

ANALYSIS OF THE EFFECT OF CALIBRATION ERROR ON LIGHT FIELD SUPER-RESOLUTION RENDERING

Kuang-Tsu Shih, Chen-Yu Hsu, Cheng-Chieh Yang, and Homer H. Chen

National Taiwan University

ABSTRACT

Light field photography, which has recently drawn considerable attention, provides novel functionalities such as refocusing and depth estimation at the same time. However, the resolution of the rendered refocus image is incomparably lower than the number of light ray samples in the light field. In this paper, we show that super-resolution can be performed to bridge the gap and that deconvolution, which is missing in most previous methods, is essential to the success of light field super-resolution rendering. In addition, we investigate the effect of four different kinds of camera calibration error on the quality of rendered images. Given an expected level of image quality, the upperbound of the camera calibration error is analyzed.

Index Terms—Light field photography, camera calibration, super-resolution, error analysis.

1. INTRODUCTION

Light field photography has drawn worldwide attention in recent years and has become an important field of research. The most attractive functionality enabled by light field photography is perhaps refocusing, which allows the user to change the focal plane after the picture is taken. In addition, light field photography can provide other useful functionalities such as scene geometry inference and extended-depth-of-field imaging. Despite that light field cameras are commercially available today, there is a serious image resolution reduction on the order of 10:1 with respect to the raw light field data.

Super-resolution can be applied in light field rendering process to address the image resolution reduction problem by selecting a proper set of projection directions in the ray space. However, the super-resolved image is often not as sharp as the image directly acquired by a high-resolution camera, see Fig. 1(a). Therefore, there is a need to improve super-resolution light field rendering.

Our paper has two main contributions. First, we introduce a deconvolution step at the end of the super-resolution process and show the essentialness of the step for light field super-resolution rendering. Without it, blurring of the rendered image is inevitable. Second, we investigate the effect of calibration error on the quality of the rendered

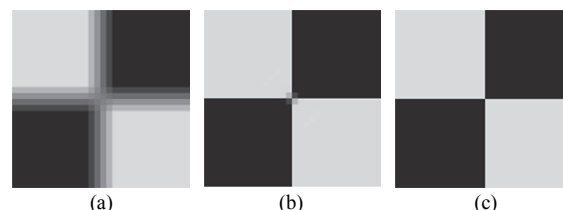


Fig. 1. Cropped refocus image rendered by (a) direct projection and (b) projection and deconvolution. (c) The ground truth image.

image and analyze the upper bound of the error for a targeted image quality.

The rest of the paper is organized as follows. We introduce the related work in Section 2 and describe the importance of the deconvolution operation in Section 3. The effect of camera calibration error on light field rendering is discussed in Section 4, and the conclusion is drawn in Section 5.

2. RELATED WORK

Light field photography dates back several decades [6], [7]. In this section, we briefly review some relevant work from the practical point of view.

The first compact hand-held light field camera is developed by Ng et al. [9], which uses Fourier slice theorem [10], [11] to describe the image formation process in frequency domain. However, because the spatial resolution is traded off for the angular resolution, the resolution of the rendered image is one order less than the number of captured light field samples.

The prototype developed by Liang et al. [1] incorporates a programmable aperture to capture the light field in multiple exposures. Unlike Ng's camera, the resolution of the rendered image is the same as the number of photo sensors in the camera. However, like Ng's camera, the resolution of the rendered image is still one order less than the total number of recorded light ray samples.

Several methods were proposed to address the resolution reduction problem. Pérez derived a set of proper projection directions in the ray space for light field super-resolution rendering [3]–[5]. However, the rendered image still appears blurry despite its large pixel count. On the other hand, Bishop et al. [12] modeled the image formation process for light field cameras using microlens arrays and rendered the super-resolution image by energy minimization.

However, it is difficult to extend the technique to other types of light field cameras such as camera array. The design by Georgiev et al. [13], [14], which is a variant of Ng's camera, changes the distance between the main lens and the microlens array (or equivalently the focal length of the main lens). Their method takes raw Bayer image as input to avoid any undesired effect of color interpolation. Since it exploits the low fill-factor of the red, green, and blue pixels, the resolution magnification ratio is limited to two.

3. LIGHT FIELD SUPER-RESOLUTION

In this section, we first describe the light field super-resolution rendering method in [4], [5], and then analyze the cause of the blur problem. We later show that the problem can be solved by adding a deconvolution step to the rendering process. We focus on the analysis of this method because, as proved in [5], it is more general than the other methods [12]–[14]. Since four-dimensional (4D) light fields are difficult to visualize, we consider two-dimensional (2D) light fields captured with a one-dimensional (1D) camera array in this section. The result can be easily generalized to 4D light field.

We explain the super-resolution method using the toy camera array shown in Fig. 2(a), which consists of 4 cameras. Suppose each camera has 7 pixels in the common field of view. Therefore, a total of 28 light rays are captured for the entire camera array. Assume the camera centers are collinear and the spacing between cameras is uniform. Define u -axis to be the line connecting camera centers and x -axis to be the line connecting the shared sample points of each camera. Consequently, the points representing the sampled rays form a rectangular grid pattern in ray space constructed by x and u , Fig. 2(b). It has been shown that refocusing is equivalent to the projection of light ray samples in ray space, and different refocus planes correspond to different projection directions [2], [10]. For example, the refocus planes 1 and 2 in Fig. 2(a) correspond to the projection directions indicated by slope m_1 and m_2 , respectively, in Fig. 2(b). Note that $m_1 = -4\Delta u/\Delta x$, where Δx is the sampling pitch on the x -axis and Δu is the distance between adjacent camera centers (Fig. 2(b)). This way, the rendered refocus image on plane 1 has four times the number of samples than the individual image captured by any single camera. On the other hand, for refocus plane 2 (the corresponding slope m_2 is $-\Delta u/\Delta x$) the number of samples does not increase. From this example, we can see that if the refocus plane is selected properly, it is possible to render refocus images with the number of samples equal to the number of light rays samples of the discrete light field. A discussion on the selection of proper refocus planes can be found in [3]–[5].

However, an image with more pixels does not necessarily imply a higher image resolution. Although choosing a proper refocus plane increases the number of

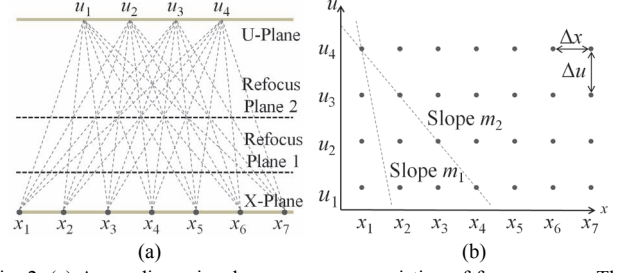


Fig. 2. (a) A one-dimensional camera array consisting of four cameras. The dashed lines represent the recorded light rays. (b) The distribution of the light rays samples captured by the camera array in (a). Slope $m_1 = -4\Delta u/\Delta x$ and $m_2 = -\Delta u/\Delta x$.

samples, the resulting image still appears blurry (Fig. 1(a)). The reason is as follows. Assume that each camera in the array has identical PSF. The recorded radiance value r_{ij} of the light ray parameterized as (x_i, u_j) is

$$\begin{aligned} r_{ij} &= \frac{1}{A \cdot \Delta x} \int_{u_j - A/2}^{u_j + A/2} \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} l(x, u) * K_o(x) \cdot dx du \\ &= \frac{1}{A \cdot \Delta x} \int_{u_j - A/2}^{u_j + A/2} l(x, u) * K(x) \Big|_{x=x_i} \cdot du, \end{aligned} \quad (1)$$

where l denotes the continuous light field, A denotes the aperture diameter, the symbol $*$ denotes the convolution operator (u is treated as a constant in the convolution), K_o denotes the optical PSF, and K denotes the overall PSF by convolving K_o with the sensitivity function of the sensor, which is assumed to be a rectangular function of width Δx . If the aperture size A is small enough, r_{ij} can be approximated by

$$r_{ij} = \frac{1}{\Delta x} l(x, u_j) * K(x) \Big|_{x=x_i}. \quad (2)$$

If the object of interest has Lambertian surface and lies on the refocus plane, the light field l has the property

$$l\left(x + \frac{u}{m}, u\right) = l(x, 0). \quad (3)$$

The variable m in (3) is the slope corresponding to the depth of the object. Combining (2) and (3), we have

$$r_{ij} = \frac{1}{\Delta x} l(x, 0) * K(x) \Big|_{x=x_i - \frac{u_j}{m}}. \quad (4)$$

For our toy example, if we project the samples in the ray space along the direction m_1 , the resulting rendered 1D refocus image is the sequence of pixels $r_{11}, r_{12}, r_{13}, r_{14}, r_{21}, r_{22}, r_{23}, r_{24}, \dots, r_{74}$. Substitute $m = m_1 = -4\Delta u/\Delta x$ into (4), we can see that the 1D image is the densely sampled version of the low-pass filtered signal $l(x, 0) * K(x)$ with sampling period $\Delta x/4$. The low pass kernel $K(x)$ is the cause of the blur in the image rendered by direct projection of the ray samples. Therefore, a deconvolution step is essential for light field super-resolution rendering to cancel the blur resulting from $K(x)$. To show the effectiveness of adding a deconvolution step, compare the blurred image by simple ray projection

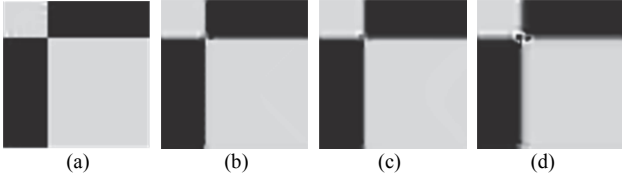


Fig. 3. Cropped super-resolved images with different σ_p : (a) $\sigma_p=0$ (ideal) (b) $\sigma_p=0.05$ (c) $\sigma_p=0.1$ (d) $\sigma_p=0.2$ (unit: pixel).

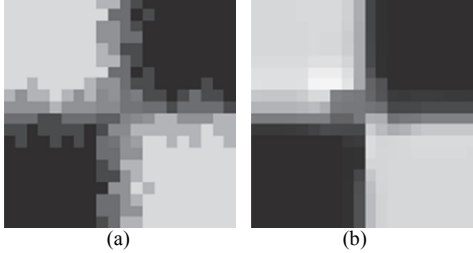


Fig. 4. The rendered image with camera position error (a) direct projection (b) super-resolved image after deconvolution.

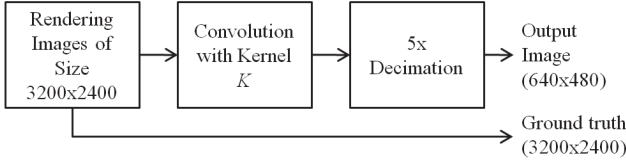


Fig. 5. The flowchart describing the procedure to generate the image and ground truth for each camera in the array.

shown in Fig. 1(a) and the super-resolved refocus image after deconvolution shown in Fig. 1(b). Compare with the ground-truth image in Fig. 1(c), the sharp edges in the super-resolved image are successfully reconstructed.

4. ANALYSIS

Although introducing a deconvolution step is capable of sharpening super-resolved refocus image, we observe that the quality of the rendered image degrades greatly when there is camera calibration error. As shown in Figs. 3 and 4, if the position estimates of the cameras are not exact, the edges in the resulting super-resolved image would definitely appear blurry (Figs. 3(b)–(d)) compared with the ideal case (Fig. 3(a)). In this section, we analyze the effect of camera calibration error on light field super-resolution rendering.

We generate discrete light field used in our analysis with the ray tracing tool POV-ray [15]. The light field is captured with a 5x5 virtual camera array. The test scene is a checker board placed at a fixed distance from the camera array. To control the PSF, we follow the procedure shown in Fig. 5 to generate the image for each camera. The resulting discrete 4D light field is of size 640x480x5x5. We use the ground truth image (generated by the procedure shown in Fig. 5) as reference and the peak signal-to-noise ratio (PSNR) as the objective quality measure. In addition, we use the method described in Section 3 to generate super-resolved images. Specifically, the light ray samples are projected along the direction corresponding to the refocus

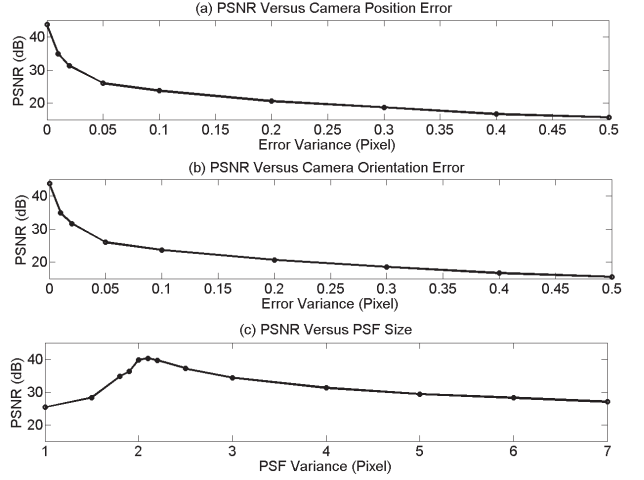


Fig. 6. The effect of calibration error on the quality of rendered image (a) Camera position error (b) Camera orientation error (c) PSF size error. For (a) and (b), the horizontal axis is σ_p^2 and σ_r^2 , respectively.

plane (aligned with the checker board in the test scene), and then the deconvolution algorithm developed by Shan [8] is applied to the resulting image. The camera calibration errors are classified into four categories: camera position error, camera orientation error, PSF size error, and PSF shape error. We analyze the effect of each error.

4.1. Camera position error

To analyze the effect of camera position error on light field super-resolution, a random vector (n_x, n_y, n_z) is added to the coordinates of the camera center (C_x, C_y, C_z) of each camera, where n_x , n_y , and n_z are independent and identically distributed (i.i.d.) Gaussian random variables with variance σ_p^2 . We plot the PSNR of the super-resolved image as a function of σ_p^2 in Fig. 6(a). For subjective evaluation, we also show cropped results in Fig. 3.

From Fig. 6(a) we can see that the camera position error greatly degrades the quality of the super-resolved image. To keep the PSNR of the super-resolved image above 24 dB, the camera position error σ_p^2 must be controlled within 0.1 pixel. As shown in Fig. 5(a), the error introduces zigzag artifact to the edges in the image constructed by direct projection of the light ray samples. Although the prior term introduced to the deconvolution algorithm alleviates the zigzag artifact, it results in blurred edge, Fig. 5(b).

4.2. Camera orientation error

In this experiment, we analyze the effect of camera orientation error. Let the estimated optical axis of each camera in the array be parallel to the vector $(0, 0, L_z)$. We assume the true optical axis is slightly tilted and is parallel to the vector (m_x, m_y, L_z) , where m_x and m_y are i.i.d. Gaussian random variables with variance σ_r^2 . In Fig. 6(b), the x -axis represents σ_r^2 and L_z is 320 (unit: pixel) in this experiment. Similar to the effect caused by position error, tilting of

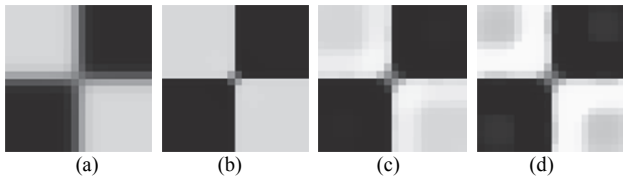


Fig. 7. Super-resolved images generated with Gaussian kernel of different variances: (a) Var = 1 (b) Var = 2 (c) Var = 3 (d) Var = 4 (unit: pixel).

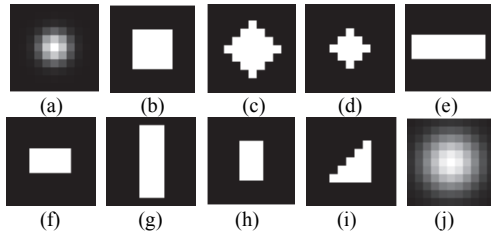


Fig. 8. PSFs used in the experiment described in Sec. 4.4: (a) true PSF, Gaussian (variance=2 pixels) (b) box (c) large diamond (d) diamond (e) long horizontal rectangle (f) horizontal rectangle (g) long vertical rectangle (h) vertical rectangle (i) triangle (j) Gaussian (variance=7 pixels).

optical axis also results in zigzag edges in the image generated by direct projection of light rays.

4.3. PSF size error

In Sec. 3, we assume that the PSF of each camera can be estimated beforehand, and hence the deconvolution is applied in a non-blind manner. However, the estimation may not be exact in real world. Here, we focus on the estimation error of the size of the PSF support. Specifically, we generate the input discrete light field using a Gaussian kernel with a variance of 2 pixels in the filtering step in Fig. 5, while Gaussian kernels of different variances are used to generate the super-resolved images in the deconvolution step. The objective quality is plotted as a function of the variance of the Gaussian kernels in Fig. 6(c), and the super-resolved results are shown in Fig. 7. We can see that the super-resolution quality degrades as the variance of the estimated PSF deviates from the true value (2 pixels). We also observe that, compared with camera position and orientation error, the PSF size error degrades the quality less. The support size (related to cut-off frequency) of the PSF affects the sharpness of the super-resolved image, but it does not produce significant zigzag artifacts like camera position error and orientation error do.

4.4. PSF shape error

Besides the PSF support size, we analyze the effect of inexact estimation of PSF shape. In this experiment, the deconvolution is applied using PSFs of different shape and the discrete light field is generated using a Gaussian kernel with variance of 2 pixels. We show deconvolution kernels that are used in this experiment in Fig. 8 and the PSNR of the super-resolved images generated with these kernels in

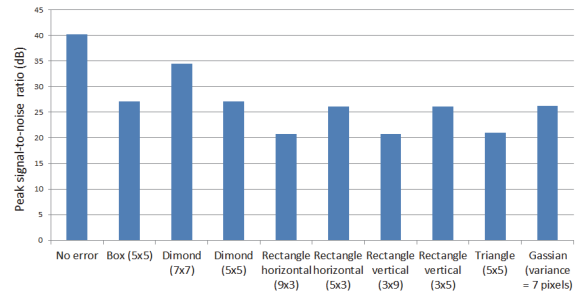


Fig. 9. PSNR of super-resolved images generated with PSFs with different shapes.

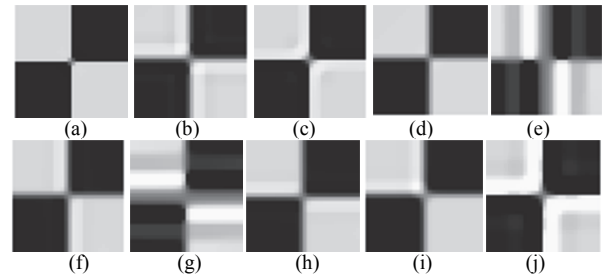


Fig. 10. Cropped super-resolved images with different shapes of gaussian PSF: (a) ideal camera PSF, Gaussian (variance=2 pixels) (b) box (c) large diamond (d) diamond (e) long horizontal rectangle (f) horizontal rectangle (g) long vertical rectangle (h) vertical rectangle (i) triangle (j) Gaussian (variance=7 pixels).

Fig. 9. For subjective evaluation, we also show a cropped portion of the super-resolved results in Fig. 10.

For those PSFs having a similar shape to the true PSF, such as the one in Fig. 10(c), the PSNR is higher. Fig. 10(d) can be regarded as an approximation of a Gaussian kernel with small variance, and the resulting PSNR is consistent to the result in Fig. 6(c). For PSF shapes that are less similar to the true PSF, such as the ones in Fig. 10(e), 10(g), and 10(i), significant artifacts appear on the edge of the super-resolved image. The quality of the super-resolved image is less sensitive to PSF shape compared with camera position and rotational error.

5. CONCLUSION

In this paper, we have shown that deconvolution is an essential step in light field super-resolution rendering and provided a quantitative analysis of the effect of various camera calibration errors on the quality of the super-resolved image. The results show that the calibration errors of camera position and orientation introduce more serious degradation to the rendering quality than the calibration error of the support size and the estimation error of the PSF shape of the camera. This work provides valuable insights into the fundamental limits of super-resolution light field rendering.

REFERENCES

- [1] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable Aperture Photography: Multiplexed Light Field Acquisition," *ACM Trans. on Graphics*, Vol. 27, No. 3, pp. 55:1–55:10, 2008.
- [2] C.-K. Liang, Y.-C. Shih, and H. H. Chen, "Light Field Analysis for Modeling Image Formation," *IEEE Trans. on Image Processing*, Vol. 20, No. 2, pp. 446–460, 2011.
- [3] F. Pérez, "Super-Resolution in Plenoptic Cameras by the Integration of Depth from Focus and Stereo," *Computer Communications and Networks, Proc. of 19th International Conference on*, pp. 1–6, 2010.
- [4] F. Pérez and J. P. Lúke, "Simultaneous Estimation of Super-Resolved Depth and All-in-Focus Images from A Plenoptic Camera," *3DTV Conference: The True Vision-Capture Transmission and Display of 3D Video*, pp. 1–4, 2009.
- [5] F. Pérez, A. Pérez, M. Rodríguez, and E. Magdaleno, "Fourier Slice Super-resolution in Plenoptic Cameras," *Computational Photography, IEEE International Conf. on*, pp. 1–11, 2012.
- [6] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational Plenoptic Imaging," *Computer Graphics Forum*, Vol. 30, No. 8, pp. 2397–2426, 2011.
- [7] M. Levoy and P. Hanrahan, "Light Field Rendering," *Proc. of the 23rd Annual Conf. on Computer Graphics and Interactive Techniques*, pp. 31–42, 1996.
- [8] Q. Shan, J. Jia, and A. Agarwala, "High-Quality Motion Deblurring from A Single Image," *ACM Trans. on Graphics*, Vol. 27, No. 3, pp. 73:1–73:10, 2008.
- [9] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with A Hand-Held Plenoptic Camera," *Computer Science Technical Report CSTR*, 2005.
- [10] R. Ng, "Fourier Slice Photography," *ACM Trans. on Graphics*, Vol. 24, No. 3, pp. 735–744, 2005.
- [11] R. Ng, "Digital light field photography," PhD dissertation, Stanford University, 2006.
- [12] T. E. Bishop, S. Zanetti, and P. Favaro, "Light Field Superresolution," *Computational Photography, IEEE International Conf. on*, pp. 1–9, 2009.
- [13] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the Focused Plenoptic Camera," *Proc. SPIE Computational Imaging*, Vol. 7873, 2011.
- [14] T. Georgiev, A. Lumsdaine, and G. Chunev, "Using Focused Plenoptic Cameras for Rich Image Capture," *Computer Graphics and Applications*, pp. 62–73, 2011.
- [15] <http://www.povray.org/>