A Comparison between Active and Passive 3D Vision Sensors: BumblebeeXB3 and Microsoft Kinect

Diana Beltran and Luis Basañez

Technical University of Catalonia, Barcelona, Spain
{diana.beltran,luis.basanez}@upc.edu

Abstract. This paper shows a performance comparison of two sensors capable of obtaining depth information using two different methods, i.e. stereo information and infrared based depth measurement. The sensors are a Bumblebee XB3 and a Microsoft Kinect, and they provide in-depth information with some advantages and disadvantages that will be presented and evaluated in this paper. The analysis compares the devices single characteristics and tests their performance.

Keywords: BumblebeeXB3, Kinect, Point Cloud, 3D measurements.

1 Introduction

At the present time there exist different techniques to acquire depth information from a scene. These techniques are generally grouped in two main categories [1]: active and passive. The active group refers to the techniques that use a controlled source of structured energy emission, such as a scanning laser source or a projected pattern of light, and a detector like a camera. A common active vision device is a laser range scanner, where an active source moves around an object in order to scan the entire object surface. These sensors are dense in 3D measurements, but most of them are limited to static environments [2]. Few years ago a new class of active depth sensing systems based on the time-of-flight (TOF) principle, has emerged [3], [4]. The operational principle of these sensors is similar to other laser scanners but their advantage is that they can capture the whole scene at the same time, enabling their use in dynamic scenes applications [5]. The disadvantages of the TOF sensors are the big price and the low resolution. A recent development in active range sensing technology is the Microsoft Kinect sensor [6] that has a good working range, low price, a reasonable resolution and a low computational cost.

On the other hand, the passive techniques do not use a specific structured source of energy in order to form an image and hence, the light source may not be directly used in the range calculation. The basic principle used in recovering 3D information is the triangulation principle. In active vision techniques, a triangle is created between the light, the object and the sensor. In passive stereo vision techniques, the triangle may be created between the object and two sensors. Many kinds of sensors have been

M.A. Armada et al. (eds.), *ROBOT2013: First Iberian Robotics Conference*,
Advances in Intelligent Systems and Computing 252,
DOI: 10.1007/978-3-319-03413-3_54,

designed to acquire depth information using specific techniques making them suitable for a specific application.

The objective of this paper is to analyze a passive and an active sensor in order to show their performance and to emphasize their differences. Some previous comparison of active and passive sensor for acquiring 3D information have been made; for instance in [7] a Kinect and a bumblebe2 are used under different light condition with the objective of appreciating the performance of each sensor under light variation; other interesting comparison can be found in [8] where the accuracy of a PMD and a stereo vision, both systems under optimal condition, is compared.

2 BumblebeeXB3 Stereo Camera

The BumblebeeXB3 is a stereo camera with three sensors that can be used with two different baselines; 24 cm and 12 cm. The 24 cm baseline allows obtaining 3D points with more accuracy in longer ranges making it useful for outdoor applications, while the narrow baseline is more suitable for indoor areas improving close range matching and minimum-range-limitation [9]. Table 1 shows the main specifications of the BumblebeeXB3 camera [10], and Fig. 1 shows an image of the BumblebeeXB3 camera.

Table 1. Main specifications of the BumblebeeXB3 camera and the Microsoft Kinect sensor

BumblebeeXB3 Camera Specifications	
Imaging Sensor	Three Sony ICX445 a/3¨ progressive scan CCD´s
	1280x960 max pixels, 3.75µm square pixel
Baseline	12cm and 24cm
Lens Focal Length	2.5mm with 100° HFVO or 3.8mm with 70°HFVO or 6mm with 50° HFVO
A/D Converter	Analog Device 12-bit analog to digital converter
Video Data Output	8 and 16-bit digital data
Frame Rates	15, 7.5, 3.75, 1.875 FPS

Microsoft Kinect Specifications	
Imaging Sensor	IR Projector. RGB Camera IR Camera
Resolution, Depth Stream	QVGA (320x240)
Resolution, Color Stream	VGA (640x480)
Frame Rate	30 FPS
Mechanical Tilt Rate	±28°
Field Of View	Horizontal: 57° Vertical: 43°
Working Range	1.2 to 3.5 meters

Fig. 1. BumblebeeXB3 camera

2.1 Operation Mode

The process followed by BumblebeeXB3 camera to obtain depth information is similar to the used by a normal stereo vision system. In general, the purpose of stereo vision is to perform range measurements based on the left and right images obtained from stereoscopic cameras. Basically, an algorithm is implemented to establish the correspondence between image features in different views of the scene and then calculate the relative displacement between features coordinates in each image [11].

A general description of the process to obtain 3D points is described next.

Camera Calibration. The first step is to calibrate the camera; the BumblebeeXB3 is accurately precalibrated by Point Grey Research for lens distortions and camera misalignment [12]. The camera calibration process is a necessary step to know the parameters that define the camera model in order to obtain scene measurements from images. The accuracy of the calibration will determine the precision of the measurements obtained from the images.

Matching. To obtain a 3D data set from a scene, it is necessary to solve a crucial problem: find corresponding points in each image; i.e., points in image A, $P(x, y)$, and image B, $P'(x', y')$, corresponding to the same point of the scene. This is a difficult process because it is possible to have areas of the scene where there is no solution (occluded areas) or where exist multiple solutions (areas without enough texture).

Reconstruction. The reconstruction is the method by which the spatial layout of a scene can be recovered from two views. Therefore, once the corresponding points are found (P, P'), the final step is to find the depth of the points $Z = bf/d$, where b is the baseline, f is the focal length and d is the disparity $d = x' - x$.

3 Microsoft Kinect

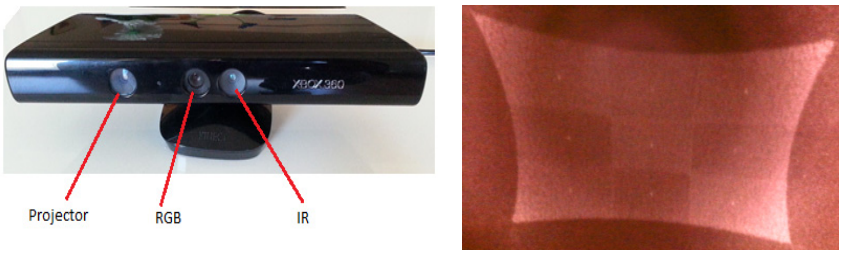
The Kinect sensor was originally designed as a game interface, but the robotics community has seen in it an interesting 3D sensor for robotics application. This sensor is being used by the researchers because of its high quality and low price. Some applications where the Kinect sensor is being used are: 3D reconstruction, face detection, slam, object detection, and augmented reality, among others.

Table 1 shows the main specifications of the Microsoft Kinect. Fig. 2a shows the sensor´s placement on the device. Each lens is associated with a camera or a projector [13]. The RGB sensor has a resolution of 640x480 pixels and a frame rate of 30 fps and the infrared (IR) camera has a resolution of 320x240 pixels and a frame rate of 30 fps.

3.1 Operation Mode

The Kinect depth is calculated by triangulation against a known pattern from the projector, (Fig 2b). The pattern is memorized at a known depth. For a new image, at

each pixel in the IR image, a small correlation window is used to compare the local pattern at that pixel with the memorized pattern. The best match gives an offset from the known depth, in terms of pixels: this is called disparity. The Kinect device performs a further interpolation of the best match to get sub-pixel accuracy of 1/8 pixel. Given the known depth of the memorized plane and the disparity, an estimated depth for each pixel can be calculated by triangulation [13].



(a) (b)

Fig. 2. (a) Kinect sensors placement, (b) Pattern projected from the Kinect

3.2 Disparity to Depth Relationship

In a normal stereo system, the cameras are usually calibrated so that the rectified images are parallel and have corresponding horizontal lines. At zero disparity, the rays from each camera are parallel, and the depth is infinite. Larger values for the disparity mean shorter distances [13].

In the Kinect case, it returns a raw disparity that is not normalized in this way, that is, a zero Kinect disparity does not correspond to infinite distance. The Kinect disparity is related to a normalized disparity by the relation:

$$d = (d_{off} - k_d)/8$$

Where d is a normalized disparity, k_d is the Kinect disparity, and d_{off} is an offset value particular to a given Kinect device. The factor 1/8 appears because the values of k_d are in 1/8 pixel units [13].

4 Experiments

In order to show the differences between the stereo camera BumblebeeXB3 and Microsoft Kinect sensor, a real experiment has been done for comparison. (The experiment has been performed during the afternoon in order to have enough external light, because the BumblebeeXB3 does not work without enough light).

For a specific scene, the aim is to obtain the disparity image and the points cloud given by the BumblebeeXB3 and the Kinect. In order to compare the same area of a scene the cameras has been placed according to the setup of Fig. 3.

Fig. 3. Kinect and BumblebeeXB3 setup

4.1 BumblebeeXB3 Images Acquisition

The Robot Operating System (ROS) has been used to perform the experiments, [14]. ROS has a package for the Bumblebee2, the previous version of the BumblebeeXB3, so it has been necessary to make some changes in this code to adapt it to the BumblebeeXB3. The adapted package allows to obtain the raw images from each of the sensors of the camera (left, center and right) and, adding the parameters obtained from the stereo_calibration ROS package, it is possible to obtain calibrated images from each stereo pair (left-center, right-center, left-right). To get the disparity image and the 3D point cloud it was necessary to use the stereo_image_proc ROS package. In Fig. 4, the left and right images acquired from the BumblebeeXB3 using its maximum resolution are showed.

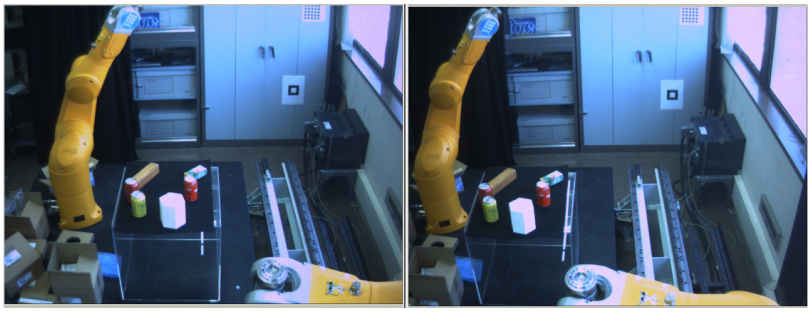


Fig. 4. Left and right images from BumblebeeXB3

To obtain the disparity image, *stereo_image_proc* uses the class *stereobm* of *opencv*. Fig. 8a shows the disparity image obtained using the left and right images of the BumblebeeXB3 (Fig. 4). The *stereobm* class computes stereo correspondence using the block matching algorithm. The class uses some parameters that define the size of disparity range for an optimal search and determine the size of the averaging window used to match pixel blocks. These parameters and the prefiltering and postfiltering parameters could be changed by the user through the *dynamic_reconfigure* ROS packages with the window showed in Fig. 5.

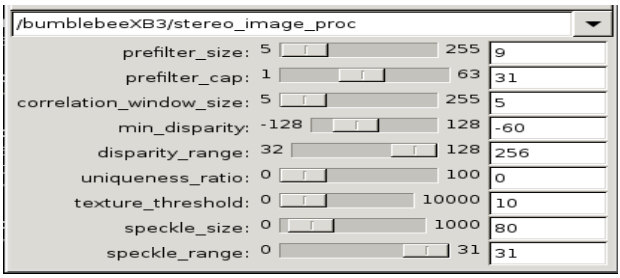


Fig. 5. Dynamic reconfigure window

4.2 Kinect Images Acquisition

For the Kinect sensor, the *Openni_camera* ROS package has been used. This package allows obtaining the raw IR image, the raw RGB image, the disparity image and the 3D point cloud. Using the *image_calibration* ROS packages it is possible to get the calibrated IR image and the calibrated RGB images (Fig. 8) and, in consequence, obtain a better disparity image (Fig. 7b) and a 3D point cloud (figure 11b).

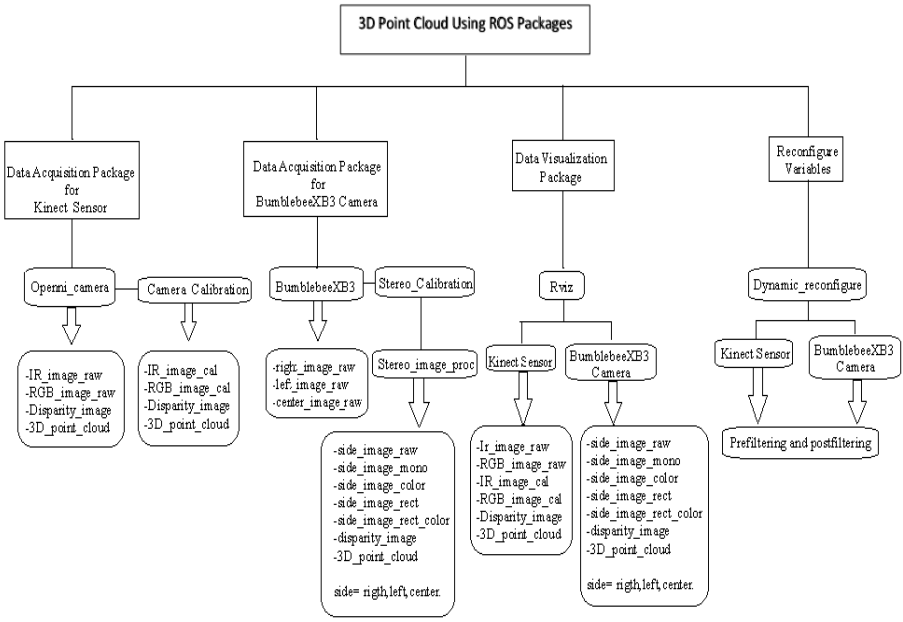
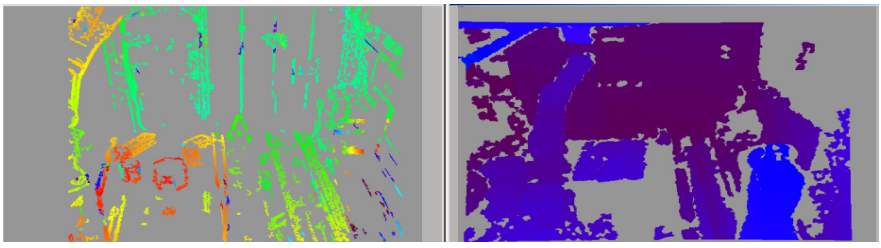


Fig. 6. Software used to obtain 3D points from BumblebeeXB3 and Kinect

5 Experimental Analysis

In order to visualize the images captured by the BumblebeeXB3 and the Kinect sensor, the *image_view* and *rviz* packages has been used. Fig. 6 shows an overview of the software used to obtain the 3D points from the two sensors.

Analyzing the disparity images from the Kinect and the BumblebeeXB3 sensors (Fig.7), the differences between them are evident. By one side the BumblebeeXB3 generates a disparity map (the colors of the disparity map indicate the distance of the objects from the camera) without information in the textureless areas, but describing the edges of the transparent objects. By the other side, the Kinect sensor generates a disparity maps with rich information in textureless areas but has problems with reflective objects.



(a) (b)

Fig. 7. (a) Disparity map from BumblebeeXB3, (b) Disparity map from Kinect

Fig. 11 presents the point cloud images obtained from the Kinect and the BumblebeeXB3 cameras. Since the disparity is proportional to the depth ($Z = bf/d$), it is coherent that in the point cloud there are not points in the areas of each image where the corresponding sensor has no information (explained above). In order to have some real distance measurements, several points from the scene have been chosen, and their real distance from each camera has been obtained with the digital laser distance meter Bosch DLE50 professional that has a range from 0.05m to 50m and a precision of ±1.5mm.

(a) (b)

Fig. 8. (a) IR image from Kinect, (b) RGB image from Kinect

Table 2. Depth measurements

Sample	Real World (cm)	Kinect (cm)	Error(cm)	BumblebeeXB3	Error (cm)
1	145,8	146,1	0,300	NaN	NaN
2	164,0	164,9	0,900	1,659	1,900
3	169,0	170,6	1,600	1,710	2,000
4	175,0	NaN	NaN	1,728	2,250
5	184,6	183,3	1,300	1,867	2,100
6	193,4	196,9	3,500	1,956	2,200
7	206,9	207,6	0,700	2,058	1,100
8	207,9	208,8	0,900	2,092	1,300
9	233,7	235,9	2,200	2,314	2,300
10	241,4	249,6	8,200	2,431	1,700
11	247,1	251,4	4,300	2,484	1,300
12	252,8	229,7	23,100	NaN	NaN
13	260,1	266,9	6,800	2,626	2,500
14	299,1	301,9	2,800	2,957	3,400
15	480,8	490,7	9,900	4,781	2,700
16	490,4	483,8	6,600	4,749	15,500

Table 2 allows a comparative analysis of the accuracy of both cameras. In some areas the BumblebeeXB3 works better than the Kinect camera and in other areas the Kinect works better than the BumblebeeXB3; it is due to the different specifications of each camera, the characteristics of the scene and the intrinsic parameters like radial distortion; in points that are around the optical center the accuracy is better, but as one moves away from the center the accuracy decreases. In point 1, the BumblebeeXB3 has no information because this point is out of the working space of the camera (using the 24cm baseline), while point 4 lies in an area where the Kinect has lost information because of an external interference. In point 16 the BumblebeeXB3 measurements has a big error because this point is in an area of low texture and at big distance away from the camera. By other hand, point 12 has a big error because this area had high external illumination and the Kinect is affected for it. In the case of the BumblebeeXB3 this point 12 is out of its field of view. In Fig. 11 it is possible to appreciate the 3D point cloud used to make the measurements of both cameras.

Fig. 9 and Fig. 10 show a dispersion of the points in the plane (x,y) (each point has the sample´s number and its corresponding error). It can observe how near or far are the points from the optical center of each 3D camera. Points near to the optical center of the camera should have less error, because of less radial distortion, but it is not the only factor that affect the quality of the measurements of a point The influence of

external light, the camera calibration and the distance from the sensor are also important factors to take into account; that is why some points, despite being close of the optical center, have higher error than others.

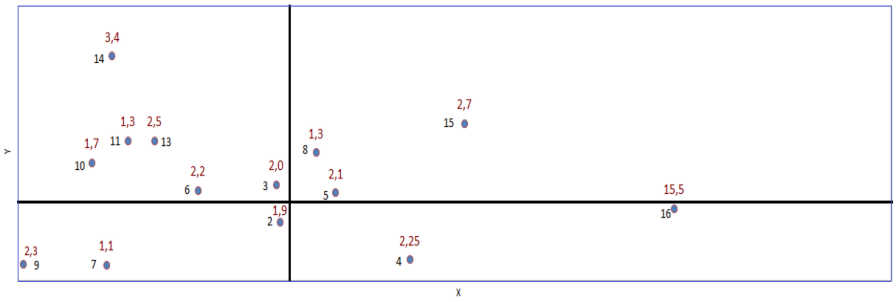


Fig. 9. Spatial distribution of the points in Bumblebee's reference frame

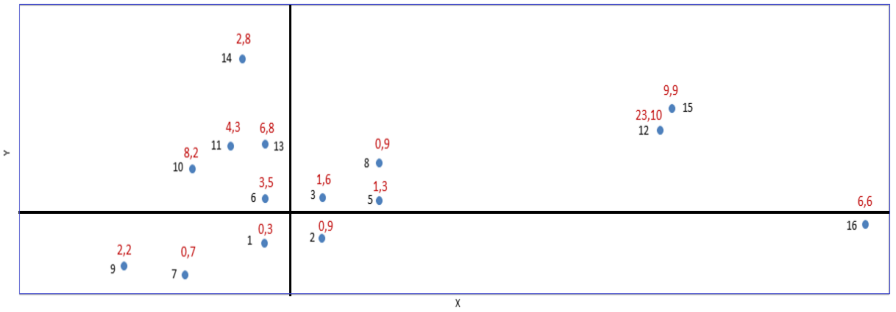
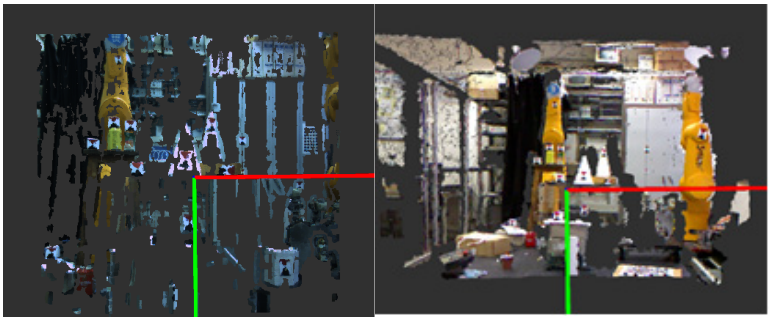


Fig. 10. Spatial distribution of the points in Kinect's reference frame



(a) (b)

Fig. 11. (a) Point cloud from BumblebeeXB3, (b) Point cloud from Kinect

6 Conclusions

The BumblebeeXB3 is a camera with high accuracy in areas with enough texture, has a high resolution and a big field of work, making it appropriate for indoor and outdoor scenes. Its disadvantages appear in textureless, low illuminated areas. By the other side, the Kinect sensor has high accuracy in textureless areas, is faster and works well in its work space (1,2-3.5 meters), its resolution is enough to obtain good 3D measurements and it is a sensor with low computer cost (because, most of the process is performed in the device chip). But it is not appropriate for outdoor applications because its sensibility to the interferences by external light, and it loses information in reflective objects. With this analyze we can conclude that these two sensor complement each other making think the fusion of their information as a future work.

References

1. Srrenivasa, K., Mada, M., Smith, L., Smith, S.: An Overvire of Passive and Active Vision Techniques for Hand -Held 3D Data Acquisition. In: Proceedings of SPIE Opto-Ireland: Optical Metrology, Imaging, and Machine Vision, vol. 4877 (2003)
2. Salvi, J., Pages, J., Batlle, J.: Pattern Codification Strategies in Structured Light Systems. Pattern Recognition 37(4), 827–849 (2004)
3. 3DV systems, z-cam, http://inition.co.uk/3D-Technologies/3dv-systems-zcam
4. MESA Imaging, http://www.mesa-imaging.ch/
5. Pmd technologies, http://www.pmdtec.com/
6. Microsoft Kinect, http://www.xbox.com/en-us/kinect/
7. Beder, C., Bartczak, B., Koch, R.: A Comparison of PMD-Cameras and Stereo-Vision for the Task of Surface Reconstruction using Patchlets. In: Proceedings of the Second International ISPRS Workshop BenCOS (2007)
8. Tarfin, S., Bugra, O., Bugra, A., Konukseven, E.: Comparison of Kinect and Bumblebee2 in Indoor Environments. Midle East Technical University (2011)
9. Point Grey, http://www.avsupply.com/Point-Grey/bumblebee-xb3.php
10. Point Grey datasheet, http://ww2.ptgrey.com/stereo-vision/bumblebee-xb3
11. Nguyen, S., Nguyen, T.H., Nguyen, H.T.: Semi-autonomous Wheelchair System Using Stereoscopic Cameras. In: 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minesota, USA (September 2009)
12. Rossu, L., Molinier, T., Akhloufi, M., Tison, Y.: Measurement of Laboratory fire Spread Experiments by Stereovision. In: Image Processing Theory, Tools and Applications. IEEE (2010)
13. Kinect_calibration, http://www.ros.org/wiki/kinect_calibration/technical
14. ROS Packages, http://www.ros.org/browse/list.php