VIRTUAL VIEW RENDERING USING SUPER-RESOLUTION WITH MULTIVIEW IMAGES

Bumsub Ham, Dongbo Min, Jinwook Choi and Kwanghoon Sohn

Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
khsohn@yonsei.ac.kr

ABSTRACT

This paper presents a new approach to solve the problem of quality degradation of a synthesized view, when a virtual camera moves forward. Interpolation techniques using only two neighboring views are generally applied when a virtual view is synthesized. Because the size of an object increases when the virtual camera moves forward, conventional methods have usually addressed this problem by interpolation techniques 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 images in multiview camera configuration. That is, this problem is solved by applying super-resolution concept which reconstructs a high resolution image from several low resolution images. Data fusion is performed by geometric warping with disparity maps of the multiple images followed by deblurring. Experimental results show that the image quality can further be improved by reducing blurring and halo effects in comparison with the interpolation method.

Index Terms— IBR (Image based rendering), 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 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 [1]. 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 a 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 [2] [3] [4].

Min et al proposed a view synthesis method considering motion parallax effect when the virtual camera moves forward [5]. 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 blurred. In multiview configuration, we prevent the quality of a virtual view from being degraded when the virtual camera moves forward. We apply super-resolution(SR) concept to IBR in order to prevent a synthesis view from being degraded. SR has been studied extensively in last two decades due to the limited sensor size and cost [6]. Many solutions have been developed: maximum- likelyhood method (ML), maximum a-posteriori methods (MAP), projection onto convex sets (POCS) and nonlinear interpolation method [7]. Our proposed 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 can be classified into two cases as follows:

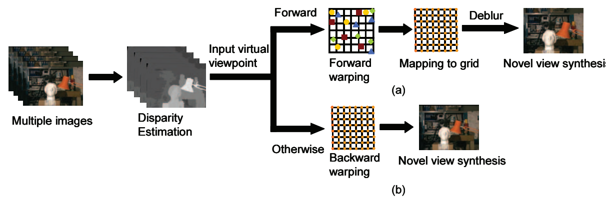


Fig. 1. Schemes for synthesizing a virtual view. (a) a virtual camera moves forward. (b) a virtual camera moves backward.

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 a previous method that is based on backward warping, holes are interpolated using only two neighboring views [5]. In addition, if we use multiple images in contrast to a previous method, backward warping has some problems as shown in Fig. 3. Therefore, a synthesized view is degraded. This is the reason why we use forward warping as shown in Fig. 1(a).

A view is synthesized by interpolation using two neighboring cameras in the previous method [5]. However, it makes objects blur and produces artifacts. Seeing this scene, a user feels uncomfortable, and loses the details of the scenes. In multiview configuration, we have several scenes captured at different positions. Thus, 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.

978-1-4244-5654-3/09/$26.00 ©2009 IEEE

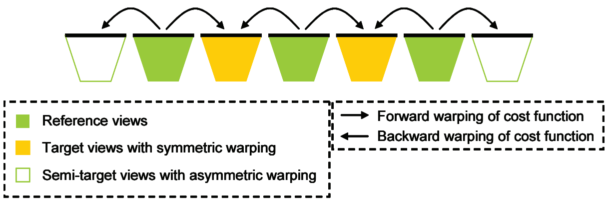


Fig. 2. semi N-view & N-depth framework.

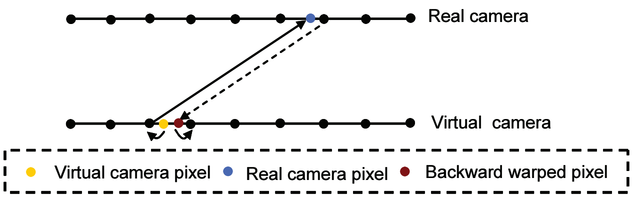


Fig. 3. The problem of backward warping.

It does not assume constant depth and fixed viewpoint in contrast to the conventional SR methods [8] [9] [10].

We propose a novel view synthesis method superior to the conventional method. Our algorithm consists of three stages: Disparity estimation, data fusion and deblurring. The main contribution of this paper is in the data fusion stage. 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. Experimental results are shown in section 3. Finally, in section 4, we conclude with a summary.

2. NOVEL VIEW SYNTHESIS

2.1. Stereo Matching

We estimate disparity by semi N-view & N-depth method proposed in [5]. In multiview camera configurations, given N views, we require N depth images 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 disparity maps of neighboring images are generally similar to each other, except in occlusion regions. Thus, we can reduce the complexity in N-view & N-depth framework by warping cost function from the reference view to the target view as shown in Fig. 2.

2.2. Geometric based forward warping

Conventional method is based on backward warping because forward warping suffers from a hole problem [5] [11] [12]. However, we directly use the pixels only that are from N-views instead of using the interpolated pixels. In this case, forward warping is more suitable than backward warping in that there is a possibility that a pixel may move slightly at inappropriate location as shown in Fig. 3 when a virtual camera moves forward. It leads to the problems at boundaries when the virtual view is synthesized.

Fig. 4 shows a movement of the virtual camera. Virtual camera moves along x-axis and z-axis, which are defined as glo_x and glo_z, respectively. 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. N-views are numbered in ascending order from left to right. Using the geometric information, that is, the disparity, we change coordinate value from image coordinate of N real cameras to that of the virtual camera. The relation between (x_i, y_i) in the i^{th} image and (x^v, y^v) in the synthesized view can be computed as follows:

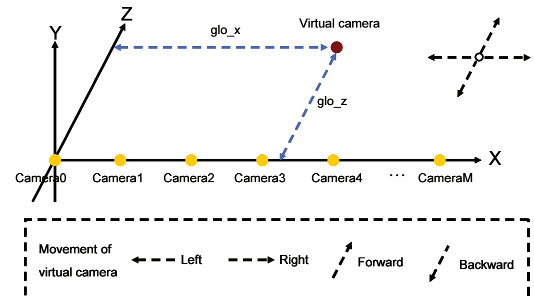


Fig. 4. Movement of virtual camera.

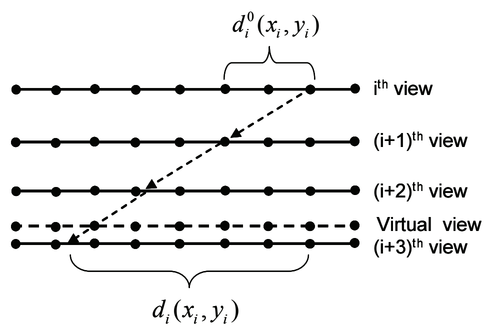


Fig. 5. Disparity refinement in multiview images

$$\begin{aligned} x^v - x_0 &= f\frac{(x_i-x_0)B_i/d_i(x_i,y_i)+T_{i,x}}{fB_i/d_i(x_i,y_i)+T_{i,z}} = \frac{(x_i-x_0)+d_i(x_i,y_i)\alpha_{i,x}}{1+d_i(x_i,y_i)\alpha_{i,z}/f} \\ y^v - y_0 &= f\frac{(y_i-y_0)B_i/d_i(x_i,y_i)+T_{i,y}}{fB_i/d_i(x_i,y_i)+T_{i,z}} = \frac{(y_i-y_0)+d_i(x_i,y_i)\alpha_{i,y}}{1+d_i(x_i,y_i)\alpha_{i,z}/f} \end{aligned} \tag{1}$$

where f and B represent a focal length and a base line distance between cameras, respectively. (x_0, y_0) is the center of the image plane and $d_i(x_i, y_i)$ is a disparity between camera i and $\lceil glo\_x \rceil$. 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 between 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. (2-1) (Eq. (2-2)) from left (right) side views 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. 5. This operation is independent of the other view, so it is possible to run in parallel.

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

The problem of forward warping is that mapping is not one-to-one correspondence. Therefore, there is a problem that the background and foreground pixels 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 an intensity to the virtual camera coordinate. Let I_i and I^v be the i^{th} reference image and a synthesized image, then $I^v(x^v, y^v) = I_i(x_i, y_i)$. However, this problem still occurs as shown in Fig. 6 although we prevent a disparity from penetrating into the foreground region. Fig. 7 shows the reason why

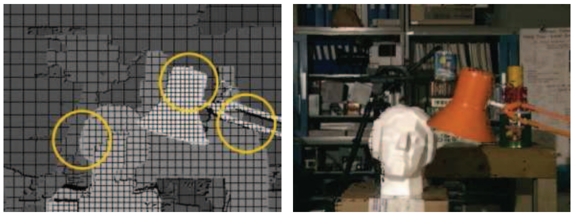


Fig. 6. Problem: Background disparity penetrates into the foreground region.

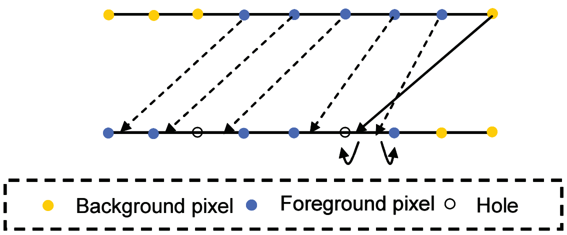


Fig. 7. Problem of disparity warping.

this problem occurs. When both background and foreground disparities are warped in similar virtual camera coordinate, we can select foreground pixel 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 fill a hole, some pixels are not assigned to 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. 7. In order to address this problem when the disparity is warped to coordinate of the virtual view, we determine whether the disparity is similar to neighboring disparities which is already filled. If they are not similar enough, the warped disparity and and intensity are eliminated.

2.3. Hole filling using backward warping

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 them. Thus, we apply backward warping with the interpolated disparity map ($Dis_i(x^v, y^v)$). Interpolated disparity map only used to change coordinate of the virtual camera to that of the real camera as in Eq. (3).

$$\begin{aligned} x_i &= (x^v - x_0)(1 - f\alpha_{i,z}Dis_i(x^v, y^v)) + x_0 \\ &\quad + \alpha_{i,x}Dis_i(x^v, y^v) \\ y_i &= (y^v - y_0)(1 - f\alpha_{i,z}Dis_i(x^v, y^v)) + y_0 \\ &\quad + \alpha_{i,y}Dis_i(x^v, y^v) \end{aligned} \tag{3}$$

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. 8. The disparity of real camera is warped to the virtual camera coordinate and appropriately replicated so that a hole does not exist. We must consider occlusion problem when a view is synthesized by backward warping. We do not have to consider this problem when forward warping is applied, because we select a large disparity instead of a small one as mentioned earlier. Thus, we define visibility function. It shows that whether a pixel in the virtual view is visible in the reference view with values of 1 when visible. By using a visibility function of the virtual view, a hole is filled as in Eq. (4).

$$I^v(x^v, y^v) = I_i(x_i, y_i)V_i(x^v, y^v) \tag{4}$$

Finally, a synthesized view is deblurred slightly by shock filter because it has some artifacts such as noise. Shock filter eliminates noise and enhances edge effectively [13].

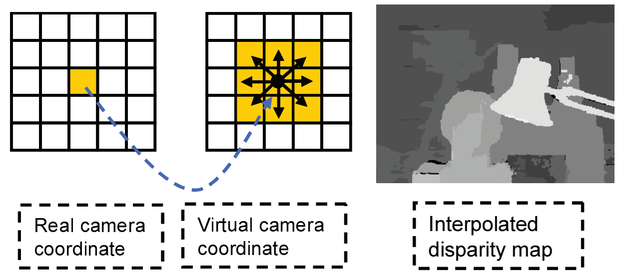


Fig. 8. How to make interpolated disparity maps.

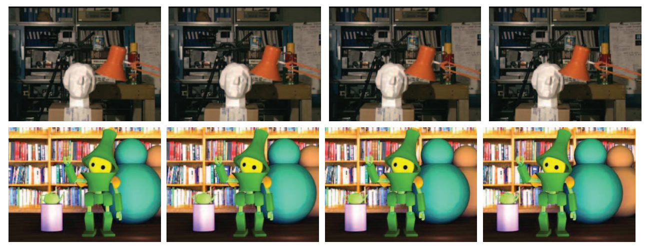


Fig. 9. Experiment data set : Tsukuba (Top)and Robot (Bottom)

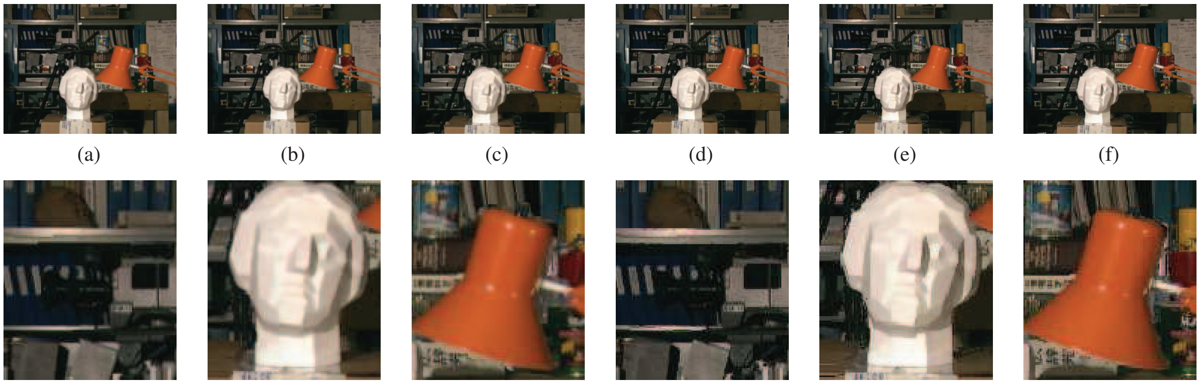
3. EXPERIMENTAL RESULTS

We used four reference views to synthesize a virtual view as shown in Fig. 9. Fig. 10 and 11 show the results of the proposed method and the conventional method. Fig. 10 and Fig. 11 (a)-(c) are the synthesized views of the conventional method. The more the virtual camera moves forward, the more artifacts such as blurring and halo effect are shown. (Please see the electrical version for better visibility.) However, our method shows sharp edges and no artifacts at boundary regions as shown in Fig. 10 and Fig. 11 (d)-(f). The second row of Fig. 10 (Fig. 11) shows cropped images of Fig. 10 (c) and Fig. 10 (f) (Fig. 11 (c) and Fig. 11 (f)), respectively. In background region, a little quality improvement was observed. In foreground region, however, we can see much better results. Fig. 12 shows another example of the synthesized view. We can see the improvement more clearly. The synthesized freeview images are available at [14].

The experiment was additionally performed. A scene was synthesized using four reference views by only applying backward warping such as in the conventional methods. However, we do not use interpolation technique in order to prevent a synthesized view from being blurred. Alternatively, we applied an algorithm which was used to fill holes in the proposed method. Fig. 13 (a) is the result for a backward warping method. Because interpolation technique was not used, there was little blurring. However, texture copying problem occurred as we showed in Fig. 3. (You can see it at the spout of a kettle.) Our method does not suffer from this problem as shown in Fig. 13 (b), because we applied both forward and backward warping properly. These experimental results show that quality improvement is due to not only using many reference views but also applying forward and backward warping properly. A performance of the proposed algorithm depends on disparity maps. Some outliers observed in the scene come from disparity error. Thus, we will investigate methods robust to the outliers as a further research.

4. CONCLUSIONS

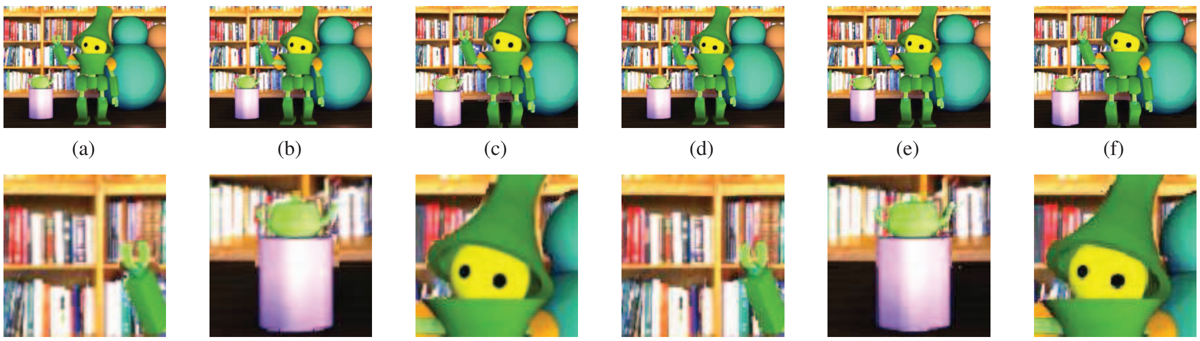
In this paper, we have proposed a novel view synthesis method that is not blurred when virtual camera moves forward. At first, we estimate a disparity by using semi N-view & N-depth framework. Each reference image is then forward warped geometrically. Because some holes exist, they are filled by backward warping using a visibility function followed by deblurring operation. We compare it with a conventional method. Experimental results show that our proposed



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Fig. 10. Results for view synthesis: Tsukuba sequence. (a)-(c) conventional method, (d)-(f) our method, (g)-(i) cropped images of (c), (j)-(l) cropped images of (f). (Please see the electrical version for better visibility.)



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

Fig. 11. Results for view synthesis: Robot sequence. (a)-(c) conventional method, (d)-(f) our method, (g)-(i) cropped images of (c), (j)-(l) cropped images of (f). (Please see the electrical version for better visibility.)

method is better than the conventional method in visual quality. As further work, we will investigate more elaborate data fusion which is robust to the outliers such as disparity error.

5. REFERENCES

[1] A. Kubota, K. Aizawa and T. Chen,"Reconstructing Dense Light Field From Array of Multifocus Images for Novel View Synthesis," *IEEE Trans. Image Processing*, vol. 16, pp. 269-279, 2007.

[2] SE. Chen and L. Williams,"View interpolation for image synthesis," *SIGGRAPH*, pp. 279-288, 1993.

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

[4] Y. Mori, N. Fukushima, T. Fujuu and M. Tanimoto, "View generation with 3D warping using depth information for FTV," *Signal Processing : Image Communication*, vol. 1-2, pp. 65-72, 2009.

[5] D. Min, D. Kim and K. Sohn, "2D/3D Freeview video generation for 3DTV System," *Proc. IEEE ICIP*, pp. 1760-1763, 2008.

[6] 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.

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

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

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

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

[11] M. Gong, JM. Selzer, 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.

(a) (b)

Fig. 12. Results for view synthesis: (a) conventional method , (b) our method.

(a) (b)

Fig. 13. Results for view synthesis: (a) backward based method , (b) our method.

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

[13] S. Osher and LI. Rudin, "Feature-Oriented Image Enhancement Using Shock Filters," *SIAM Journal on Numerical Analysis*, vol. 27, pp. 919-940, 1990.

[14] http://diml.yonsei.ac.kr/∼bumseop