VIRTUAL VIEW RENDERING USING MULTIPLE STEREO IMAGES

Bumsub Ham, Dongbo Min, Kwanghoon Sohn

School of Electrical and Electronic Engineering
Yonsei University
Seoul, Korea
E-mail: khsohn@yonsei.ac.kr

ABSTRACT

This paper represents a new approach which addresses quality degradation of a synthesized view, when a virtual camera moves forward. Generally, interpolation technique using only two neighboring views is used when a virtual view is synthesized. Because a size of the object increases when the virtual camera moves forward, most methods solved this by interpolation in order to synthesize a virtual view. However, as it generates a degraded view such as blurred images, we prevent a synthesized view from being blurred by using more cameras in multiview camera configuration. That is, we solve this by applying super-resolution concept which reconstructs a high resolution image from several low resolution images. Therefore, data fusion is executed by geometric warping using a disparity of the multiple images followed by deblur operation. Experimental results show that the image quality can further be improved by reducing blur in comparison with interpolation method.

Keywords: IBR, virtual view rendering, super resolution, forward warping, backward warping.

1. INTRODUCTION

Freeview TV is a next generation TV system that enables a user change viewpoint freely as if he or she is there. Conventional TV is characterized by word "passivity" in that a user only sees a fixed viewpoint. In contrast, freeview TV is an active system in that this gives a user freedom of selecting viewpoint. Due to the limited number of acquisition equipment, view synthesis that belongs to image based rendering (IBR) category is one of the key techniques for freeview TV.

Many methods have been proposed to synthesize a novel view. IBR can be classified into three categories according to the geometric information used to synthesize views: rendering without geometry, rendering with explicit geometry and rendering with implicit geometry. Light field and lumigraph that belong to the first category capture multiple images densely enough to synthesize a view so that aliasing does not occur[8]. Because it samples multiple images, it can render a view without accurate geometry. However, it suffers from acquiring and storing data due to oversampling. Constructing a complete 3D model needs explicit geometry information, that is, true depth map or 3D coordinate. The system constructs a complete 3D model from 2D image set and renders the view which a user wants to see. However, finding explicit geometry information is difficult and time consuming task.

The third category, rendering with implicit geometry, is a method that synthesizes a view using sparsely sampled images. Generally, it consists of three stages: disparity estimation, image warping and view interpolation. A number of view interpolation methods have been proposed [10] [11]. Min et al proposed a view synthesis method considering motion parallax effect when the virtual camera moves forward [1]. They synthesized views by bilinear interpolation using two neighboring images when the virtual camera does so that holes cannot exist. However, this causes a virtual view to be blurred as shown in fig. 2 (a). In the remainder of this paper, we call this method conventional view synthesis.

In multiview configuration, we prevent the quality of a virtual view from being degraded when the virtual camera moves forward. We apply super-resolution concept [9] to IBR in order to prevent a synthesis view from being degraded. Super-resolution(SR) has been studied extensively in last two decades due to the limited sensor size and cost [12] [13]. Many solutions have been developed: maximum- likelyhood method (ML), maximum a-posteriori methods (MAP), projection onto convex sets (POCS) and nonlinear interpolation method. Our algorithm is motivated by nonlinear interpolation. As shown in fig. 1, we estimate a disparity between views. Using this information, each reference view is geometrically warped to the virtual camera viewpoint according to the virtual camera position. It is to be classified into two cases:

1. A virtual camera moves backward.
   Because the object of a virtual view gets smaller, the quality of a virtual view is not degraded in spite of using two neighboring views as shown in fig. 1 (b).
2. A virtual camera moves forward.
   Because we consider motion parallax effect, a size of the foreground object gets bigger than a size of the background object does when the virtual camera moves forward. Therefore, relatively many pixels are warped in background region and holes exist in foreground region. If we use conventional method that is based on backward warping, holes are interpolated using only two neighboring views. In addition, if we use multiple images in contrast to conventional method, backward based warping has some problems as shown in fig. 4 (b). Therefore, a synthesized view is degraded. This is a reason why we use forward based warping as shown in fig. 1 (a).

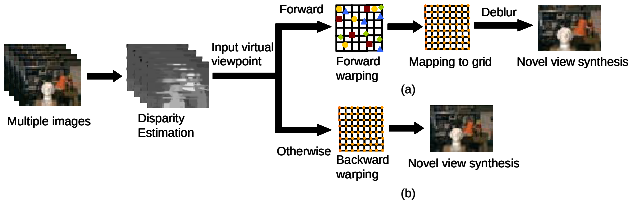


Fig. 1: Schemes for synthesizing a virtual view. (a) A virtual camera moves forward. (b) A virtual camera moves backward.

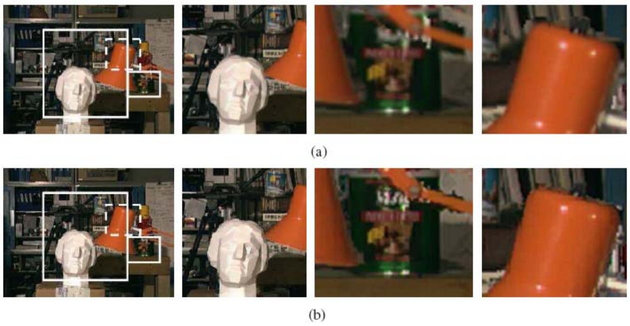


Fig.2: Conventional view synthesis (a) and our method (b).

A view is synthesized by interpolation such as bilinear using two neighboring cameras in conventional method. However, this makes objects blur and produces artifacts as shown in fig. 2(a). (The blurring is shown in the solid lined-rectangular, especially at object boundary and the halo effect are shown in the dashed lined-rectangular.) Seeing this scene, a user feels uncomfortable, and loses the detail of the scenes.

In multiview configuration, we have several scenes captured at different positions. Therefore instead of using only two neighbor images, we can use more than two images which are captured at different viewpoints in multiview system. In order to prevent a synthesized view from being degraded, we apply super-resolution concept instead of using interpolation techniques. It also does not assume constant depth and fixed viewpoint in contrast to the conventional super-resolution [6] [7] [14].

We suggest a novel view synthesis method superior to the conventional method as shown in fig. 2 (b). Our algorithm consists of three stages: Disparity estimation, data fusion and deblur. The main contribution is the data fusion. Using multiple images, we reconstruct an image whose quality is not degraded when a virtual camera moves forward.

The organization of this paper is as follows: Stereo matching used in this paper and generalized forward warping using multiple image are described in section 2. Hole filling method based on backward warping is also described. Experiment results are described in section 3. Finally, in section 4, we conclude with a summary.

2. NOVEL VIEW SYNTHESIS

2.1 Stereo Matching

We estimate a disparity by [1] named semi N-view & N-depth. In multiview camera configurations, given N views, we require N depth images as shown in fig. 3(a) for view synthesis. One of the problems in stereo matching is that it needs huge computational load. However, each reference image is estimated in same manner though the disparity maps of neighboring images are generally similar to each other, except occlusion regions. Therefore we can reduce the complexity in N-view & N-depth framework by warping of cost function from reference view to target view as shown in fig. 3(b).

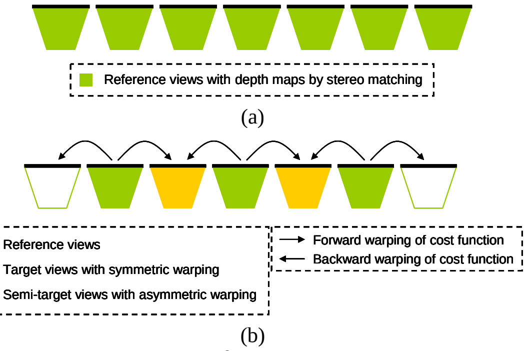


Fig. 3: N-view & N-depth framework(a) and semi N-view & N-depth framework(b).

2.2 Geometric based forward warping

Conventional method is based on backward warping because forward warping suffers from a hole problem [4] [5]. However, we use only pixels that are from N-views instead of using the interpolated pixels as shown in fig. 5. In this case, forward warping is suitable than backward warping in that there is a possibility that a pixel slightly moves at inappropriate location as shown in fig. 5 (b) when a virtual camera moves forward. It leads to problems when the virtual view is synthesized, especially the boundary. Fig. 6 shows the movement of the virtual camera. The glo_x and glo_z are the virtual camera location along x-axis and z-axis, respectively. Virtual camera moves along x-axis and z-axis. The y-axis movement is limited. Given M cameras, we first symmetrically choose N-views from the location of the virtual camera along x-axis. The N-views are numbered in ascending order from the left most view.

Using the geometric information, that is, the disparity, we change coordinate from image coordinate of N cameras to that of the virtual camera. A point in the synthesized view is computed from a disparity of the i^{th} view as follows:

$$x^v - x_0 = f\,\frac{(x_i - x_0)B_i/d_i(x_i,y_i) + T_{i,x}}{fB_i/d_i(x_i,y_i) + T_{i,z}} = \frac{(x_i - x_0) + d_i(x_i,y_i)\alpha_{i,x}}{1 + d_i(x_i,y_i)\alpha_{i,z}/f}$$

$$y^v - y_0 = f\,\frac{(y_i - y_0)B_i/d_i(x_i,y_i) + T_{i,y}}{fB_i/d_i(x_i,y_i) + T_{i,z}} = \frac{(y_i - y_0) + d_i(x_i,y_i)\alpha_{i,y}}{1 + d_i(x_i,y_i)\alpha_{i,z}/f} \quad (1)$$

where f represents a focal length and B represents a base line. To simplify the notation, we normalize the translation, $(T_{i,x}/B_i, T_{i,y}/B_i, T_{i,z}/B_i) = (\alpha_{i,x}, \alpha_{i,y}, \alpha_{i,z})$ and set baseline to 1.0. We only estimate a disparity from neighboring views. If a disparity between i and $i+1$ view is x, a disparity between i and $i+2$ can be approximately $2x$, because we assume that baseline is equal. Since disparity influences view synthesis, it is computed as in Eq. (2) from left and right side views respectively in order to find a refined disparity. We represent a disparity from i to $i+1$ as $d_i^0(x_i, y_i)$ and a disparity from i to $\lceil glo\_x \rceil$ as $d_i(x_i, y_i)$ as shown in fig. 7. This operation is independent of the other view, so it is possible to run parallel. The problem of the forward warping is that the mapping is not one-to-one

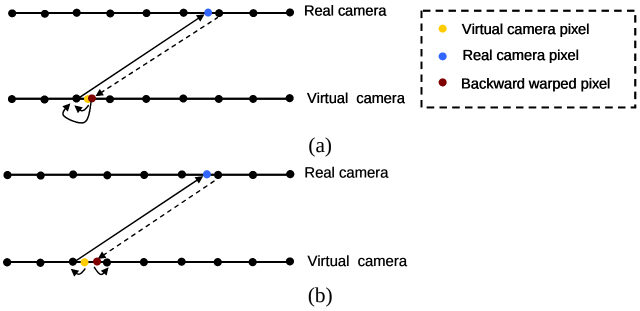


Fig. 4: Several case of backward warping.

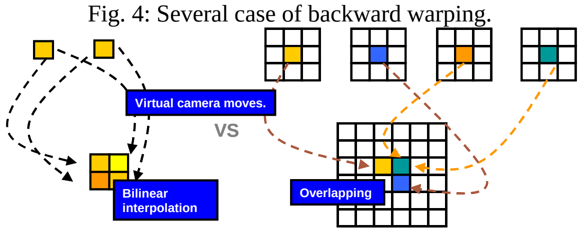


Fig. 5: Bilinear interpolation and our method.

correspondence. Therefore, there is a problem that the background pixel and the foreground pixel are mapped into the same point of the virtual view at the object boundary.

If the warped disparity is larger than the disparity which is already filled, we select the former so that the background disparity cannot penetrate into the foreground region followed by warping intensity to the virtual camera coordinate as in Eq. (3).

$$d_i(x_i,y_i)=\begin{cases} d_i^0(x_i,y_i)+\sum\limits_{k=i+1}^{\lfloor glo\_x \rfloor} d_k(x_k,y_k) \\ \quad x_k = x_{k-1}-d_{k-1}^0(x_{k-1},y_{k-1}) \\ d_i^0(x_i,y_i)+\sum\limits_{k=\lceil glo\_x \rceil}^{i-1} d_k(x_k,y_k) \\ \quad x_k = x_{k+1}+d_{k+1}^0(x_{k+1},y_{k+1}) \end{cases} \tag{2}$$

Visibility function (V) is also updated. It shows that whether a pixel in the virtual view is visible in the reference view with values of 1 when visible. However, forward warping by visibility function may have problem that the background disparity penetrates into the foreground regions as shown in fig. 8.

$$\begin{aligned} I^v(x^v,y^v) &= I_i(x_i,y_i) \\ D_i(x^v,y^v) &= d_i(x_i,y_i) \end{aligned} \tag{3}$$

Fig. 9 shows why this problem occurs. When both background and foreground disparity are warped in similar virtual camera coordinate, we can select foreground pixel as in fig. 9 (a) because a disparity of the foreground is larger than the one in the background. However, when the virtual camera moves forward, the size of an object gets bigger. Because we do not use interpolation in order to filling a hole, some pixel is not assigned at the object boundary. Therefore, when there is no warped pixel in virtual camera at object boundary and the background disparity is warped near the hole, it is filled by the background disparity as shown in fig. 9 (b). Therefore, when the disparity is warped to coordinate of the virtual view, it is decided that whether the disparity is similar to neighboring disparities which is already filled. If they are dissimilar exactly, the warped disparity is eliminated.

2.3 Hole filling using backward warping

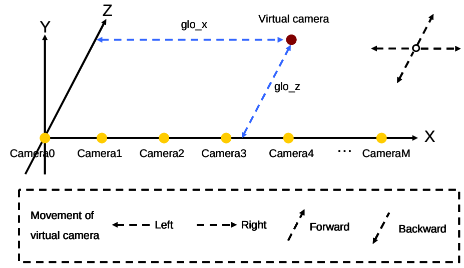


Fig. 6: Movement of virtual camera.

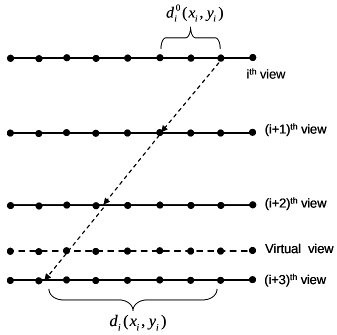


Fig. 7: Disparity refinement in multiview images.

We do not apply any interpolation techniques to synthesize a virtual view. Because the result of forward warping has more holes when camera moves near the object, we have to fill it. Therefore, we apply backward warping with interpolated disparity map ($Dis_i(x^v,y^v)$). Interpolated disparity map used only to change coordinate of the virtual camera to that of the real camera as in Eq. (4).

$$\begin{aligned} x_i &= (x^v-x_0)(1-f\alpha_{i,z}Dis_i(x^v,y^v))+x_0+\alpha_{i,x}Dis_i(x^v,y^v) \\ y_i &= (y^v-y_0)(1-f\alpha_{i,z}Dis_i(x^v,y^v))+y_0+\alpha_{i,y}Dis_i(x^v,y^v) \end{aligned} \tag{4}$$

Interpolated disparity maps are made as follows. When an image of real camera is warped to the coordinate of the virtual camera, an image grid is scaled up. Then the disparity is appropriately replicated as shown in fig. 10. The disparity of real camera is warped to virtual camera coordinate and appropriately replicated so that a hole does not exist. By using visibility function of the virtual view, a hole is filled as in Eq. (5).

$$I^v(x^v,y^v) = I_i(x_i,y_i)V_i(x^v,y^v) \tag{5}$$

Finally, a synthesized view is deblurred slightly by shock filter [2] [3].

3. EXPERIMENT RESULTS

We use four reference views to synthesize a virtual view. A view of the first camera is shown in fig. 11. Fig. 13 shows several results in comparison with bilinear interpolation. Normalized weighted interpolation is a method that disparities and color values are forward warped, and a hole is filled by weighted interpolation using a distance from pixel grid.

When virtual camera moves forward slightly, except the bilinear interpolation, others are similar as shown in fig. 13 and fig. 14.(The first row) However, the more the virtual camera moves, the more degradation is observed as shown in the third row of fig. 13 and fig. 14. As you can see, bilinear interpolation is blurred and brings to halo effect at object boundary.

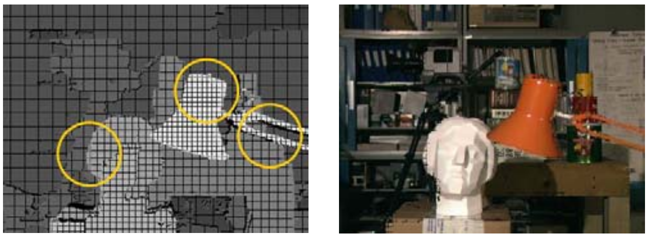


Fig. 8: Background disparity prnetrates into the foreground.

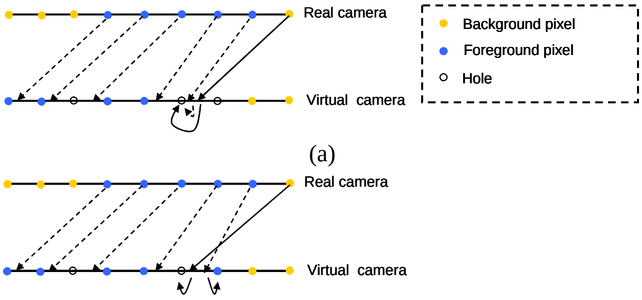


Fig. 9: Problem of disparity warping.

Normalized weighted interpolation has a little blur. However since both methods are based on the interpolation techniques, PNSR is low in comparison with backward warping and our method. The result of backward warping shows a problem as shown in fig. 4(b). Our method shows a sharp and no artifact such as halo effect at boundary region.

Since the ground truth of Tsukuba sequence does not exist, we calculate PSNR as in table 1 with Robot sequence rendered 3D MAX as shown in fig. 12. As we expected, forward based method is superior to other methods both PSNR results and visual quality.

While visual quality of normalized weighted method is better than bilinear interpolation, we see that PSNR result of the former is lower than the latter. It may be that relatively many outliers exist as compared with other method. Also, overall PSNR gain is mild in comparison with visual quality. This may be that the pixels are slightly shifted. Another reason that may account for this is some outlier. We intend to further explore this in order to propose more elaborate data fusion method that shows good results both PSNR and visual quality.

Table 1: PSNR results. B.I: bilinear interpolation, N.W.I: normalized weighted interpolation, F.W: forward warping, B.W: backward warping.

PSNR	Before Deblur	After Deblur
B.I	19.72dB	19.62dB
N.W.I	18.33dB	18.48dB
B.W	20.06dB	20.22dB
F.W+B.W	20.32dB	20.59dB

4. CONCLUSIONS

In this paper, we have proposed a novel view synthesis that is not blurred when virtual camera moves forward. First, we estimate a disparity by semi N-view & N-depth. Second, each reference images are forward warped geometrically and fused using visibility function. Because a hole exists, it is filled by applying backward warping. Finally, we apply deblur algorithm based on shock filter. We compare it with conventional bilinear interpolation, backward based

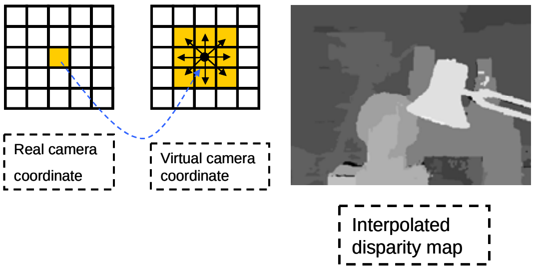


Fig. 10: How to make interpolated disparity maps.

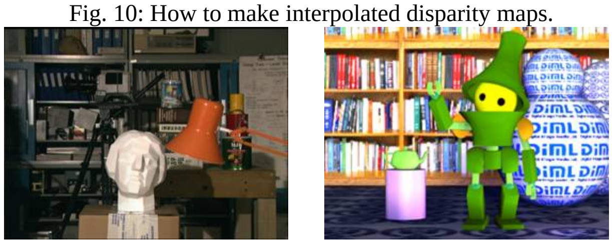


Fig. 11: Reference view.

Fig. 12: Ground truth image.

warping methods and normalized weighted interpolation. Experiment results show that our method is better than others both visual quality and PSNR. In further works, we will investigate more elaborate date fusion which is robust to disparity error and outliers.

5. REFERENCES

[1] D. Min and D. Kim and K. Sohn, "2D/3D FREEVIEW VIDEO GENERATION FOR 3DTV SYSTEM," ICIP 2008.

[2] S. Osher and LI. Rudin, “Feature-Oriented Image Enhancement Using Shock Filters,” SIAM Journal on Numerical Analysis Vol. 27. pp. 919-940, 1990.

[3] L. Alvarez and L. Mazorra, "Signal and Image Restoration Using shock Filters and Anisotropic Diffusion," SIAM Journal on Numerical Analysis Vol. 31. pp. 590-605, 1994.

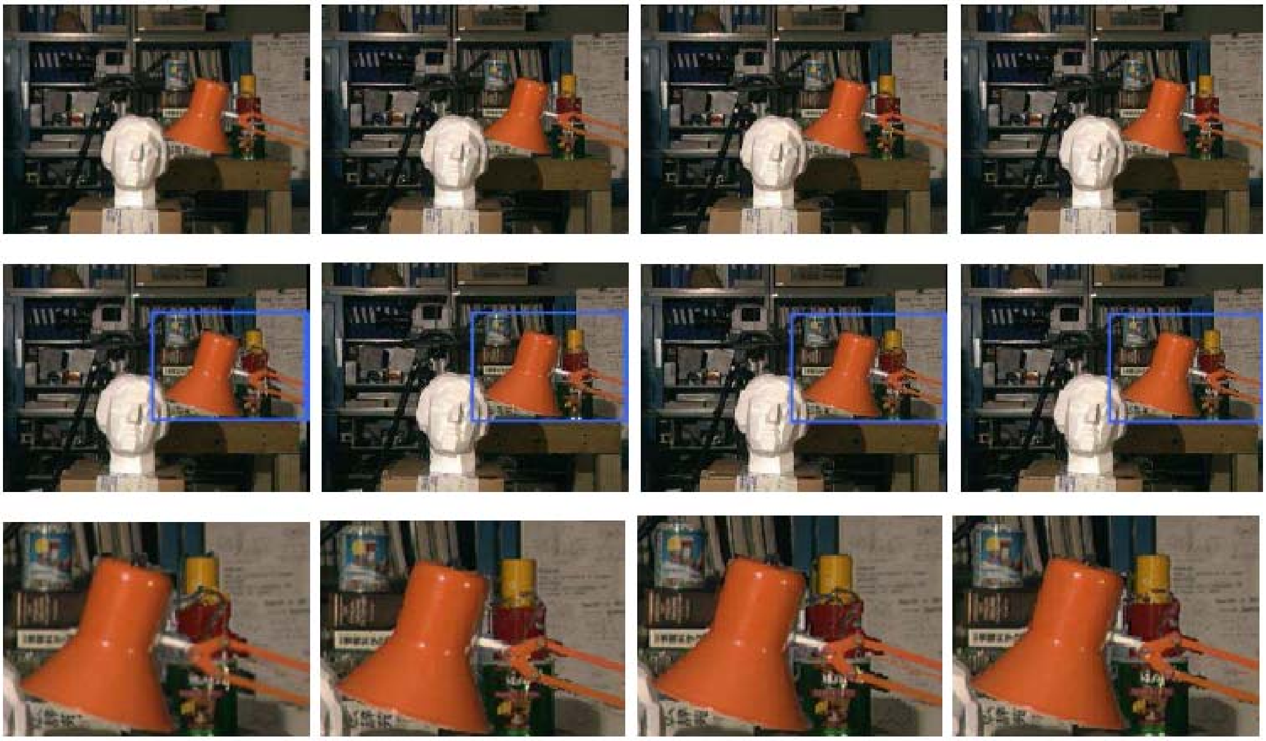
[4] M. Gong and JM. Selzer and C. Lei and YH. Yang, "Real-time Backward Disparity-based Rendering for Dynamic Scenes using Programmable Graphics Hardware,." Graphics Interface 2007. Vol. 234. pp 241-248, 2007.

[5] K. Connor and I. Reid, "Novel View Specification and Synthesis," In Proc. 10th British Machine Vision Conference pp 243-252, 2002.

[6] X. Li, "Super-Resolution for Synthetic Zooming," EURASIP Journal on Applied Signal Processing pp 1-12, 2006.

[7] K. Kimura and T. Nagai and H. Nagayoshi and H. Sako, "Simultaneous Estimation of Super-Resolved Image and 3D Information using Multiple Stereo-Pair Images," ICIP pp 417-420, 2007.

[8] A. Kubota and K. Aizawa and T. Chen, "Reconstructing Dense Light Field From Array of Multifocus Images for Novel View Synthesis," IEEE



(a) (b) (c) (d)

Fig. 13: Results for view synthesis: Tsukuba sequence. (a) bilinear interpolation, (b) Normalized weighted interpolation, (c) backward warping, and (d) our method. (To check the difference definitely, see the electrical version.)

(a) (b) (c) (d)

Fig. 14: Results for view synthesis: Robot sequence. (a) bilinear interpolation, (b) Normalized weighted interpolation, (c) backward warping, and (d) our method. (To check the difference definitely, see the electrical version.)

Trans. Image Processing Vol. 16. pp 269-279, 2007.

[9] S. Park and M. Park and M. Kang, "Super-Resolution Image Reconstruction : Technical Overview," Signal Processing Magazine Vol. 20. pp 21-36, 2003.

[10] SE. Chen and L. Williams, "View interpolation for image synthesis," SIGGRAPH. pp279-288, 1993.

[11] SM. Seitz and CR. Dyer., "Physically-Valid View Synthesis by Image Interpolation," ICCV. pp.18-25, 1995.

[12] M. Elad and Y. Hel-Or, "A Fast Super-Resolution Reconstruction Algorithm for pure Translational Motion and Common Space-Invariant Blur," IEEE Trans. Image Processing Vol.10.pp 1187-1193, 2001.

[13] S. Farsiu and MD. Robinson and M. Elad and P. Milanfar, "Fast and Robust Multiframe Super Resolution," IEEE Trans. Image Processing Vol.13. pp 1327-1344, 2004.

[14] S. Knorr and M. Kunter and T. Sikora, "Super-Resolution Stereo and Multi-View Synthesis from Monocular Video Sequences," 3DIM Sixth International Conference pp 55-64, 2007.