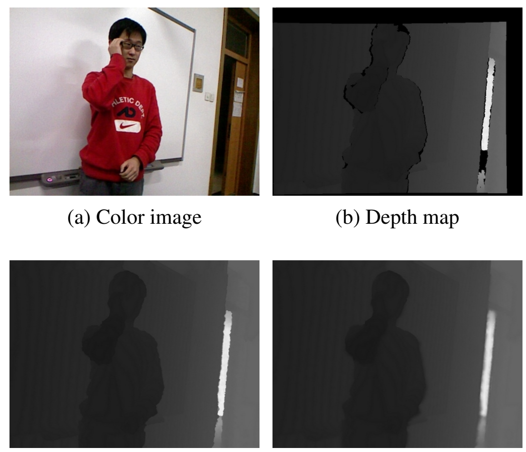
21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Guided Inpainting and Filtering for Kinect Depth Maps

Junyi Liu, Xiaojin Gong, Jilin Liu
Department of Information Science and Electronic Engineering
Zhejiang University, Hangzhou, Zhejiang, 310027, China
{*shshzaa, gongxj, liujl*}@*zju.edu.cn*

Abstract

Depth maps captured by Kinect-like cameras are lack of depth data in some areas and suffer from heavy noise. These defects have negative impacts on practical applications. In order to enhance the depth maps, this paper proposes a new inpainting algorithm that extends the original fast marching method (FMM) to reconstruct unknown regions. The extended FMM incorporates an aligned color image as the guidance for inpainting. An edge-preserving guided filter is further applied for noise reduction. To validate our algorithm and compare it with other existing methods, we perform experiments on both the Kinect data and the Middlebury dataset which, respectively, provide qualitative and quantitative results. The results show that our method is efficient and superior to others.



(a) Color image (b) Depth map

(c) Inpainting result (d) Filtering result

Figure 1: Our result of real-world Kinect data

1. Introduction

Depth information is quite useful for many computer vision applications. Among many depth sensors, the Kinect [7] gains a great success because of its low price and relatively high resolution. The Kinect sensor consists of an infrared light projector, a depth image CMOS sensor and a color image CMOS sensor. It captures real-time depth maps using a light coding technique. However, the device is unable to correctly estimate depth data in some cases due to the limit of working distance, occlusions, reflective surfaces, or relative surface angles. This leads to missing regions and unstable boundaries in depth maps. Fig. 1(a) and Fig. 1(b) show a typical pair of images captured by a Kinect, including a color image and its aligned depth map. The latter contains invalid depth regions near occluded boundaries and outside the door. Moreover, the alignment also causes large invalid areas along the image borders.

Although a lot of applications based on the Kinect have been released, research about the depth map refinement is deficient. Filling holes in depth maps is similar to the inpainting technique in optical images to a certain degree. Criminisi et al. [2] propose an image inpainting method using the exemplar-based texture synthesis and structure propagation. Telea [10] uses the fast marching method (FMM) to reconstruct damaged portions of an image. When applying the methods that are originally designed for optical images to depth maps, results are limited to only using the depth information to fill holes in a visually plausible way.

With the purpose of acquiring fine depth boundaries of objects, researchers also propose some methods taking advantage of color images. Some related work [1, 6] uses a joint bilateral filter [8] to enhance depth maps. Dolson et al. [3] present an accelerated high-dimensional bilateral filtering approach for upsampling sparse laser range data. In addition, the work of He et al. [4] demonstrates that a guided filter operates rapidly and avoids gradient reversal artifacts which may be caused by the bilateral filter. However, the depth enhancements only relying on filtering techniques do not work well when unknown regions are large.

978-4-9906441-1-6 ©2012 IAPR

In order to take advantage of available color images and meanwhile deal with big unknown regions better, in this paper we propose a novel inpainting method. It extends the efficient and effective fast marching method [10] so that a color image can be incorporated as the guidance while filling holes in a depth map. We hence refer to it as a guided FMM. Once the depth map is inpainted by our guided FMM, we further apply the guided filter [4] to refine it. At the end, we perform experiments on both the Kinect data and the Middlebury dataset [9] to obtain qualitative and quantitative evaluations and comparisons.

2. Guided Inpainting

In this section, we demonstrate our guided fast marching method for depth inpainting. We first introduce a color-based inpainting model to estimate the depth of a pixel on the boundary of unknown regions. Then we present how to determine the order of depth propagation by the guided FMM. Since our algorithm is inspired by Telea's work that is based on the FMM , several comparisons will be presented.

2.1. Inpainting Model

Like other inpainting literatures, we use Ω to represent a region to be filled where a Kinect cannot estimate depth values. What's more, the boundary of Ω is denoted by $\partial\Omega$. Fig. 2 illustrates the principle of inpainting one pixel on $\partial\Omega$. Considering a small area $\mathcal{B}(p)$ whose depth values are known around pixel p, the depth value $D(p)$ can be estimated by the weighted average of first order approximations:

$$D(p) = \frac{\sum_{q\in\mathcal{B}(p)} \omega(p,q)[D(q) + \nabla D(q)(p-q)]}{\sum_{q\in\mathcal{B}(p)} \omega(p,q)}. \quad (1)$$

Here $\nabla D(q)$ indicates the depth gradient at pixel q, $\omega(p,q)$ is the weighting function. We design $\omega(p,q)$ as a product of the following four terms:

$$\omega_{dst}(p,q) = \frac{1}{||p-q||^2}$$
$$\omega_{dir}(p,q) = \frac{p-q}{||p-q||} \cdot \nabla T(p)$$
$$\omega_{lev}(p,q) = \frac{1}{1+||T(p)-T(q)||}$$
$$\omega_{col}(p,q) = \exp(-\frac{||I(q)-I(p)||^2}{2\sigma_c^2}),$$

where I indicates the aligned color image and T is a distance map which will be detailed later. $\omega_{dst}(p,q)$ decides the contribution of pixels according to geometric distances to p. $\omega_{dir}(p,q)$ ensures pixels closer to the normal direction achieve higher contributions. $\omega_{lev}(p,q)$ is the level set term to make pixels closer to the contour through p have higher weight. $\omega_{col}(p,q)$ makes pixels having similar color to $I(p)$ contribute more than others.

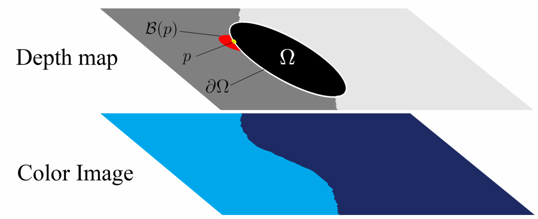


Figure 2: Inpainting model

In contrast to Telea's method, we introduce the color term $\omega_{col}(p,q)$ so that the weighting function $\omega(p,q)$ is able to incorporate color information for depth inpainting. This term is designed based on an assumption that neighboring pixels similar in color are likely to have similar depth values. With this term, it is possible for us to get fine depth edges in unknown regions.

2.2. Depth Propagation

As emphasized by Criminisi et al. [2], when propagating depth from $\partial\Omega$ to Ω, the order by which the inpainting procedure takes highly influences the resulting quality. Therefore, in this work we modify the propagation order in Telea's FMM for making better use of the guided color image.

Let us first introduce Telea's algorithm based on the FMM. It first sets $T=0$ for the pixels in known regions, then progressively generates a distance map T while marching into Ω satisfying $||\nabla T|| = 1$. Telea's algorithm decides the propagation order by choosing a pixel with the minimum T in each iteration. Nevertheless, since the order is only determined by the distance to the boundary, this algorithm may break edges as Fig. 3(e).

In order to use an aligned color image as the guidance, a color-similarity term around the pixel is calculated by

$$C(p) = \frac{1}{|\mathcal{N}(p)|}\sum_{q\in\mathcal{N}(p)} \exp(-\frac{||I(q)-I(p)||^2}{2\sigma_c^2}), \quad (2)$$

where $\mathcal{N}(p)$ is the neighboring window of p, including the pixels with both known and unknown depth values, and $|\mathcal{N}(p)|$ is the number of pixels. The color-similarity term assigns a high priority to a pixel having similar color around it. In other words, this term makes the

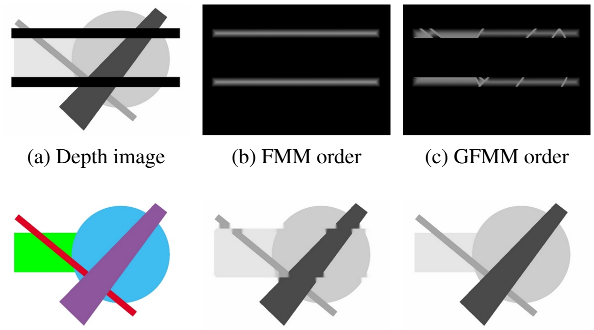

(a) Depth image (b) FMM order (c) GFMM order
(d) Color image (e) FMM result (f) GFMM result

Figure 3: Inpainting results of synthetic images

pixels near edges estimated later, thus trying to achieve fine edges. With this modification term, we decide the proceeding order by

$$E = (1 - \lambda)T + \lambda(1 - C), \tag{3}$$

where the scalar λ is a weighting factor between the distance term and the color-similarity term. Our depth propagation algorithm drives the inpainting order by choosing a pixel with the minimum E in each iteration.

Fig. 3 illustrates the power of our proposed method on a synthetic color image and the corresponding depth map. The two black rectangles in the depth map indicate regions to be inpainted. Considering darker pixels get inpainted earlier, (b) and (c) demonstrate that the guided FMM (GFMM) estimates the depth near edges at last while the original FMM determines the inpainting order by the distance to the boundary. (e) and (f) show that our method preserves the edges better than Telea's method.

3. Guided Filtering

The guided inpainting process makes every pixel in unknown regions get a proper depth value, however, the inpainted depth map still suffers from noise and unstable boundaries. Izadi et al. [5] use the bilateral filter (BF) to refine depth maps. The work by Wasza et al. [11] applies the guided filter (GF) [4] to accomplish real-time denoising on the GPU. Considering the efficiency of the guided filter, we choose it to refine depth maps. Here we take the aligned color image as the guidance to filter the inpainted depth map.

The guided filter is proposed as a time-efficient edge-preserving smoothing operator. It is based on the assumption that the output image has a local linear model to the guidance image in a small window, which means

$$\widehat{D}(p) = a_k I(p) + b_k, \tag{4}$$

where $I(p)$ is the color data regarded as the guidance and $\widehat{D}(p)$ is the output depth value. Besides, a_k and b_k are coefficients of the linear model. In a certain window, a_k and b_k are solved by

$$\min_{(a_k, b_k)} \sum_p (a_k I(p) + b_k - D(p))^2 + \varepsilon {a_k}^2, \tag{5}$$

where ε adjusts the filtering effects and $D(p)$ indicates the input depth data. Since one pixel is covered in different windows, the guided filter computes the average of $a_k I(p) + b_k$ for all the windows to achieve the output depth map. Fig. 1(d) shows the guided filtering result on the depth map after inpainting.

4. Experimental Results

In our experiments, we set weight $\lambda = 0.99$ described in Eq. 3. We validate our approach and compare it with other methods using both real-world Kinect data and the Middlebury dataset [9]. Fig. 4 presents the former results. There are some missing regions caused by particular surfaces such as human hair and chairs in the depth map. Here we compare four methods including FMM, FMM + GF, GFMM, and GFMM + GF. The results based on the FMM are unsatisfactory because of only using depth information. Our guided inpainting method fills holes while preserving the edges in unknown regions. Furthermore, the results after guided filtering show the improvement of edges and reduction of noise. For the Kinect real data (320×240) about 17000 points to inpaint, our inpainting algorithm takes less than 300ms on an Intel Core 2 CPU @ 2.33 GHz and 2 GB memory PC.

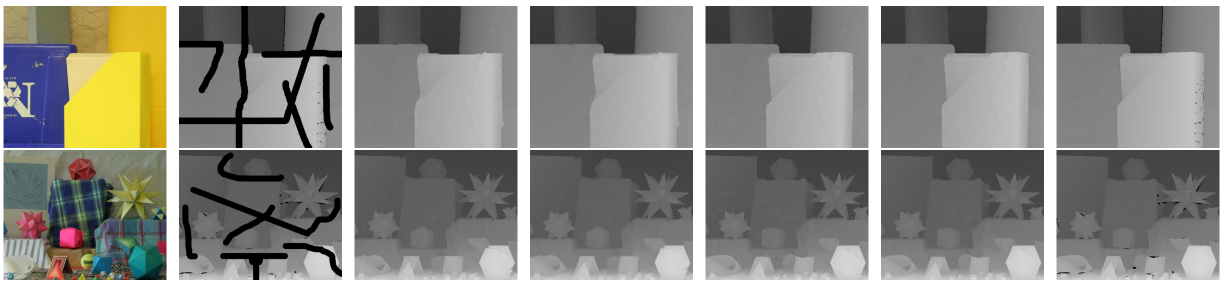
Since the ground truth depth for real-world scenes cannot be obtained easily, we test these approaches on the Middlebury stereo dataset whose ground truth depth is available, in order to acquire a quantitative comparison. To generate the artificial defective depth map, we first add the gaussian noise ($\sigma = 5$) to the ground truth depth, then we draw some black areas which indicate no depth value. Using Middlebury dataset, the depth map results and the RMSE against the ground truth are shown in Fig. 5 and Table 1, respectively. We can easily see that the quantitative performance of our approach which combines the guided FMM inpainting and the guided filter is the best. More experimental results are available online.

5. Conclusions

In order to enhance defective depth maps captured by Kinect-like cameras, we have proposed a novel in-

(a) Color (b) Depth (c) FMM (d) FMM + GF (e) GFMM (f) GFMM + GF

Figure 4: Results of different methods on Kinect data



(a) Color (b) Depth (c) FMM (d) FMM + GF (e) GFMM (f) GFMM + GF (g) Ground Truth

Figure 5: Results of different methods on the Middlebury dataset

Table 1: Results of RMSE against the ground truth on the Middlebury dataset.

Methods	Plastic	Moebius
FMM	6.87	5.23
FMM + GF	4.33	2.72
GFMM	4.93	4.97
GFMM + GF	1.94	2.42

painting method based on the guided FMM to reconstruct unknown regions. In addition, we incorporate the guided filter to improve the depth quality. Experimental results show that our method outperforms other existing methods in terms of both visual quality and RMSE. We believe that our method can provide better inputs to the applications based on the Kinect.

6. Acknowledgements

This research work was supported in parts by the National Natural Science Foundation of China via grants 61001171, 60534070, 90820306, and Chinese Universities Scientific Fund.

References

[1] M. Camplani and L. Salgado. Efficient spatio-temporal hole filling strategy for kinect depth maps. *Three-Dimensional Image Processing (3DIP) and Applications II*, page 82900E, 2012.

[2] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *Image Processing, IEEE Transactions on*, 13(9):1200–1212, sept. 2004.

[3] J. Dolson, J. Baek, C. Plagemann, and S. Thrun. Upsampling range data in dynamic environments. In *CVPR*, pages 1141–1148, 2010.

[4] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV*, pages 1–14, 2010.

[5] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. A. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. J. Davison, and A. W. Fitzgibbon. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *UIST'11*, pages 559–568, 2011.

[6] S.-Y. Kim, J.-H. Cho, A. Koschan, and M. Abidi. Spatial and temporal enhancement of depth images captured by a time-of-flight depth sensor. In *ICPR*, pages 2358–2361, aug. 2010.

[7] Microsoft. *Xbox 360 Kinect Sensor Manual*, 2010.

[8] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, 2004.

[9] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *CVPR*, pages 1–8, 2007.

[10] A. Telea. An image inpainting technique based on the fast marching method. *journal of graphics, gpu, and game tools*, 9(1):23–34, 2004.

[11] J. Wasza, S. Bauer, and J. Hornegger. Real-time preprocessing for dense 3-d range imaging on the gpu: Defect interpolation, bilateral temporal averaging and guided filtering. In *ICCV Workshops*, pages 1221–1227, nov. 2011.