28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

FREE VIEWPOINT IMAGE GENERATION WITH SUPER RESOLUTION

Norishige FUKUSHIMA, Yutaka ISHIBASHI

Graduate School of Engineering, Nagoya Institute of Technology

ABSTRACT

In this paper, we propose a method of free viewpoint image generation with super resolution. In the conventional approaches, such as nearest neighbor and linear interpolation, the synthetic image on zoomed virtual view tends to have low resolution, because the reference images do not have enough textures. To overcome this problem, we reconstruct the image with super resolution. Super resolution can generate higher image resolution than the input image one, and then we combine super resolution with free viewpoint image generation. In the experiment, we use a camera array which contains 11 x 11 aligned cameras and use 4 x 4 cameras subset per pixel to reconstruct image by means of super resolution. The experimental results show that synthesized image in the effective range has about 4.5 dB higher PSNR than ones created by the nearest neighbor and 2.5 dB higher than ones created by the linear interpolation.

Index Terms— Image Based Rendering, Super Resolution, Ray Space, Light Field, Free Viewpoint Image

1. INTRODUCTION

In the last two decades, IBR (Image Based Rendering) grows popularly as photo-realistic rendering method. Light field rendering [1] and Ray space [2] have been known as 3D representation method in the context of IBR and both methods can render free viewpoint image. According to the plenoptic sampling theory [3], however, the rendering range where synthesized image should have enough quality is limited. If the generated view is out of range, it is aliased, i.e. ghosted or blurred. The method of [4] which combines depth estimation and ray space can render free viewpoint image in full range without aliasing. However there still exists a problem. If virtual camera is zoomed, the image looses sharpness because the resolution of reference images is insufficient.

Now, from first addressing in [5], super resolution which enhances the image resolution from multiple low resolution images is well studied too. The reference [5] proposes frequency domain approach and also spatial domain approaches are developed, such as iterative back projection [6] and bilateral total variation [7].

The spatial domain super resolution has high affinity to the free viewpoint image generation. Therefore we will combine free view generation with super resolution method.

Organization of this paper is as follows. Section 2 shows the method of free viewpoint image generation and describes the limitation of the resolution. Section 3 explains super resolution view generation method. Section 4 reveals the efficacy of proposing method by simulated experiments. Finally, we conclude this paper in section 5.

2. FREE VIEWPOINT IMAGE GENERATION AND ITS RESOLUTION

In this section, we describe how free viewpoint image is generated from multi view images, and its limitation on image resolution.

Free viewpoint image means that a user can change viewpoint freely; not only intermediate view of camera but forward and backward view. In this paper, we generate this free viewpoint image by using multi view images which are captured by aligned camera array (see Figure 1). This data is called light field or ray space.

If depth map information on desired view is given, the ray or pixel $\mathbf{r}=(u,v)$ in virtual view, where (u,v) is pixel position, is represented by weighed linear interpolation of rays in neighborhood views I_i, $i=\{1,2,\ldots,N\}$, according to the disparity information. Here, i is the index of view and N is the total number of views.

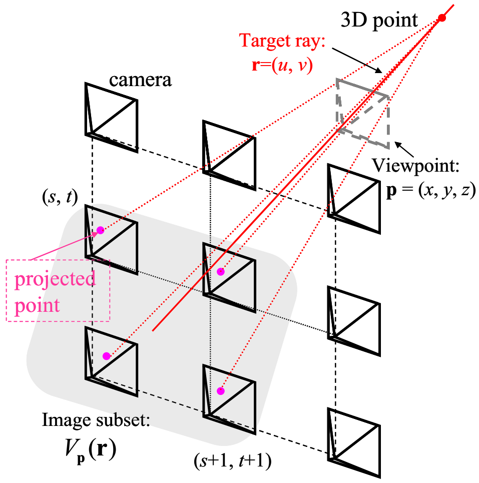

Figure 1 Ray interpolation from 3x3 camera array.

978-1-4244-7133-1/10/$26.00 ©2010 IEEE

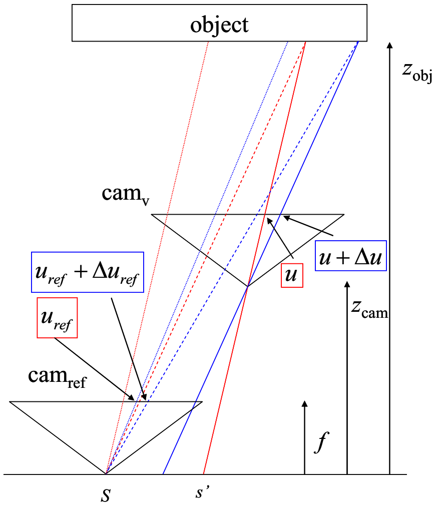

Figure 2 Relationship among the object, virtual camera cam_v and reference camera cam_{ref}.

The equation of the ray in the free viewpoint image $I_{\mathbf{p}}$ on viewpoint $\mathbf{p} = (x, y, z)$ is written as

$$I_{\mathbf{p}}(\mathbf{r}) = \sum_{i\in V_{\mathbf{p}}(\mathbf{r})} w_i(\mathbf{r}) I_i(\mathbf{r}+\mathbf{d}_{\mathbf{p},i}(\mathbf{r})),$$

where $w_i(\mathbf{r})$ is a weighted function depending on position of ray, and $\mathbf{d}_{\mathbf{p},i}(\mathbf{r})$ is a shifting vector to match the points which indicate same object among required views, and are computed from depth information easily. If the pointing pixel is sub-pixel, we naturally use linear interpolation or nearest neighbor interpolation. $V_{\mathbf{p}}(\mathbf{r})$ is the subset of cameras which is required. This subset is defined by the ray space [2] method. In the case of using 4-neighbourhood views for ray computation, the subset becomes as follow;

$$s = \lfloor x+uz \rfloor, t = \lfloor y+vz \rfloor,$$

$$V_{\mathbf{p}}(\mathbf{r}) = \left\{ \begin{pmatrix} s \\ t \end{pmatrix} \begin{pmatrix} s+1 \\ t \end{pmatrix} \begin{pmatrix} s \\ t+1 \end{pmatrix} \begin{pmatrix} s+1 \\ t+1 \end{pmatrix} \right\},$$

where s and t are horizontal and vertical camera indexes in the camera array, respectively. Figure 1 shows an example of 3 x 3 camera array case. The weighting factor w_i depends on the distance from neighborhood views and normalized by

$$\sum_{i\in V\mathbf{p}(\mathbf{r})} w_i(\mathbf{r}) = 1.$$

Now, we explain where is the reference ray in the reference image at (s,t) by Figure 2. At first, we define the reference ray $\mathbf{r}$+ $\mathbf{d}_{\mathbf{p},i}(\mathbf{r})$ as (u_{ref}, v_{ref}). These values become as follow by using the geometrical relation in Figure 2 (vertical relationship is omitted because of space limitations).

$$u_{ref} = u - \frac{z_{cam}}{z_{obj}} u + \frac{sf}{z_{obj}}, v_{ref} = v - \frac{z_{cam}}{z_{obj}} v + \frac{tf}{z_{obj}},$$

where z_{cam} is z value from the camera array plane to the virtual camera and z_{obj} is z value from the camera array plane to the object. f means focal length of the cameras. Then if z_{obj} is independent of $\mathbf{r}$, i.e. the object is plane and parallel to camera array, derivation of vector $\mathbf{r}_{ref}$, which means sampling pitch of reference ray, becomes as follow;

$$\Delta\mathbf{r}_{ref} = \Delta(u_{ref}, v_{ref}) = \frac{z_{obj} - z_{cam}}{z_{obj}} \Delta\mathbf{r}.$$

This equation indicates that sampling pitch of reference ray depends only on z_{obj} and z_{cam}. In the case of zoomed virtual camera, sampling pitch of reference ray becomes smaller than one of target ray. In this condition, rendering image resolution is decreased. Therefore up-sampling is required.

3. SUPER RESOLUTION

Super resolution reconstruction produces a high resolution image from a set of low resolution images. In this section, we introduce the method of free view point image generation with super resolution.

Let us summarize super resolution model. An ideal signal X(p, q) is observed by a camera as noisy image Y(u, v). This flow is expressed by this matrix representation:

$$Y_k = D_k H_k F_k X + \sigma_k (k = 1,2,\ldots,N),$$

where index of k indicates the number of image. X and Y_k is matrix representation of ideal signal and the kth noisy image. F_k means the kth image geometric motion operator between X and the kth frame of Y. H_k represents blurring effects caused by the combination of lens, motion and atmosphere on the kth image. D_k is down sampling operator of the kth image from high resolution image to low resolution one. σ_k represents the system noise on image and N is the number of available images. Thus, estimated high resolution image $\hat{X}$ will be generally obtained by the following function, which minimizes the L_2 norm. The norm is the distance between images of low resolution and degraded image of high resolution;

$$\hat{X} = \arg\min_{\underline{X}} \sum_{k=1}^{N} \| D_k H_k F_k X - Y_k \|^2 .$$

However, super resolution reconstruction is ill-posed problem, so that there are an infinite number of images which satisfy this assumption. To obtain a stable solution, we use regularization term which compensates the lost information with prior information about general high resolution image, e.g. sharpness of image. Above equation is modified;

$$\hat{X} = \arg\min_{X} \left[\sum_{k=1}^{N} \| D_k H_k F_k X - Y_k \|^2 + \lambda\gamma(X) \right],$$

where γ is a function which represents pre-known information, such as high pass filter and is balanced by the parameter λ. To combine this super resolution method and free viewpoint image generation, we regard reference image of forwarding view for free viewpoint image as low resolution image. The term of super resolution ray interpolation is;

$$\hat{I}\mathbf{p}(\mathbf{r}) = \underset{I_{\mathbf{p}}(\mathbf{r})}{\arg\min} \sum_{i\in V_{\mathbf{p}}(\mathbf{r})} w_i(\mathbf{r}) \| h * I_{\mathbf{p}}(\mathbf{r} + \mathbf{d}_{\mathbf{p},i}(\mathbf{r})) - I_i(\mathbf{r}) \|^2 , \\ + \lambda\gamma(I_{\mathbf{p}}(\mathbf{r}))$$

where $\hat{I}_{\mathbf{p}}(\mathbf{r})$ is estimated high resolution ray, and $I_{\mathbf{p}}(\mathbf{r})$ is ideal high resolution ray. h is blurring PSF (Point Spread Function) for image sampling, unfocused lens and atmosphere and is convoluted into ideal high resolution image . We use BTV (Bilateral Total Variation) method [7] based on MAP (Maximum A Posteriori) estimation as regularization term. $I_i(\mathbf{r})$, $w_i(\mathbf{r})$ and $\boldsymbol{d}_{\mathbf{p},i}$ are same as free viewpoint equation. Now, we rewrite this pixel domain super resolution function into matrix representation:

$$\hat{I}_{\mathbf{p}} = \underset{I_{\mathbf{p}}}{\arg\min} \sum_{i\in V} \mathbf{W}_i \| \mathbf{HF}_i I_{\mathbf{p}} - I_i \|^2 + \lambda\gamma(I_{\mathbf{p}}) .$$

In this equation, we use V which indicates set of all images instead of the subset, so that we control valid region on reference image I_i by filling the elements of weighting matrix $\mathbf{W}_i$ with zero. The matrix $\mathbf{F}_i$ represents the geometric motion operator of the ith image, which is showed in super resolution, and this matrix is build by the vectors $\mathbf{d}_{\mathbf{p},i}(\mathbf{r})$. $\mathbf{H}$ is blurring operator, and we assume that the blur kernel is independent of viewpoint and pixel position in this paper.

We use steepest descent to find the estimated free viewpoint image for this minimization problem,

$$I_{\mathbf{p}}^{k+1} = I_{\mathbf{p}}^{k} + \beta\left\{\sum_{i\in V} \mathbf{W}_i \mathbf{F}_i^T \mathbf{H}^T (\mathbf{HF}_i I_{\mathbf{p}} - I_i) + \lambda\gamma(I_{\mathbf{p}}^{k})\right\},$$

where β is scalar parameter which indicates step size of gradient to creep estimated image and $I_{\mathbf{p}}^{k}$ represent kth iterated image. T means transpose operation. The regularization term is defined as follow;

$$\gamma(I) = \sum_{l=0}^{P}\sum_{m=0}^{P} \alpha^{m+l} | I - \mathbf{S}_v^l \mathbf{S}_u^m I | ,$$

where matrices $\mathbf{S}_v^{\,l}$ and $\mathbf{S}_u^{\,m}$ shift I by m pixel horizontally and by l pixel vertically respectively. The parameter α ($0 < \alpha < 1$) spatially decreases the regularization effect and P is local block window size. Initial image whose parameter is k=0 is linear interpolation result in this paper.

4. EXPERIMENTAL RESULTS

In the experiment, we have simulation by using computer graphics (CG) image rendered by POV-Ray. Figure 3 shows the experimental condition which contains camera array and subjects. Figure 4 is the snapshot of central camera on the camera array. In this environment, we use a sphere with an image of Lenna as a texture, and slanted plane with an image of Mandrill as a texture as well. There are 11 x 11 cameras and each camera distance is 0.4. The field of view of all cameras is 25 degree and the image resolution is 640 x 480 with RGB color space. The distance between the nearest sphere and the camera array is set to 20.

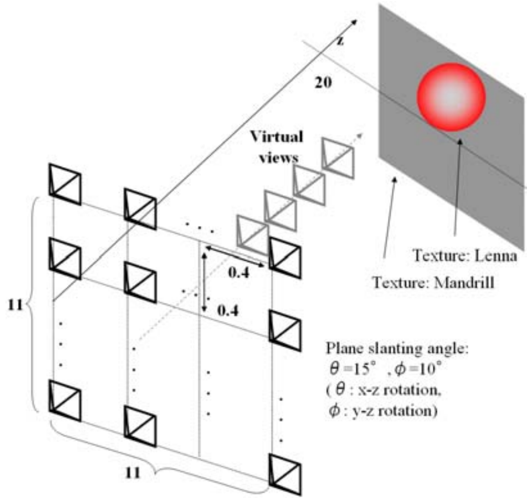

Figure 3 Experimental environment: Camera array, virtual camera and objects.

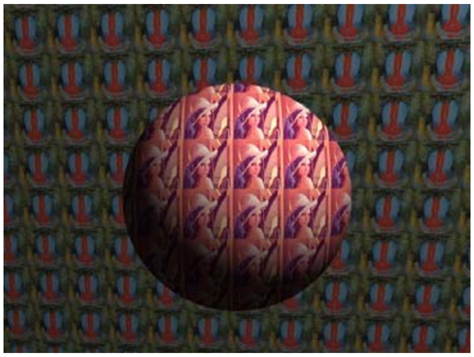

Figure 4 Rendering image of the central camera in the camera array generated by POV-Ray.

In the experiment, we have rendered free viewpoint images by three kinds of methods; nearest neighbor, linear interpolation and proposed super resolution method.

Each method has been evaluated by PSNR (Peak Signal to Noise Ratio) and reference images for PSNR measurement are generated by CG rendering. In this experiment, we use the following parameters; $\beta = 0.2, \lambda = 0.1, \alpha = 0.6$ and the number of iteration is 20. We have dealt the depth information with known, so that we have used computer generated depth information on the CG rendering. This precision of depth information has been under quarter pixels.

Figure 5 shows PSNR of rendering image versus viewpoint z. z=0 means that virtual camera on the camera array and z=20 means that the camera is on the nearest subject in this experiment case. This figure indicates the proposed method has highest PSNR within all range.

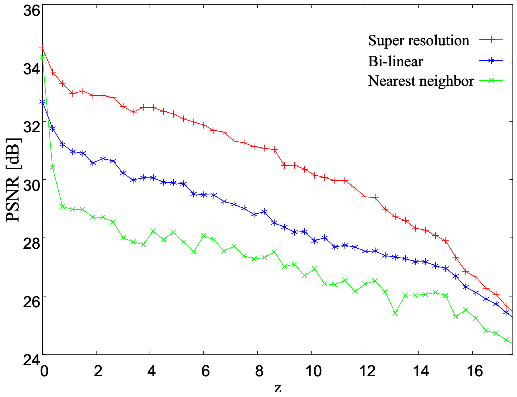

Figure 5 PSNR of conventional and proposed method

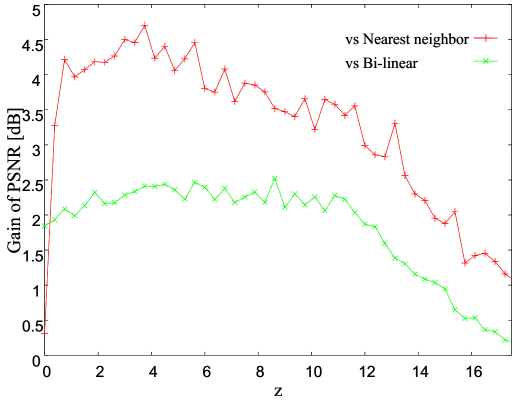

Figure 6 Improvement of PSNR which compares to Super resolution with conventional methods

Figure 6 shows the gain of PSNR between super resolution and bi-linear interpolation / nearest neighbor. Forwarding the virtual camera, the advantage of the proposed method is decreasing. The results show that synthesized image in the effective range has about 4.5 dB higher PSNR than the nearest neighbor and 2.5 dB higher than the linear interpolation. When the camera position z is around 15, which means 75% distance from camera array to the nearest object, the advantage which compare with linear interpolation becomes almost lost. Figure 7 shows ideal image and rendering result of all methods. Subjectively speaking, the proposed method can reconstruct edge information and sharpen the image.

5. CONCLUSION

In this paper, we introduced the super resolution free viewpoint image generation method from multi view images captured by camera array. Our experimental results showed that the proposed method has the highest image quality among conventional methods, which are nearest neighbor and linear interpolation. Objective measurement of PSNR showed that synthesized image in the effective range has about 4.5 dB higher PSNR than the nearest neighbor and 2.5 dB higher PSNR than the linear interpolation respectively. Our method has valid range from slightly forwarding z position to 75% distance from camera array to nearest object.

Applying this method to the natural images instead of CG images, improvement of noise-robustness is required and also sub-pixel depth estimation method is required. We will research these topics as future works.

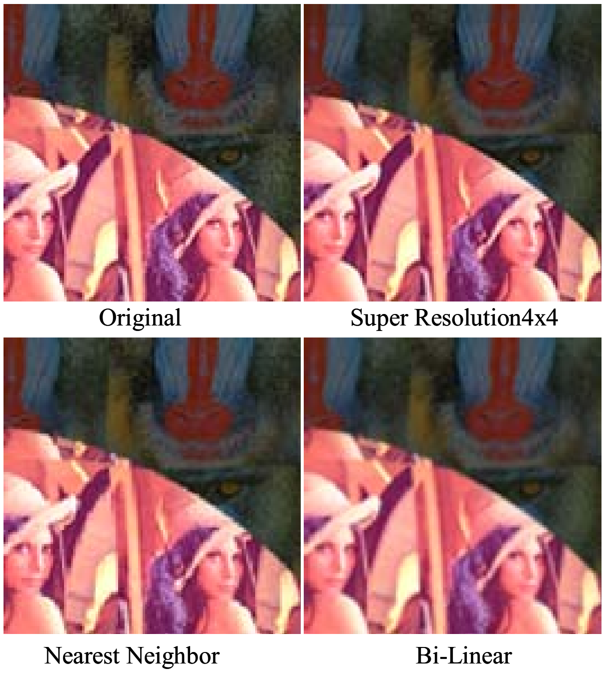

Figure 7 Synthesized images at **p** = (2.2, 2.2, 6)

This work was partly supported by the Grand-In-Aid for Young Scientists (B) of Japan Society for the Promotion of Science under Grant 22700174 and The Hori Information Science Promotion Foundation.

11. REFERENCES

[1] M. Levoy, P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH’96, pp. 31–42, 1996.

[2] T. Fujii, T. Kimoto, M. Tanimoto, “Ray Space Coding for 3D Visual Communication,” PCS'96, pp. 447-451, 1996.

[3] J. X. Chai, S. C. Chan, H. Y. Shum, X. Tong, “Plenoptic sampling,” Proc. ACM SIGGRAPH’00, pp.307–318, 2000.

[4] N. Fukushima, T. Yendo, T. Fujii, M. Tanimoto, “Free Viewpoint Image Generation Using Multi-Pass Dynamic Programming,” Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XIII, 6490A-59, 2007.

[5] T. S. Huang and R. Y. Tsai, “Multi-frame image restoration and registration,” Adv. Comput. Vis. Image Process., vol. 1, pp. 317–339, 1984.

[6] M. Irani and S. Peleg, “Improving resolution by image registration,” CVGIP: Graph. Models Image Process., vol. 53, pp. 231–239, 1991.

[7] S. Farsiu, D. Robinson, M. Elad, P. Milanfar, “Fast and Robust Multiframe Super Resolution,” IEEE Trans. Image Processing vol. 13, issue 10, pp. 1327–1344, 2004.