

Colored 3D surface reconstruction using Kinect sensor

GUO Lian-peng (郭连朋)*, CHEN Xiang-ning (陈向宁), CHEN Ying (陈颖), and LIU Bin (刘彬)

Department of Optical and Electronic Equipment, Equipment Academy of PLA, Beijing 101400, China

(Received 19 January 2015)

©Tianjin University of Technology and Springer-Verlag Berlin Heidelberg 2015

A colored 3D surface reconstruction method which effectively fuses the information of both depth and color image using Microsoft Kinect is proposed and demonstrated by experiment. Kinect depth images are processed with the improved joint-bilateral filter based on region segmentation which efficiently combines the depth and color data to improve its quality. The registered depth data are integrated to achieve a surface reconstruction through the colored truncated signed distance fields presented in this paper. Finally, the improved ray casting for rendering full colored surface is implemented to estimate color texture of the reconstruction object. Capturing the depth and color images of a toy car, the improved joint-bilateral filter based on region segmentation is used to improve the quality of depth images and the peak signal-to-noise ratio (PSNR) is approximately 4.57 dB, which is better than 1.16 dB of the joint-bilateral filter. The colored construction results of toy car demonstrate the suitability and ability of the proposed method.

Document code: A **Article ID:** 1673-1905(2015)02-0153-4

DOI 10.1007/s11801-015-5013-2

There has been an enormous trend to employ new technology for visual media creation in recent years, creating a demand for high-quality 3D surface reconstruction of real-world objects^[1,2]. Low-cost consumer-grade RGB-D sensor, such as Microsoft's Kinect^[3,4], provides an attractive alternative to expensive laser scanners in many different application areas, such as manufacturing verification, augmented reality and object recognition^[5]. Kinect was primarily designed for natural interaction in a computer game environment, and the characteristics of the Kinect's data, which is capturing depth and color images simultaneously, have attracted the attention of researchers from other fields, including mapping and 3D modeling^[6-8]. The recent KinectFusion system^[9,10] demonstrates the Kinect's ability to quickly and accurately produce detailed 3D geometric reconstruction of an indoor scene, which fuse all of the depth data streamed from Kinect sensor into a single global implicit surface model of environment in real time. However, the reconstruction doesn't have color texture. A solution of this problem is using an additional voxel grid consisting of the color data to estimate the color texture of the reconstruction object.

In this paper, based on colored truncated signed distance fields and ray casting, a novel full colored 3D surface reconstruction method which effectively fuses information from both depth and color image got from Microsoft Kinect is proposed and demonstrated. To remove lost and noisy pixels of the depth images in Microsoft Kinect, an improved joint-bilateral filter based on region segmentation that takes advantage of the region map is presented to facilitate the filtering process. An

incremental update of the camera pose between two subsequent frames is estimated using a variant of the iterative closest point (ICP) algorithm. Given a valid camera pose, the registered data are integrated into a global model to achieve a surface reconstruction via colored truncated signed distance fields. Finally, for rendering surface of reconstruction object to users, an improved ray casting is presented to generate views of the implicit surface. A full colored 3D surface reconstruction implemented through simulation demonstrates that the proposed method can achieve a full colored 3D surface reconstruction. In addition, all the algorithms which are used in this paper and the achieved results are described and discussed in the follows.

The algorithmic core and the main novelty of our method are the improved joint-bilateral filter based on region segmentation to remove lost and noisy pixels and the colored truncated signed distance fields using an additional voxel grid consisting of the color data to estimate the color texture of the reconstruction object. Fig.1 shows an overview of the proposed full colored 3D surface reconstruction method. The procedure can be subdivided into five main stages.

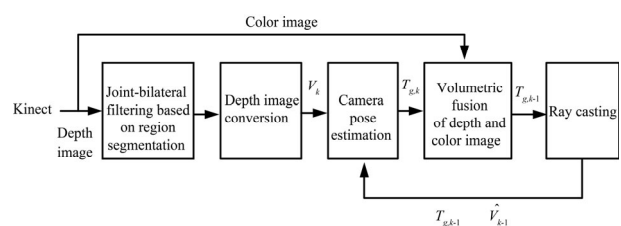


Fig.1 Workflow of the reconstruction method

* E-mail: kakagordon@126.com

Kinect is a structured light 3D scanner^[11], consisting of a depth sensor and a color camera. The depth sensor is a composite device consisting of an IR projector projecting light patterns to the space and an IR camera receiving the reflected light. Due to the sensor systematic errors, the measurement setup such as lighting condition and the properties of the object surface, the quality of the Kinect depth image is degraded by lost and noisy pixels. Joint bilateral filter (JBF)-based depth hole filling method^[12] is often used to remove the noisy pixels and fill the lost pixels. However, the properties of noise at different regions are not the same, which poses obstacle to the filtering process using the only filtering parameters and hence degrades the preservation of depth image structures. Therefore, we present an improved joint-bilateral filter based on region segmentation that takes advantage of the region map to facilitate the filtering process. Specifically, noise properties with the regions are utilized to determine the filtering parameters corresponding to the involved pixels.

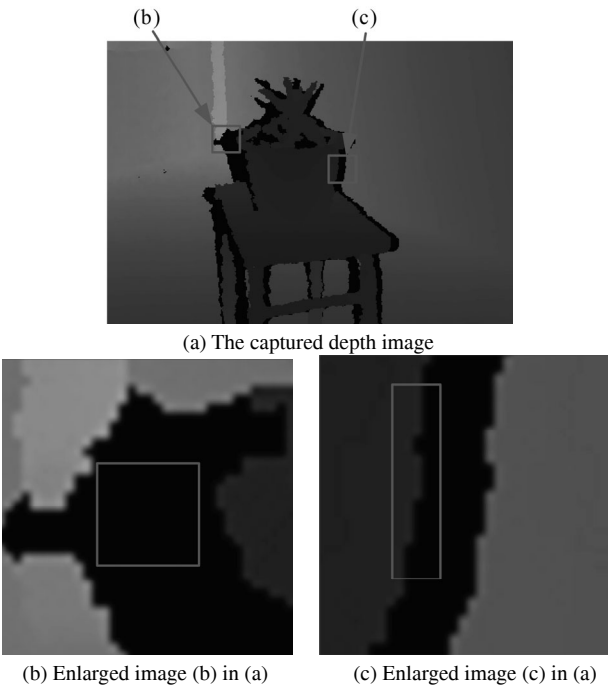


Fig.2 Depth image captured by Kinect

The depth image must be converted from the image coordinates into 3D points in the coordinate space of the camera. As shown in Fig.1, given the intrinsic calibration matrix K of the Kinect's infrared camera, a separate pixel $u=(x, y)$ in the input depth image $D_i(u)$ is projected to the camera's coordinate space as $v_i(u)=D_i(u)K^{-1}[u, 1]$. This results in a point cloud V_i in the camera's coordinate space^[8,9].

Iterative closest point (ICP) is a popular method for alignment of three-dimensional points cloud. For our current needs, the camera pose $T_{g,k}$ of the current frame can be estimated by the method of a real-time variant of ICP^[13].

The pipeline of the volumetric fusion of the depth im-

age for geometrical reconstruction is to create a truncated signed distance field by shooting rays and computing the signed distance $f_i(u)$ of each point u to the nearest depth surface along the line of the sight to the Kinect. Since we are interested in recovering the color of the object as well, this paper estimates a color texture using an additional voxel grid:

$$c_i(u) \rightarrow [RGB_i(u), W_i(u)], \quad (1)$$

where the $RGB_i(u)$ stores the red, green and blue components of the color element at point u , and the $W_i(u)$ stores the weight. The colored truncated signed distance field $[f_i, c_i]$ consisting of signed distance f_i and color volume c_i is used to reconstruct the geometry and coloring similar to Ref.[14]. When camera pose has been estimated, each voxel grid stores a running average of signed distance and color.

As the surface is located at zero crossing in the signed distance function, the marching cubes algorithm in Ref.[15] is applied to extract the corresponding triangle mesh. The color texture of the voxel grid is estimated from color image by running average, and the color texture of vertices is computed by tri-linear interpolation.

As shown in Fig.3, for illustrating the effectiveness of the method proposed in this paper, the Kinect is placed in a fixed location to observe the toy car (25 cm×15 cm×15 cm) mounted on a manual turntable. The turntable is then spun through a full rotation, capturing 500 frames of depth images and color images correspondingly. Running time for reconstruction is measured on a standard PC with an Intel(R) CPU 2.6 GHz and 4 GB RAM memory.



Fig.3 The experimental scene

The Kinect sensor can measure the depth data from 0.8 m to 4.0 m, and quality of the depth image is degraded by lost and noisy pixels. The relation between the random error of depth information and the distance to the sensor can be measured by a plane fitting test similar to Ref.[4]. Fig.4 shows the depth error against the distance to the Kinect. It can be seen that the errors increase drastically from 0.3 mm at 0.5 m distance to about 4 cm at 3.5 m distance. Therefore, the pixel-based joint bilateral filter suffers difficulties in an appropriate selection of the filtering parameters due to the disturbance of noise, which poses obstacles to the filtering process in dealing with the preservation of the depth image structures. What's more, using the single filtering parameter can also lead to unsatisfactory results in denoising. Based on

the segmentation of the depth images, an improved joint-bilateral filter that takes advantage of the region map is presented to facilitate the filtering process. Fig.5(a) is an illumination of the Kinect depth image, Fig.5(b) is the filtered depth image obtained by the method of joint bilateral filter, and Fig.5(c) is the filtered depth image obtained by the joint-bilateral filter based on region segmentation. Comparing Fig.5(b) and Fig.5(c), it is observed that the joint-bilateral filter based on region segmentation can improve the denoising performance especially on the preservation of structure information. What's more, the PSNR gain of the joint-bilateral filter based on region segmentation is approximately 4.57 dB, greater than 1.16 dB of the joint-bilateral filter.

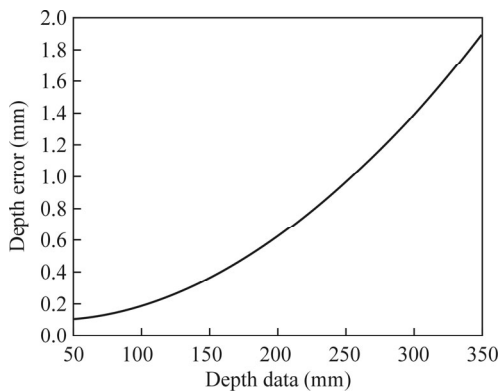


Fig.4 Depth errors at different distances

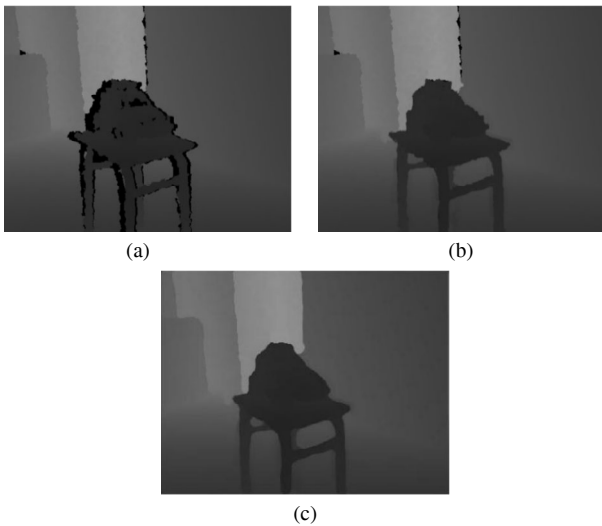


Fig.5 The filtered depth images: (a) Raw depth image; (b) Based on the joint-bilateral filter; (c) Based on our proposed method

After estimating the pose of the Kinect using the variant of ICP and converting the depth image into a global coordinate space, a colored truncated signed distance field is created to integrate the depth and color information. Then, a straight-forward implementation of the marching cubes algorithm is used to extract the triangle mesh, as shown in Fig.6.

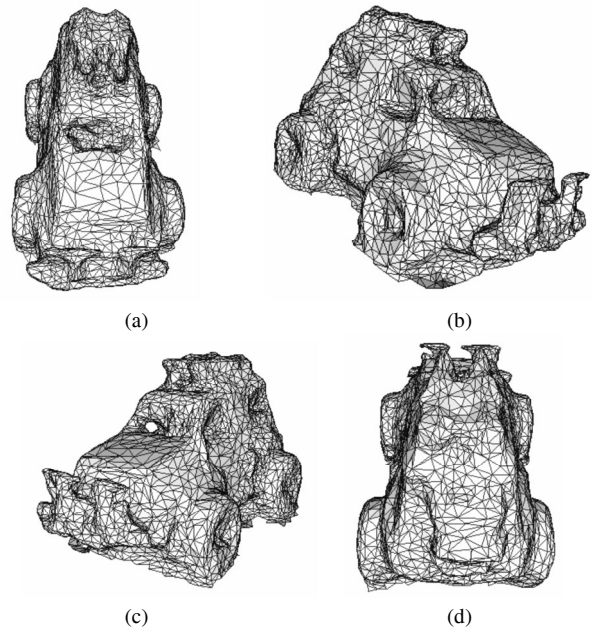


Fig.6 Geometric 3D surface reconstruction results

After extracting the geometric surface from the colored truncated signed distance fields, the color texture of the voxel grid is estimated from c_i and the reconstruction surface of the toy car is shown in Fig.7. The results show that the method presented in this paper can not only resolve inherent lost and noisy pixel problem of Kinect, but also render color for the reconstruction object, demonstrating the usefulness and ability of it.

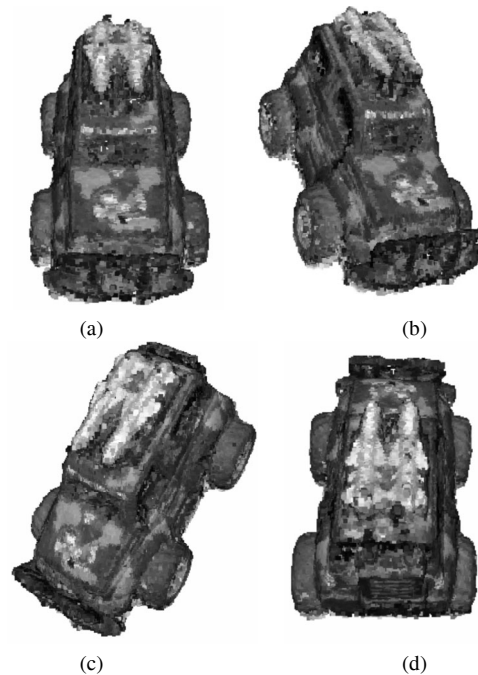


Fig.7 Full colored 3D surface reconstruction results

A novel full colored 3D surface reconstruction method using Microsoft Kinect is presented in this paper based on volumetric depth image fusion. An improved joint-

bilateral filter based on region segmentation is presented in this method to improve the quality of depth images, which preserves the structure of depth data when improving the denoising performance. For recovering the color of the object, this paper estimates a color texture using an additional voxel grid $c_i(\mathbf{u})$, creates a colored truncated signed distance field and improves the ray casting algorithm. Experimental results demonstrate the usefulness and ability of our method in full colored 3D surface reconstruction. Future work will include bundle adjustment over the colored truncated signed distance fields and a more efficient fusion method in computation.

References

- [1] Cui Y., Schuon S., Thrun S., Stricker D. and Theobalt C., *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1039 (2013).
- [2] Rashmita Khilar, S. Chitrakala and SurenderNath SelvamParvathy, *3D Image Reconstruction: Techniques, Applications and Challenges*, IEEE International Conference on Optical Imaging Sensor and Security, 1 (2013).
- [3] Freedman B., Shpunt A., Machline M and Arieli Y., *Depth Mapping Using Projected Patterns*, United States Patent No.8493496, 2012.
- [4] Cheng Zhang, Hairong Yang, Hong Cheng and Sui Wei, *Journal of Optoelectronics·Laser* **24**, 805 (2013). (in Chinese)
- [5] Lai K., Bo L., Ren X. and Fox D., *Sparse Distance Learning for Object Recognition Combining RGB and Depth Information*, IEEE International Conference on Robotics and Automation (ICRA), 4007 (2011).
- [6] Khoshelham K. and Elberink S. O., *Sensors* **12**, 1437 (2012).
- [7] Herbst E., Henry P., Ren X. and Fox D., *Toward Object Discovery and Modeling via 3-D Scene Comparison*, IEEE International Conference on Robotics and Automation (ICRA), 2623 (2011).
- [8] Menna F., Remondino F., Battisti R. and Nocerino E., *Proceedings of SPIE* **8085**, 80850G (2011).
- [9] Newcombe R. A., Izadi S., Hilliges O. and Molyneaux D., *KinectFusion: Real-time Dense Surface Mapping and Tracking*, 10th IEEE International Symposium on Mixed and Augmented Reality, 127 (2011).
- [10] Izadi S., Kim D., Hilliges O., Molyneaux D., Newcombe R., Kohli P. and Fitzgibbon A., *KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera*, Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, 559 (2011).
- [11] Camplani Massimo and Luis Salgado, *Proceedings of SPIE* **8920**, 82900E (2012).
- [12] Matyunin S., Vatolin D., Berdnikov Y. and Smirnov M., *Temporal Filtering for Depth Maps Generated by Kinect Depth Camera*, 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 1 (2011).
- [13] Rusinkiewicz, Szymon, Olaf Hall-Holt and Marc Levoy, *ACM Transactions on Graphics (TOG)* **21**, 438 (2002).
- [14] Curless Brian and Marc Levoy, *A Volumetric Method for Building Complex Models from Range Images*, Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, 303 (1996).
- [15] Lorensen William E. and Harvey E. Cline, *ACM Siggraph Computer Graphics* **21**, 163 (1987).