PERFORMANCE ANALYSIS OF RECONSTRUCTION-BASED SUPER-RESOLUTION FOR CAMERA ARRAYS

Kuang-Tsu Shih and Homer H. Chen

National Taiwan University, Taipei, Taiwan

ABSTRACT

In this paper, we look into the possibility of combining multiple images captured by a camera array into a high-resolution image by reconstruction-based super-resolution. Specifically, we analyze the effect of camera array parameters, including the number of cameras, pixel size, and sampling interval, on the quality of the super-resolved image. Our experimental results show that the pixel size of the cameras is critical to the success of resolution enhancement and that only limited resolution gain is achievable if the pixel of the cameras is not sufficiently small. Our analysis serves as a guide to the design of high-resolution camera arrays.

Index Terms— Computational photography, camera array, super-resolution, light field

1. INTRODUCTION

An increasing number of electronic devices are equipped with more than one camera to achieve better image quality and more novel functions. For example, dual cameras are used in smartphone to widen the dynamic range and provide the synthetic aperture function. Among the specifications of an imaging system, resolution is often the top one. Therefore, it is natural to ask if it is possible to combine the images captured by a camera array into one high-resolution sharp image. For example, if we have two 10-megapixel cameras, can we generate a 20-megapixel sharp image?

This question is practically important. Most existing high-resolution imaging systems are costly because of the use of expensive optics and sensors, and hence people turn to computational photography for a more economical solution. Camera array, as one possible solution, has high potential to be implemented at low cost. However, there is little literature on resolution-related performance analysis of camera array. Therefore, we are motivated to investigate it, with the hope that the analysis can provide significant insight for the design of high-resolution camera arrays.

Analyzing the resolution of a general camera array is not a trivial task because it depends on a large number of parameters. These parameters can be roughly divided into two groups: hardware-related ones, which describe the way the camera array is configured, and software-related ones, which describe the algorithm that fuses the images captured by individual cameras. In this work, we focus on camera arrays composed of identical co-planar cameras with parallel optical axes. In addition, we assume that the field of view (FOV) of the cameras overlaps by a large portion.

Our contribution is twofold. First, we provide an analysis of the effect of the camera array parameters on the performance of super-resolution. Second, a reconstruction-based super-resolution algorithm is developed for images captured by a camera array that cannot be registered through a global parametric warping adopted in most previous work.

2. RELATED WORK

2.1. Camera Array

Wilburn et al. have done a series of studies [2]–[4] of camera array. Their camera array is re-configurable and consists of one hundred cameras. By changing the configuration, the camera array can achieve many different objectives. Particularly, to enhance resolution, the camera array is configured so that only a small portion of the field of view (FOV) of each camera overlaps with others. The captured images are then stitched together using global homographic warping and multi-band blending. However, there is no way to estimate the scene depth for regions where the FOVs do not overlap, and the resolution gain for regions where the FOVs overlap is still an issue.

The high-resolution camera array built by Venkataraman et al. [7] is closely related to our work. On the hardware side, great effort was spent on optics design to broaden the modulation transfer function (MTF) of the lens. On the software side, a super-resolution algorithm and a depth estimation algorithm were developed to generate high-resolution image and depth map. However, no analysis was provided to show how the performance of super-resolution algorithm varies with the parameters of the camera array.

2.2. Analysis of Super-Resolution

We review two important pieces of work on performance analysis of super-resolution reconstruction. Baker and Kanade [8] examined the limit of super-resolution from a mathematical point of view. Under the assumption that the low-resolution images can be registered by pure translation,

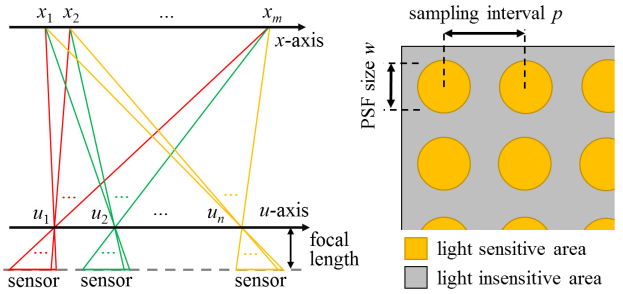

Fig. 1. An illustration of the generation of input images. The left figure illustrates the relation between the images captured by a camera array and the 4D light field data, while the right figure illustrates the sensor plane of a camera.

they reached the conclusion that the performance of super-resolution algorithm has an upper bound and gets inevitably worse when the magnification factor is larger. Likewise, Lin and Shum [13] derived an explicit formula of the performance bound of super-resolution reconstruction using the perturbation theory in linear algebra and under the assumption that the relation between the low-resolution images is purely translation.

Our work is distinct from previous work in two aspects. First, we do not assume a pure translation relation between images because it is impractical for real applications. Instead, the disparity map of the scene is used to register the images. Second, our analysis focuses on camera array. Specifically, we investigate the impact of the parameters associated with the camera array on the quality of the output image. Therefore, we believe that the conclusions drawn from our analysis are more direct and helpful for guiding the design of camera arrays.

3. SUPER-RESOLUTION FOR CAMERA ARRAY

An experiment is designed to investigate the effect of the camera array parameters on the quality of the reconstructed image. A quick sketch of the experiment is as follows. First, we generate a series of low-resolution images. Next, a super-resolved image is computed from the low-resolution images. Finally, we observe how the quality of the super-resolved image varies with the parameters under consideration using the peak signal-to-noise ratio (PSNR) as the quality metric. Three parameters are considered in this experiment: the number of cameras n, the size of point spread function (PSF) w, and the sampling interval p (i.e. pixel pitch). An illustration of the parameters is shown in Fig. 1.

3.1. Assumptions

We make the following assumptions in the experiment. First, all objects in the scene are diffusive (i.e. Lambertian). This assumption is equivalent to the constant brightness assumption widely adopted in previous work [10], [12]. Second, all cameras in the camera array are ideal pin-hole cameras. This assumption rules out possible resolution degradation caused by defocus and diffraction. Third, camera calibration is assumed to be error free. In other words, the information about the position and orientation of the cameras is available and exact. Fourth, the ground-truth depth map is available for every camera. Finally, we assume that scene depth is constant within one pixel [5].

3.2. Input Images

We use a 4D light field [11] to simulate the images captured by a camera array. Let $L(x, y, u, v)$ and $I_{uo,vo}(x, y)$, respectively, denote the 4D light field of a scene and the image captured by the camera centered at (u_o, v_o). From the 2D analogy shown in Fig. 1, we can see that $I_{uo,vo}(x, y) = L(x, y, u_o, v_o)$. To simulate light ray aggregation and discrete sampling, we convolve $I_{uo,vo}(x, y)$ with the PSF $K(x, y)$ and then subsample the image with sampling interval p. Here we assume that $K(x, y)$ is a pillbox function with diameter w (readers are referred to our previous work [6] for the effect of PSF shape on super-resolution). Note that, since the FOV of each camera is fixed, increasing p is equivalent to reducing the number of pixels.

Let $I_1(x, y)$, $I_2(x, y)$, …, and $I_n(x, y)$ each denote the high-resolution image that falls on the sensor of each camera in the camera array. After light ray aggregation and sampling, we get n low-resolution images $I_1[i, j]$, $I_2[i, j]$, …, and $I_n[i, j]$, which serve as the input images in our experiment. Specifically,

$$I_t[i,j] = I_t(x,y) * K(x,y)\big|_{(x,y)=(ip,jp)} + N[i,j], \qquad (1)$$

where t =1, 2, …, n, is the image index, " $*$ " denotes convolution, and $N[i, j]$ denotes white Gaussian noise. Here we use parentheses and square brackets to distinguish between high-resolution signal and low-resolution signal. Also, the summation of high-resolution signals is denoted by the integral symbol.

3.3. Super-Resolution

Here we describe how we compute the super-resolved image from the low-resolution images captured by a camera array. First, let us define the symbols to be used in the derivation. Let $D_t[i, j]$ denote the depth of pixel $[i, j]$ of the t-th camera, t = 1, 2, …, n. From the basic geometry, we know that the horizontal disparity $\Delta x_{tr}[i, j]$ of pixel $[i, j]$ of the t-th camera with respect to the r-th camera is

$$\Delta x_{tr}[i,j] = f(u_t - u_r)/D_t[i,j], \qquad (2)$$

where f is the focal length and u_t and u_r, respectively, are the u-coordinate of the center of the t-th and the r-th cameras. Likewise, the vertical disparity is

$$\Delta y_{tr}[i,j] = f(v_t - v_r)/D_k[i,j], \qquad (3)$$

where v_t and v_r, respectively, are the v-coordinate of the center of the t-th and the r-th cameras. With the assumptions that the scene depth is constant within a pixel and that the scene is Lambertian, we have

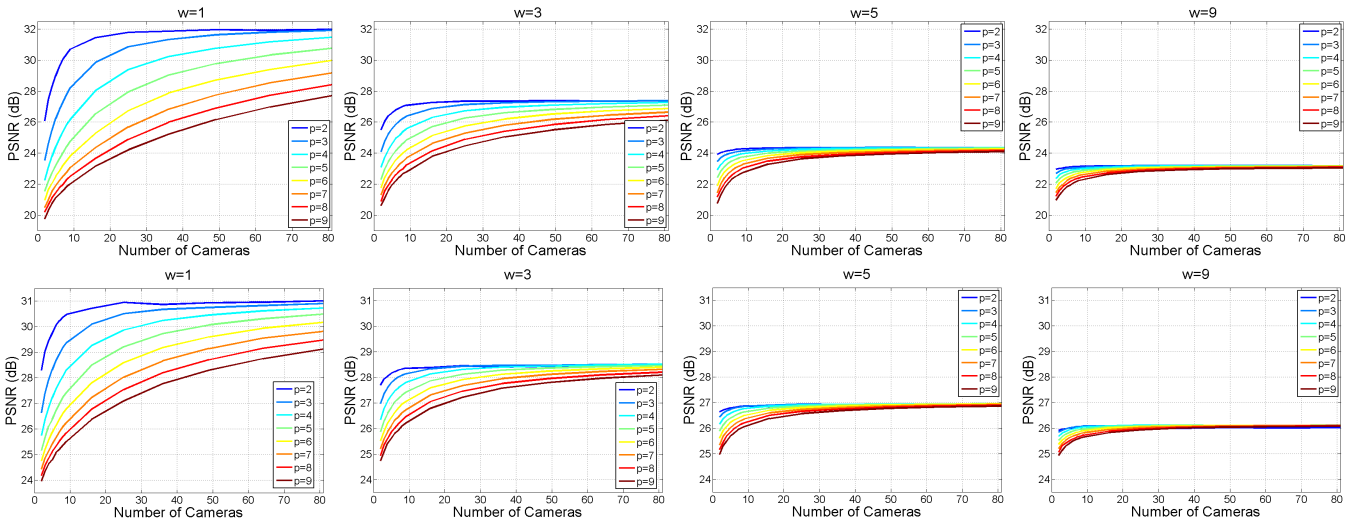

Fig. 2. The PSNR value of the super-resolved image plotted as a function of the number of cameras n. Figures in the top row are the results for the light field data "horses," while those in the bottom row are the results for the light field data "medieval."

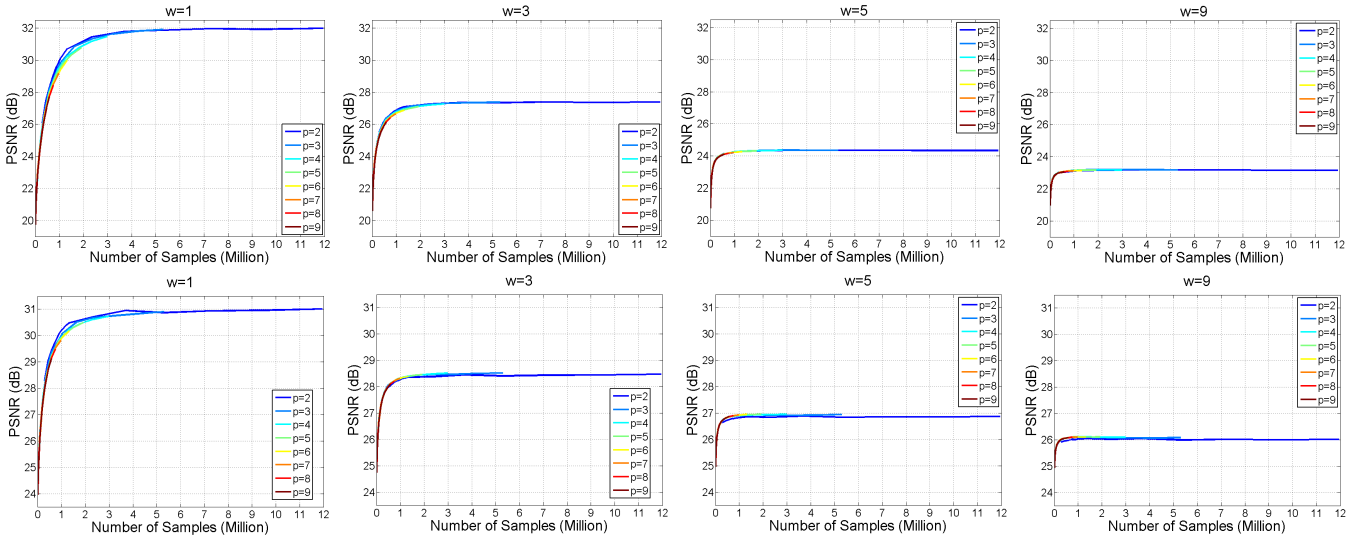

Fig. 3. The PSNR value of the super-resolved image plotted as a function of the number of total samples. Figures in the top row are the results for the light field data "horses," while those in the bottom row are the results for the light field data "medieval."

$$I_t(x,y) = I_r(x + \Delta x_{tr}[i,j], y + \Delta y_{tr}[i,j]) \tag{4}$$

for all (x, y) within the circle centered at (i, j) with diameter w.

We are now ready to derive the target function in the optimization. First, a camera is selected as the reference camera. Our goal is to compute the high-resolution version of the image captured by the reference camera. Let the r-th camera be the reference camera and $I_r(x, y)$ be the desired high-resolution image. We relate an arbitrary camera to the reference camera as follows:

$$\begin{aligned} I_t[i,j] &= I_t(x,y) * K(x,y)\big|_{(x,y)=(ip,jp)} \\ &= \int I_t(x',y') K(x-x', y-y') dx'dy' \Big|_{(x,y)=(ip,jp)} \\ &= \int I_r(x' + \Delta x_{tr}[i,j], y' + \Delta y_{tr}[i,j]) \\ &\qquad K(x-x', y-y') dx'dy' \Big|_{(x,y)=(ip,jp)} \\ &= \int I_r(x'',y'') K(x + \Delta x_{tr}[i,j] - x'', y + \Delta y_{tr}[i,j] - y'') \\ &\qquad dx''dy'' \Big|_{(x,y)=(ip,jp)} \\ &= \int I_r(x'',y'') K(x-x'', y-y'') dx''dy'' \Big|_{(x,y)=(ip+\Delta x_{tr}[i,j], jp+\Delta y_{tr}[i,j])} \\ &= I_r(x,y) * K(x,y)\big|_{(x,y)=(ip+\Delta x_{tr}[i,j], jp+\Delta y_{tr}[i,j])}. \end{aligned} \tag{5}$$

From (5), it is clear that $I_t[i, j]$ can be treated as samples of the blurred version of the desired high-resolution image at a shifted position. Therefore, to solve for $I_r(x, y)$, we construct the objective function as follows:

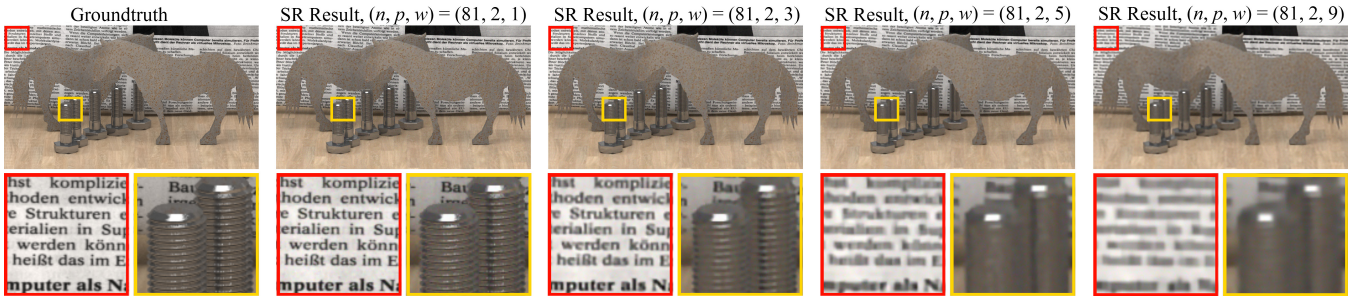

Fig. 5. The super-resolved images generated with different parameter sets and the ground truth high-resolution image for the dataset "horses."

$$\arg\min_{I(x,y)} \sum_{t=1}^{n} \sum_{i,j} (I_t[i,j] - I(x,y) * K(x,y)\big|_{(x,y)=(ip+\Delta x_{tr}[i,j],\, jp+\Delta y_{tr}[i,j])})^2. \quad (6)$$

To solve (6), we introduce a new variable $I_b(x, y)$, split the objective function, and add a regularization term,

$$\arg\min_{I(x,y),I_b(x,y)} \sum_{t=1}^{n} \sum_{i,j} (I_t[i,j] - I_b(x,y)\big|_{(x,y)=(ip+\Delta x_{tr}[i,j],\, jp+\Delta y_{tr}[i,j])})^2 + \int (I_b(x,y) - I(x,y) * K(x,y))^2 dxdy + \lambda \left\| \nabla I(x,y) \right\|_1^{\alpha}, \quad (7)$$

where we set α=0.8 and λ=0.02.

We solve (7) iteratively. In the first round, we fix $I(x, y)$ to be the up-sampled version of $I_r[i, j]$ and solve for $I_b(x, y)$ using the least-square optimization technique [9]. After $I_b(x, y)$ is obtained, we fix $I_b(x, y)$ and solve for $I(x, y)$ using the non-blind deconvolution technique [1]. The resulting $I(x, y)$ is then treated as a constant in the next round to solve for $I_b(x, y)$ and so on. The algorithm stops when the incremental difference of $I(x, y)$ is below a predetermined threshold or when the maximum number of iterations is reached.

4. EXPERIMENTAL RESULT

We apply the super-resolution algorithm to a number of 4D light fields, two of which are shown in Fig. 2. Two observations are made. First, the PSNR of the super-resolved image does not increase boundlessly with the number of cameras. In fact, it seems to converge to a constant value, which we call the PSNR ceiling. Second, the convergence speed and the PSNR ceiling are largely determined by the PSF size w. Specifically, for a smaller w, the PSNR ceiling is higher and the convergence speed is slower. This is more obvious if we use the total number of samples (i.e. the number of pixels of each camera multiplied by the number of cameras n) as the horizontal axis. We can see from Fig. 3 that, if the PSF size w is large, increasing the number of samples (either by increasing the number of cameras n or by decreasing the pixel pitch p) has very limited gain of super-resolution quality. To investigate how the PSNR ceiling varies with w, we plot the PSNR ceiling versus w for four 4D light field data in Fig. 4. We can see that the PSNR ceiling drops monotonically as w increases. For the reader to subjectively evaluate the super-resolution quality, we show super-resolved images in Fig. 5. It is clear that super-resolution cannot recover image details when w is large.

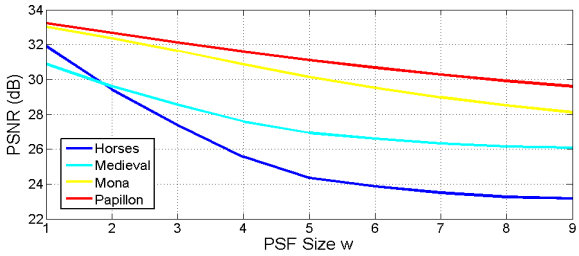

Fig. 4. The PSNR ceiling plotted as a function of pixel width w for the four datasets used in our experiment.

To sum up, our experimental results suggest that, for a camera array with overlapped FOV, the resolution gain that can be acquired through computation is very limited unless the pixel size of the cameras, which is the lower bound of w, is sufficiently small. Consequently, if a camera array consists of cameras whose sensors have a nearly 100% fill factor (i.e. $w \fallingdotseq p$), the achievable resolution gain is expected to be low. Therefore, to design a high-resolution imaging solution using a camera array with overlapped FOV, it is necessary to reduce the light sensitive area of the cameras.

Let us close the discussion with a few remarks. First, the experiment assumes perfect optics and image registration. Hence, the actual achievable resolution for real world applications is expected to be lower. Second, although the regularization term (7) adopted in our algorithm is not as sophisticated as that in some state-of-the-art super-resolution algorithms, it is sufficient to make our point because what we care is the gain brought about by adding cameras rather than the gain brought about by prior knowledge.

5. CONCLUSION

In this paper, we have investigated the effect of camera array parameters on the performance of reconstruction-based super-resolution. We have demonstrated that increasing the number of cameras or the sampling density does not necessarily improve the quality of super-resolved images. Our experimental results show that, to improve the quality of super-resolved images, it is necessary that the cameras have a sufficiently small pixel size and a sharp PSF. Therefore, to design a high-resolution imaging solution using a camera array, the light sensitive area of the cameras needs to be reduced as a tradeoff.

6. REFERENCES

[1] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 70:1–70:9, 2007.

[2] B. Wilburn, "High performance imaging using arrays of inexpensive cameras," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., 2004.

[3] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 765–776, 2005.

[4] B. Wilburn, N. Joshi, V. Vaish, M. Levoy and M. Horowitz, "High-speed videography using a dense camera array," in *CVPR*, pp. 294–301, 2004.

[5] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Transactions on Graphics*, vol. 34, no. 2, pp. 16:1–16:19, 2015.

[6] K.-T. Shih, C.-Y. Hsu, and H. H. Chen, "Analysis of the effect of calibration error on light field super-resolution rendering," IEEE Int. Conf. Acoustics, Speech, Signal Process., pp. 534-538, May 2014.

[7] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: an ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 166:1–166:13, 2013.

[8] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[9] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University press, 2004.

[10] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[11] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," *International Symposium on Vision, Modeling and Visualization*, pp. 1–8, 2013.

[12] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field super-resolution," *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–9, 2009.

[13] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based super-resolution algorithms under local translation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, 2004.