Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Volume 22, No. 1, ISSN 1726-9679
ISBN 978-3-901509-83-4, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2011
Make Harmony between Technology and Nature, and Your Mind will Fly Free as a Bird
Annals & Proceedings of DAAAM International 2011

EVALUATION OF KINECT DEPTH SENSOR FOR USE IN MOBILE ROBOTICS

KEFER, M[artin] & KUBINGER, W[ilfried]

Abstract: *To autonomously navigate through unstructured environment mobile robots rely mainly on vision systems. Commonly used are stereo vision, laser and infrared sensors. This paper deals with an evaluation of the Kinect depth sensor and compares it with stereo vision. Also the depth resolution, the accuracy and the repeatability have been determined throughout several experiments based on the SDK of Code Laboratories (CL) for Windows 7. Our experimental results show that the Kinect has a maximum depth resolution of 1.3[mm/color value] and a maximum accuracy of 0.02%. Also, depth density has been assessed and compared to stereo vision images, concluding several benefits the Kinect brings to the field of mobile robotics. Finally, suggestions are being made for several applications for use in mobile robotics*

Key words: *kinect, mobile robots, depth sensing*

1. INTRODUCTION

Mobile robots are designed to autonomously navigate through their environment for different fields of application; whether it is on the ground, in the water or in the air. In order to achieve such ambitious tasks they rely on various sensors to extract meaningful information from the environment. Vision sensors such as stereo cameras are often used for mobile robots since they encounter unforeseen characteristics in the environment as they move around. Additionally, they provide the robot with 3D data since they imitate the human sense of vision (Siegwart & Nourbakhsh, 2004). In early November 2010 the Kinect sensor became available for consumers and quickly managed its way into mobile robotic research. It soon was clear the Kinect will provide the robotic community with never-before-seen depth sensing to an unrivaled price.

This paper deals with the evaluation of the Kinect and with the prospect of using this sensor for future research in mobile robotics. Benchmarking the Kinect and comparing it to a stereo vision camera based on various experiments and scenarios have been performed. Furthermore, the results have been assessed in view of the deployment with autonomously navigating robots. As a consequence, the possibilities in use are presented and the system's limitations are concluded.

2. DEPTH SENSING

Numerous methods have been proposed by scientists in the past couple of years, whereas the most promising are described in this section.

The principle of time of flight (TOF) is based on capturing the light scattered from objects. The reflected intensity is a function of the distance to these objects since the light intensity is modulated to enable reliable distance calculations. The amount of light intensity is a function of speculation, diffusivity of the object's surface, the shutter speed and the distance to the object (Um et al., 2011).

Structured light coding is based on finding the correspondence between image pixels and the projected light pattern illuminating the scene. The main challenges are related to the choice of light pattern and coding strategy. These strategies can either be multiplexing, direct coding or spatial neighborhood coding (Albitar et al., 2007).

Stereo vision is based on the human sense of seeing, which uses two cameras in a specified position to each other. The principle is to identify certain pixels in the left and right image of both cameras. Since the left and right image are slightly different, the feature found in both images is linked to different pixels. This difference is the basis to calculate the actual depth throughout triangulation (Humenberger et al., 2010), (Bleyer et al., 2011).

The Kinect technology is based on speckle-pattern depth mapping which is a method of infrared light intensity based depth measurement. It is similar to structured light coding but uses cross-correlation with speckled infrared light projections on objects in the scene. This means that the primary speckle pattern is compared to the pattern on the reference surface which enables the depth calculation throughput cross-correlation or decorrelation (Um et al., 2011).

3. MICROSOFT KINECT

The Kinect weighs 564.5g and its dimensions are 73mm in height, 283mm in width and 72.8mm in depth. It consists of a depth projector, a depth receiver, an RGB camera with VGA resolution, a LED for basic interaction, two audio microphone arrays with noise cancellation technology, a 3D accelerometer and a motor to tilt it ±28°. Its field of view is, according to Microsoft (2011), 43° vertical by 57° horizontal and its operating range is 1.2 to 3.5 meters. The most characteristic feature in the depth image is an effect called shadow. This effect is derived from the fact that objects conceal the light pattern which, as a consequence, cannot be projected on certain areas behind an object. This shadow is illustrated in Figure 1.

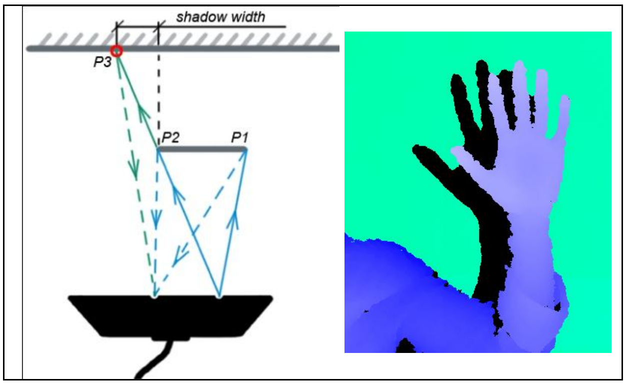

Fig. 1. Kinect Shadow

The outline and the surface of the hand is detected properly and represented in the illustration by an object (P1, P2). The blue beam passing P2 can in theory be prolonged to the wall where it would be reflected at P3. The area between P3 and P2 results in the shadow where no depth data can be obtained.

4. EXPERIMENTAL RESULTS

4.1 Depth resolution and accuracy

In order to find out about the exact depth resolution an experiment has been performed with the support of a test platform and a measuring tape. The distance between platform and Kinect was constantly altered. As a result a maximum resolution of 1.3mm/color value could be determined. The accuracy was basically determined over the same procedure and could be determined with a maximum relative uncertainty of 0.02%.

4.2 Repeatability

In order to assess the repeatability of the system the Kinect was viewing a static scene where several pixels had been addressed. Then the motor was tilting the head randomly before it was set to starting position again. Then the addressed pixel information was obtained and evaluated. The results for best and worst case are shown in Figure 2.

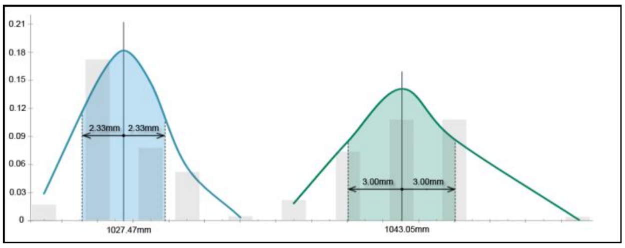

Fig. 2. Repeatability

4.3 Benchmarking: Kinect vs Stereo Vision

For benchmarking a stereo vision camera provided from AIT Austrian Institute of Technology (Humenberger et al., 2010) was put in various scenarios under the same light conditions as the Kinect. The object of interest was a platform made of wood and additionally assembled with several objects of different materials. Such as black foam, Styrofoam, a white tile, a collection of small colored tiles and black and white buttons of felt. This was done to test different materials and surfaces a mobile robot might encounter during operation. In addition, a reflective, translucent plate was used as well as a mirror. In the first scenario the vision systems were facing the platform which was in turn facing a window during broad daylight, excluding direct sunlight which caused high contrast light conditions. In the second scenario the composition was reversed, meaning the sensors were facing the window under the same light conditions as in scenario one but, consequently, with less contrast. In a third scenario a reflective plate was put in front of the sensors in order to conclude about its behavior.

As it turned out in the scenarios, the stereo system lacks depth information which was assessed through a value called depth density. This value puts pixels with valid depth information in relation to all pixels including black ones, representing no depth information. The benchmark of the scenario results is presented in Tab. 1.

	Kinect	AIT Stereo Vision
Scenario 1	~93%	~62%
Scenario 2	~95%	~18%
Scenario 3	~96%	~37%

Tab. 1. Benchmarking results

Our experimental results show an objective value when considering the object of interest, the platform, only. However, the Kinect manages to cope with any light condition as long as it is indoors. Outdoors, it seems the infrared light of the sun disturbs the sensor in a way it does not retrieve any 3D information at all. The depth density being constantly high above 90% is compelling and, to our opinion, sufficient for the navigation of indoor mobile robots.

Issues arose in case the Kinect encounters reflections or mirrors. In both cases, the desired behavior would be to detect plain surfaces for a mirror or a translucent plate. Unfortunately, the reflections do not disturb the sensor's acquisition and mirrors do also reflect infrared light as any other light. Consequently, the Kinect actually returns a virtual image of the scene, when encountering a mirror, instead of a plain surface.

5. CONCLUSION AND FUTURE WORK

This paper was written before Microsoft released their Kinect SDK with the capability to retrieve depth information. To obtain depth with CL SDK a correlation between colored pixels of the video feed and real measurements was developed. This was done empirically with a test platform and a measuring tape which naturally comprises uncertainties. However, a maximum depth resolution of 1.3[mm/color value] has been investigated and a maximum accuracy of 0.02% which means an actual deviation at a distance of 5m of ±1mm.

Also, the video stream could not be addressed directly which led to some workarounds to extract depth information, reducing the real-time performance of the sensor significantly.

The Kinect is already in use, for instance applied to a search and rescue robot at Warwick University. Because of that and the evaluation results, the Kinect might be a game changer and, eventually, will outmaneuver stereo cameras supporting indoor robot navigation. Further evaluation steps might be to attach the Kinect to different mobile robots and to evaluate its functionality in a dynamic environment. This would include object recognition, path planning and obstacle avoidance. Also commands by gestures or spoken words could be implemented to create some sort of master-slave relationship between a human and a robot. This might have some value for show applications or robotic advertisement to the public.

6. ACKNOWLEDGEMENTS

The project has been funded by the Austrian Security Research Programme KIRAS - an initiative of the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit).

7. REFERENCES

Albitar, C., Graebling, P., Doignon, C., 2007. *Robust Structured Light Coding for 3D Reconstruction*. Available at: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp ?arnumber =4408982 , *Accessed 09 June 2011*

Bleyer, M., Rother, C., Kohli, P., Scharstein, D., Sinha, S., 2011. *Object Stereo— Joint Stereo Matching and Object Segmentation.* Available at: http://research.microsoft.com/apps/pubs/default.aspx?id=147300, *Accessed 04 June 2011*

Humenberger, M., Zinner, Ch., Weber, M., Kubinger, W. & Vincze, M., 2010. *A fast stereo matching algorithm suitable for embedded real-time systems*. Computer Vision and Image Understanding, 114(11), pp.1180-1202

Microsoft, 2011. *Programming Guide: Getting Started with the Kinect for Windows SDK Beta.* [pdf] Available at: http://research.microsoft.com/en-us/um/redmond/projects /kinectsdk/guides.aspx, *Accessed 17 June 2011*

Siegwart, R., Nourbakhsh, I. R., 2004. *Autonomous Mobile Robots.* Cambridge: Massachusetts Institute of Technology.

Um, D., Ryu, D., Kal, M. J., 2011. *Multiple Intensity Differentiation for 3D Surface Reconstruction with Mono-Vision Infrared Proximity Array Sensor.* IEEE Sensors Journal [e-journal] PP(99), Available through: IEEE Sensors Council , *Accessed 09 June 2011*