RECONSTRUCTING A SUPER-RESOLUTION IMAGE FOR DEPTH-VARYING SCENES

Ami YOKOYAMA Akira KUBOTA Yoshinori HATORI

† ‡ Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology
Yokohama, Japan
E-mail: † yokoyama.a.aa@m.titech.ac.jp ‡ {kubota/hatori}@ip.titech.ac.jp

ABSTRACT

In this paper, we present a novel method for reconstructing a super-resolution image using multi-view low-resolution images captured for depth varying scene without requiring complex analysis such as depth estimation and feature matching. The proposed method is based on the iterative back projection technique that is extended to the 3D volume domain (i.e., space + depth), unlike the conventional super-resolution methods that handle only 2D translation among captured images.

Keywords: Improving resolution, Iterative Backward Projection, Ghosting artifacts, Multi-view images

1. INTRODUCTION

Recently, many electronic image products(e.g. digital camera, HDD recorder, television) deal with high-resolution contents because of development of CMOS or CCD sensor process. However, there are some cases that high-resolution camera can not be used, for example, traffic system, space satellite and etc. In such cases, cheap and low power consumption camera that captures low-resolution image is needed to obtain high-resolution image. In this paper, we propose a novel method to reconstruct a high-resolution image from a set of low-resolution images by image processing.

Super resolution can be broadly classified as:

1) Super resolution from one low-resolution image, which considers control conditions, for example continuance of spectrum.

2) from a set of low-resolution images, which we explain in detail bellow.

In reconstructing high-resolution image from a set of low-resolution images, how a high-resolution image is degraded to low-resolution images, is most important. Basically, the model consists of "pixel shift", "PSF (the Point Spread Function)", and "down sampling." For the single depth scene, we only consider 2D shift. However, scene has varying depth in general. Then we have to consider the complex warping for each pixel. So far, it's usually achieved with depth map. Through the estimation of depth, the problem deals with single depth method on each estimated depth. However, we propose a new model that considers shift, PSF, down sampling all together. This model can reconstruct high-resolution image wherever the object exists.

2. THE PROPOSED METHOD

2.1 Iterative Backward Projection

Irani et al. [1] proposed IBP (Iterative Backward Projection) for reconstructing a high-resolution image from a set of low-resolution images. It's the method for single depth scene and based on the optimization problem. First, a set of low-resolution images $\{c_j\}$ is acquired with real cameras (j is number of camera, $j = 1, 2, \ldots, M$). Second, a high-resolution image f is assumed. Third, a set of low-resolution images $\{\hat{c}_j\}$ is simulated through the assumed degradation model. $\{c_j\}$ and $\{\hat{c}_j\}$ are compared at each corresponded pixel. The difference between them is used to update f. When the difference is minimized, a high-resolution image f is finally reconstructed.

2.2 Reconstruction method for depth varying scene

IBP has been used in many methods. But most methods deal with a single depth scene. Thus we propose a novel method for depth varying scene. The diagram of the proposed method is shown in Fig.1. We improve the algorithm of IBP in the depth domain.

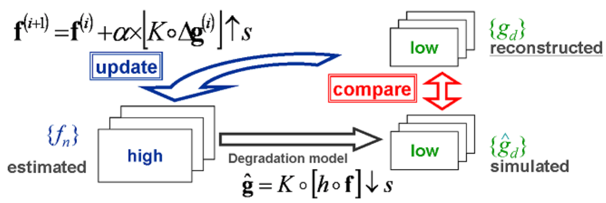


Fig. 1: Diagram of the proposed method

Since our algorithm deals with all depths, we write the depth image as vector of

$$\boldsymbol{g} = \left[\begin{array}{cccc} g_1 & g_2 & \ldots & g_N \end{array} \right]^T, \tag{1}$$

where $\{g_d\}$ is a set of low-resolution images, and d is number of depth ($d = 1, 2, \ldots, N$). $\{f_n\}$ is a set of high-resolution images, each f_n is the texture image at depth n, and n is number of depth ($n = 1, 2, \ldots, N$). The high-resolution image f is reconstructed by

$$f = f_1 + f_2 + \ldots + f_n + \ldots + f_N = \sum_{n=1}^{N} f_n. \quad (2)$$

$\{\hat{g}_d\}$ is a set of low-resolution images simulated through the degradation model. In this paper, we propose the degradation model that considers shift, PSF, down sampling all together. In the following, we explain the algorithm in detail.

In the proposed method, first, we assume N depths z_d in the scene. Depending on each depth, we generate virtual view image g_d from captured images $\{c_j\}$:

$$g_d = \frac{1}{M} \sum_{j=1}^{M} c_j(c \quad {}_{j,d}), \quad (3)$$

where ${}_{j,d}$ is the disparity on c_j corresponding to z_d. Now, we assume that cameras are arranged in the same intervals. Next, we assume a set of high-resolution depth images $\{f_n\}$ by

$$\boldsymbol{f}^{(0)} = [h \circ \boldsymbol{g}] \uparrow s, \quad (4)$$

where h is Gaussian filter , and $[\,] \uparrow$ is up-sampling operator.

Then, our degradation model is defined as

$$\hat{g}_d = k_{d1} * [h * f_1] \downarrow s + \ldots + k_{dN} * [h * f_N] \downarrow s$$

$$= \sum_{n=1}^{N} k_{dn} * [h * f_n] \downarrow s; k_{dn} = \delta(d = n), \quad (5)$$

where k_{dn} is the ghost operator that depends on depth, and $[\,] \downarrow$ is down-sampling operator. Thus, it can be written by

$$\hat{\boldsymbol{g}} = \boldsymbol{K} \circ [h \circ \boldsymbol{f}] \downarrow s, \quad (6)$$

where

$$\boldsymbol{K} = \begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1N} \\ k_{21} & k_{22} & \ldots & k_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ k_{N1} & k_{N2} & \ldots & k_{NN} \end{bmatrix} \; k_{dn} = 1(d = n). \quad (7)$$

The update process is expressed as

$$\boldsymbol{f}^{(i+1)} = \boldsymbol{f}^{(i)} + \alpha \quad [\boldsymbol{K} \circ \boldsymbol{\Delta g}^{(i)}] \uparrow s, \quad (8)$$

where i is number of iteration, α is a positive coefficient, and $\triangle g$ is the difference between g_d and $\hat{g}_d$. The update process is done iteratively. Our proposed method can reconstruct high-resolution image, independent of depth of the objects.

3. EXPERIMENTAL SIMULATIONS

In this section, we tested the performance of the proposed method.

3.1 Reconstruction image from 7 cameras

First, we made $320 \quad 240$ images by POV-ray, computer graphic software. Then we filtered and down-sampled those images through the same degradation process in the algorithm. The scene for experiment is a plane rotated $45°$ at depth 34. We mapped standard image "lenna" successively on the plane. We arranged a camera at the origin and 6 cameras on the circumference with the radius of 1. The set of input images are $160 \quad 120$ for the experiment of 2 times and $80 \quad 60$ for the experiment of 4 times. A input image set consists of 7 images. The view point of output image is set at the origin. The resolution is $320 \quad 240$. Number of depth is 5, and number of iteration is 100. Then, the generated depth images g_d are shown in Fig.2, where Fig.2 (a) is the nearest image, and Fig.2 (e) is the furthest image. We can realize that the clear part moves from right(near) to left(far).

The results of experiment are shown in Fig.3. Fig.3 (a) is the grand truth image by POV-ray, Fig.3 (b) is the 2 times up-sampled image by B-spline method (PSNR 21.10[dB]), Fig.3 (c) is the 4 times up-sampled image by B-spline method (PSNR 18.02[dB]), Fig.3 (d) is the reconstructed image (2 times up-sampled) by the proposed method (PSNR 21.99[dB]), and Fig.3 (e) is the reconstructed image (4 times up-sampled) by the proposed method (PSNR 16.88[dB]). The PSNR of Fig.3 (d) is improved compared with B-spline method and the image is much sharper.As shown in Fig.3 (e) the proposed method decreases PSNR and the reconstructed image suffers from wavy artifacts.

3.2 Reconstruction image from 13 cameras

Next, we increased number of cameras up to 13. Other 6 cameras were put on the circumference with the radius of 2. Then, the results are shown in Fig.4. Fig.4 (a) is the grand truth image by POV-ray, Fig.4 (b) is the 2 times up-sampled image by B-spline method (PSNR 21.10[dB]), Fig.4 (c) is the 4 times up-sampled image by B-spline method (PSNR 18.02[dB]), Fig.4 (d) is the reconstructed image (2 times up-sampled) by the proposed method (PSNR 22.22[dB]), and Fig.4 (e) is the reconstructed image (4 times up-sampled) by the proposed method (PSNR 18.87[dB]). Fig.4 (d) achieved better PSNR than the result of experiment from 7 cameras. PSNR of Fig.4 (e) is better than that of Fig.3 (e).

3.3 Discussion

In the experiment for 2 times, the proposed method improved PSNR by 1.83[dB] compared with B-spline method, and we subjectively confirmed that the image was reconstructed better too. In the experiment for 4 times, the reconstructed image by our method improved PSNR by 2.25[dB] compared with B-spline method, This suggests that increasing camera number leads to suppressing artifacts.

4. CONCLUSIONS

In this paper, we proposed the novel algorithm to reconstruct a high-resolution image from a set of low-resolution images for depth varying scene. We showed improvement of PSNR and subjective evaluation.
We will perform experiment using real images for more complex scene.

5. REFERENCES

[1] Michal Irani et al. "Improving Resolution by Image Registration," ICVGIP: GRAPHICAL MODELS AND IMAGE PROCESSING, Vol.53, No.3, May, pp. 231–239, 1991.

[2] Akira Kubota, Keita Takahashi, Kiyoharu Aizawa, and, Tsuhan Chen, "All-Focused Light Field Rendering," Eurographics Symposium on Rendering (EGSR2004), pp. 235–242, 2004.

(a)

(b)

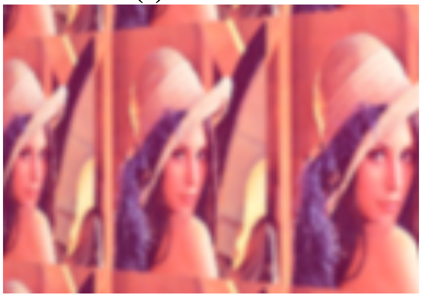
(c)

(d)

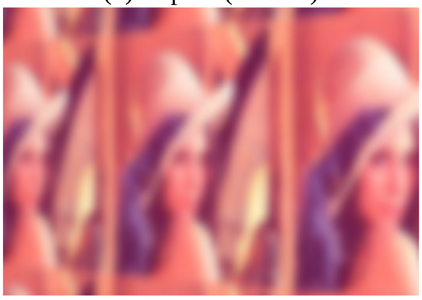
(e)

Fig. 2: Generated depth images g_d

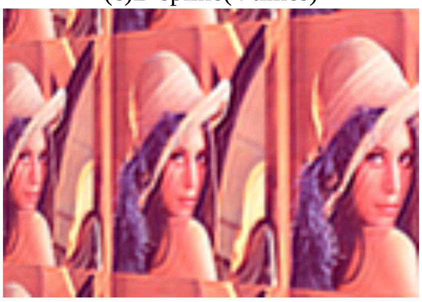
(a)Grand truth



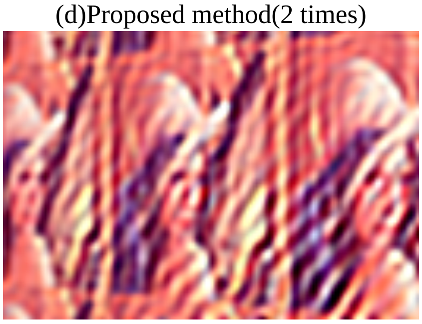
(b)B-spline(2 times)



(c)B-spline(4 times)



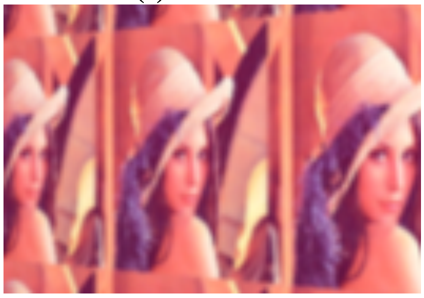
(d)Proposed method(2 times)



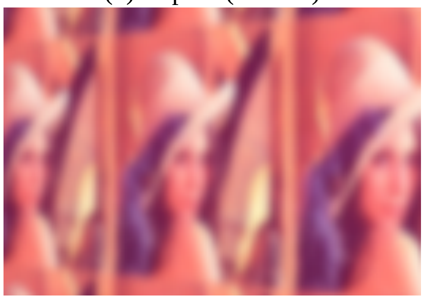
(e)Proposed method(4 times)

Fig. 3: Simulation Results with 7 cameras

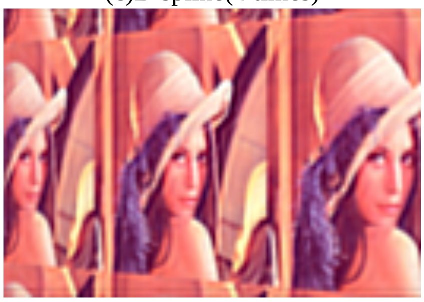
(a)Grand truth



(b)B-spline(2 times)



(c)B-spline(4 times)



(d)Proposed method(2 times)

(e)Proposed method(4 times)

Fig. 4: Simulation Results with 13 cameras