Shape Preserving RGB-D Depth Map Restoration

Wei Liu[1], Haoyang Xue[1], Yun Gu [1], Qiang Wu[2], Jie Yang[1], and Nikola Kasabov[3]

[1]The Key Laboratory of Ministry of Education for System Control and Information Processing
Shanghai Jiao Tong University, Shanghai, China
[2]School of Computing and Communications, University of Technology, Sydney, Australia
[3]Knowledge Engineering and Discovery Research Institute, Auckland University of Technology Auckland, New Zealand

Abstract. The RGB-D cameras have enjoined a great popularity these years. However, the quality of the depth maps obtained by such cameras is far from perfect. In this paper, we propose a framework for shape preserving depth map restoration for RGB-D cameras. The quality of the depth map is improved from three aspects: 1) the proposed *region adaptive bilateral filter* (RA-BF) smooths the depth noise across the depth map adaptively, 2) by associating the color information with the depth information, incorrect depth values are adjusted properly, 3) a *selective joint bilateral filter* (SJBF) is proposed to successfully fill in the holes caused by low quality depth sensing. Encouraging performance is obtained through our experiments.

Keywords: depth map restoration, joint bilateral filter, diffusion, Kinect

1 Introduction

Recent years, growing attention has been paid to the RGB-D images. In particular, the great success of low cost structured-light camera such as Kinect [1] has brought lots of RGB-D based applications like gaming [2], and new research area such as object recognition [9].

However, due to the simple depth measuring mechanism, the quality of the obtained depth map is far from perfect and mainly suffers from three problems: 1) *Invalid pixels* which do exist but are not sensed by the depth sensor, i.e. zero depth values or close to zero. In our work, a pixel is classified as an invalid pixel once its depth value is zero. The rest of the pixels on the depth map are *valid pixels*. Invalid pixels always form "holes" (black regions) on the depth map. 2) *Region various noise* on the original depth map-the noise on the area close to depth edges is much heavier than that away from the depth edges. We use *depth edges* to denote the edges on the depth map. Though the noise normally

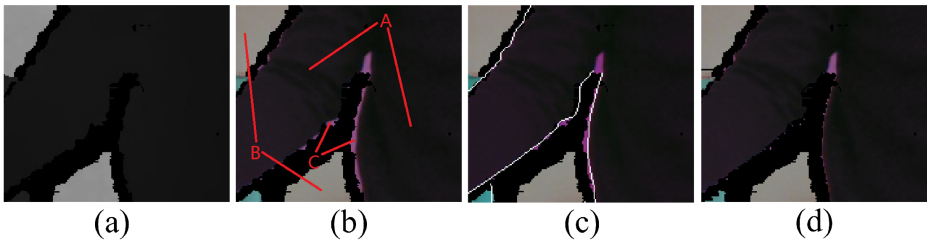

Fig. 1. (a) original depth map, (b) and (c) incorrect pixels (in pink) illustration, (d) incorrect pixels corrected by the proposed method.

follows the quadratic law both in theory and experiments [11] [5], in our work, it is shown that region various noise is more obvious and this is also confirmed by [14]. 3) *Incorrect pixels* which have incorrect depth values (different from noise). Incorrect pixels exist along some regions (but not all the regions) of depth edges on the original depth map. Figure 1(b) illustrates incorrect pixels: regions in black labeled with "A" are foreground, regions in grey labeled with "B" are background, and regions in pink labeled with "C" are fake foreground, i.e. incorrect pixels.

The goal of the restoration is to restore a noise free depth map while the holes are properly filled in without altering the shape of the objects on the depth map. Depth map restoration has two categories: 1) restoration based on the depth information only and 2) restoration based on both the depth and color information. For each category, the operation can be carried out based on the information of current frame or based on the information of multiple frames. Restoration in [10] utilized a cascade of two modified median filters based on the depth information of current frame, while the method in [8] restored the depth map with normalized convolution and the guided filter taking the depth information of multiple frames into account. However, both methods in [10] [8] produced results of limited quality because only depth information was considered. Methods in [14] [15] [3] [13] took both the depth information and color information of current frame into account for restoration and promising results were shown in their papers. Especially, the method in [14] introduced the concept of depth layer and produced results of state of art performance. Methods in [5] [6] [4] took the depth and color information of multiple frames into account for the restoration. However, these methods were mainly designed for scenes with static background and dynamic foreground. Though the approach in [12] based on the motion analysis and the non-causal spatial-temporal median filter could handle dynamic scene, it is time consuming and cannot produce accurate restoration. The hole issue appearing on depth map is mainly caused by invalid pixels. There are two kinds of holes. The first kind is the *small holes* which can be properly filled in by considering the depth information in the neighboring area and the corresponding color information provided by the RGB-D camera. The second kind is the *large holes* which normally appear along the regions of

depth edges on the original depth map. Applying the methods in [15] [3] [13] for filling in such holes may cause jagged or blurring depth edges or extra incorrect pixels on the restored depth map.

Previous work more focused on smoothing the noise and properly filling the holes [10] [8] [15] [3] [13]. However, all these methods above could not well maintain the object shape during the process of noise smoothing and hole filling. Moreover, these methods did not have a solution to correct depth values of incorrect pixels.

To tackle the constraints in the current methods, this paper contributes a new framework for depth map restoration, which considers not only depth information plus color information but also the depth discontinuity information in order to preserve the object shape. First, we propose a *region adaptive bilateral filter* (RA-BF) to smooth the noise. Then we correct the error depth values of the incorrect pixels with the help of both the depth and color information. Finally a novel *selective joint bilateral filter* (SJBF) is proposed to properly fill in the holes.

2 The Proposed Method

Our method consists of three steps: region various noise smoothing followed by incorrect pixels correction plus depth discontinuity map refinement, and finally holes filling.

In the following sections, we use *depth discontinuities* to denote the positions in the real world where distance between the objects and the depth sensor changes. The map of depth discontinuities is different from the map of depth edges defined in Section 1 mainly because of the holes and the incorrect pixels. The depth edge map of a completely accurate depth map is the same with the depth discontinuity map. In fact, we use depth discontinuities to refer the edges on the depth discontinuity map in the following sections. Uppercase letter with a subscript denotes either the pixel or the value of the pixel at the position indexed by the subscript. The uppercase letter with a hat and a subscript denotes the evaluated value of the pixel at position indexed by the subscript.

2.1 Region Various Noise Smoothing with Region Adaptive Bilateral Filter (RA-BF)

In our work, noise smoothing is carried out only on the regions of valid pixels on the original depth map. This operation is formulated as Equation (1):

$$\hat{D}_i = \frac{1}{Z_i} \sum_{D_j \in \mathcal{N}_i} D_j \cdot e^{-\left(a \cdot \frac{|D_i - D_j|^2}{f(\theta)} + b \cdot |i-j|^2\right)} \tag{1}$$

where $\mathcal{N}_i$ ($D_j \in \mathcal{N}_i$) is the valid pixels set in the $w \times w$ patch on the original depth map , in which the center is D_i, and Z_i is a normalization constant which is the sum of the coefficient of D_j in Equation (1), and a, b are also constant

values, the region adaptive term $f(\theta)$ is designed to consider the region various noise where θ is the perpendicular distance of D_i to its nearest depth edge on the original depth map. $f(\theta)$ is defined as Equation (2):

$$f(\theta) = \begin{cases} C_1, & \theta \leq r; \\ C_2, & otherwise \end{cases} \tag{2}$$

where constants $C_1 > C_2$. This means the smooth strength for D_i is larger if D_i is closer to the depth edges. Otherwise, the smooth strength is smaller. This can be implemented by firstly dilating the edge map obtained from the original depth map with a *disk* basic element of radius r. Then the pixels within the dilated edges are smoothed with $f(\theta) = C_1$. Otherwise, we set $f(\theta) = C_2$.

2.2 Incorrect Pixels Correction & Depth Discontinuity Map Refinement

Depth discontinuities can be described using edge information on the depth map. However, edges on the original depth map cannot well describe the depth discontinuities mainly due to the holes and incorrect pixels on the depth map. Figure 2(b) shows the edge map obtained from the original depth map. According to [15] [3] [7], depth discontinuities often simultaneously appear at the same locations on a depth map and the corresponding color image. We initialize the depth discontinuity map as follows: the edge map obtained on the depth map is processed by dilate operation, and the output is combined with the edge map of the corresponding color image through AND binary operation. The result of AND operation is regarded as the initial depth discontinuity map. Figure 2(c) shows an initial depth discontinuity map. It is shown that some edge points on the initial depth discontinuity map are not correct. We call these incorrect edge points *introduced texture edges*. It is observed that introduced texture edges exist inside flat depth areas. We design Equation (3) to remove these fake depth discontinuities.

$$g(E_i) = sgn\left(\max_{j \in \Omega_i} \left|D_j^{\Delta}\right| - T_{\Delta}\right), i \in \Lambda \tag{3}$$

where Λ is the set of coordinates on the initial depth discontinuity map where the values of the binary pixels equal one. E_i represents the pixel at position i on the initial depth discontinuity map. $sgn(\cdot)$ is a sign function where $sgn(x) = 1$ for $x \geq 0$ and $sgn(x) = 0$ for $x < 0$. T_{Δ} is a given threshold. D_j^{Δ} is the Laplacian of the smoothed depth map which is obtained from Section 2.1. Ω_i is the neighboring area of position i. In fact, Ω_i is a straight line that is perpendicular to the edge on the initial depth discontinuity map and across the position i. The length of $\Omega_i$is determined by$[-R, R]$.$-R$ and R mean we consider D_j^{Δ} in Ω_i on two sides of the edge.

We compute Equation (3) for all E_i $(i \in \Lambda)$ on the initial depth discontinuity map pixel by pixel. If $g(E_i) = 0$, E_i belongs to the introduced texture edge and is removed. If $g(E_i) = 1$, E_i belongs to the depth discontinuity and is kept. And then we get the *refined depth discontinuity map*. The refined depth discontinuity

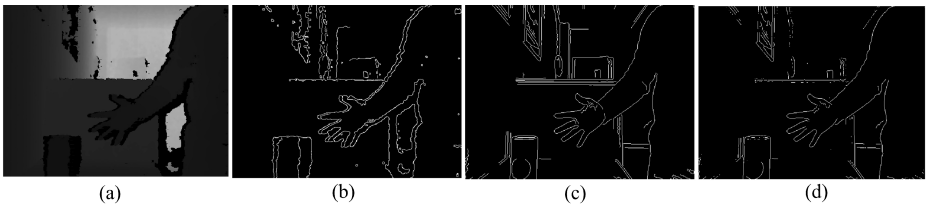

Fig. 2. (a) original depth map, (b) depth map discontinuity map, (c) initial depth discontinuity map, (d) refined depth discontinuity map.

map is regarded as an approximation of the ideal depth discontinuity map. Figure 2(d) shows the refined depth discontinuity map. It is shown that most introduced texture edges have been removed.

For all $g(E_i) = 1$, if there exists incorrect pixels around the position i (incorrect pixels only exist along *some* regions of depth edges), we further correct the incorrect pixels on the smoothed depth map obtained in Section 2.1. Figure 1(c) shows incorrect pixels (regions in pink) when we draw the refined depth discontinuity map on the smoothed depth map. If we have $k = \arg\max_{j \in \Omega_i} |D_j^{\Delta}|$, then the correction will be performed as $D_s \leftarrow D_k$ for all $s \in \Omega_i$, and s is between i and k. D_s is the depth value of the pixel on the smoothed depth map at the same position where D_s^{Δ} lies. In this way, we refine the initial depth discontinuity map and obtain a depth map with the incorrect pixels corrected at the same time. The obtained depth map is denoted as *refined depth map*. Figure 1(d) illustrates the refined depth map of Figure 1(a).

2.3 Holes Filling with Selective Joint Bilateral Filter (SJBF)

We fill the holes with the help of color information as well as the refined depth discontinuity map obtained in Section 2.2. Each time we only fill the invalid pixels with at least one valid pixel in the 8-neighborhood. This is to gradually diffuse the valid pixels into the holes . We start the diffusion at all holes simultaneously. A diffusion process at a hole area is terminated once it meets either valid pixels or the edge defined in the refined depth discontinuity map. The whole diffusion process will be terminated till all the diffusion processes meet valid pixels or the edges. The filling is implemented as Equation (4):

$$\hat{D}_i = \frac{1}{Z_i'} \sum_{D_j \in \mathcal{SN}_i} D_j \cdot e^{-\left(c \cdot \sum_{z \in C} \left|I_i^z - I_j^z\right|^2 + d \cdot |i-j|^2\right)} \tag{4}$$

where Z_i' is a normalization constant which is the sum of the coefficient of D_j in Equation (4), C is the index set of color channels in RGB color space and I_i^z represents the value of the z channel of pixel i, the *selected neighbors* $\mathcal{SN}_i$ of invalid pixel D_i is the set of valid pixels in the $w \times w$ patch on the depth map,

in which the center is D_i. $\mathcal{SN}_i$ is defined as Equation (5):

$$\mathcal{SN}_i = \left\{ D_j \left| e^{-\left(\tilde{D}_i - D_j\right)^2} \geq T \right. \right\} \tag{5}$$

where$\tilde{D}_i$ is the average value of the valid depth pixels in the 8-neighborhood of D_i, $T \in [0, 1]$ is a given threshold that represents how similar are the valid pixels in $\mathcal{SN}_i$ to D_i .

Unlike joint bilateral filter (JBF) which utilizes all the valid pixels in the $w \times w$ patch centered at D_i for filling, SJBF uses Equation (5) to *select* the neighbors in the patch first, i.e. the main difference between SJBF and JBF is the final neighbors used for hole filling. For the small holes which mainly lie on small flat regions on the depth map, SJBF is very similar to JBF because the neighbor selection based on Equation (5) seldom eliminates any valid pixels in the $w \times w$ patch because of similar depth on the small flat area. While for the large holes, SJBF can outperform JBF especially when the depth discontinuity passes through the patch and the corresponding colors of the valid pixels on two sides of the depth discontinuity are similar. Without neighbor pixels selection as in SJBF, JBF treats valid pixels on two sides of the depth discontinuity equally and causes blurring or jagged edge or even incorrect pixels on the restored depth map. However, the neighbor selection based on Equation (5) in SJBF can select proper neighbors: most valid pixels on the same side of the depth discontinuity with D_i are kept because of their similar depth values with $\tilde{D}_i$, while most of the valid pixels on the other side of the depth discontinuity are eliminated due to the obvious difference between their depth values and $\tilde{D}_i$. Thus the properly selected neighbors can successfully avoid the problems of JBF as mentioned above when filling in large holes.

3 Experiments

We compare our proposed method with [15] [3]. The data was produced by Kinect [1]. All the three methods restore the depth map with the help of the depth and color information of current frame. [15] only considered the noise and the invalid pixels while it did not take the incorrect pixels problem into account. [3] proposed a region-adaptive JBF taking the advantages of the edge information in the corresponding color image as structure guidance for adaptive support region selection. Figure 3 shows the experiment results. Figures in the first column are original depth maps. Their corresponding color images are shown in the second column. Figures in the third, fourth and fifth column show the restored depth maps by [3], by [15] and by our proposed method respectively. The first two rows in Figure 3 show results of two testing cases. The results by [3] can well preserve the shape of objects mainly due to taking advantages of the edge information in the corresponding color image. However, the restored depth map clearly 'copies' edges from the color image. In Figure 3(c1), clear depth edge exists between the ceiling and the wall on the left while there is no depth change in this region. The situation is the same with the door. Obviously, blurring edges exist in Figure 3

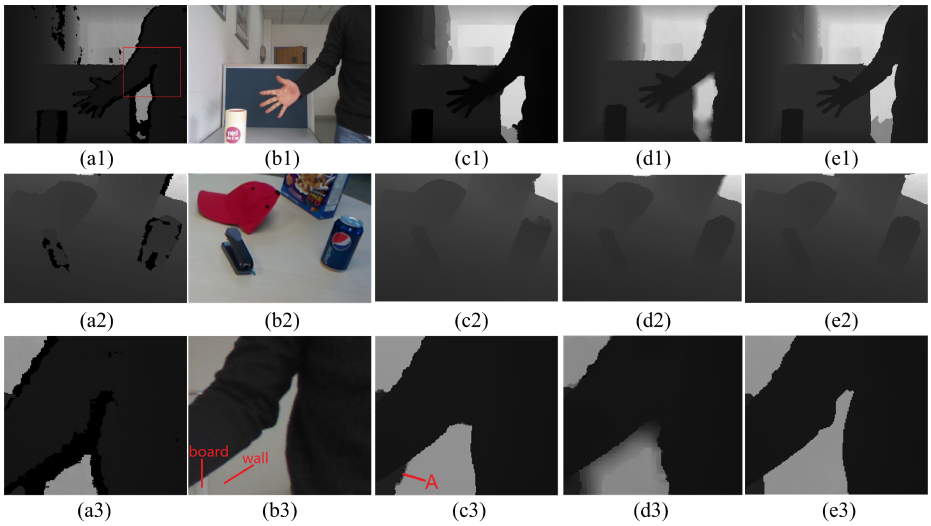

Fig. 3. original depth maps (column 1), corresponding color images (column 2), results by [10] (column 3), results by [9] (column 4), and results by our proposed method (column 5).

(d1) and (d2), while there are also jagged depth edges on the top of the board behind the hand in Figure 3 (d1) and around the hat in Figure 3 (d2). Both results by [3] and [15] introduce incorrect pixels in the region between the elbow and the chest and the region on the top of the beverage can.

To further explain the experiment results, we zoom in the region inside the red box shown in Figure 3(a1) and show them in the third row in Figure 3. The variants of JBF in [15] [3] only consider color information together with spatial correlation and do not select the neighbors when filling in the holes. As labeled in Figure 3(b3), pixels at two sides of the depth discontinuity of the wall and the board have similar colors. Thus the result by [3] introduces incorrect pixels (labeled with "A"), while the result by [15] has blurring edges on the restored depth map. Our SJBF can well handle this case because it takes the advantage of Equation (5) to select the neighbors first. Thus our result has no incorrect pixels like [3] and blurring edges like [15]. Results by [15] [3] also have extra incorrect pixels in the region between the elbow and the chest because they do not correct the incorrect pixels before the holes filling while our method does (illustrated in Figure 1(d)). It is shown that our result has much fewer incorrect pixels than the other two results. Additionally, the region various noise is also well smoothed by the proposed RA-BF according to our experiment results.

4 Conclusion

In this paper, we analyze the problems on the depth map obtained by the RGB-D camera: the holes formed by invalid pixels, the region various noise and the

incorrect pixels. RA-BF is proposed to smooth the noise. Then incorrect pixels are corrected and a refined depth discontinuity map is obtained at the same time . Finally the holes are properly filled in using SJBF with the help of the refined depth discontinuity map. The experiment demonstrates that the proposed method can greatly improve the quality of the original depth maps.

References

1. Microsoft corporation. kinect for xbox 360
2. A. Bleiweiss, e.a.: Enhanced interactive gaming by blending full-body tracking and gesture animation. In: ACM SIGGRAPH ASIA 2010 Sketches. p. 34. ACM (2010)
3. C. Chongyu, e.a.: A color-guided, region-adaptive and depth-selective unified framework for kinect depth recovery. In: Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on. pp. 007–012. IEEE (2013)
4. Camplani, M., Mantecon, T., Salgado, L.: Accurate depth-color scene modeling for 3d contents generation with low cost depth cameras. In: Image Processing (ICIP), 2012 19th IEEE International Conference on. pp. 1741–1744. IEEE (2012)
5. Camplani, M., Mantecón, T., Salgado, L.: Depth-color fusion strategy for 3-d scene modeling with kinect (2013)
6. Camplani, M., Salgado, L.: Efficient spatio-temporal hole filling strategy for kinect depth maps. In: IS&T/SPIE Electronic Imaging. pp. 82900E–82900E. International Society for Optics and Photonics (2012)
7. Diebel, J., Thrun, S.: An application of markov random fields to range sensing. In: Advances in neural information processing systems. pp. 291–298 (2005)
8. Jakob, W., Sebastian, B., Joachim, H.: Real-time preprocessing for dense 3-d range imaging on the gpu: defect interpolation, bilateral temporal averaging and guided filtering. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. pp. 1221–1227. IEEE (2011)
9. L. Kevin, e.a.: A large-scale hierarchical multi-view rgb-d object dataset. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on. pp. 1817–1824. IEEE (2011)
10. M. Andrew, e.a.: Enhanced personal autostereoscopic telepresence system using commodity depth cameras. Computers & Graphics 36(7), 791–807 (2012)
11. M. Fabio, e.a.: Geometric investigation of a gaming active device. In: SPIE Optical Metrology. pp. 80850G–80850G. International Society for Optics and Photonics (2011)
12. M. Sergey, e.a.: Temporal filtering for depth maps generated by kinect depth camera. In: 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011. pp. 1–4. IEEE (2011)
13. Q. Fei, e.a.: Structure guided fusion for depth map inpainting. Pattern Recognition Letters (2012)
14. Shen, J., Cheung, S.C.S.: Layer depth denoising and completion for structured-light rgb-d cameras. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1187–1194. IEEE (2013)
15. Y. Jingyu, e.a.: Depth recovery using an adaptive color-guided auto-regressive model. In: Computer Vision–ECCV 2012, pp. 158–171. Springer (2012)