IRCITCS-265
An Improved Depth Image Inpainting

Chieh-Fu Hsieh[a*] ,Chi-Hsiao Yih[b] and Ching-Tang Hsieh[c]

[a,b,c] Department of Electrical Engineering, Tamkang University

New Taipei, Taiwan, R.O.C.

[a]Hsiehchiehfu@gmail.com

Abstract

In recent years, the price of depth camera became low, so that researchers can use depth camera to do more application. For computer vision, depth images can provide more useful information. However, generally there are some problems in depth image, such as holes, incomplete edge, and temporal random fluctuations. Conventional inpainting approach must rely on color image and it cannot be processed in real time. Therefore, this paper proposes a real time depth image inpainting method. First, we use background subtraction and mask filter to patch up the no-measured pixels, and then using the relationship between successive depth images to remove temporal random fluctuations. Finally, using erosion and dilation smooth the edge. Experimental results outperform than traditional one.

Keyword: Depth camera, depth image, inpainting,

1. Introduction

Depth camera plays an important role in the development of the computer vision. There are three kinds of depth camera: binocular stereo vision camera (Trucco and Verri, 1998), time of flight (ToF) (Xu et al., 1998; Oggier et al., 2005), and structured light camera (Batlle et al., 1998; Scharstein and Szeliski, 2003). Binocular stereo vision camera is similar to biological vision using the binocular disparity to calculate the depth information, but the image of the stereo camera are color images, and a high sensitive brightness, would increase the difficulty of calculating triangulation measurement. ToF is monocular camera emitting infrared pulse to the object surface, and estimating the depth information by calculating the reflection time of infrared pulse to the camera. Because of simply calculation and single viewpoint, ToF camera can achieve high frame rates, and do not have holes problems in depth image. The major drawback of the ToF is low resolution. Structured light cameras, which is currently common depth camera Kinect, Microsoft released in November 2010, has cheap price, higher resolution than the ToF, and receiving the depth image and color image at the same time. The disadvantage of Kinect is that the depth information could not be measured for some material, which resulted in the holes problem.

The depth image obtained by depth camera has some problems. Holes problem happened on object overlapping, reflective material surface, and large depth difference. Edges of the object were broken and incomplete. Continuous depth images would produce temporal random fluctuations. Therefore, many researchers devote to repairing the depth image (Vzquez et al., 2006; Cheng et al., 2008; Tam et al., 2004; Zhang and Tam, 2005; Tam and Zhang, 2004). There are two common approaches to remove the holes problem. One is to fill the holes by using the neighbor depth information such as interpolation, extrapolation (Vzquez et al., 2006), mirroring of background information (Cheng et al., 2008). The other is to preprocess the depth image before 3D model reconstruction so that the reconstructed 3D model could reduce or remove the hole such as Symmetric Gaussian smoothing filter (Tam et al., 2004) or Asymmetric Gaussian smoothing filter (Zhang and Tam, 2005; Tam and Zhang, 2004). Smoothing filter can reduce the holes problem of depth image will disappear after, but it also generates the geometric distortion. For geometric distortion problems, some researchers use the local region smoothing filter (Lee and Ho, 2009; Lee and Effendi, 2011) to improve, but these approaches would destroy the original depth information.

For Kinect holes filling methods, some researchers use non-local filter to re-estimate the region without depth information. (Qi et al., 2013) For depth edge modification, median filter and Joint-Bilateral Filter (Silberman and Fergus, 2011; Camplani and Salgedo, 2012) are common approaches to correct the edge. Because Kinect has the both depth image and color image, calibrating depth image with color image is effective (Herrera et al., 2011; Zhang and Zhang, 2011), but it takes of computing time. Therefore, how to effectively eliminate the holes and retain the original depth information is the focus of this paper.

The rest part of this paper is organized as follows. The proposed approaches would be described in section 2 and the experiment results are present in section 3. Conclusion is drawn in section 4.

2. Proposed Inpainting Method

Flow chart of the proposed inpainting method is on below. (Fig. 1)

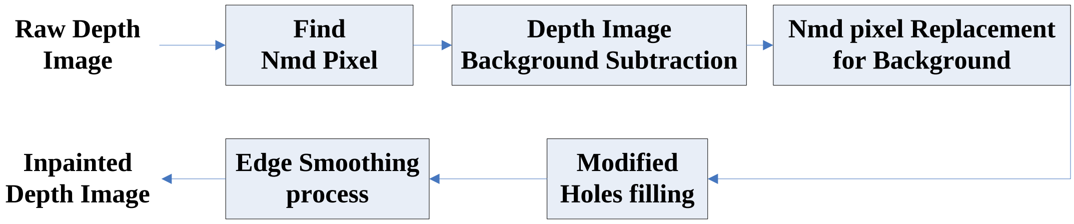


Fig.1 Flow chart of the proposed method

2.1 Find nmd pixel

Nmd pixels (Fig. 2) are no-measured pixels, that Kinect can't obtain the depth information in part of the scene. It usually appears on the surface of special materials, the edges of the objects, and the large difference of depth information. In order to fill these holes easily, we need to mark the location of nmd pixels.

(a) (b)

Fig.2 (a) is raw depth image, (b) is nmd pixels

2.2 Depth image background subtraction

When an object enters the scene of Kinect, depth information on the location of the object will change obviously. Therefore, using the background subtraction can get the foreground of the scene. In this paper, to fill the holes we use different methods which are presented in next step.

(a) (b)

Fig.3 (a) is background image, (b) is background depth image

2.3 Nmd-pixel replacement for BG

After the objects enter the scene, holes will be generated holes surrounding the objects. We can use depth information from background to replace these holes around the object such as the red rectangle region given in Fig. 4. The background depth information replacement is more reliable than the prediction method presented in next step. The holes in the background and on the surface of the object will be filled in the next step.

Fig.4 replacement for background

2.4 Modified hole filling method

We modify the holes filling method used in Fei Qi's method (Qi et al., 2013). The unknown depth information d(**p**) (Qi et al., 2013) use the first order Taylor approximation to estimate the unknown pixels. The two factors of first order Taylor approximation are gradient and geometric distance. However, the geometric distance has been considered to weighting function, and the gradient factor which represents edge information in image is processed after holes filling, so that we modify the unknown pixel d(**p**) to the near pixels **q** around the position **p**.

$$\mathrm{d}(\mathbf{p}) = \sum_{q \in N(p)} w(\boldsymbol{p}, \boldsymbol{q}) d(\boldsymbol{q}) \tag{1}$$

d(**p**) is the value of unknown depth pixel, d(**q**) is the value of depth pixel around the position **p**, N(**p**) is the range of the mask which is shown in figure 5 at the center p, w(**p**,**q**) is the weighting function based on two parameters, geometric distance and depth similarity. The form of the mask is not square mask, but the similar asterisk mark shaped mask such as fig. 5. The depth image is usually smoother and less complex texture than the color image. Therefore we use the similar asterisk mark shaped mask to reduce computing time instead of square mask. On the edge of depth information filling, we use weighting function to control what depth information to fill in, how to assign weighting function will be presented in equation (4) and (5).

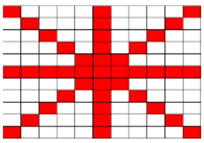


Fig.5 similar asterisk mark shaped mask

Weighting function

$$w'(\boldsymbol{p},\boldsymbol{q}) = \exp\left(-\frac{w_1(\boldsymbol{p},\boldsymbol{q})}{2h_1^2} - \frac{w_2(\boldsymbol{p},\boldsymbol{q})}{2h_2^2}\right) \tag{2}$$

w_1 is the parameter of geometric distance, w_2 represents depth similarity, h_1 and h_2 are two elements via the experimental result[13]. The normalized weighting function is obtained in equation (3).

$$w(\mathbf{p},\mathbf{q}) = \frac{w'(\boldsymbol{p},\boldsymbol{q})}{\sum_{\boldsymbol{q}\in N(\boldsymbol{p})} w'(\boldsymbol{p},\boldsymbol{q})} \tag{3}$$

Equation (4) is the geometric distance between position **p** and **q**.

$$w_1(\boldsymbol{p},\boldsymbol{q}) = \sqrt{(\mathbf{p}-\mathbf{q})^2} \tag{4}$$

It means that the farther distance from the position **p** the lower weighting value is given, whereas the closer distance the higher weighting value. When we fill the holes at the non-edge region, the geometric distance can provide a precise estimated value. At the edge region, the second parameter depth similar is more reliable than geometric distance.

$$w_2 = \left(1 - \frac{d(\boldsymbol{q})}{d_{max}(\boldsymbol{p})}\right) \tag{5}$$

w_2 is the depth similar, $d_{max}(\boldsymbol{p})$ is the maximum depth information in the mask at center **p**. At the edge region, the holes usually generates at the region of further distance from the camera. Thus, we give greater weighting to depth value where having the greater depth information, so that the filled value will close to the actual depth information. The edge of the object would not be dilated.

2.5 Edge smoothing process

After holes filling, edge smoothing is the last step. Most researchers use median filter or weighted median filter for smoothing, however it needs large size of mask for effective smoothing in depth image, and time consumption become large. In order to improve computing time, in this paper, we smooth the edges with two procedures.

In successive depth image, the difference between frame and frame may generate the temporal random fluctuation, and most of them happened at the edge of object. Therefore, in this paper we use subtraction between frame and frame to obtain the location of random fluctuations region, and then use the stable background depth image to replace the unstable region, thereby obtain a stable edge depth image. In this step we can roughly reduce the exaggerating wrong edge information owing to the temporal random fluctuation. The detail smoothing edge will be present in next procedure.

We use the erosion and dilation method to correct the edge in detail. The erosion and dilation part are the region of large depth information. Typically, the edge obtained by the depth image will be larger than the actual edge, so that we do the erosion and dilation on large depth information region will make the estimated edge closer to the actual edge of the object.

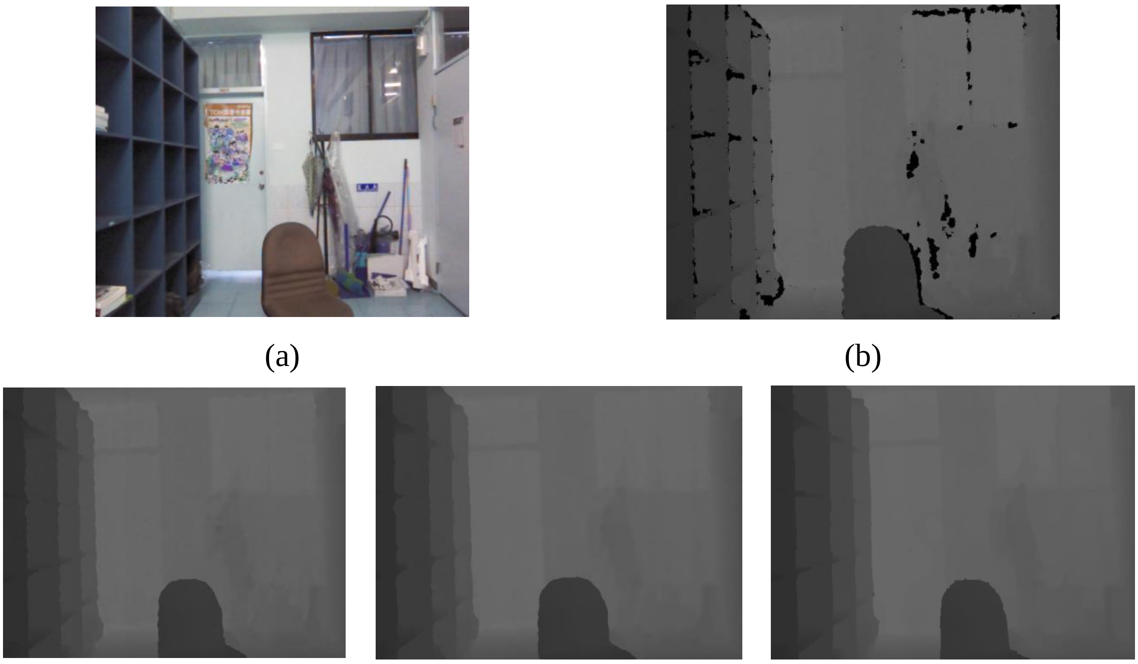
3. Experimental Result

In this section we present the results obtained with the proposed inpainting method, in Sec. 3.1 is focus on the comparison of the proposed method and other method. Sec. 3.2 is aim to the comparison of computing time.

3.1 Comparison

The experimental environment is indoor. The depth image is captured using Microsoft Kinect. Experimental platform is Window 7, CPU is Core2 3.00GHz, and RAM is 4.00GB.

In Fig. 4 we present the three different depth image inpainting approaches: Fig. 6(a) is color image, Fig. 6(b) is raw depth image, non-local filter [13] (Fig. 6(c)), joint-bilateral filter (Camplani and Salgedo, 2012) (Fig. 6(d)) and the proposed method (Fig. 6(e)). The result show that these three methods are able to fill the holes from raw depth image effectively, and the difference of effect is insignificant.



(a) (b)

(c) (d) (e)

Fig. 6 (a) is color image, (b) is raw depth image, (c) is inpainted by (Qi et al., 2013), (d) is inpainted by (Camplani and Salgedo, 2012), (e) is proposed method

3.2 Computing time

Table 1 show that the proposed method has greatly progress in computing time. The proposed method is faster than other two methods, and the frame per second is 0.398. The other methods use the median or joint-bilateral filter to modify edge. The mask size of filter is 11*11.

Table 1 computation comparison

Number of frames	Fei Qi	M. Camplani	Proposed method
30	32.238s	31.971s	14.292s
60	59.296s	60.257s	25.226s
90	91.259s	88.624s	35.782s

4. Conclusion

The proposed method is improved depth image inpainting, and only relying on depth information. The color image would be greatly affected by the brightness, and would not be used in the shadows. The experimental result show that comparing with other methods using color image, the effect is not much difference, however the computing time has greatly progress. In future works, the inpainting depth image can be applied to 3D object modeling, or complex environment object tracking.

5. Acknowledgement

This paper is supported by National Science Council of Taiwan, and the Project Number is 102-2410-H-032-060.

6. Reference

Emanuele Trucco and Alessandro Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, Upper Saddle River, New Jersey, 1998.

Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, T. Ringbeck, L. Nachrichtenverarbeitung,, S.t. Gmbh, K. Straße,. Smart pixel n photonic mixer device (PMD) New system concept of a 3D-imaging camera-on-a-chip. Proc Int. Conf. on Mechatron Machine Vision, 1998, 259–264.

T., Oggier, T., Bu¨ttgen, B., Lustenberger, F., Becker, G., Ru¨egg, B., Hodac, A., Swissranger SR3000 and first experiences based on miniaturized 3D-ToF cameras., in: In Proc. of the First Range Imaging Research Day at ETH Zurich, 2005.

J. Batlle, E. Mouaddib, J. Salvi, Recent progress in coded structured light as a technique to solve the correspondence problem: a survey Pattern Recognition 31, 1998, 963– 982.

D. Scharstein, R. Szeliski, High-accuracy stereo depth maps usingstructured light, in: Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol.1.43, 2003, pp. 195-202.

C. Vzquez, W. J. Tam, and F. Speranza, Stereoscopic Imageing Filling Disoccluded Areas in Depth Image-Based Rendering, Proceedings of SPIE, Vol.6392, 2006

C. M. Cheng, S. J. Lin, S. H. Lai, and J. C. Yang, Improved Novel View Synthesis from Depth Image with Large Baseline, Proceedings of the 19th International Conference on Pattern Recognition (ICPR), pp. 1-4, 2008.

W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, Smoothing Depth Maps for Improved Stereoscopic Image Quality, in Proc. International Society for Optical Engineering Conf. Three-Dimensional TV, Video, and Display III, vol. 5599, Oct. 2004, pp162-172.

L. Zhang and W. J. Tam, Stereoscopic Image Generation Based on Depth Images for 3DTV, IEEE Trans. Broadcast., vol. 51, Jun. 2005, pp. 191-199.

W. J. Tam and L. Zhang, Non-Uniform Smoothing of Depth Maps before Image-Based Rendering, in Proc. of ITCOM'04, vol. 5599, Oct. 2004, pp. 173-183.

S. B. Lee and Y. S. Ho, Discontinuity-Adaptive Depth Map Filtering for 3D View Generation, in IMMERSCOM Proceedings of the 2nd International Conference on Immersive Telecommunications, ICST, 2009, pp. 1-6.

P. Lee and Effendi, Nongeometric Distortion Smoothing Approach for Depth Map Preprocessing, IEEE Transactions on Multimedia, vol. 13, No. 2, 2011, pp.246-254.

Fei Qi, J. Han, P. Wang, G. Shi, and Fu Li, Structure Guided Fusion for Depth Map Inapinting, Pattern Recognition Letters, Vol. 34, 2013, pp. 70-76.

N. Silberman and R. Fergus, Indoor Scene Segmentation Using a Structured Light Sensor, in Proc. Int. Conf. Comput. Vision Workshop 3-D Representation Recognition, 2011, pp.601–608.

M. Camplani and L. Salgedo, Efficient Spatio-Temporal Hole Filling Strategy for Kinect Depth Maps, Proc. SPIE, Three-Dimensional Processing (3DIP) and Applications II, Feb. 2012.

D. Herrera, J. Kannala, and J. Heikkila, Accurate and practical calibration of a depth and color camera pair, in Proc. CAIP, 2011, pp. 437–445.

C. Zhang and Z. Zhang, Calibration between depth and color sensors for commodity depth cameras, in Proc. IEEE ICME, 2011, pp. 1–6.