Improved Novel View Synthesis from Depth Image with Large Baseline

Chia-Ming Cheng[1], Shu-Jyuan Lin[1], Shang-Hong Lai[1], Jinn-Cherng Yang[2]
Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan[1]
Industrial Technology Research Institute, Taiwan[2]
cmcheng@cs.nthu.edu.tw

Abstract

In this paper, a new algorithm is developed for recovering the large disocclusion regions in depth image based rendering (DIBR) systems on 3DTV. For the DIBR systems, undesirable artifacts occur in the disocclusion regions by using the conventional view synthesis techniques especially with large baseline. Three techniques are proposed to improve the view synthesis results. The first is the preprocessing of the depth image by using the bilateral filter, which helps to sharpen the discontinuous depth changes as well as to smooth the neighboring depth of similar color, thus restraining noises from appearing on the warped images. Secondly, on the warped image of a new viewpoint, we fill the disocclusion regions on the depth image with the background depth levels to preserve the depth structure. For the color image, we propose the depth-guided exemplar-based image inpainting that combines the structural strengths of the color gradient to preserve the image structure in the restored regions. Finally, a trilateral filter, which simultaneous combines the spatial location, the color intensity, and the depth information to determine the weighting, is applied to enhance the image synthesis results. Experimental results are shown to demonstrate the superior performance of the proposed novel view synthesis algorithm compared to the traditional methods.

1. Introduction

The 3DTV display [1] provides the observer with realistic 3D viewing experience. Conventional stereoscopic displays deliver two video streams for left and right eyes, respectively, and the parallax generates the sense of depth. More recently, the autostereoscopic display extends the content to wider baseline and more viewpoints, up to 21 distinct views. To simultaneously deliver so many video streams requires extremely large bandwidth. A more compact format, video plus

(a) (b) (c)

Figure 1. Illustrations of the proposed method compared with previous methods. (a) DIBR. (b) image inpainting; (c) our solution.

depth, uses a single stream and the corresponding depth map, and the device-independent format enables to synthesize multiple views in real-time.

For the new view synthesis, the DIBR systems [2][3] established a framework that consists of the following three steps: preprocessing of the depth image, image warping, and hole filling. The first step can shrink the holes after image warping. The second step computes the displacement from the depth and warps the image to the new viewpoint. The last step is to restore the holes, i.e. the disocclusion regions. Based on this framework, related techniques can be divided into two approaches. The first approach is the pre-filtering of the depth image, such as average filtering, Gaussian filtering [3], and asymmetric filtering [4], to eliminate or reduce the holes. This approach has the advantage of simple implementation, efficiency, and smooth synthesized image. However, the over-smoothing of the depth image usually leads to unrealistic sense of depth, especially for the disocclusion region for the boundary with sharp depth

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

changes when the new viewpoint corresponds to large baseline.

The other approach is to recover the disocclusion regions. The linear interpolation and extrapolation has been widely accepted as a straightforward and efficient way for filling in small holes, but artifacts appear more obvious when the holes become larger. Chen et al. [5] proposed an edge-dependent interpolation to preserve the edge structure to reduce the artifacts. In addition, image inpainting [6] has been demonstrated as one of the most powerful tools for interpolating holes in an image. However, directly applying the existing methods may also cause artifacts, as shown in figure 1 (b). The main reason is that the disocclusion regions normally involve sharp depth changes, but the conventional image inpainting techniques only consider the planar image information.

The demand of wide-baseline view synthesis for 3DTV systems, which is an inevitable trend for future development, is the main motivation for this research work. We integrate the depth information with the image intensity for restoring the disocclusion regions. The first step of the proposed algorithm is to estimate the depths for the pixels inside the holes. Then the depth-guided image inpainting is applied to the color image for texture synthesis in the disocclusion regions. Lastly, we employ a trilateral filter that incorporates the spatial, intensity, and depths to enhance the final result. In our experiments, we apply the proposed algorithm to the image-plus-depth examples with fine depth map captured by depth sensor and with rough depth map recovered by stereo matching and show its superior performance compared to previous methods.

2. Preliminary processing

2.1 Depth Image Preprocessing by Bilateral Filtering

Before introducing the main techniques to recover the depth and color images, let us illustrate the preliminary process to obtain the input, i.e. the warped depth image and the warped color image. The process consists of two major steps, i.e. preprocessing of the depth map and image warping.

For the preprocessing step, we employ the bilateral filter [7] to enhance the depth image. Compared with the conventional averaging or Gaussian filter, the bilateral filter preserves the sharp depth changes in conjunction with the intensity variation, which results in consistent boundary of the distinct depth levels between the depth and color images. Moreover, the smoothing of depths in the neighborhood of similar color pixels alleviates the sensitivity of depth noises on the warped image.

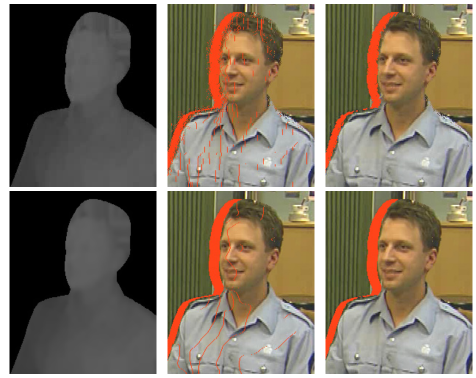

Figure 2. The preliminary process. The first column shows the depth images; the second column is the warped image; the last column presents the result after filling the small holes. The first row shows the results from origin depth map, and the second row demonstrates the results from the pre-processed depth map.

Figure 2 depicts an example for this case. The bottom row shows the results from the depth map after the preprocessing with the bilateral filtering, compared to the top row, which shows the results from the original depth map. Through this example, it is clear to see the improvement on the boundary between foreground and background as well as the reduction of depth noise effect to the hair region of the warped image.

The bilateral filter associates the spatial locations and intensity variation in a local window, formulated as follows,

$$B(u+\delta u, v+\delta v) = f_s(\delta u, \delta v) \cdot f_I\left(I(u,v), I(u+\delta u, v+\delta v)\right) \quad (1)$$

where (u,v) denotes the center of a local sliding window in image coordinate, and $(\delta u, \delta v)$ indicates the displacement vector of the local window relative to the center. The first term denotes the spatial locations that contribute to the center in terms of the geometric distances:

$$f_s(\delta u, \delta v) \equiv \exp\left(-\frac{1}{\sigma_s}\sqrt{\|\delta u\|^2 + \|\delta v\|^2}\right) \quad (2)$$

The second term indicates the intensity variation in color space, designed as follows,

$$\begin{aligned} & f_I\left(I(u,v), I(u+\delta u, v+\delta v)\right) \\ & \equiv \exp\left(-\frac{1}{\sigma_I}\sqrt{\sum_{c\in\{R,G,B\}} \left\|I_c(u+\delta u, v+\delta v) - I_c(u,v)\right\|^2}\right) \end{aligned} \quad (3)$$

where I_c denotes the color in RGB space with each channel normalized to [0,1]. Note that the standard deviation (σ_s, σ_c) balances the two terms. For all of our experiments given in this paper, we apply the bilateral filter with a 5×5 window size and the setting $(\sigma_s = 1.0, \sigma_c = 0.25)$.

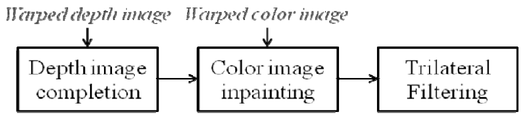

Figure 3. The flow diagram of the proposed view synthesis algorithm for the disocclusion regions.

2.2 Image warping

The image warping is to transform the pixel-wise depth into the corresponding displacement for a novel viewpoint, and reproduce the color in the new image. This image warping is based on the new parallel viewpoints [8] and can be formulated as follows:

$$s_k = s_0 + k\frac{\alpha Z}{d+Z} \quad (4)$$

where s_k and s_0 denote the horizontal coordinates of the new and original viewpoints, respectively, k is a signed number with the positive/negative values indicating the degrees of right/left shift of the new viewpoint, α integrates the length of baseline and the unit transformation from real world to the display, and d denotes the viewing distance between the observer and the screen. Note that the depth Z is obtained from the pre-processed depth map. We refer the details of this warping to [8].

2.3 Filling in the small holes

We define the small holes as the ones that can be filled by a 3×3 median filter [9]. That is, a vacant pixel is filled with the median value of the surrounding filled pixels, and it remains empty when more than half of the surrounding pixels are still vacant. The last column of Figure 2 depicts an example of filling in the small holes.

3. Image restoration for the disocclussion regions

From the warped depth and color images, we first recover the disocclusion regions of the depth image. Then we propose the depth-guided inpainting for the color image. Lastly, we apply a trilateral filtering to enhance the final result. Figure 3 shows the flow diagram of the proposed method, and the following subsections describe the details.

3.1 Depth Image Restoration

The disocclusion region occurs around the boundary of sharp depth changes due to the occluded regions reappear from the new viewpoint. Therefore, the criterion used to recover the depth is to extend the occluded depth, i.e. the background, to the disocclusion regions. In addition, to preserve the depth structure, we further employ the image inpainting [10] to propagate the structure of the depth image for complicated background.

Figure 4. Depth-guided inpainting after 100, 500, 1000 iterations.

3.2 Depth-Guided Color Image Inpainting

For recovering the disocclusion regions of the color images, we develop the depth-guided exemplar-based algorithm. The depth information constraints the search ranges for the texture matching. On the other hand, the color image provides the information to determine the priority for preserving the image structure.

The candidates to be filled in are the vacant pixels connected to the background, and their structure strengths are defined by the partial image gradients of a local window, e.g. a 5×5 window B in our examples, from the surrounding pixels, formulated into the vertical and horizontal components as follows,

$$\bar{g}_{\perp} \equiv \sum_{(\delta u,\delta v)\in B} \rho(\delta u,\delta v)\,\bar{\omega}_{\perp}(\delta u,\delta v)\,\bar{g}_{\perp}(\delta u,\delta v) \quad (5)$$

$$\bar{g}_{//} \equiv \sum_{(\delta u,\delta v)\in B} \rho(\delta u,\delta v)\,\bar{\omega}_{//}(\delta u,\delta v)\,\bar{g}_{//}(\delta u,\delta v) \quad (6)$$

where ρ is a Boolean variable to indicate pixel vacancy, i.e. 1 for filled pixel and 0 for vacant pixel. The weights are proportional to the inverse geometric distance, as following:

$$\bar{\omega}(\delta u,\delta v) \equiv \frac{(\delta u,\delta v)}{\sqrt{\|\delta u\|^2+\|\delta v\|^2}} \quad (7)$$

The magnitude of the strength is $\left(\|\bar{g}_{\perp}\|^2+\|\bar{g}_{//}\|^2\right)^{1/2}$ and the orientation is $\tan^{-1}\left(\bar{g}_{\perp}/\bar{g}_{//}\right)$. Iteratively, the algorithm determines the maximal strength of the candidates as the first priority to be filled, and search the best match according to the orientation in the background regions. Figure 4 gives an example to depict the intermediate results of different iterations.

3.3 Trilateral Filtering

The trilateral filter utilizes the spatial and depth information to filter the color image that enhances the final result of the synthesized image. Compared to equation (1), this filter has an additional term, i.e. depth, and is formulated as follows,

$$T(u+\delta u,v+\delta v)=f_s(\delta u,\delta v)\cdot f_I\left(I(u,v),I(u+\delta u,v+\delta v)\right)\cdot f_d\left(d(u,v),d(u+\delta u,v+\delta v)\right) \quad (8)$$

The depth term is defined as follows,

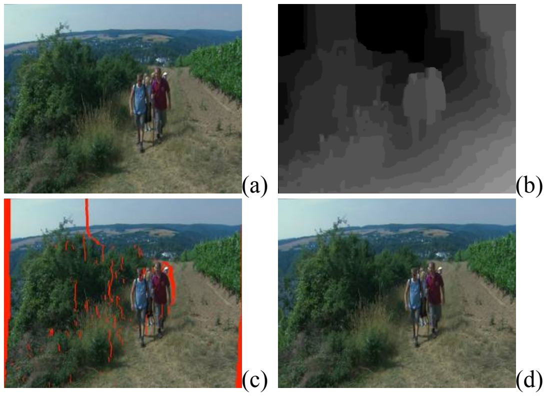

Figure 5. The proposed method applied to the depth image obtained by stereo matching. (a) The color image. (b) The depth image. (c) The disocclusion regions from the new viewpoint. (d) The final result.

$$f_d\left(d\left(u,v\right),d\left(u+\delta u,v+\delta v\right)\right) \equiv \exp\left(-\frac{1}{\sigma_d}\left\|d\left(u+\delta u,v+\delta v\right)-d\left(u,v\right)\right\|\right) \quad (9)$$

Note that we denote the normalized depth value in range of [0,1]. For all of our experiments shown in this paper, we applied the trilateral filter with a 7×7 window size and the parameter setting $(\sigma_s = 1.0, \sigma_c = 0.25, \sigma_d = 0.15)$.

4. Experimental Results

For subjective assessment of the novel view synthesis, we compare the proposed algorithm with the conventional DIBR [3] and fast image inpainting [10] and show the results in Figure 1 and 6. Figure 5 gives the result by applying the proposed algorithm to the image associated with the depth map obtained by stereo matching algorithm. For quantitative analysis, we used the stereo image shown in Figure 6 to calculate the PSNR of the synthesized images, as shown in Table 1.

Table 1. The PSNR of the restored synthesized image as shown in Figure 6.

Method	*DIBR [3]*	*Inpainting [10]*	*Before tri-filtering*	*Trilateral filtering*
PSNR	29.2869	29.0043	29.1884	29.7571

4. Conclusions

In this paper, we presented an improved solution for novel view synthesis of large baseline based on DIBR systems for 3DTV. Three main techniques, i.e. the bilateral filtering for the preprocessing of the depth image, the depth-guided color image inpainting, and the trilateral filtering, were developed to improve the quality of the synthesized image from a new viewpoint. The experimental results for both objective and subjective assessment demonstrate the superior

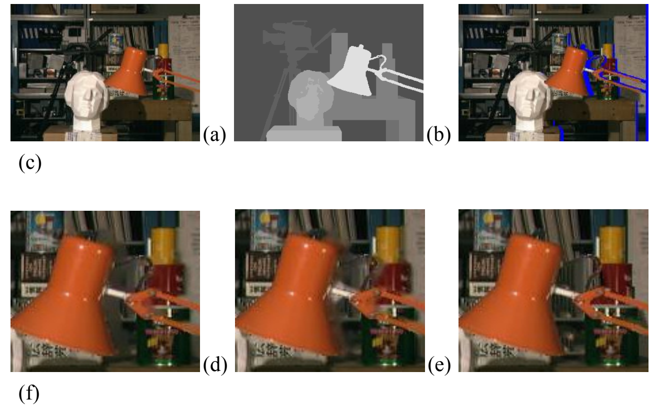

Figure 6. The synthesized novel views for objective comparisons. (a) Input left image, (b) input depth image, (c) the warped image of a new view, and the final synthesized images by (d) DIBR interpolation, (e) image inpainting, and (f) the proposed method.

performance of the proposed algorithm compared to the previous methods.

References

[1] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. Kopylow, "A survey of 3DTV displays: techniques and technologies", IEEE Trans. on circuits and systems for video technology, 2007, 17(11), pp. 1647-1658.

[2] A. Redert, M. Op de Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, and P. Surman. "ATTEST- advanced three-dimensional television system techniques", IEEE Proc. 3DPVT, 2002, pp. 313-319.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), the compression and Transmission for s New Approach on 3DTV", SPIE Proc. Stereoscopic Displays and Virtual Reality Systems XI, 2004, pp. 93-104.

[4] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3DTV", IEEE Tranc. broadcasting, 2005, 51(2), pp. 191-199.

[5] W.-Y. Chen, Y.-L. Chang, S.-F. Lin, L.-F. Ding, and L.-G. Chen, "Efficient depth image based rendering with edge dependent depth filter and interpolation", IEEE Proc. Multimedia and Expo, Jul 2005, pp. 1314-1317.

[6] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting", ACM Proc. SIGGRAPH 2000, pp.417-424.

[7] C. Tomasi and R. Manduchi. "Bilateral filtering for gray and color images", IEEE Proc. ICCV, 1998, pp. 839-846.

[8] R-P.M. Berretty, F.J. Peters, and G.T.G. Volleberg. "Real time rendering for multiview autostereoscopic", SPIE Proc. Stereoscopic Displays and Virtual Reality Systems XIII, 2006, V.6055, pp.208-219.

[9] T. S. Huang, G. J. Yang, and G.Y. Tang, "A fast two-dimensional median filtering algorithm", IEEE Trans. Acoustics, Speech, and Signal Processing, 1979, 27(1):13-18.

[10] M. M. Oliveira, B. Bowen, R, McKenna, and Y.-S. Chang, "Fast Digital Image Inpainting", Proc. Inter. Conf. Visualization, Imaging and Image Processing, 2001, pp. 261-266.