Experimental Characterization of Two Generations of Kinect's Depth Sensors

Rizwan Macknojia, Alberto Chávez-Aragón, Pierre Payeur, Robert Laganière
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
[rmack102, achavez, ppayeur, laganier]@uottawa.com

***Abstract*—This paper presents an analysis and experimental study of the Kinect's depth sensor. Particular attention is given to resolution, quantization error and random distribution of depth data. The effects of color and reflectance characteristics of the object are also analyzed. The study examines two versions of Kinect sensors, one dedicated to operate with the Xbox 360 video game console and the more recent Microsoft Kinect for Windows version.**

***IndexTerms*—RGB-D imaging, depth measurement, Kinect for Xbox, Kinect for Windows.**

I. INTRODUCTION

Since its introduction in 2010, the Kinect sensor has earned extensive popularity in gaming and natural user interface (NUI) applications. The Kinect sensor is a multi-view system, containing one RGB camera, one infrared (IR) camera and one infrared laser projector. Kinect is capable to collect depth information for each pixel mapped into a color image, which opens the door to a great variety of applications from 3D modeling to robotic vision. Two different directions of research have attracted special attention using Kinect technology: *i*) the investigation of the technology behind the device, including the analysis of its properties, performance, and comparison with other similar devices, and *ii*) the development of applications for Kinect in fields such as robotics, user interfaces, and medicine, among others. This investigation was motivated by the fact that originally the sensor was launched as a peripheral for the Xbox 360 video game console and very limited information about the sensor technology itself and its performance in other contexts was provided by the manufacturer.

From the first release which was dedicated to operate in conjunction with the Xbox 360 video game console, lots of research has been carried out on NUI applications. As a result, Microsoft recently introduced a second generation of the Kinect sensor, which looks like the original Kinect for Xbox 360 sensor, but is designed to work at closer range. This paper examines the mathematical model and quality of depth data provided by both the *Kinect for Xbox 360* sensor and the new Microsoft *Kinect for Windows* version. Additionally, a mathematical model to convert raw depth data into real world depth measurements is introduced.

The paper is organized as follows. Section 2 describes related work. Section 3 explains the operation of a Kinect's depth sensor. Section 4 presents an analysis of the depth resolution, XY resolution and the quantization error of the Kinect's depth sensor. Section 5 reports on experimental results and evaluation. Finally, Section 6 presents some conclusions and future work.

II. RELATED WORK

In 2010 Microsoft introduced the *Kinect for Xbox 360* sensor as an affordable and real-time source for medium quality textured 3D data dedicated to gesture detection and recognition in game controller. Since then, numerous researchers have recognized the potential of this RGB-D imaging technology, especially because of its speed of acquisition, and attempted to integrate it in a broad range of applications. More recently, in 2012, a second generation of Kinect sensors, the *Kinect for Windows*, was introduced with slightly different characteristics, which is meant to better support the research community.

Among the numerous examples of applications for the Kinect technology that rapidly appeared in the literature, Zhou *et al*. [1] propose a system capable of scanning human bodies using multiple Kinect sensors arranged in a circle. Maimone and Fuchs [2] present a real-time telepresence system with head tracking capabilities based on a set of Kinect sensors. They also contribute an algorithm for merging data and automatic color adjustment between multiple depth data sources. An application of Kinects in the medical field for position tracking in CT scans is proposed by Noonan *et al*. [3]. They track the head of a phantom by registering Kinect depth data to high resolution CT template of a head phantom. Rakprayoon *et al*. [4] use a Kinect sensor for obstacle detection of a robotic manipulator.

The Kinect measures the depth with the help of an IR projector and an IR camera. Therefore, a frame of reference for depth measurement is defined with respect to the IR camera. The color and IR cameras in the Kinect are separated by a fixed baseline. In order to merge data collected from different Kinect sensors, various approaches have also been proposed for simultaneous calibration of Kinect's sensors [2][5][6][7][8].

The depth data of the Kinect sensor is also known to suffer from quantization noise [9][10], that increases as the distance to the object increases. The resolution also decreases with the distance [10]. The depth map may also contain occluded and missing depth areas mainly due to the physical separation between the IR projector and the IR camera, and to the inability to collect sufficient IR signal reflection over some types of surface. These missing values can however be approximated by filtering or interpolation [2][11].

The work presented here consists of an experimental characterization of Kinect's RGB-D sensors. This evaluation is conducted to formally determine the quality of data that can be expected from the sensor and the limitations of the sensing technology for integration in advanced applications where relatively accurate 3D reconstructions are required over objects with a large volume, which motivates the use of rapid RGB-D sensor like Kinect. The evaluation covers Kinect's quantization error, random distribution of data, as well as the response to color and reflectance characteristics of the depth sensor. Also, an experimental comparison between the first two generations of Kinect RGB-D sensors, *Kinect for Xbox 360* and *Kinect for Windows*, is presented.

III. OPERATION OF KINECT'S DEPTH SENSOR

The Kinect sensor operates on the principle of structured light range sensors. The structured pattern projected by a Kinect unit is within the infrared spectrum and cannot be seen by the human eye. The pattern is predefined and fixed. It is generated by projecting optical radiation through a transparency containing the engraved pattern [12]. The infrared camera contained in the Kinect captures the projected light pattern and compares it against the predefined reference pattern [12]. Fig. 1 shows the reference pattern created on a flat surface, as captured by the Kinect's IR camera. As can be noticed, the intensity of the IR radiation is greater in the center of the structured light map created on the surface when compared with the corner areas.

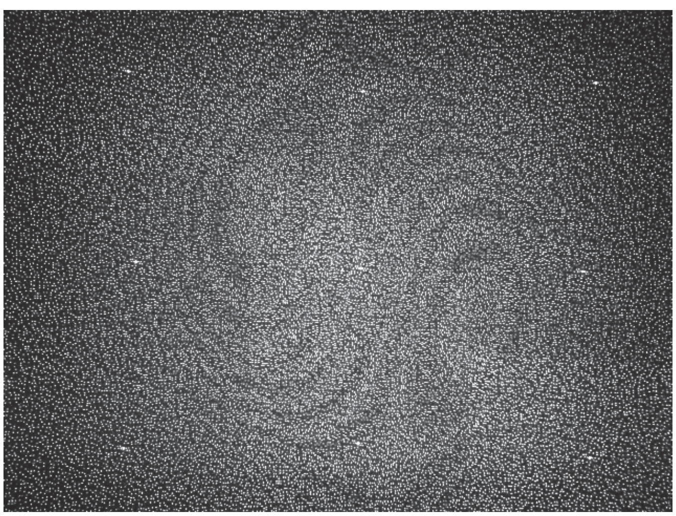

Fig. 1. Structured light pattern projected on an object.

The depth calculation is performed through a triangulation process illustrated in Fig. 2. The depth, Z, of a point, P, on an object point is expressed from the perspective center of the infrared camera. The Z axis of the IR camera is orthogonal to the image plane and passes through the optical center in the direction of the object. The IR projector is aligned with the X axis of the IR camera. The distance between the optical center of the IR camera and that of the IR projector is defined by the baseline, b. The reference plane in the schematic diagram of Fig. 2 corresponds to the projected IR image shown in Fig. 1 when no object is present in the scene. This reference image contains a predefined pattern which consists of tiny dots imaged in the IR spectrum and is stored in the firmware of the sensor during the manufacturing process. The introduction of an object in the field of view of the Kinect depth sensor deforms the reference pattern. The resulting shift of the dots in the pattern from the reference image is measured as disparity in the image plane of the IR camera. The shift essentially takes place along the X axis because the IR camera and the IR projector are parallel and only translated along the X axis, with a separation defined by the baseline, b. Eq. (1) defines the relationship between the distance to the object, Z, and the raw disparity, d.

$$Z = \frac{fb}{d} \tag{1}$$

where Z is the depth of the object from the optical center of the IR camera in the direction of the Z axis (in meters), b is the horizontal baseline between the IR camera and the IR projector (in meters), f is the focal length of the IR camera (in pixels), and d is the raw disparity (in pixels).

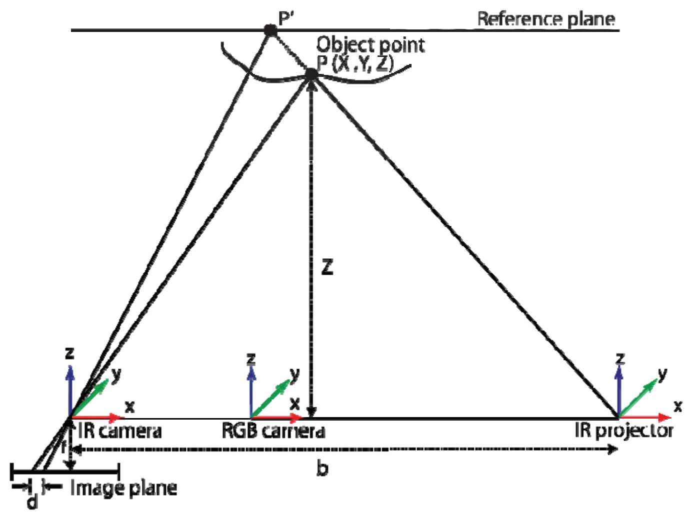

Fig. 2. Depth and disparity relationship used for depth estimation.

In order to estimate disparity, the Kinect sensor calculates the correlation between points on the reference pattern, P', and points on the deformed pattern, P, produced by the object in the scene. The correlation is performed using a window of size 9x9 or 9x7 pixels [8][9]. The results of the matches found by correlation are stored as a raw disparity image. However, Kinect does not return the raw disparity image. Instead it provides the normalized disparity, d', between 0 and 2047 (11 bit integer). Eq. (2) defines the relationship between the raw disparity, d, and the normalized disparity, d', which is finally transposed in Eq. (1) to estimate the depth of a given point marked by the IR pattern over the surface of the object, as shown in Eq. (3).

$$d = md' + n \tag{2}$$

$$Z = \frac{fb}{md' + n} \tag{3}$$

where m and n are the factors for denormalization. The typical values for those parameters are: m = 0.125, n = 135.25, b = 7.5cm and f = 580 pixels [8].

IV. Theoretical Depth Resolution And Quantization

For both versions of the Kinect sensor, the depth resolution depends on the resolution of the disparity image. Kinect returns an 11 bit disparity output. Therefore, it contains 2048 levels of disparity in principle. Nevertheless, *Kinect for Xbox 360* mostly returns disparity values between 400 and 1050, which restrict the output between 50cm and 850cm. Below the minimum range the IR pattern looks like a bright spot with no regular pattern to establish correlation between the reference pattern and the actual IR pattern. The disparity value of 2047 is returned where no disparity can be estimated because the object is out of the sensor's depth of field or no correlation between the reference and the actual IR marks can be established in some areas of the scene. The range of disparity value for *Kinect for Windows* is not known because it is only supported by Microsoft SDK and a hardcoded limitation is applied to the final output. The output range is between 40cm and 300cm for *Kinect for Windows*. The depth resolution of a Kinect sensor decreases non-linearly as the distance between the object and the measuring device increases. The relationship between depth resolution and the distance can be calculated using Eq. (3), where Z is the actual distance in the real world and d' provides the normalized disparity in the range (0-2047). The relationship between those variables is shown in Fig. 3. Kinect returns 350 disparity levels between 0.5 m and 1 m (closest operational range), while only 23 disparity levels are available to quantify depth between 2.5 m and 3 m (furthest operational range).

The distance between two possible consecutive depth values determines the quantization error. It is obvious from Fig. 3 that the quantization step size increases with distance because of the reduction in the depth resolution. The difference between each consecutive value is plotted in Fig. 4. These values demonstrate the quadratic relationship between the distance and the quantization step. The best fit quadratic curve, *Poly*, is also shown in Fig. 4. The quantization step, q, as a function of distance, Z, is defined in Eq. (4), which is obtained by fitting a quadratic polynomial on the data.

$$q(Z) = 0.302\, Z^2 - 0.056\, Z + 0.0307 \tag{4}$$

where Z is in meters and returns the quantization step, q, in cm. Since the valid operating region is typically between 0.5m and 3.0m, the corresponding quantization error on depth measurements ranges between 0.07cm and 2.58cm.

The XY resolution (related to the density of depth measurements over the depth map) depends on the resolution of the depth image. The depth image has a fixed size of 640x480 pixels, therefore the resolution of the points projected on the XY plane also depends on the distance between the object and the Kinect sensor. The distance between two consecutive pixels in a real world is plotted in Fig. 5 for the depth range between 50cm and 800cm. These values demonstrate the linear relationship between the quantization step size in (X, Y) and the depth, which can be defined as follows:

$$q_x(Z) = q_y(Z) = \frac{Z}{f} \tag{5}$$

where Z is the depth of a pixel and f is the focal length of the IR camera. An object of size 40x40cm is mapped by 149769 points at 60 cm away from the sensor, while the same object supports only 6084 points if located at 3m.

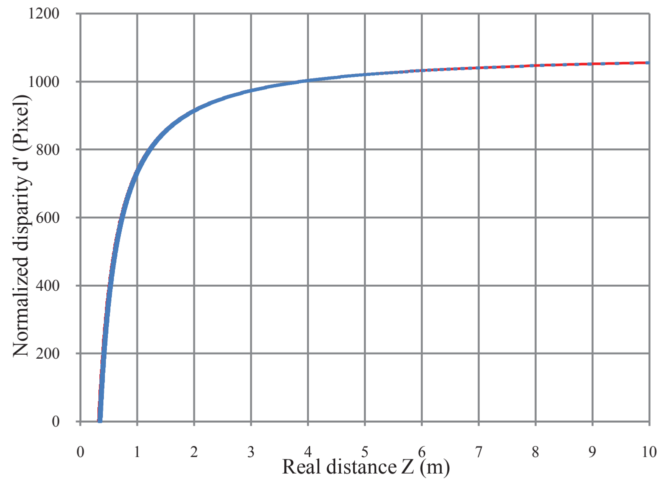

Fig. 3. Kinect normalized disparity for various distances.

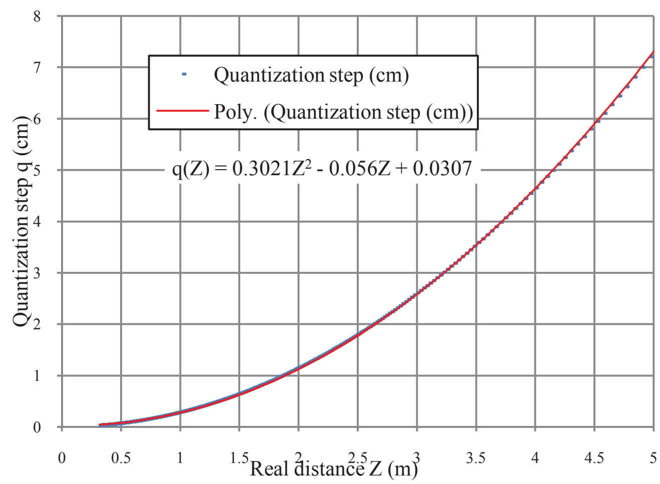

Fig. 4. Quantization step size for various distances and best quadratic curve fit.

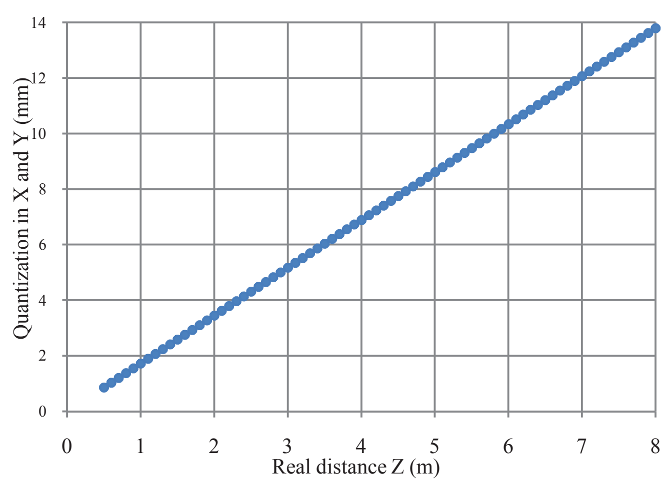

Fig. 5. Quantization step size in X and Y for various distances.

V. Experimental Evaluation

Experiments were performed respectively with a *Kinect for Windows* and a *Kinect for Xbox 360* in order to evaluate the performance of both generations of the RGB-D sensor and compare with the theoretical expectations detailed in section IV. The Microsoft SDK is used to capture the data from the *Kinect for Windows*, while OpenNI is used with *Kinect for Xbox 360*. There are two operating modes for *Kinect for Windows*, i.e. the default mode and the near mode. *Kinect for Windows* works exactly as *Kinect for Xbox 360* in the default mode, but in the near mode it provides depth data as close as 40cm. However, Microsoft SDK also limits the range of Kinect data. With this software package, the Kinect sensor range in default mode is between 80cm and 400cm, while in near mode the range is between 40cm and 300cm. Therefore the data is only available over a given range for the experiments. Recently OpenNI added preliminary support for the *Kinect for Windows*, but only in default mode. OpenNI does not impose any limit on the output and provides data between 50cm and 850cm for both Kinect devices. All the experiment are performed using the default parameters used by either OpenNI or Microsoft SDK

A. Depth evaluation

To study the depth evaluation of the two generations of Kinect sensors, each Kinect sensor is placed on a plane surface in front of a wall. The Kinect IR camera is positioned parallel to the wall. Then 3D points are captured over the plane. The Kinect is moved repeatedly between 0.4m to 3.5m away from the wall, with an interval of 10cm. At each interval 100 depth images are captured and the whole process is repeated 5 times. The setup of the experiment is shown in Fig. 6.

The recorded depth measurements at each distance not only contain the exact depth of the plane but also some depth values with slight errors. Fig. 7 shows the depth measurements distribution for a wall located 3m in front of the *Kinect for Xbox 360*. Although the Kinect's IR camera plane is kept parallel to the planar target to minimize the error, the depth exhibits significant variations. The distribution of points over specific slices clearly shows the impact of the quantization error, which is estimated at about 2.5cm at 3m. The quantization step size observed on each recorded value is plotted in Fig. 8 for both generations of Kinect sensors. The quantification error on both Kinect sensors approximately follows the same quadratic curve defined in Eq. (4), with a slight overestimation of the quantization error being perceptible for the *Kinect for Xbox 360* version.

In the above experiment, 500 depth images were captured for each interval. The standard deviations of the distribution is calculated for each interval and plotted in Fig. 9. The standard deviation of both Kinect sensors increases fairly linearly within 1.0m and becomes more erratic beyond that distance. This distribution is random but its magnitude dependent on quantization error. Therefore, it is also proportional to the square of distance from the object. Furthermore, *Kinect for Windows* does not compromise the quality of data below 50cm in near mode, which allows for slightly closer acquisition. These experiments demonstrate that it is preferable to operate both types of Kinect sensors within a 2m distance from the object to ensure coherence of measurements with a maximum standard deviation of around 1cm.

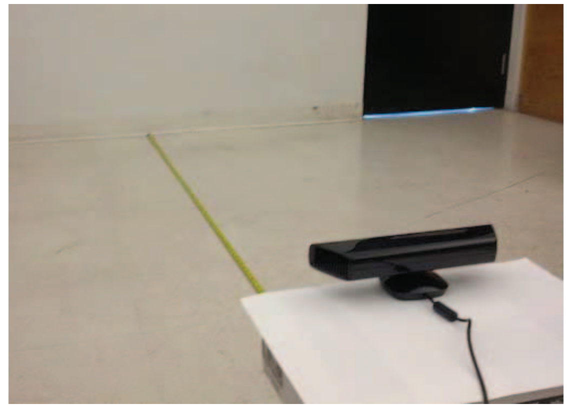

Fig. 6. Setup for depth evaluation.

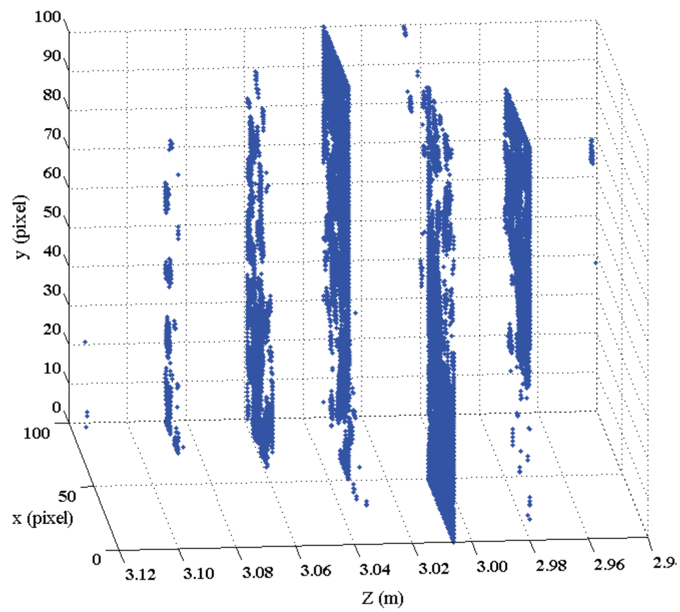

Fig. 7. Distribution of depth values of Kinect for Xbox 360 over a plane at 3m.

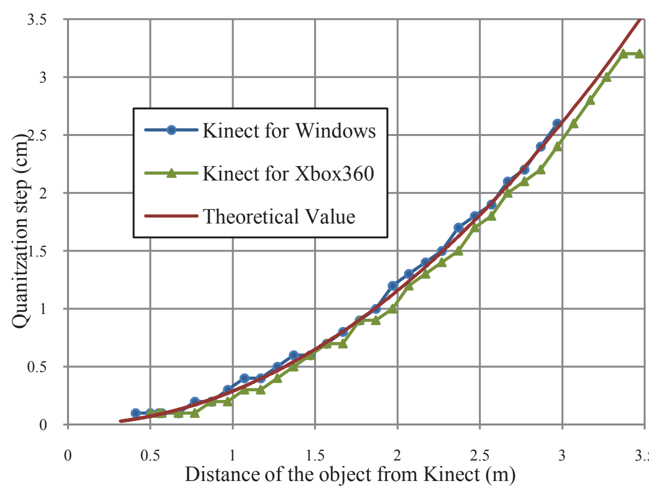

Fig. 8. Experimental quantization step size as a function of distance.

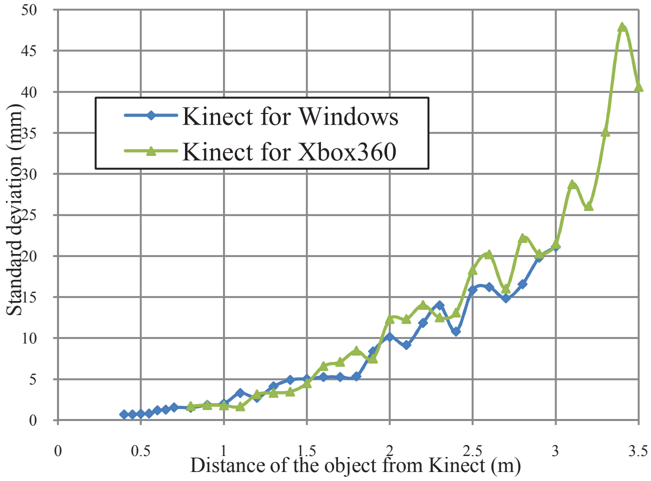

Fig. 9. Standard deviation of depth with distance.

B. Sensitivity to object's color and reflectance characteristics

The way the Kinect sensor responds to various colors and reflectance characteristics when creating depth maps is also a recognized issue with this technology. The second set of experiments conducted aims at determining the range of capabilities of the sensor when operating on objects with various colors and reflectance characteristics. Since Kinect uses a structured pattern of IR light that needs to reflect back into the IR camera, the amount of IR energy reflected toward the IR camera, and the sharpness of the imaged IR pattern, directly influence the density and the accuracy of the 3D reconstruction.

Fig. 10 illustrates the response of the IR projection on a scene containing a white wall, a black door and various other objects with different colors and textures. The Kinect is placed 2m away from the door. In the first scene, the Kinect sensor is positioned parallel to the door. The whole scene is illuminated by the IR projector pattern under standard fluorescent indoor lighting. We notice that the top and bottom parts of the door are not properly imaged in the disparity map, as shown by black pixels (missing points) in the lower part of Fig. 10(a). These regions correspond to areas over which most of the IR energy is absorbed by the surface of the object, as shown in the IR image in the middle part of Fig. 10(a).

Testing with a different configuration, shown in Fig. 10(b), the Kinect sensor is shifted and rotated, such that the door is off centered in the image plane. The black door is therefore imaged over a section of the IR pattern where the IR radiation is not as intense as in the center of the projected pattern, as shown in Fig. 1. In this case the entire door is missing in the range image, as shown in Fig. 10(b), except for the signs attached on the door which are more reflective.

The evaluation was further refined by studying the impact of objects' color and reflectance characteristics over the operating field of view of the Kinect sensor. To further evaluate the operational field of view of the sensor with respect to the distance of the object from the Kinect sensor, black and white objects were used. The black color is known to absorb more radiation, while white objects tend to reflect a larger fraction of the energy that they receive.

(a) (b)

Fig. 10. Color response of Kinect depth camera on a black door. (a) Color, IR and depth images respectively, when Kinect is aligned parallel to the door, (b) Color, IR and depth images respectively, when Kinect's pose is modified.

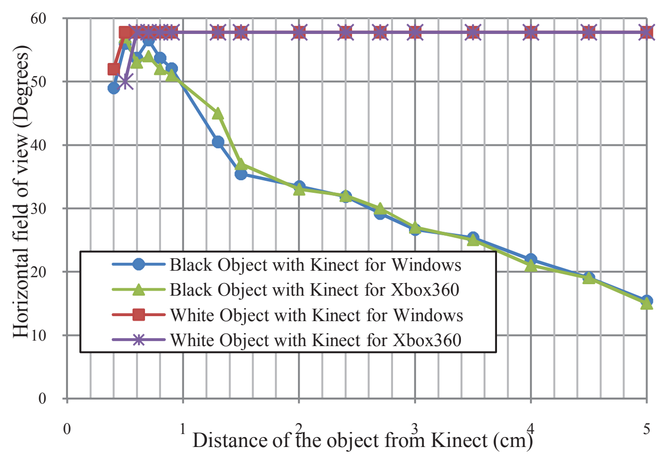

Fig. 11. Response of the Kinect sensor over surfaces with different colors and reflectance characteristics. Angles show the effective area seen by the depth sensor.

An experiment was performed using both generations of Kinect sensors. As before, in each case the Kinect sensor was placed on a flat surface and aligned parallel to a white wall and a black door. Range data was collected at different distances by moving the Kinect sensors between 0.4m and 5m. The Kinect depth sensor's field of view typically covers 57 degrees horizontally and 43 degrees vertically. However, black surfaces appear to further limit the viewing angles. In this experiment, it was found that the portion of the door visible in the depth image is further reduced as the distance between the sensor and the door increases.

The horizontal angular field of view within which a portion of the door appears in the depth image is calculated for each interval and is plotted in Fig. 11. Conversely, white objects keep being imaged over the complete field of view of 57 degrees for both Kinect sensors independently from the distance. As black objects absorb the IR energy and its intensity is further reduced as objects move farther away from the sensor, it results in some clipping over such dark and absorbing objects. Therefore the horizontal viewing angle is progressively reduced, as shown in Fig. 11. A similar characteristic is observed in the vertical direction.

Finally, to investigate the response of the Kinect sensor over different types of material with complex reflectance characteristics, depth maps were captured over a vehicle in an underground garage with artificial ambiant lighting. Fig. 12 presents two of the views of the vehicle with the corresponding reconstructions. The windshield and lateral windows of the vehicle are entirely missing in the depth map because the IR energy passes through the transparent surfaces or is deflected in other directions. However, the rear window of the vehicle, which is made of tinted glass, is partially captured. The transparent headlamp is not captured while the rear lamp reflector, which partially returns the IR pattern back to the sensor, is successfully measured apart from some holes in the reconstructed model. The front wind deflector, made of black shiny plastic gets only partially reconstructed in the 3D model, depending on the relative orientation of the sensor to the sections of the curved deflector. Finally, all of the main areas of the vehicle body and wheels, including dark rubber tires, are very accurately reconstructed, even over narrow roof supporting beams and highly curved bumpers areas. It is also noticeable that the sensor handles well the shiny painted surface of the vehicle in spite of some glare from the overhead lighting.

Fig. 12. Two views of a vehicle (top) and corresponding 3D reconstructions (bottom), with missing depth areas depicted in white.

VI. Conclusion

In this paper, we have presented an experimental study of the Kinect's depth sensor, and compared the performance of two generations of Kinect devices. The results exhibit the main characteristics of the operating conditions for the respective devices. The depth quantization error is found to be similar for both devices and increases quadratically with the distance of the object from the device. The standard deviation of depth measurements shows that the quality of data collected with the *Kinect for Windows* version does not deteriorate between 40cm and 50cm, a range that is not accessible with the *Kinect for Xbox 360* version. These experiments also demonstrate that the best functional operating range for both devices can be up to 2m with standard deviation and quantization error bounded within about 1cm. Finally, both Kinect generations tend to severely degrade performance by not providing substantial depth information over black or dark color objects with low reflectance characteristics. Proper perception of depth over low reflectance objects is confined to the center of the field of view and rapidly degrades with the increase in the distance from the object. Furthermore Kinect sensors should be positioned parallel to such absorbing object surfaces to provide reliable depth measurements. Finally, an experimental demonstration was made that the Kinect technology can perform fairly well to acquire relatively accurate 3D reconstructions over objects with a large volume, non transparent but shiny surfaces, and highly curved shapes.

References

[1] J. Zhou, L. Liu, Z. Pan and H. Yan, "Scanning 3D Full Human Bodies Using Kinects," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, no. 4, pp. 643 - 650, 2012.

[2] A. Maimone and H. Fuchs, "Encumbrance-free Telepresence System with Real-time 3D Capture and Display using Commodity Depth Cameras," in *IEEE International Symposium on Mixed and Augmented Reality*, pp.137 - 146, 2011.

[3] P. J. Noonan, T. F. Cootes, W. A. Hallett and R. Hinz, "The Design and Initial Calibration of an Optical Tracking System using the Microsoft Kinect," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 3614 - 3617, 2011.

[4] P. Rakprayoon, M. Ruchanurucks and A. Coundoul, "Kinect-based Obstacle Detection for Manipulator," in *IEEE/SICE International Symposium on System Integration*, pp. 68 - 73, 2011.

[5] K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, A. Scholzk and M. Magnor, "Markerless Motion Capture using Multiple Color-Depth Sensors," in *Proc. Vision, Modeling and Visualization*, pp. 317 – 324, 2011.

[6] N. Burrus, "RGBDemo," [Online]. Available: http://labs.manctl.com/rgbdemo/. [Accessed 10 07 2012].

[7] G. Mark, "Kinect/3D Scanner Calibration Pattern," [Online]. Available: http://www.thingiverse.com/thing:7793. [Accessed 12 07 2012].

[8] K. Konolige and P. Mihelich, "Technical description of Kinect calibration," [Online]. Available: http://www.ros.org/wiki/kinect_calibration/technical. [Accessed 2 07 2012].

[9] J. Smisek, M. Jancosek and T. Pajdla, "3D with Kinect," in *IEEE International Conference on Computer Vision Workshops*, pp. 1154 - 1160, 2011.

[10] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data," in *ISPRS Workshop on Laser Scanning*, pp. 1437 - 1454, 2011.

[11] S. Matyunin, D. Vatolin, Y. Berdnikov and M. Smirnov, "Temporal Filtering for Depth Maps Generated by Kinect Depth Camera," in *3DTV Conference on The True Vision - Capture, Transmission and Display of 3D Video*, pp. 1 - 4, 2011.

[12] B. Freedman, A. Shpunt, M. Machline and Y. Arieli, "Depth Maping Using Projected Patterns". United States Patent US 2010/0118123 A1, May 2010.