respect to camera i; and D, R denote the decimation and blurring operator, respectively. By substituting (2) into (1), and since $K_i \cdot y_i = K_i \cdot o_i$, we can have

$$(\mathbf{1} - K_i) \cdot y_i + K_i \cdot M_i x_b = b_i. \quad (4)$$

By concatenating all K_i, y_i, M_i and b_i for all i as K, y, M and b, we can rewrite (5) into a compact form as follows:

$$(\mathbf{1} - K) \cdot y + K \cdot M x_b = b. \quad (5)$$

K, x_b and b are estimated with the following algorithm:

Algorithm I

Initialize K^0 and x_b^0.

Step 1: $(\mathbf{1} - K^t) \cdot y + K^t \cdot M x_b^t = b^{t+1}$. (6)

Step 2: $\Re\{b^{t+1}\} = x_b^{t+1}$. (7)

Step:3 $\mathbb{N}\{x_b^{t+1}\} = \{K^{t+1}\}$. (8)

Repeat Step 1-3 with the new K and x_b until converged.

In **Algorithm I**, K^t, x_b^t and b^t are the estimate of K, x_b and b at iteration t. $\Re\{.\}$ is a super-resolution operator that takes on an array of low resolution images to generate a high resolution estimate. In this paper, the LR deconvolution method with the Huber prior is adopted for the implementation of $\Re\{.\}$. It will be described in detail in Section III. $\mathbb{N}\{.\}$ is our proposed "seeds growth" method which will be described in Section IV. It makes use of the current estimated x_b to generate a better estimation of the mask. Note that the relationship between the mask K_i for different cameras i is not a perspective transformation, since different camera may capture different side of the occlusion object. Hence they have to be estimated one-by-one separately.

III. Adding Penalty Term to LR Deconvolution

Although there are many efficient super-resolution algorithms in the literature, here we follow the approach in [17] to adopt the LR deconvolution to implement the operator $\Re\{.\}$ as follows:

$$x_b^{t+1} = \Re\{b\} = \mathrm{diag}(x^t) M^T \left(\mathrm{diag}(M x_b^t)\right)^{-1} b. \quad (9)$$

It is firstly due to its simplicity. Besides, different from other super-resolution methods, we find it is the super-resolved version of synthetic aperture method that can blur the occlusion in the scene and keep the background recognized due to the mismatched homography as defined in M (the homography defined in M is for the background), while other methods may destroy the background. However, similar to many maximum likelihood super-resolution methods, direct application of the LR deconvolution can introduce observable high frequency noise. So a prior is often added to the iterative process. One popular choice is the Huber prior [16], which is constructed by using the Huber function defined as follows:

$$\rho(x,\alpha) = \begin{cases} x^2 & |x| \le \alpha \\ 2\alpha|x| - \alpha^2 & |x| > \alpha \end{cases} \ . \quad (10)$$

where α is a free parameter. One can see that the function is quadratic in the center and linear in the trails. We can use it as a prior function to our iterative process as follows:

$$p(x) = \frac{1}{Z} exp\{-\nu \textstyle\sum_c \rho(d_c x, \alpha)\} \ , \quad (11)$$

where z is a normalization constant; ν is the prior strength which is often chosen empirically. $d_c x$ measures the image gradient at the position and direction defined in the parameter set c as follows:

$$\begin{aligned} d_{m,n,1}x &= x_{m,n-1} - 2x_{m,n} + x_{m,n+1} \\ d_{m,n,2}x &= 0.5x_{m+1,n-1} - x_{m,n} + 0.5x_{m-1,n+1} \\ d_{m,n,3}x &= x_{m-1,n} - 2x_{m,n} + x_{m+1,n} \\ d_{m,n,4}x &= 0.5x_{m-1,n-1} - x_{m,n} + 0.5x_{m+1,n+1} \end{aligned} \ . \quad (12)$$

It can be shown that the Huber prior can be combined with the LR deconvolution as follows:

$$x_b^{t+1} = \mathrm{diag}(x^t)\left(M^T\left(\mathrm{diag}(M x_b^t)\right)^{-1} b - \nu Q^T \textstyle\sum_c \rho'(Q x_b^t, \alpha)\right) \quad (13)$$

where the matric Q performs the operation of d_c. By suitably selecting the parameter ν and α, the high frequency noise generated by the LR deconvolution can be suppressed while keeping the image's edge structure. Since (13) is also iterative, we combine Step 1 and 2 of **Algorithm I** as follows:

Algorithm II

Initialize K^0 and x_b^0.

Step 1: $x_b^{t+1} =$ (14)

$$\mathrm{diag}(x_b^t)\left(M^T\left(\mathrm{diag}(M x_b^t)\right)^{-1}[(\mathbf{1} - K^t) \cdot y + K^t \cdot M x_b^t] - \nu Q^T \textstyle\sum_{c \in C} \rho'(Q x_b^t, \alpha)\right).$$

Step:2 $\mathbb{N}\{x_b^{t+1}\} = \{K^{t+1}\}$ (15)

Repeat Step 1 and 2 with the new K until converged.

In our experiments, the initial x_b^0 is obtained by using the traditional synthetic aperture method which focuses on the background. K^0 is all zero vector.

IV. SEED GROWTH METHOD FOR LEARNING THE MASK

To estimate the mask, we introduce in this section a new algorithm called the "seeds growth" method. For the proposed algorithm, we firstly use the SIFT to detect the feature points in the scene; then use the RANSAC algorithm to estimate the homography based on the feature points. We denote it as H_b. Assume that the background feature points are more than the occlusion ones. The estimated H_b should fit for the background but not the occlusion. Then those points fitting for H_b are excluded and RANSAC is done again to find the homography H_o. Among the rest, those feature points fitting for H_o belong to the occlusion. The feature points for the background and occlusion can thus be separated.

We treat those feature points of the occlusion as "seeds" for finding the mask. For each seed, we initially set a small window centered at it. Then the four borders of each window will grow iteratively and stop until it just reach the occlusion boundary. Then all windows are merged together to form the mask. More specifically, the windows are grown based on a probability function as follows:

$$P\left(s_1|l_{q,i}, l_{q,x_{b,i}^t}\right) = \begin{cases} 1, & if\ d\left(l_{q,i}, l_{q,x_{b,i}^t}\right) < \text{length}\left(l_{q,i}\right)\cdot TH^t \\ \exp\left(-\frac{d\left(l_{q,i}, l_{q,x_{b,i}^t}\right) - length\left(l_{q,i}\right)\cdot TH^t}{2\cdot length\left(l_{q,i}\right)\cdot\sigma^2}\right), & if\ d\left(l_{q,i}, l_{q,x_{b,i}^t}\right) \geq \text{length}\left(l_{q,i}\right)\cdot TH^t \end{cases}, \quad (16)$$

where $l_{q,i}$ refers to the pixels of the qth border of a window in the image taken by camera i of the array; $l_{q,x_{b,i}^t}$ refers to the pixels at the same position of an image transformed from the estimated background x_b^t according to the homography with respect to the position of camera i. The function $\text{length}\left(l_{q,i}\right)$ gives the number of pixels in $l_{q,i}$ and $d\left(l_{q,i}, l_{q,x_b^t}\right)$ is just the l_1-norm of the difference between $l_{q,i}$ and l_{q,x_b^t}. TH^t is a threshold determined using the k-Nearest Neighbors approach with the feature points positions on the background to estimate the difference at the occlusion rims between x_b^t and y_i. Due to the page limitation, it cannot be described in detail here. $P\left(s_1|l_{q,i}, l_{q,x_b^t}\right)$ is the possibility that a window border has grown beyond the estimated occlusion region according to the difference between the estimated background and the observed images. The growth rate is inversely proportional to the probability as follows:

$$S_q = S_{max}\cdot\left[1 - P\left(s_1|l_{q,i}, l_{q,x_b^t}\right)\right]. \quad (17)$$

where S_q denotes the growth step size for the qth border of the window, and S_{max} is the maximum growth we allowed. So in each iteration, the qth border of the window will move outward in its normal direction by S_q pixels. The length of the two adjacent borders will also be adjusted accordingly. When all windows stop growing, they are combined to form the mask K in (16). Three examples are shown in Fig.2. For objects with rectangular shape, the masks can fit the occlusions very well. For objects with irregular shape, we can still roughly get the mask of the occlusion object, which is sufficient as shown in the results evaluation in section V.

V. EXPERIMENTS AND EVALUATION

As mentioned before, we do not find an existing approach that can achieve exactly the same as the proposed algorithm. So our comparison is only based on evaluation of the proposed algorithm and compare with the synthetic aperture method and the LR deconvolution with Huber prior. In the experiments, a high resolution image of the background is first taken and used as the ground truth (Fig.3(f)). Then the low resolution images with occlusions are acquired by the camera array and then enlarged with interpolation to the same size as the high resolution image. An example is shown in Fig.3(b). Such interpolated images are used in the synthetic aperture method. The result is shown in Fig.3(c). It can be seen that although the synthetic aperture method can "see through" the occlusion to a certain extent, it leaves a big shadow covering the background. Besides, it cannot super-resolve low resolution images; the resulting image is as blur as the original ones. When testing the LR deconvolution with Huber prior method, we use the low resolution images directly captured by the camera array. As mentioned before, the LR deconvolution method can give a result similar to the synthetic aperture method. It can also "see through" the occlusion (see Fig.3(d)). Besides, it can super-resolve the images to generate a clear high resolution image. The Huber prior further improves the image by removing the high frequency noise. However, the shadow left is quite annoying. Fig.3(e) shows the result of the proposed algorithm. It can be seen that we can almost perfectly reconstruct the background in high resolution compared with the ground truth (Fig.3(f)) without any occlusion. The only artifact appears at the lower part of the image which is indeed the table in the scene. It is because, in the initialization step, the homography is estimated based on the background plane. So the pixels of the desk, which is a bit closer to the camera array, cannot be perfectly reconstructed. Fig. 4 shows some intermediate results during iterations.

The base model can also be extended to the case of multiple occlusion objects. Fig.5(a) and (b) shows the result of removing two occlusion objects at different depths. It is indeed achieved by adding another occlusion feature point separation step at the initialization stage. The proposed algorithm can also recover the background plane with higher resolution well. Similar result is achieved for irregular shape objects as shown in Fig.5(c) and (d). Although the mask derived as shown in Fig.2 is not very accurate, we can still remove the occlusion object and super-resolve the background. Fig. 6 shows another example. It shows that the proposed algorithm can perform equally well for different backgrounds. Due to the page limitation, we cannot include more results.

VI. CONCLUSION

In this paper we have presented a novel framework which can super-resolve the background and automatically remove occlusions and if any, at the same time. In the proposed algorithm, we utilize the different parallax of the background and the occlusion object for separating their feature points. Based on them, we developed a new "seed growth" method for estimating the occlusion masks. The super-resolution is

carried out based on the LR deconvolution method with the modifications of including the Huber prior and the estimated occlusion mask. Experimental results show that the proposed algorithm can well remove the occlusion and super-resolve the background at the same time. To the best of our knowledge, there is no existing work that achieves exactly the same function.

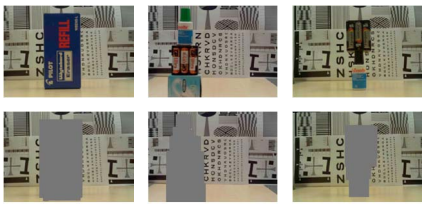
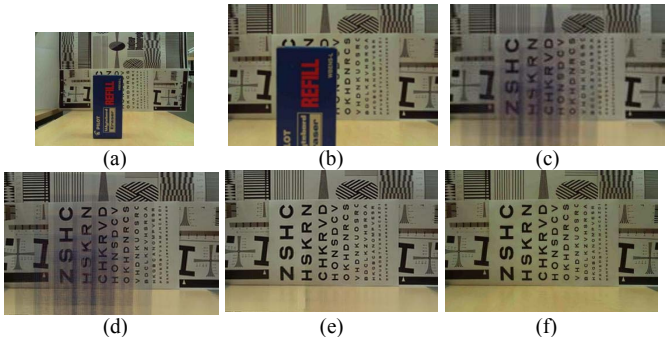


Fig.2. Original pictures (top row) and their corresponding mask estimation results (bottom row)

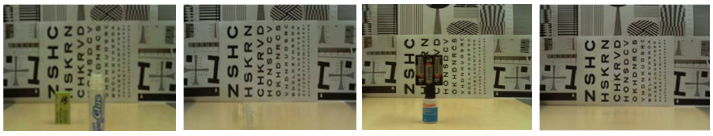


(a) (b) (c)

(d) (e) (f)

Fig.3. (a) Original low resolution image captured by the center camera of the array. (b) High resolution image interpolated from (a). Results of using (c) the synthetic aperture method; (d) the LR deconvolution plus Huber prior; and (e) the proposed algorithm. (f) High resolution ground truth.

Fig. 4. From left to right: low resolution image, the initial result we input, the result after 5 iterations,the result after 10 iterations and the final result.



(a) (b) (c) (d)

Fig.5. (a) Two occlusion objects. (c) Irregular shape object. (b) and (b) Results of using the proposed algorithm on the scenes in (a) and (c), respectively.

Fig. 6. The occluded low resolusion image (left); result of using the porposed algorithm (middle); the original high resolution ground truth (right).

ACKKNOWLEDGEMENT

This work is supported by the Hong Kong Polytechnic University under student account number RU9P.

REFERENCES

[1] K. H. Ho, W. K. Cheuk, and Daniel P. K. Lun, “Content-based scalable H.263 video coding for road traffic monitoring”, IEEE Trans. Multimedia, vol. 7, no. 4, pp. 615-623, 2005.

[2] Budianto and Daniel P. K. Lun, “Inpainting for Fringe Projection Profilometry Based on Geometrically Guided Iterative Regularization”, IEEE Trans. Image Processing, vol. 24, no. 12, pp. 5531-5542, 2015.

[3] A. Isaksen, L. McMillan and S. J. Gortler, “Dynamically reparameterized light fields” in Proc. Conf. Computer graphics and interactive techniques, pp. 297-306, 2000.

[4] B. Wilburn, N. Joshi, V. Vaish, Eino-Ville Talvala, Emilio Antunez, Adam Barth, et al, “High performance imaging using large camera arrays”. ACM Trans. Graphics (TOG), vol. 24, no. 3, pp. 765-776, 2005.

[5] M. Levoy, B. Chen, V. Vaish, Mark Horowitz, Ian McDowall and Mark Bolas, “Synthetic aperture confocal imaging”, ACM Trans. Graphics (TOG), vol. 23, no.3, pp. 825-834, 2004

[6] V. Vaish, R. Szeliski, C. L. Zitnick, Sing Bing Kang and Marc Levoy, “Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures”, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 2: pp. 2331-2338, 2006.

[7] Tao Yang, Yanning Zhang, Rui Yu, Xiaoqiang Zhang, Ting Chen and Lingyan Ran, “Simultaneous active camera array focus plane estimation and occluded moving object imaging”. Image and Vision Computing, vol. 32, no. 8: pp. 510-521, 2014.

[8] N. Joshi, S. Avidan, W. Matusik and D. J. Kriegman, “Synthetic aperture tracking: tracking through occlusions”, in Proc. IEEE Conf. Computer Vision (ICCV), pp. 1-8, 2007.

[9] R. Eshel, Y. Moses, “Homography based multiple camera detection and tracking of people in a dense crow”, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2008.

[10] Tao Yang, Yanning Zhang, Xiaomin Tong, Xiaoqiang Zhang and Rui Yu, “Continuously tracking and see-through occlusion based on a new hybrid synthetic aperture imaging model”, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp.3409-3416, 2011.

[11] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, "Image inpainting", in Proc. Conf. Computer graphics and interactive techniques, pp. 417-424, 2000.

[12] Yonatan Wexler , Eli Shechtman, and Michal Irani. "Space-time completion of video." IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 3, pp. 463-476, 2007.

[13] A. Criminisi, P. Patrick, and T. Kentaro, "Region filling and object removal by exemplar-based image inpainting." IEEE Trans. Image Processing, vol.13. no. 9, pp. 1200-1212, 2014.

[14] A.J. Patti, M.I. Sezan, and A.M. Tekalp, “Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time,” IEEE Trans. Image Processing, vol. 6, no. 8, pp. 1064-1076, 1997.

[15] Peyman Milanfar, Super-Resolution Imaging, CRC Press, 2011.

[16] R. R. Schulz and R.L. Stevenson, “Extraction of high-resolution frames from video sequences,” IEEE Trans. Image Processing, vol. 5, no.6, pp. 996-1011, 1996.

[17] G. Carles, J. Downing, A. R. Harvey, “Super-resolution imaging using a camera array”. Optics letters, vol. 39, no. 7, pp. 1889-1892, 2014.

[18] D. G. Lowe, “Distinctive image features from scale-invariant keypoints” International journal of computer vision, vol. 60, no.2, pp. 91-110, 2004.

[19] M. A. Fischler, R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”. Comm. of the ACM, vol. 24, no. 6, pp. 381-395, 1981.