

# Theoretical Analysis of Multi-view Camera Arrangement and Light-Field Super-Resolution

Ryo Nakashima<sup>1</sup>, Keita Takahashi<sup>2</sup>, and Takeshi Naemura<sup>1</sup>

<sup>1</sup> The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

<sup>2</sup> The University of Electro-Communications, 1-5-1 Chofugaoka,  
Chofu-shi, Tokyo 182-8585, Japan

**Abstract.** We analyzed a light-field super-resolution problem in which, with a given set of multi-view images with a low resolution, the 3-D scene is reconstructed with a higher resolution using super-resolution (SR) reconstruction. The arrangement of the multi-view cameras is important because it determines the quality of the reconstruction. To simplify the analysis, we considered a situation in which a plane is located at a certain depth and a texture on that plane is super-resolved. We formulated the SR reconstruction process in the frequency domain, where the camera arrangement can be independently expressed as a matrix in the image formation model. We then evaluated the condition number of the matrix to quantify the quality of the SR reconstruction. We clarified that when the cameras are arranged in a regular grid, there exist singular depths in which the SR reconstruction becomes ill-posed. We also determined that this singularity can be avoided if the arrangement is randomly perturbed.

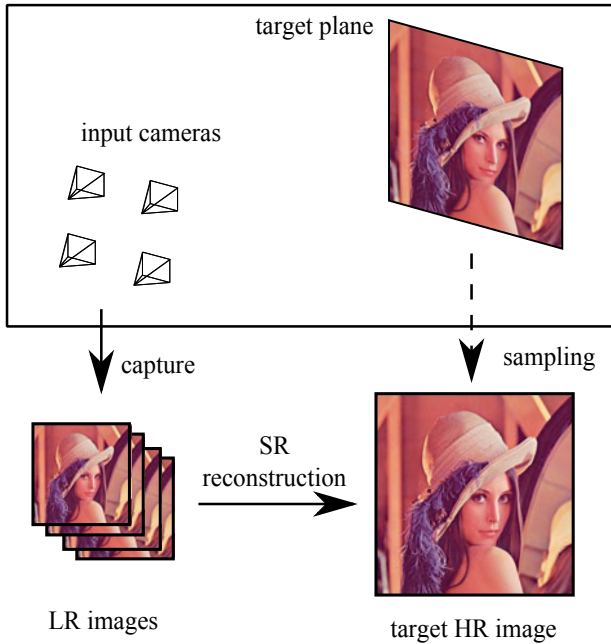
**Keywords:** multi-view cameras, super-resolution, camera arrangement, condition number.

## 1 Introduction

The reconstruction of a 3-D scene from multi-view images is a challenging problem and currently the focus of active research. To improve the quality of reconstruction, recent methods [3,4,9] use the framework of super-resolution (SR) reconstruction, which is a process of restoring an underlying high-resolution (HR) image from multiple low-resolution (LR) images. The quality of SR reconstruction is determined by the number, the point spread function (PSF), and the arrangement of cameras. The last factor, which has rarely been discussed, is the main focus of this paper.

The arrangement of the cameras determines disparities (pixel shifts) between the camera images given a certain depth. These disparities affect the stability of the SR reconstruction; for example, the SR reconstruction is ill-posed if all the disparities are integers. Therefore, the cameras should be arranged in such a way to avoid this ill-posed situation and improve the well-posedness of the SR reconstruction.

The purpose of this study is to analyze the relation between the arrangement of cameras and the well-posedness of an SR reconstruction. A general framework



**Fig. 1.** Framework of LFSR

in which a high-resolution 3-D scene is reconstructed from multi-view images with a lower resolution is referred to as light-field super-resolution (LFSR). To simplify the analysis, we consider a situation in which a target plane is located at a certain depth and a texture on that plane is super-resolved, as illustrated in Fig. 1. Although we only consider a single plane at a certain depth, our analysis is applicable to general scenes with multiple objects placed at various depths because the target plane can be placed at an arbitrary depth. Our theoretical model is constructed in the frequency domain, where the camera arrangements are independently expressed as a matrix in the image formation model of the SR reconstruction. The condition number of that matrix is used to measure the well-posedness of the SR reconstruction. We determined that when the cameras are placed on a regular grid, some depths are singular, meaning that the SR reconstruction at these depths is ill-posed. Singular depths can be avoided by randomly perturbing the cameras, which is a key finding in our study.

This paper is organized as follows. Section 2 introduces related works. We formulate the SR reconstruction process in Sect. 3, followed by some descriptions of the condition number in Sect. 4. In Sect. 5, we evaluate specific camera arrangements based on our theory. Section 6 concludes the paper.

## 2 Related Works

SR reconstruction generally consists of two steps [5]: the registration of LR images and the reconstruction of an HR image from the registered LR images.

In this paper, we assume that registration has been done in advance and hence we focus on the reconstruction step.

The numerical performance of SR reconstruction is mainly affected by three factors: the number of LR images, subpixel shifts between LR images, and the PSF of LR images. The first and second factors are closely related; if we have a greater number of images, we are more likely to have more varied subpixel shifts, resulting in more stable SR reconstruction. However, SR reconstruction is ill-posed if all the pixel shifts are integers, no matter how many images are available.

The pixel-shifts factor has been analyzed in several other works. Robinson et al. [6] evaluated the numerical performance of SR reconstruction using the Cramér-Rao lower bound and demonstrated that reconstruction quality is maximized when the sampling points of the LR images are evenly distributed. Champagnat et al. [2] used Monte Carlo simulations to analyze the quality of SR reconstruction when fractional parts of shifts are distributed uniformly in 0–1 pixel. They found that the reconstruction quality with random pixel shifts is moderate on average and comparable to that of optimal pixel shifts. In this study, we also analyzed the pixel shifts, but they were bounded by the camera arrangement and the depth of the scene in our problem. We used Monte Carlo simulations to analyze the arrangement of the cameras because analytical optimization is difficult for our problem.

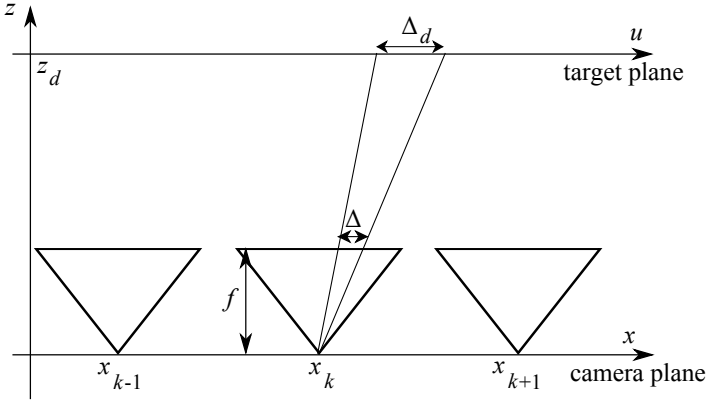
The PSF factor is also studied using the condition number, which is widely used in linear algebra to measure the well-posedness of linear equations. Baker et al. [1] analyzed box-shaped PSFs and discovered a relation between the condition number and the magnification ratio. Tanaka et al. [8] derived condition numbers for general space-invariant PSFs assuming that an infinite number of LR images are available. Inspired by these works, we also used the condition number as a measure of the well-posedness of the SR reconstruction, although we focused on the arrangement of cameras rather than the PSF.

### 3 Formulation of Super-Resolution Reconstruction

We formulated an SR reconstruction in the frequency domain. Our formulation is equivalent to [7], although some parameters were rearranged to fit to our problem.

#### 3.1 Configuration

See Fig. 2 for the configuration. Let  $(x, y, z)$  be the spatial coordinate. We assume that  $K$  cameras that capture LR images are placed on the camera plane at  $z = 0$ . The position of the  $k$ -th camera is denoted as  $(x_k, y_k, 0)$ . We also assume that all the cameras have the same focal length, pixel size, and PSF. A target plane is placed at  $z = z_d$  in parallel to the camera plane. The goal of the SR reconstruction is to obtain a texture on the target plane with a resolution higher than the input LR images. We assume that the magnification ratio is 2, but our analysis can easily be extended to more general cases.



**Fig. 2.** Configuration analyzed in this study

Let  $(u, v)$  be the image coordinate on the target plane. The texture on it is denoted by  $h(u, v)$  as a continuous 2-D signal. The HR image, which we want to synthesize by SR reconstruction, is denoted by  $g_H(u, v)$ . The LR image captured by the  $k$ -th camera is denoted as  $g_{L,k}(u, v)$ . Both  $g_H(u, v)$  and  $g_{L,k}(u, v)$  are the discrete signals sampled from  $h(u, v)$ . The pixel pitches are written as  $\Delta$  and  $\Delta/2$  for the LR and HR images, respectively.

### 3.2 Image Formation Model

The  $k$ -th LR image  $g_{L,k}(u, v)$  is generated by sampling the light-rays on the focal plane. This process is equivalent to sampling the continuous texture on the target plane  $z = z_d$  with intervals  $\Delta_d(z_d)$ , where  $\Delta_d(z_d)$  is defined as

$$\Delta_d(z_d) = \frac{z_d}{f} \Delta, \tag{1}$$

where  $f$  is the focal length of the cameras, as shown in Fig. 2. Note that  $\Delta_d(z_d)$  depends on the depth of the target plane  $z_d$ . To simplify the notations, we abbreviate  $\Delta_d(z_d)$  as  $\Delta_d$  in this section.

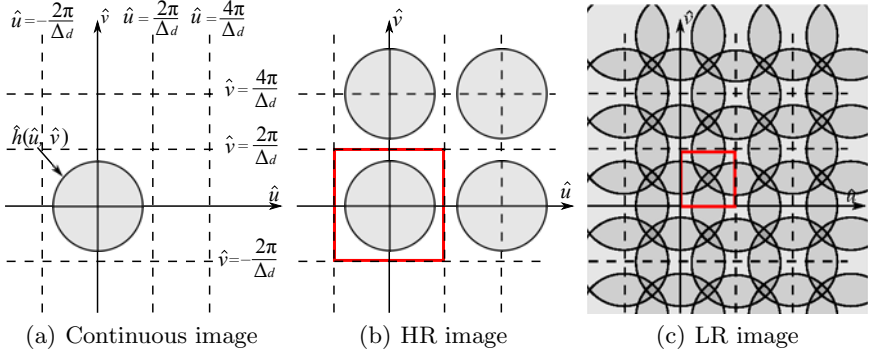
Using  $\Delta_d$ , the  $k$ -th LR image  $g_{L,k}(u, v)$  is defined as

$$g_{L,k}(u, v) = (h(u, v) * b_L(u, v)) \delta_{\Delta_d}(u - x_k, v - y_k) + n_k(u, v), \tag{2}$$

where  $*$  denotes convolution,  $b_L(u, v)$  is a camera PSF, and  $n_k(u, v)$  is the observation noise.  $\delta_{\Delta'}(u, v)$  represents the sampling grid that is defined as

$$\delta_{\Delta'}(u, v) = \sum_{(m,n) \in \mathbb{Z}} \delta(u - m\Delta', v - n\Delta'), \tag{3}$$

where  $\delta(u, v)$  is the Dirac delta function.



**Fig. 3.** Frequency spectra of continuous, HR, and LR images

Since the resolution of the HR image is double that of the LR images, the sampling interval of the HR image is  $\Delta_d/2$ . Therefore, the HR image  $g_H(u, v)$ , whose origin is set to  $(u, v) = (0, 0)$ , is defined as

$$g_H(u, v) = (h(u, v) * b_H(u, v)) \delta_{\Delta_d/2}(u, v), \quad (4)$$

where  $b_H(u, v)$  denotes a PSF of the HR image.

### 3.3 Super-Resolution in the Frequency Domain

Assume that the underlying continuous image  $h(u, v)$  is band-limited within  $(-2\pi/\Delta_d, 2\pi/\Delta_d)$ . In other words, the sampling interval of the HR image satisfies the Nyquist condition. This situation is illustrated in Fig. 3(a). The circular region in the figure represents the spectral support of the underlying continuous image  $\hat{h}(\hat{u}, \hat{v})$ , where  $\hat{\cdot}$  denotes the frequency-domain representation of the corresponding symbol.

Here, we want to obtain the Fourier transform of (4) and (2). First, we obtain the Fourier transform of (3) as

$$\hat{\delta}_{\Delta'}(\hat{u}, \hat{v}) = \frac{4\pi^2}{\Delta'} \sum_{\{m, n\} \in \mathbb{Z}} \delta\left(\hat{u} - \frac{2m\pi}{\Delta'}, \hat{v} - \frac{2n\pi}{\Delta'}\right). \quad (5)$$

This equation represents a spectral replication in the frequency domain caused by discretization. For the case of the HR image, where  $\Delta' = \Delta_d/2$ , the repeating cycle is  $(4\pi/\Delta_d, 4\pi/\Delta_d)$ , as shown in Fig. 3(b). Since the original signal  $\hat{h}(\hat{u}, \hat{v})$  is band-limited within  $(-2\pi/\Delta_d, 2\pi/\Delta_d)$ , no overlapping occurs in the frequency domain. Consequently, for the frequency  $\hat{u}, \hat{v} \in (-2\pi/\Delta_d, 2\pi/\Delta_d)$ ,  $\hat{g}_H(\hat{u}, \hat{v})$  can be written as

$$\hat{g}_H(\hat{u}, \hat{v}) = \frac{16\pi^2}{\Delta_d^2} \hat{h}(\hat{u}, \hat{v}) \hat{b}_H(\hat{u}, \hat{v}). \quad (6)$$

Meanwhile, for the case of the LR image, where  $\Delta' = \Delta_d$ , the repeating cycle is  $(2\pi/\Delta_d, 2\pi/\Delta_d)$ . As shown in Fig. 3(c), four spectral components overlap in the range  $\hat{u}, \hat{v} \in (0, 2\pi/\Delta_d)$ . Therefore, for this range,  $\hat{g}_{L,k}(\hat{u}, \hat{v})$  is described as

$$\hat{g}_k(\hat{u}, \hat{v}) = \frac{4\pi^2}{\Delta_d^2} \hat{h}(\hat{u}, \hat{v}) \hat{b}_L(\hat{u}, \hat{v}) * \sum_{m,n \in \{0,1\}} \delta\left(\hat{u} - \frac{2m\pi}{\Delta_d}, \hat{v} - \frac{2n\pi}{\Delta_d}\right) e^{-j(x_k \hat{u} + y_k \hat{v})} + \hat{n}(\hat{u}, \hat{v}). \quad (7)$$

Using (6), we obtain

$$\hat{g}_k(\hat{u}, \hat{v}) = \frac{1}{4} \hat{g}_H(\hat{u}, \hat{v}) \frac{\hat{b}_L(\hat{u}, \hat{v})}{\hat{b}_H(\hat{u}, \hat{v})} * \sum_{m,n \in \{0,1\}} \delta\left(\hat{u} - \frac{2m\pi}{\Delta_d}, \hat{v} - \frac{2n\pi}{\Delta_d}\right) e^{-j(x_k \hat{u} + y_k \hat{v})} + \hat{n}(\hat{u}, \hat{v}). \quad (8)$$

Equation (8) can be rearranged into a linear equation:

$$\hat{\mathbf{g}}_L = \hat{\mathbf{W}} \hat{\mathbf{g}}_H + \hat{\mathbf{n}}, \quad (9)$$

where  $\hat{\mathbf{g}}_L$  represents the spectra of all LR images and  $\hat{\mathbf{g}}_H$  represents the four overlapping components of the HR image:

$$\hat{\mathbf{g}}_L = \begin{pmatrix} \hat{g}_{L,1}(\hat{u}, \hat{v}) \\ \hat{g}_{L,2}(\hat{u}, \hat{v}) \\ \vdots \\ \hat{g}_{L,K}(\hat{u}, \hat{v}) \end{pmatrix}, \hat{\mathbf{g}}_H = \begin{pmatrix} \hat{g}_H(\hat{u}, \hat{v}) \\ \hat{g}_H(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v}) \\ \hat{g}_H(\hat{u}, \hat{v} - \frac{2\pi}{\Delta_d}) \\ \hat{g}_H(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v} - \frac{2\pi}{\Delta_d}) \end{pmatrix}. \quad (10)$$

$\hat{\mathbf{W}}$  represents the image formation model, expressed as

$$\hat{\mathbf{W}} = \hat{\mathbf{M}} \hat{\mathbf{B}}, \quad (11)$$

$$\text{where } \hat{\mathbf{M}} = \begin{pmatrix} 1 & e^{-2j\pi x_1/\Delta_d} & e^{-2j\pi y_1/\Delta_d} & e^{-2j\pi(x_1+y_1)/\Delta_d} \\ 1 & e^{-2j\pi x_2/\Delta_d} & e^{-2j\pi y_2/\Delta_d} & e^{-2j\pi(x_2+y_2)/\Delta_d} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-2j\pi x_K/\Delta_d} & e^{-2j\pi y_K/\Delta_d} & e^{-2j\pi(x_K+y_K)/\Delta_d} \end{pmatrix}, \quad (12)$$

$$\hat{\mathbf{B}} = \text{diag} \begin{pmatrix} \hat{b}_L(\hat{u}, \hat{v}) / \hat{b}_H(\hat{u}, \hat{v}) \\ \hat{b}_L(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v}) / \hat{b}_H(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v}) \\ \hat{b}_L(\hat{u}, \hat{v} - \frac{2\pi}{\Delta_d}) / \hat{b}_H(\hat{u}, \hat{v} - \frac{2\pi}{\Delta_d}) \\ \hat{b}_L(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v} - \frac{2\pi}{\Delta_d}) / \hat{b}_H(\hat{u} - \frac{2\pi}{\Delta_d}, \hat{v} - \frac{2\pi}{\Delta_d}) \end{pmatrix}. \quad (13)$$

$\hat{\mathbf{M}}$  is a  $K \times 4$  matrix and represents the camera arrangement. The  $k$ -th row of  $\hat{\mathbf{M}}$  corresponds to the position of the  $k$ -th camera. We refer to this matrix as a camera arrangement matrix.  $\hat{\mathbf{B}}$  is a  $4 \times 4$  matrix and represents the PSF. We call this matrix a PSF matrix.  $\hat{\mathbf{n}}$  represents the observation noises.

$$\hat{\mathbf{n}} = \begin{pmatrix} \hat{n}_1(\hat{u}, \hat{v}) \\ \hat{n}_2(\hat{u}, \hat{v}) \\ \vdots \\ \hat{n}_K(\hat{u}, \hat{v}) \end{pmatrix}. \quad (14)$$

The SR reconstruction is formulated as the problem of estimating  $\hat{\mathbf{g}}_H$  given  $\hat{\mathbf{W}}$  and  $\hat{\mathbf{g}}_L$  in (9), where  $\hat{\mathbf{W}}$  determines the well-posedness.

### 4 Condition Number of the SR Reconstruction

Given a coefficient matrix of a linear equation, the condition number of the matrix determines the stability of the solution. When the condition number is low, the linear equation is well-posed and is robust to noises. In contrast, if the number is high, the system is ill-posed and is sensitive to noises. The linear system is singular when the condition number is infinite.

The condition number of  $\hat{\mathbf{W}}$  is defined as

$$\text{cond}(\hat{\mathbf{W}}) = \|\hat{\mathbf{W}}\| \cdot \|\hat{\mathbf{W}}^+\| = \sqrt{\frac{\lambda_{\max}(\hat{\mathbf{W}}^* \hat{\mathbf{W}})}{\lambda_{\min}(\hat{\mathbf{W}}^* \hat{\mathbf{W}})}}, \tag{15}$$

where  $\|\cdot\|$  denotes operator norm,  $*$  denotes conjugate transpose,  $^+$  denotes Moore-Penrose pseudoinverse, and  $\lambda_{\max}(\hat{\mathbf{W}}^* \hat{\mathbf{W}})$  and  $\lambda_{\min}(\hat{\mathbf{W}}^* \hat{\mathbf{W}})$  are the maximum and minimum eigenvalues of  $\hat{\mathbf{W}}^* \hat{\mathbf{W}}$ , respectively.

The condition number gives the upper bound of the relative errors as

$$\frac{\|\hat{\mathbf{e}}\|_2}{\|\hat{\mathbf{g}}_H\|_2} \leq \text{cond}(\hat{\mathbf{W}}) \frac{\|\hat{\mathbf{n}}\|_2}{\|\hat{\mathbf{g}}_L\|_2}, \tag{16}$$

where  $\hat{\mathbf{e}}$  is the estimation error of  $\hat{\mathbf{g}}_H$ . This equation shows that the condition number can be used to estimate the reconstruction quality.

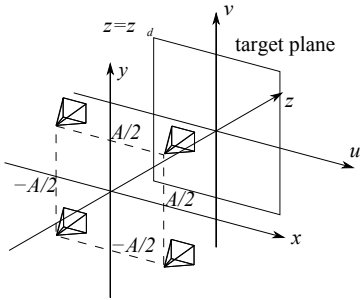
A key feature of our formulation is that  $\hat{\mathbf{W}}$  is expressed as the product of  $\hat{\mathbf{M}}$  and  $\hat{\mathbf{B}}$ , as shown in (11). This enables us to evaluate the camera arrangement by using the condition number of  $\hat{\mathbf{M}}$  separately from the PSFs represented by  $\hat{\mathbf{B}}$ . The condition number of  $\hat{\mathbf{W}}$  is upper-bounded by the condition numbers of  $\hat{\mathbf{M}}$  and  $\hat{\mathbf{B}}$  as

$$\begin{aligned} \text{cond}(\hat{\mathbf{W}}) &= \|\hat{\mathbf{M}}\hat{\mathbf{B}}\| \cdot \|(\hat{\mathbf{M}}\hat{\mathbf{B}})^+\| \\ &\leq \left(\|\hat{\mathbf{M}}\| \cdot \|\hat{\mathbf{B}}\|\right) \left(\|\hat{\mathbf{M}}^+\| \cdot \|\hat{\mathbf{B}}^{-1}\|\right) \\ &= \left(\|\hat{\mathbf{M}}\| \cdot \|\hat{\mathbf{M}}^+\|\right) \left(\|\hat{\mathbf{B}}\| \cdot \|\hat{\mathbf{B}}^{-1}\|\right) \\ &= \text{cond}(\hat{\mathbf{M}}) \cdot \text{cond}(\hat{\mathbf{B}}) \end{aligned} \tag{17}$$

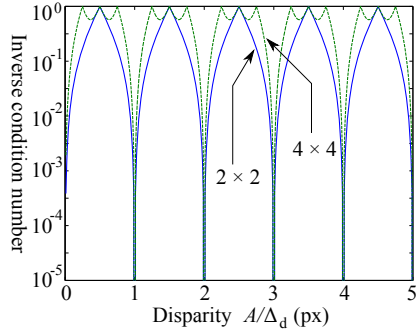
using sub-multiplicativity of the operator norm. We also use the inverse condition number for convenience.

### 5 Analyses of Camera Arrangements

In this section, we analyze some specific camera arrangements using the condition number. In subsection 5.1, we analyze regular grid arrangements and show that the condition number becomes infinite at periodic depths. In subsection 5.2, we analyze grid-and-perturbation arrangements using Monte Carlo simulation.



**Fig. 4.** A grid arrangement of 4 cameras



**Fig. 5.** Inverse condition number of a grid arrangement

### 5.1 Analysis on Regular Grid Arrangement

Assume that four cameras are placed on a  $2 \times 2$  regular grid. Let the camera positions be  $(\pm A/2, \pm A/2, 0)$ , where the distance between the cameras is  $A$ , as illustrated in Fig. 4.

Using (12), the camera arrangement matrix  $\hat{\mathbf{M}}$  is written as

$$\hat{\mathbf{M}} = \begin{pmatrix} 1 & e^{-j\pi A/\Delta_d} & e^{-j\pi A/\Delta_d} & e^{-2j\pi A/\Delta_d} \\ 1 & e^{-j\pi A/\Delta_d} & e^{j\pi A/\Delta_d} & 1 \\ 1 & e^{j\pi A/\Delta_d} & e^{-j\pi A/\Delta_d} & 1 \\ 1 & e^{j\pi A/\Delta_d} & e^{j\pi A/\Delta_d} & e^{2j\pi A/\Delta_d} \end{pmatrix}, \quad (18)$$

whose condition number (see appendix for derivation) is

$$\text{cond}(\hat{\mathbf{M}}) = \frac{1 + \left| \cos \pi \frac{A}{\Delta_d(z_d)} \right|}{1 - \left| \cos \pi \frac{A}{\Delta_d(z_d)} \right|}. \quad (19)$$

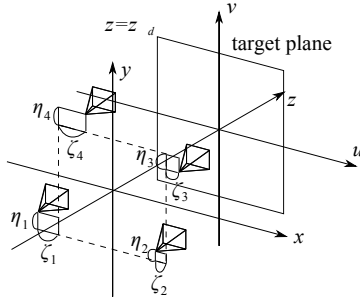
We also analyzed a case in which 16 cameras were arranged on a  $4 \times 4$  regular grid. The condition number of  $\hat{\mathbf{M}}$  (see appendix for derivation) is

$$\text{cond}(\hat{\mathbf{M}}) = \frac{1 + \left| \cos \frac{2\pi A}{\Delta_d(z_d)} \cos \frac{\pi A}{\Delta_d(z_d)} \right|}{1 - \left| \cos \frac{2\pi A}{\Delta_d(z_d)} \cos \frac{\pi A}{\Delta_d(z_d)} \right|}. \quad (20)$$

Figure 5 shows the inverse condition number of the camera arrangement matrix for the regular grid arrangement of 4 cameras and 16 cameras. The horizontal axis represents the value of  $A/\Delta_d(z_d)$ . Note that  $A/\Delta_d(z_d)$  is inversely proportional to the depth of the target plane  $z_d$  and corresponds to the disparity between the input LR images.

As shown in the figure, the inverse condition number takes zero at periodic depths where  $A/\Delta_d(z_d)$  is an integer. When the target plane is located at these





**Fig. 6.** An example of grid-and-perturbation arrangements of 4 cameras

depths, the sampling points of all input cameras coincide with each other. We refer to these depths as singular depths. It should be noted that the singular depths exist regardless of the number of cameras as long as they are arranged in regular grids.

It is obvious that the inverse condition number takes the maximum value at periodic depths, where the disparity is a half-integer in the case of the  $2 \times 2$  grid and a quarter-integer in the case of the  $4 \times 4$  grid. These depths, where the SR reconstruction is the most stable, are referred to as the best depths.

To summarize, when the cameras are arranged in a regular grid, SR reconstruction becomes ill-posed at some depths yet well-posed at other depths. This situation is undesirable in terms of reconstructing an entire 3-D scene.

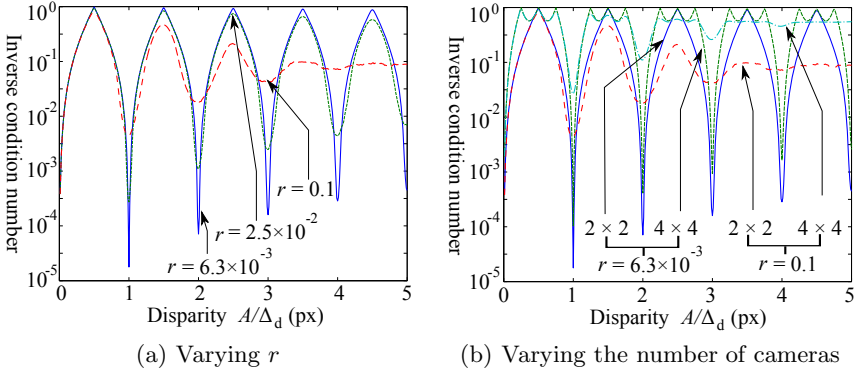
### 5.2 Analysis on Grid-and-Perturbation Arrangement

The periodic structure of the condition number along  $A/\Delta_d$  comes from the regularity of the camera arrangement. Thereby, randomizing the camera arrangement should decrease the periodicity and might be helpful to avoid the singular depth problem. In this subsection, we analyze a case where the camera arrangement is randomly perturbed from the regular grid.

For this case, analytical derivation of the condition number is difficult, so we used Monte Carlo simulations. We randomly generated many camera arrangements and numerically computed the condition numbers of the camera arrangement matrices  $\hat{\mathbf{M}}$ .

**Monte Carlo Simulations.** The number of cameras was set to either 4 or 16. The cameras were shifted from the  $2 \times 2$  or  $4 \times 4$  regular grid arrangements, as shown in Fig. 6. The shift of the  $k$ -th camera is denoted as  $(\zeta_k, \eta_k, 0)$ , where  $\zeta_k$  and  $\eta_k$  were independently sampled from the uniform distribution over  $(-rA, rA)$ .  $r > 0$  is the parameter that defines the range of the distribution and was used to control the randomness of the camera arrangement.

We exponentially varied  $r$  from  $10^{-10}$  to 1 and generated 1000 shifts for each  $r$  value. We then numerically computed the condition number of each arrangement for the range of  $0 < A/\Delta_d(z_d) \leq 10$ .

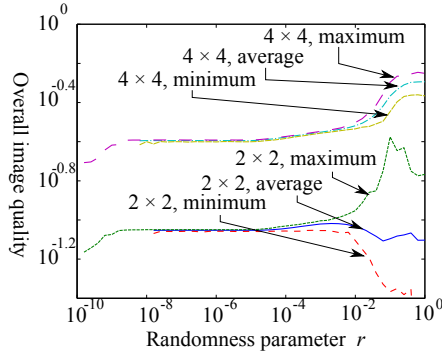


**Fig. 7.** Inverse condition number of a grid-and-perturbation arrangement

**Relation between Depth and Condition Number.** For each  $r$  value, we computed the geometric average of the inverse condition numbers over 1000 arrangements. The results with  $r = 6.3 \times 10^{-3}$ ,  $2.5 \times 10^{-2}$ , and  $0.1$  are shown in Fig. 7(a). Note that the horizontal axis is  $A/\Delta_d(z_d)$ . When we set  $r > 0$ , the inverse condition number became higher than zero for the depths where they were singular with the original regular grid arrangement. As  $r$  increased, the inverse condition number also increased at these depths. For instance, when  $A/\Delta_d(z_d) = 2$ , the inverse condition numbers for  $r = 6.3 \times 10^{-3}$ ,  $2.5 \times 10^{-2}$ , and  $0.1$  were  $10^{-4.5}$ ,  $10^{-2.5}$ , and  $10^{-1.5}$ , respectively. Meanwhile, as  $r$  increased, the inverse condition number decreased at best depths. For instance, when  $A/\Delta_d(z_d) = 3.5$ , the inverse condition numbers for  $r = 6.3 \times 10^{-3}$ ,  $2.5 \times 10^{-2}$ , and  $0.1$  were  $1$ ,  $10^{-0.5}$ , and  $10^{-1}$ , respectively. This tendency indicates that there is a trade-off between the improvement at the singular depths and the decline at the best depths. Randomizing the arrangement is likely to flatten the performance of the inverse condition number over the depths. This trade-off is discussed in more detail in the next subsection.

We also analyzed the relation between the number of cameras and the inverse condition number. Figure 7(b) shows the results when the number of cameras was 4 or 16 and  $r = 6.3 \times 10^{-3}$  or  $0.1$ . The grid-and-perturbation arrangement was effective for both 4 and 16 cameras. As a whole, the inverse condition numbers with 16 cameras were larger than those with 4 cameras.

**Relation between Randomness and Overall Image Quality.** As mentioned above, randomizing the camera arrangements raised the inverse condition numbers at singular depths, but lowered them at best depths. Therefore, we introduced a new measure, referred to as overall image quality, which is the geometric average of the inverse condition numbers over the range of the entire 3-D scene. The range was set to  $0 < A/\Delta_d(z_d) \leq 5$  in this experiment. Note that the overall image quality is zero for regular grid arrangements, since there are singular depths.



**Fig. 8.** Relation between randomness parameter  $r$  and overall image quality

We tested 1000 random arrangements for each  $r$ . Figure 8 shows the relation between the randomness parameter  $r$  and the overall image quality. The maximum, average, and minimum values of the overall image quality are plotted in this figure. Note that the vertical axis is logarithmic.

The average and minimum values are not plotted on the left side of the graph because their values were zero due to the machine precision. This indicates that when  $r$  is very small and the arrangement is very close to the regular grid arrangement, some depths become nearly singular. Therefore,  $r$  should not be very small.

When the number of the cameras was four, the overall image quality gradually increased when  $r < 10^{-3}$  but gradually decreased when  $r > 10^{-3}$ . Improvements around the singular depth and degradation around the best depths seem to have balanced around  $r = 10^{-3}$ . When the number of the cameras was 16, the overall image quality monotonically increased in the range of  $10^{-10} \leq r \leq 1$ . This result indicates that randomizing the camera arrangement is more effective when more cameras are used.

The difference between the maximum and minimum overall image quality became bigger as  $r$  increased. This tendency indicates that a very large  $r$  should be avoided to control the overall image quality. It should also be noted that the worst overall image quality is still more than zero when  $r > 10^{-8}$ . Therefore, we can expect the overall image quality to improve by randomizing the camera arrangement even in a worst-case scenario.

## 6 Conclusion

In this paper, we considered the arrangement of multi-view cameras for light-field super-resolution. We formulated an SR reconstruction in the frequency domain and derived the relation between the camera arrangement and the stability of the SR reconstruction using the condition number. Based on this relation, we showed that the singular depths, where the reconstruction becomes ill-posed, periodically appear in the case of regular grid arrangements. We also revealed

that randomizing the camera arrangement can prevent the singular depths and improve the stability of the SR reconstruction for the entire 3-D scene. Our future work is to verify the correctness of our theoretical analysis by experiments. We also plan to analyze more general camera arrangements.

**Acknowledgments.** This research is supported by the Strategic Information and Communication R&D Promotion Programs (SCOPE) of the Ministry of Internal Affairs and Communications, Japan.

## References

1. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis Machine Intelligence* 24(9), 1167–1183 (2002)
2. Champagnat, F., Besnerais, G.L., Kulcsár, C.: Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework. *Journal of the Optical Society of America A* 26(7), 1730–1746 (2009)
3. Fukushima, N., Ishibashi, Y.: Free viewpoint image generation with super resolution. In: *Picture Coding Symposium*, pp. 1–4 (2010)
4. Mudenagudi, U., Gupta, A., Goel, L., Kushal, A., Kalra, P., Banerjee, S.: Super Resolution Of Images of 3D Scenecs. In: Yagi, Y., Kang, S.B., Kweon, I.S., Zha, H. (eds.) *ACCV 2007, Part II. LNCS*, vol. 4844, pp. 85–95. Springer, Heidelberg (2007)
5. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine* 20(3), 21–36 (2003)
6. Robinson, D., Milanfar, P.: Statistical performance analysis of super-resolution. *IEEE Transactions on Image Processing* 15(6), 1413–1428 (2006)
7. Takahashi, K., Naemura, T., Tanaka, M.: Rate-distortion analysis of super-resolution image/video decoding. In: *International Conference on Image Processing* (2011)
8. Tanaka, M., Okutomi, M.: Theoretical analysis on reconstruction-based super-resolution for an arbitrary PSF. In: *IEEE Computer Vision and Pattern Recognition*, vol. 2, pp. 947–954 (2005)
9. Tung, T., Nobuhara, S., Matsuyama, T.: Simultaneous super-resolution and 3d video using graph-cuts. In: *IEEE Computer Vision and Pattern Recognition*, pp. 1–8 (2008)

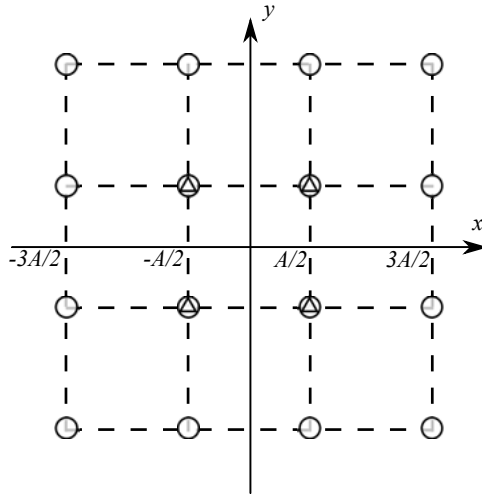
## Appendix: The Condition Number of Regular Grid Arrangements

See Fig. 9 for the configuration.

### A. $2 \times 2$ Cameras

Here, we derive (19). First, we compute  $\hat{\mathbf{M}}^* \hat{\mathbf{M}}$ , which is written as

$$\hat{\mathbf{M}}^* \hat{\mathbf{M}} = 4 \begin{pmatrix} 1 & \alpha & \alpha & \alpha^2 \\ \alpha & 1 & \alpha^2 & \alpha \\ \alpha & \alpha^2 & 1 & \alpha \\ \alpha^2 & \alpha & \alpha & 1 \end{pmatrix}, \quad (21)$$



**Fig. 9.** Configuration. The triangles and circles represents the positions of the cameras in  $2 \times 2$  and  $4 \times 4$  regular grid arrangements, respectively.

where

$$\alpha = \frac{e^{j\pi A/\Delta_d} + e^{-j\pi A/\Delta_d}}{2} \tag{22}$$

$$= \cos \frac{\pi A}{\Delta_d}. \tag{23}$$

By analytically solving the eigenequation of  $\hat{\mathbf{M}}^* \hat{\mathbf{M}}$ , we obtain

$$\lambda = 4(1 \pm |\alpha|)^2, 4(1 - |\alpha|^2). \tag{24}$$

Note that  $\lambda = 4(1 - |\alpha|^2)$  is a double root.

Since  $0 \leq |\alpha| \leq 1$ , these eigenvalues satisfy the relation

$$4(1 - |\alpha|)^2 \leq 4(1 - |\alpha|^2) \leq 4(1 + |\alpha|)^2. \tag{25}$$

The left equality holds when  $|\alpha| = 0, 1$ , and the right equality hold when  $|\alpha| = 0$ . Using (25), the maximum and minimum eigenvalues are

$$\lambda_{\max}(\hat{\mathbf{M}}^* \hat{\mathbf{M}}) = 4(1 + |\alpha|)^2, \tag{26}$$

$$\lambda_{\min}(\hat{\mathbf{M}}^* \hat{\mathbf{M}}) = 4(1 - |\alpha|)^2. \tag{27}$$

By substituting (23), (26), (27) into (15), we obtain (19).

**B.  $4 \times 4$  Cameras**

We derive (20). Similar to the previous derivation, we compute eigenvalues of  $\hat{\mathbf{M}}^* \hat{\mathbf{M}}$ . In this case,  $\hat{\mathbf{M}}^* \hat{\mathbf{M}}$  becomes

$$\hat{\mathbf{M}}^* \hat{\mathbf{M}} = 16 \begin{pmatrix} 1 & \alpha & \alpha & \alpha^2 \\ \alpha & 1 & \alpha^2 & \alpha \\ \alpha & \alpha^2 & 1 & \alpha \\ \alpha^2 & \alpha & \alpha & 1 \end{pmatrix}, \quad (28)$$

where

$$\begin{aligned} \alpha &= \frac{e^{3j\pi A/\Delta_d} + e^{j\pi A/\Delta_d} + e^{-j\pi A/\Delta_d} + e^{-3j\pi A/\Delta_d}}{4} \\ &= \cos \frac{2\pi A}{\Delta_d} \cos \frac{\pi A}{\Delta_d}. \end{aligned} \quad (29)$$

The form of  $\hat{\mathbf{M}}^* \hat{\mathbf{M}}$  is the same as (21) except for the value of  $\alpha$  and multiplication by a constant value. Therefore, the maximum and minimum eigenvalues are similarly computed as

$$\lambda_{\max}(\hat{\mathbf{M}}^* \hat{\mathbf{M}}) = 16(1 + |\alpha|)^2, \quad (30)$$

$$\lambda_{\min}(\hat{\mathbf{M}}^* \hat{\mathbf{M}}) = 16(1 - |\alpha|)^2. \quad (31)$$

By substituting (29), (30), (31) into (15), we obtain (20).