Uncalibrated Synthetic Aperture for Defocus Control

Natsumi Kusumoto Shinsaku Hiura Kosuke Sato
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531 Japan
http://www-sens.sys.es.osaka-u.ac.jp/e-index.html

Abstract

Exaggerated defocus can not be created with an ordinary compact digital camera because of its tiny sensor size, so it is hard to take pictures that attract a viewer to the main subject. On the other hand, there are many methods for controlling focus and defocus of previously taken pictures. However, most of these methods require purpose-built equipment such as a camera array to take pictures. Therefore, in this paper, we propose a method to create images focused at any depth with arbitrarily blurred background from the set of images taken by a handheld compact digital camera moved randomly. Using our method, it is possible to produce various aesthetic blurs by changing the size, shape or density of the blur kernel. In addition, we confirm the potential of our method through a subjective evaluation of blurred images created by our system.

1. Introduction

Appropriate blurring of background or non-principal objects is important for most photographs to attract viewers' attention to the main subject. Also the defocus has an influence on the impressions of pictures such as tender feelings, so controlling the amount of defocus is an important technique for most professional photographers. On the other hand, imaging devices have been shrinking as digital cameras have grown popular, and it is hard to design a lens as large as that of a film-based camera. In general, the size of the aperture is proportional to the size of the imaging device when we keep both field of view and f-number the same. Therefore, the size of the imaging device is essential to the flexibility of defocus control, but large imaging devices comparable to the film are very expensive while the quality of the image from a point-and-shoot digital camera is sufficient for most purposes.

The process of adjusting focus and defocus is another issue of photography. Since the relationships among the aperture, distance, field of view and defocus are nonlinear and difficult to model, photographers often adjust these parameters to control the amount of defocus by trial and error. More intensive techniques for defocus control such as the combination of shift, tilt and swing of the lens can only be used by well-trained photographers using an uncommon camera with traditional bellows. Therefore, many researchers have tackled the "refocus" technique to control focus and defocus after taking photographs.

The lens is a device to arrange all incoming rays onto a 2-D image plane. Since each incoming ray has four degrees of freedom(direction and translation of a line), which part of the rays are collected to each pixel on a two dimensional plane is a problem. Hence, the most straightforward way to refocus is to sample all of the incoming rays separately. Isaksen[4] has addressed various merits of refocusing techniques as an application of multi-camera array, and nowadays several camera arrays are available commercially such as ProFUSION25[12].

The light field can be sampled using a single lens with a modified image sensor. Ng[9] altered the camera by inserting a micro lenslet array just in front of the image plane. Then each pixel has a correspondence not only with the viewing direction but also the viewpoint on the aperture of the lens. In other words, the light field captured by the camera can be reassembled to form various images with different focusing depths, amounts of defocus or slight changes of the viewpoint. However, the size of equipment which determines the area of sensing surface limits the amount of defocus in principle, if we use blur generating methods based on the rearrangement of light field. In this sense, the light field captured by a single-lens camera does not magnify the radius of defocus which is possible with the original camera. Fortunately, in most cases, types of photographs that require significant amount of defocus are a still life photos, so we could replace the camera array by multiple shots with a camera on numerically controlled translation stages[6] except for the portraits. This is better than camera arrays in respect of the cost and flexibility, but still very specialized and not widely available for general camera users.

The aesthetic aspect of defocus, which is also commonly referred to as "bokeh", is also an important consideration

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

to leverage its potential. When we use high-quality lenses without evident aberration, the image of a defocused point light source is similar to the shape of the aperture which is usually just a circle with uniform transparency inside. However, such a shape of point spread function is actually not ideal, because it has a hard edge around the circle and high spatial frequency components are still preserved after blurring. Therefore, some solutions for this problem have been offered by lens manufacturers. For instance, the DC-Nikkor lens made by Nikon has a function to change the property of spherical aberration slightly to soften the defocus for either background or foreground. Minolta and Sony have also provided STF lens for an extraordinarily smooth defocus[10] by incorporating a concentric graded filter to absorb the light passing through the peripheral of the aperture. However, such lenses are not cheap, and only telephoto models are available on the market.

In consequence, we already have lots of previous studies to control focus and defocus, but special equipment is always necessary and closely associated to each method. Furthermore, the aesthetic aspect of the defocus is not well considered in most methods for refocusing. Therefore, in this paper, we present a method to control defocus which does not require special equipment; we simply take several photos of the scene using a handheld digital camera. We can control the distance and tilt angle of the in-focus plane, and even the aesthetic aspect and amount of the defocus can be arbitrarily changed afterwards.

The **contributions** of this paper are as follows.

- We propose a method to create photographs with strong defocus using an ordinary digital camera. Special equipments, such as camera arrays, computer-controlled translation stages or light field cameras are not necessary.
- The depth and angle of the focused plane can be arbitrary controlled after taking photographs. Also the depth of field can be tuned afterwards as if the aperture of the lens is changed.
- Aesthetic tuning of the defocus is also possible by changing the shape of point spread function. We confirm the effect of various PSFs through subjective evaluations.

2. Uncalibrated Synthetic Aperture

Figure 1 illustrates the process of defocus control using our method. At first, several photos of the static scene are taken by handheld digital camera with unknown slight motion. The images do not need to have a shallow depth of field, therefore we can use any kind of digital camera with a small image sensor. Additionally, the locations of viewpoints need to be neither aligned nor known.

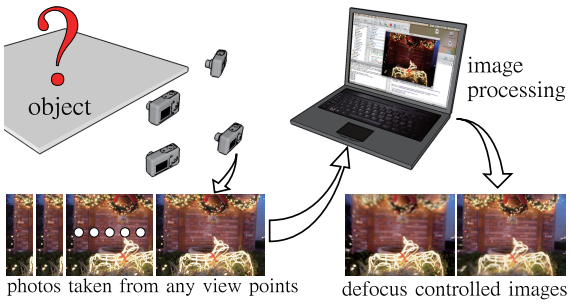


Figure 1. Overview of our proposed system.

Synthetic aperture is a refocusing technique that averages many images captured at various viewpoints to from an image as if taken by a lens with a larger aperture. If one wants to change the virtual focused distance, each image is only necessary to be shifted to register the corresponding point at the same pixel. However, the images taken by an arbitrarily moving camera do not satisfy the presumed condition that the optical axes of all cameras are parallel to each other. The followings are the undesirable effects or difficulties caused by the randomness.

- The appearance of the scene is affected by not only translation but rotation of the camera, therefore the parallel translation of the image is no longer sufficient to register all the points on the in-focus plane.
- The distribution of the viewpoints is cluttered and sparse causing unattractive bokeh.
- The locations and orientations of the camera when each shot taken are totally unknown.

To solve these problems, we have incorporated additional processes of deformation and interpolation of the image. The computational processes and necessary user operations of our method are illustrated in Figure 2.

2.1. Defocusing by Synthetic Aperture

As described above, a lens is a device to form a 2-D image by arranging all incoming rays through the aperture. Therefore, it is possible to simulate the function of the lens by accumulating the intensities of incoming rays captured separately. Several parameters such as focal length or focused distance determine the geometric relationships of which rays have to be converged to a single pixel.

Figure 3 illustrates the relationships among the aperture, distance and circle of confusion. Figure 3(A) shows the condition when the point p_1 on the object is just in focus on the image plane. If we have another point p_2 for which distance from the lens is different from p_1 has an image to satisfy the thin lens law $\frac{1}{a} + \frac{1}{b} = \frac{1}{\tilde{a}} + \frac{1}{\tilde{b}}$ as shown in Figure 3(B). If we place a small sub-aperture which is on

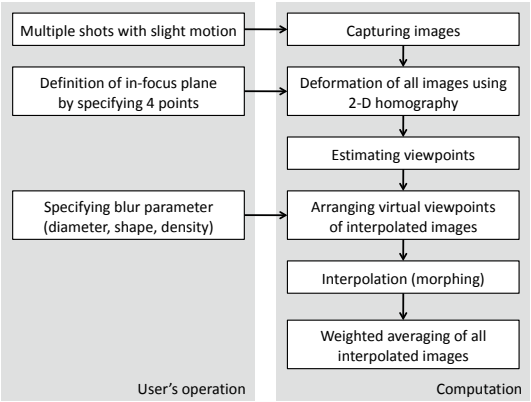


Figure 2. User interactions and corresponding computations for refocusing by uncalibrated synthetic aperture.

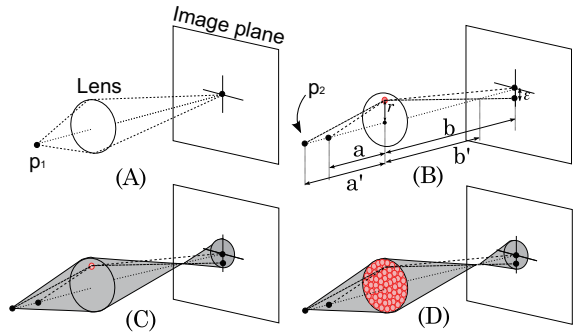
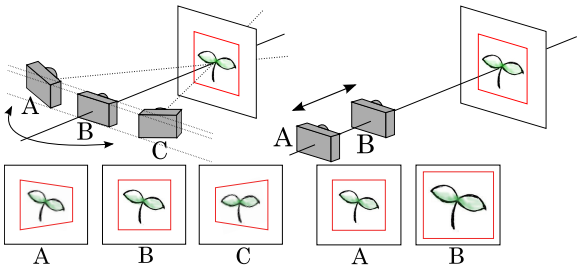


Figure 3. Principle of synthetic aperture photography illustrated with the relationships between sub-aperture and disparity.

the plane of the original aperture with a displacement r from the optical axis, we have a corresponding disparity ε on the image. Note that the disparity on the image ε is linearly proportional to the location of the sub-aperture r. On the other hand, an actual lens creates certain radius of circle of confusion as shown in 3(C) for point p_2. Therefore, we can simulate the effect of defocus by summing up all the images taken at the sub-apertures which fill the entire part of the original aperture. In the latter part of this paper, we call this a virtual aperture.

2.1.1 Correction of image deformation

If we use well calibrated camera arrays or gantries, we can expect that all optical axes of the cameras are parallel. In this case, disparities between any two images are proportional to the length of baseline which is the distance between two viewpoints. Moreover, the disparities of all points are uniform when the scene is just a plane perpendicular to the optical axis. Such characteristics are similar to the effect of the sub-aperture described in section 2.1, and the focused distance can be easily changed by translating



(a)rotation of the optical axis (b)translation to the optical axis

Figure 4. Deformation caused by the motions of handheld camera.

images to register the corresponding points on the focused plane at a same pixel.

Contrary to the well aligned cases described above, images taken by handheld camera no longer satisfy the condition of parallelism among epipolar lines. As shown in Figure 4, fluctuation of the camera posture changes the observed shape on a plane, and also the translation of the camera towards the optical axis causes undesirable image magnification effect. For both cases, the disparities between two images are no longer uniform, but still keep a constraint called a 2-D homography.

If all viewpoints are coplanar, we can calculate each homography matrix using the image rectification technique[7] from enough number of corresponding points on the images without any user interactions. However, we can not assume that all viewpoints are on the same plane, therefore we use a naive technique to calculate each homography using four reference points on a plane in the scene. In other words, the operator is required to specify four points on a reference image which is arbitrarily selected, and the system automatically finds corresponding points from the rest of the images and then calculates the homography. This operation also works as a definition of the in-focus plane in the scene, because any images can be registered to the reference image without any disparity using the calculated homography if the point is just on the plane the user specified. If the point in the image is out of the focused plane, this process preserves a certain disparity which will be a source of defocus.

The plane to be focused is not necessary to be perpendicular to the optical axis, but it could be rotated using four points on the tilted plane. The result is just the same as the image taken by the view camera using Scheimpflug rule.

2.2. Interpolation for smooth defocus

Since we have a limited number of pictures, the averaged image after registration has stair-like artifacts as shown in Figure 5. Therefore, we interpolate among images taken from sparse viewpoints. Actually this process is very similar to the view-interpolation technique [1, 2], and also the merit of interpolation for refocus has been already ad-

Figure 5. Synthetic defocus without interpolation causes stair-like artifacts.

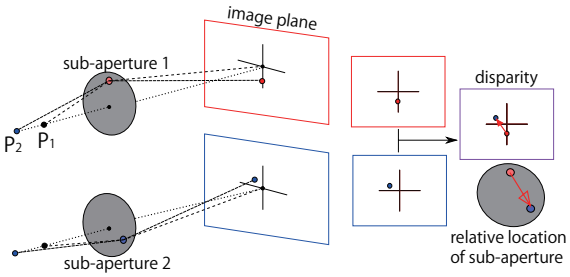


Figure 6. Relationship between sub-aperture and disparity.

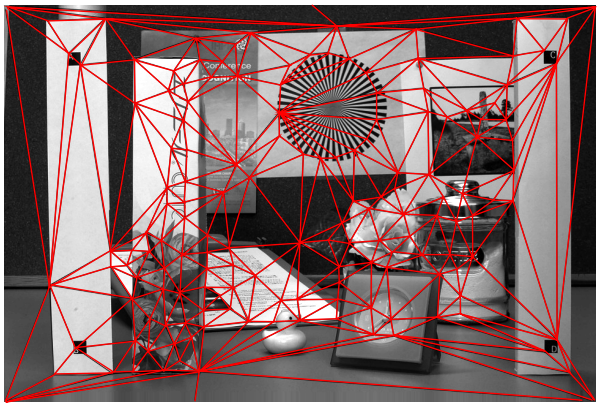


Figure 7. One example of Delaunay triangulation of the scene.

dressed by Georgeiv[3]. However, in our case, locations of the viewpoints are unknown but very close to each other, so it is very hard to apply sophisticated self-calibration techniques such as projective reconstruction or Plane+Parallax method[11] . We therefore use simple technique described below to estimate the distribution of the viewpoints.

Once the location of each viewpoint is estimated, we can place dense virtual viewpoints on regular grid for interpolation, and we can make the blur smooth.

2.2.1 Estimation of viewpoint locations

It is impossible to measure the actual location of each viewpoint with metrics, because there is no reference object which gives us the scale of the scene. However, absolute location of viewpoints are not always necessary, because what we have to do is to investigate the distribution of the viewpoints and decide where new viewpoints for interpolation must be placed in a virtual aperture. We have implemented the projective reconstruction technique, but it is not enough stable because the camera motion is so small. Therefore, we use the average of disparities of feature points to roughly estimate the location of sub-apertures on 2-D plane.

To estimate the location of viewpoint, we use registered images following the process described in section 2.1.1. Then feature points are detected from the reference image, and correspondences are found in the remaining images. Generally, the vector of the disparity is proportional to the relative location of sub-apertures as shown in Figure 6, so we calculate the average vector of all disparities to estimate the relative location of viewpoints. The proportion of the displacement of viewpoint to the disparity may be changed according to the depth to the feature point from the camera, so the distribution of the viewpoints may be reversed in some cases. Fortunately, it does not matter, because we are not interested in the shape of the scene, and reversed distribution of the viewpoints is also sufficient for our purpose as described below.

2.2.2 View interpolation

To generate images from a new viewpoint, we use the popular view interpolation method called morphing. Since all images are already registered on the in-focus plane using the method described in section 2.1.1, this stage must use a depth dependent technique, or the resultant image is not smoothed. Fortunately, actual depth to the scene from the camera is not necessary because the disparity observed on the images is sufficient to generate the interpolated images.

As described in section 2.2.1, we already have a number of feature points with correspondences through all the images, so Delaunay triangulation is applied to the 2-D distribution of the feature points as shown in Figure 7. Also the distribution of the viewpoints are processed with Delaunay triangulation to decide which images are used to interpolate the image at each viewpoint. As shown in Figure 8, three images are selected for each virtual viewpoint which is in a triangle with the vertices of the three viewpoints to

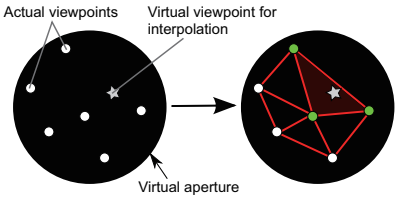


Figure 8. Selection of three images for virtual viewpoints.

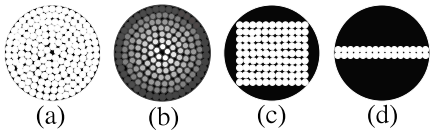


Figure 9. Tuning of PSF by weighted sub-aperture.

minimize the undesirable effect caused by occlusion. Afterwards, each triangle patch on the image is deformed using the internal division of disparities of its vertices and averaged to generate the interpolated image.

2.3. Aesthetics of the defocus with various PSFs

In general, the defocused image of a point light source (called the Point Spread Function (PSF)) is observed as a circle with uniform intensity, because the attenuation ratio for any light rays passing through the aperture is constant. Since this PSF has a hard edge around it, intensity discontinuities still remain in the defocused images. In theory, the circular blur kernel with uniform intensity has a frequency response expressed as a Bessel function which has similar characteristics to a sinc function with periodic zero responses and phase reversals. From the aesthetic point of view, such discontinuities are not pleasing, and the blur should be extremely smooth without noticeable edges. Therefore, several attempts have been made as Minolta's STF lens[10] with concentric graded filter element inside, but it requires specially made optics. So we have incorporated a function to control the PSF of the virtual aperture by setting a certain weight for each interpolated image. As shown in Figure 9, we can soften the edges of defocus by using the weight (b) instead of normal lens effect (a). We can also control the shape of the PSF by changing the arrangement of sub-aperture as (c) or (d) for further artistically creative photographs[5].

3. Experiments

In this section we will show several results including viewpoint estimation, computationally generated defocus and ability to control of point spread function. We used OpenCV library to find feature point, find correspondences among them, and evaluate Delaunay trianglations.

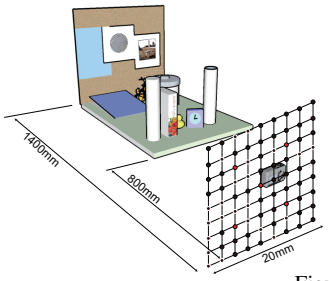


Figure 10. Arrangement of actual viewpoints.

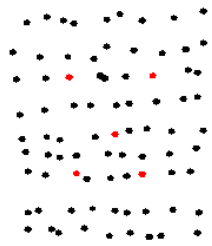


Figure 11. Estimated relative location of viewpoints.

Figure 12. Five input images for experiment.

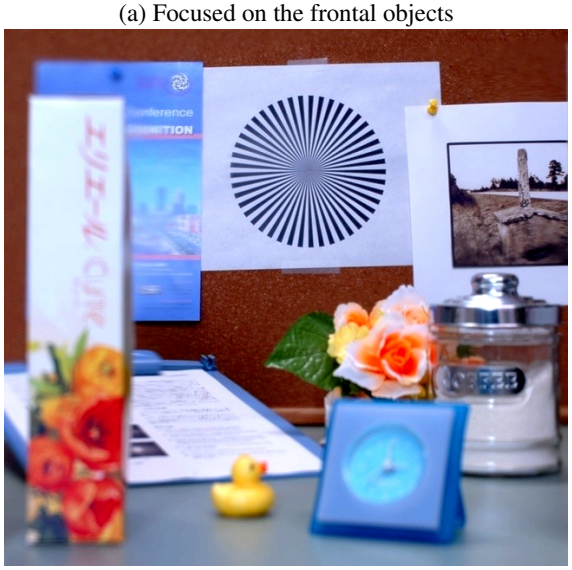
3.1. Evaluation of estimated viewpoint location

We first show a result of viewpoint location estimation. We used a Nikon D200 digital SLR camera with Nikkor 35mm F2D lens mounted on Manfrotto 303SPH tilt-slide platform. We took photos from 81 viewpoints as shown in Figure 10 with slight 2.5mm steps, therefore overall travel of the camera is within a square 20mm on a side. Figure 11 shows the relative distribution of viewpoint locations estimated by the method described at section 2.2.1. Note that the distance from the camera to the scene is much larger than the displacement of the camera, and also the rigidity and preciseness of the sliding platform is very poor.

3.2. Computationally generated defocus

Using the same data set described above, we generated defocused images using our proposed system. As shown in Figure 12, we used only 5 images which are taken at the center and four corners of a square 10mm on a side, shown as red dots in Figure 10 and 11. Figure 13 shows two results with different focused depths where 100 images are interpolated in the square, and then averaged using the circular PSF as Figure 9(a). For the first Figure 13(a) case, we used fiducials nearby the four corners to focus on the frontal object. It is very intuitive to define the focused plane in the scene. Of course we can also specify the focused plane by user interaction which is used for Figure 13(b) case.

(a) Focused on the frontal objects

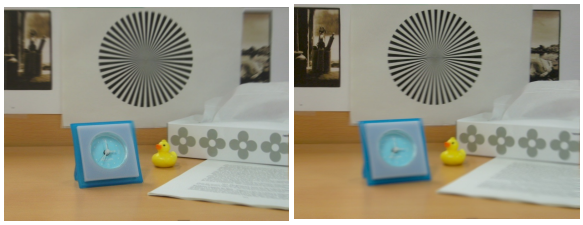


(b) Focused on the background

Figure 13. Images with generated defocus from sparse viewpoints.

3.3. Defocusing using handheld camera

For the case of experiment shown in section 3.2, all optical axes of the translated camera are parallel each other, though we never provided the actual viewpoint locations to the system. Therefore we have to show another result using a handheld point-and-shoot digital camera. Here we took only 3 photos using Panasonic DMC-FX8 camera without any kind of tripod. Figure 14 shows the results for we focusing on the frontal object and the background, confirm that smooth and strong defocus can be generated using such very common equipment and situations.



(a) Focused on the clock (b) Focused on the background

Figure 14. Generated images using three shots by handheld camera.

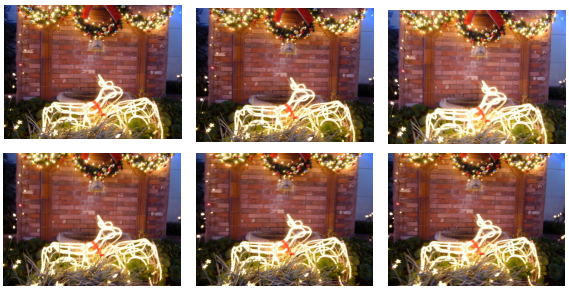


Figure 15. Six input images used for experiment.

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

Figure 16. PSFs used for the experiment.

3.4. Controlling the Aesthetic aspect of defocus

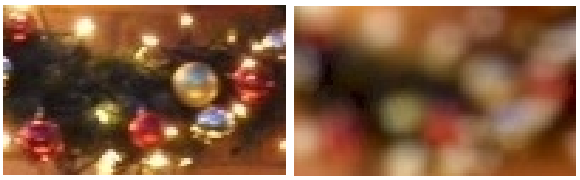
In this section we will show the results of controlling PSF to change the aesthetics of defocus called "bokeh." Again we used Panasonic DMC-FX8 camera without a tripod to take the pictures of a store display which has many bright light sources. Input pictures are shown in Figure 15 and the results with four types of PSFs as Figure 16 are shown in Figure 17. You could see the evident differences in blurred point light source. In general, the weighted circular PSF is the best for most cases, but the subjective strength of the blur is less than the one with uniform circular PSF because the effect of the peripheral sub-aperture is weaker.

4. Evaluation

To confirm the qualitative characteristics and beneficial aspect of our proposed system, we evaluated it from both objective and subjective points of view using frequency domain analysis and user studies.

4.1. Evaluation in frequency domain

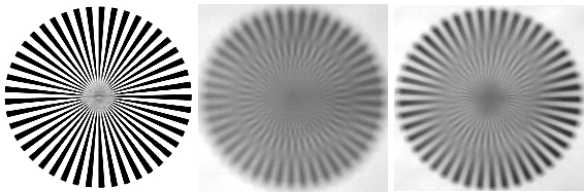
In this section we confirm the effect of various PSFs by using of the knowledge of frequency domain analysis. If



(a)input image (b)Circular PSF

(c)Rectangular PSF (d)Weighted circular PSF

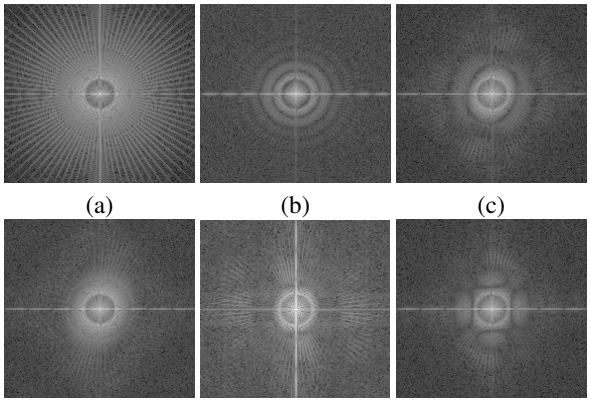
Figure 17. Enlarged pictures of the defocused images.



(a) Original (b) F2 (c)Circlar

(d)WC-4 (e)WC-6 (f)Square

Figure 18. Generated images used for frequency domain analysis.



(a) (b) (c)

(d) (e) (f)

Figure 19. Evaluation in frequency domain.

we use normal defocus with uniform circular PSF, zero response or sign reversal will be observed for certain frequencies. For our experiment, we used six types of PSF as listed below.

(a) Target chart
(b) Actual F2 lens
(c) Circular PSF with relative radius 4
(d) Weighted circular PSF with relative radius 4
(e) Weighted circular PSF with relative radius 6
(f) Square PSF with relative sides 4

The input data set is the same as Section 3.2(Figure 12), and we used the cropped region of the star target which has broad-band spatial frequency components for every direction. The results are shown in Figure 18 and also Figure 19 in the frequency domain. Note that the black circle at the center of frequency domain image is caused by the limited size (radius) of the star target. Also the radial lines on the frequency domain is caused by the limited number of wedges in star target. In Figure 18(b)(c), zero responses and sign reversals are evident, and we see the periodic zero responses in Figure 19(b)(c). In contrast, results with a weighted circular PSF in Figure 18(d) have no evident artifacts and also the frequency response in Figure 19(d) is well behaved. The larger radius of defocus has strong low-pass effects as shown in Figure 19(e), but we see some artifacts in Figure 18(e) which are caused by the limited number of interpolated images. In addition, we have confirmed the effect of rectangular PSF. Figure 18(f) shows the horizontal and vertical sign reversal and the zero responses in Figure 19(f) as well.

4.2. Subjective evaluation

If we concerned with the aesthetic aspect of defocus, we should do subjective evaluation for generated images. We show photo-quality prints of several generated images to 23 subjects with age from 20 to 56 years. Again the data set is the same as section 3.2(Figure 12), and blurred with parameters (A) - (F) just like as listed at section 4.1. The image without interpolation (G)(Figure 5) was also shown to the subjects. A region with visible artifact just around the white box is undisclosed. The subjects answered to the questionnaire with 5-point scale of images' naturalness, beauty, depth feel and preference.

Figure 20 shows the average of scores given by all subjects. Overall, images defocused by actual lens have high scores, though pan-focused image (A) lacks the depth feel. On the other hand, image (D) generated by weighted circular PSF with r=4, which means the relative radius of aperture is set to 4, had a high score as well as the image defocused by the actual lens. Results with larger blur (E) had lower scores because some artifacts are visible in the result. It is also evident that the apodization of PSF provides better results by comparing (C) to (D). Of course, interpolation is very important, or the resultant image seems to be strange.

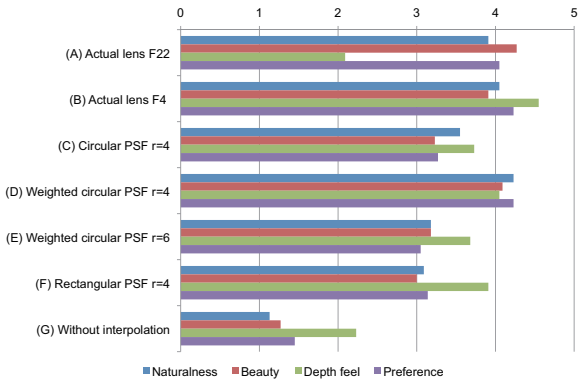


Figure 20. Results of subjective evaluation for generated images.

5. Discussion

The images generated with our proposed method still have some artifacts. Most of the causes of such undesired effects are due to the process of interpolation. As shown in Figure 21, correspondences of the feature points are not always correct as seen in example (c), and resulting in false duplicated images (b). Also Delaunay triangulation is problematic(Figure 22), because the edges of an in-focus object are often blurred by the triangles which step over the occluding contour.

To solve these problems, we have to estimate correct and dense depth maps from the images. Current system uses a KLT tracker which handles feature points only, so the depth information for edges or textures are hard to obtain. We are now trying to improve the depth recovery method using the epipolar constraint which is using a self-calibration technique such as projective reconstruction [8].

6. Conclusions

We have presented a method to control the defocus arbitrarily using several photos taken by a handheld digital camera. The depth and orientation of the focused plane can be changed after taking the photographs, and the aesthetics and the amount of the defocus can be tuned as well. We confirm the performance of our system from both objective and subjective aspects, and indicate the potential and ability of uncalibrated synthetic aperture framework, on some artifacts and limitations are left in current system.

References

[1] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen. Unstructured lumigraph rendering. In *ACM SIGGRAPH*, pages 425–432, 2001.

[2] S. E. Chen and L. Williams. View interpolation for image synthesis. In *ACM SIGGRAPH*, pages 279–288, 1993.

(a) (b) (c)

Figure 21. Artifacts on the interpolated images cased by the wrong corresponcences of feature points. (a) interpolated image, (b)ground truth taken at the viewpoint to be interpolated, (c) disparities calculated by KLT tracker.

(a) (b) (c)

Figure 22. Blurry edges caused by the triangles stepping over the occluding contours. (a) input image, (b) generated image with blurry edges, (c) Delaunay triangulation stepping over the occluding contour.

[3] T. Georgeiv, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala. Spatio-angular resolution tradeoffs in integral photography. In *In Eurographics Symposium on Rendering*, pages 263–272, 2006.

[4] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *ACM SIGGRAPH*, pages 297–306, 2000.

[5] D. Lanman, R. Raskar, and G. Taubin. Modeling and synthesis of aperture effects in cameras. In *Int. Symp. on Computational Aesthetics in Graphics, Visualization, and Imaging*, 2008.

[6] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. *ACM SIGGRAPH*, 2000.

[7] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. *Proc. IEEE CVPR*, 1:1125, 1999.

[8] S. Mahamud and M. Hebert. Iterative projective reconstruction from multiple views. In *Proc. IEEE CVPR*, pages 430–437, 2000.

[9] R.Ng, M.Levoy, M.Bredif, G.Duval, M.Horowitz, and dssP.Hanrahan. Light field photography with a hand-held plenoptic camera. *Stanford Tech Report CTSR*, pages 2005–02, 2005.

[10] Sony. Stf lens. http://www.ecat.sony.co.jp/dslr/lens/lens.cfm?PD=24685.

[11] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. *CVPR 2004*, 1:I–2–I–9 Vol.1, 2004.

[12] ViewPLUS. Profusion25. http://www.viewplus.co.jp/products/profusion25/index-e.html.