Depth Study of Kinect Sensor

[1]Miss. Bhagyashree Tingare , [2]Dr. Prasadu Peddi, [3]Dr. Prashant Kumbharkar

[1]Research Scholar, Department of Computer Science and Engineering, JJTU, Jhunjhunu, Rajasthan, India.
[2]Guide, Associate Professor, Department of Computer Science and Engineering, JJTU, Jhunjhunu, Rajasthan, India.
[3]Co-Guide, Professor, Department of Computer Science and Engineering, JSPM, Pune, Maharashtra, India.

Abstract : **Recent advances in 3D depth cameras such as Microsoft Kinect sensors (www.xbox. com/en-US/kinect) have created many opportunities for multimedia computing. Kinect was built to revolutionize the way people play games and how they experience entertainment. With Kinect, people are able to interact with the games with their body in a natural way. The key enabling technology is human bodylanguage understanding; the computer must first understand what a user is doing before it can respond. This has always been an active research field in computer vision, but it has proven formidably difficult with video cameras. The Kinect sensor lets the computer directly sense the third dimension (depth) of the players and the environment, making the task much easier. It also understands when users talk, knows who they are when they walk up to it, and can interpret their movements and translate them into a format that developers can use to build new experiences. Kinect's impact has extended far beyond the gaming industry. With its wide availability and low cost, many researchers and practitioners in computer science, electronic engineering, and robotics are leveraging the sensing technology to develop creative new ways to interact with machines and to perform other tasks, from helping children with autism to assisting doctors in operating rooms. Microsoft calls this the Kinect Effect.**

***Index Terms*—Microsoft Kinect sensor, 3D Sensor**

I. The Kinect Sensor

A)Introduction

Until recently computers had a very restricted view of the world around them, and users had very limited ways of communicating with computers. Over the years, computers have acquired cameras and audio inputs, but these have been used mostly for unrecognized input; computers can store and play such content, but it has been very difficult to make computers understand input in these forms.

For example, when people hear a sound, they can make judgments about the distance and direction of the sound source relative to their own position. Until recently, computers had more trouble making such judgments. Audio information from a number of microphones does provide considerable information about the distance and direction of the audio source, but determining this information is difficult for programs to do. Similarly, a video picture provides an image of the environment for the computer to analyze, but a computer has to work very hard to extract information about the objects in pictures or video because an image shows a flat, two-dimensional representation of a three-dimensional world.

Kinect changes all this. The Kinect sensor bar contains two cameras, a special infrared light source, and four microphones. It also contains a stack of signal processing hardware that is able to make sense of all the data that the cameras, infrared light, and microphones can generate. By combining the output from these sensors, a program can track and recognize objects in front of it, determine the direction of sound signals, and isolate them from background noise.

*B)*Getting Inside a Kinect Sensor

To get an idea of how the Kinect sensor works, Person could take one apart and look inside. (Don't do that. There are many reasons why taking Person Kinect apart is a bad idea: it's hard to do, Person will invalidate Person warranty, and Person might not be able to restore it to working condition. But perhaps the best reason not to take it apart is that I've already done it for Person!)

Figure 1 shows a Kinect sensor when it is "fully dressed."

Figure 1 A Kinect sensor.

Figure 2 shows a Kinect with the cover removed. Person can see the two cameras in the middle and the special light source on the left. The four microphones are arranged along the bottom of the sensor bar. Together, these devices provide the "view" the Kinect has of the world in front of it.

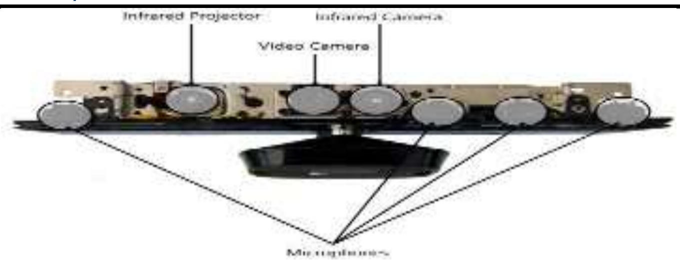

Figure 2 A Kinect sensor unwrapped.

Figure 3 shows all the hardware inside the Kinect that makes sense of the information being supplied from all the various devices.

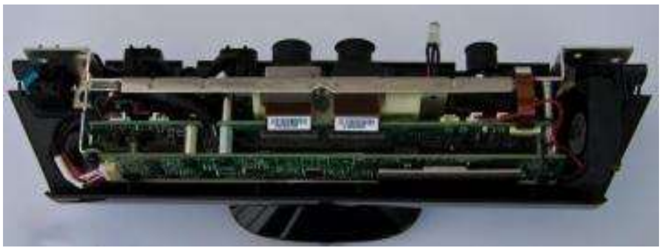
Figure 3 The Kinect sensor data processing hardware.

To make everything fit into the slim bar form, the designers had to stack the circuit boards on top of each other. Some of these components produce quite a bit of heat, so a tiny fan that can be seen on the far right of Figure 3 sucks air along the circuits to keep them cool. The base contains an electric motor and gear assembly that lets the Kinect adjust its angle of view vertically.
Now that Person have seen inside the device, Person can consider how each component helps the Kinect do what it does, starting with the "3D" camera.

C) The Depth Sensor

Kinect has the unique ability to "see" in 3D. Unlike most other computer vision systems, the Kinect system is able to build a "depth map" of the area in front of it. This map is produced entirely within the sensor bar and then transmitted down the USB cable to the host in the same way as a typical camera image would be transferred—except that rather than color information for each pixel in an image, the sensor transmits distance values.
Person might think that the depth sensor uses some kind of radar or ultrasonic sound transmitter to measure how far things are from the sensor bar, but actually it doesn't. This would be difficult to do over a short distance. Instead, the sensor uses a clever technique consisting of an infrared projector and a camera that can see the tiny dots that the projector produces.
Figure 4 shows the arrangement of the infrared projector and sensor.


Figure 4 The Kinect infrared projector and camera.

The projector is the left-hand item in the Figure 4. It looks somewhat like a camera, but in fact it is a tiny infrared projector. The infrared camera is on the right side of Figure 4. In between the projector and the camera is an LED that displays the Kinect device status, and a camera that captures a standard 2D view of the scene. To explain how the Kinect sensor works, I'll start by showing an ordinary scene in my house. Figure 5 shows my sofa as a person (okay, a camera) might see it in a room.

Figure 5 My sofa.

In contrast, Figure 6 shows how the Kinect infrared sensor sees the same view.

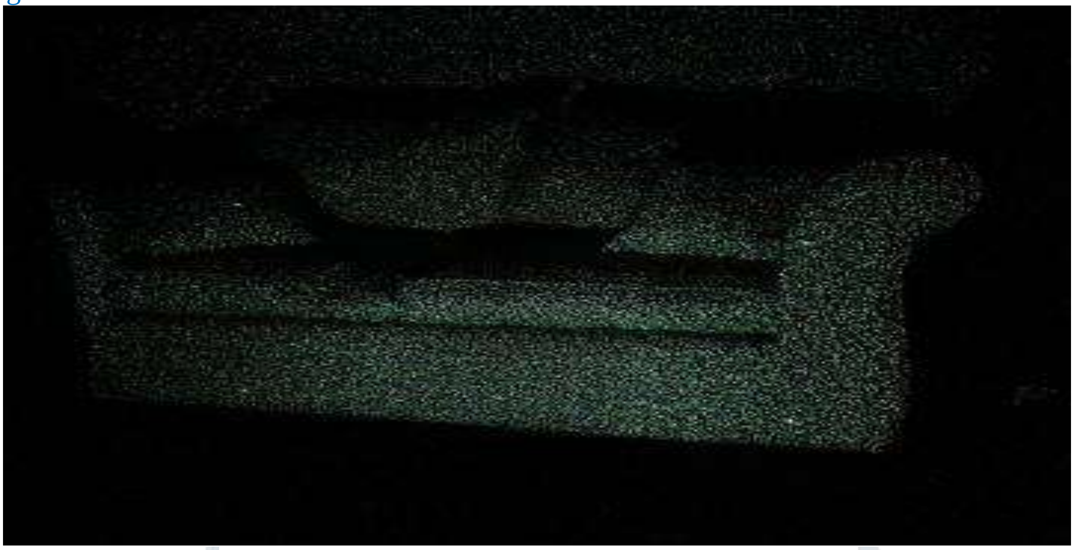

Figure 6 The sofa as the Kinect infrared sensor sees it.

The Kinect infrared sensor sees the sofa as a large number of tiny dots. The Kinect sensor constantly projects these dots over the area in its view. If Person want to view the dots Personself, it's actually very easy; all Person need is a video camera or camcorder that has a night vision mode. A camera in night vision mode is sensitive to the infrared light spectrum that the Kinect distance sensor uses.

Figure 6, for example, was taken in complete darkness, with the sofa lit only by the Kinect. The infrared sensor in the Kinect is fitted with a filter that keeps out ordinary light, which is how it can see just the infrared dots, even in a brightly lit room. The dots are arranged in a pseudo-random pattern that is hardwired into the sensor. Person can see some of the pattern in Figure 7.

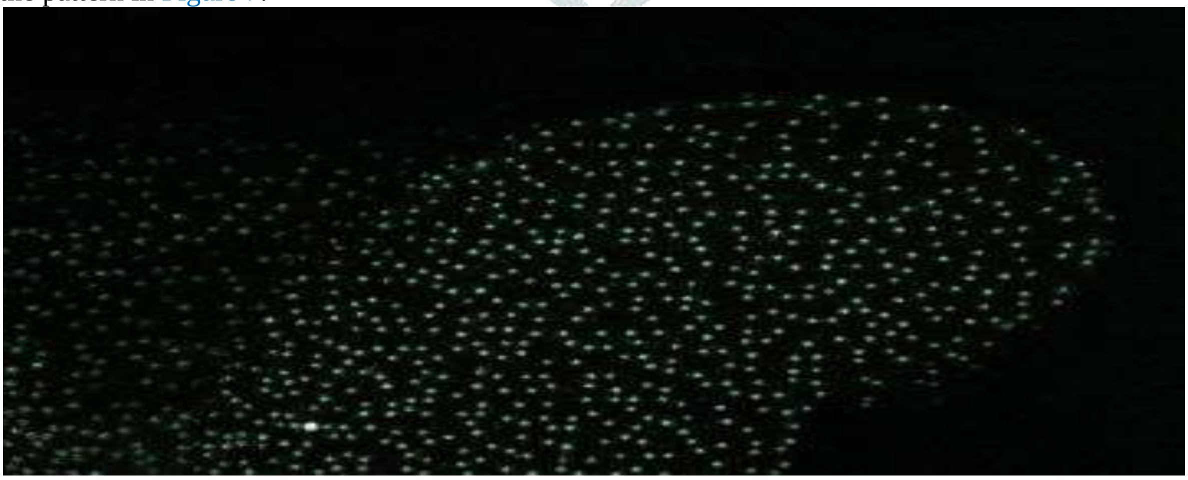

Figure 7 The dot pattern on the sofa arm.

A pseudo-random sequence is one that appears to be random, but it is actually mechanically generated and easy to repeat. What's important to remember here is that the Kinect sensor "knows" what the pattern looks like and how it is drawn. It can then compare the image from the camera with the pattern it knows it is displaying, and can use the difference between the two to calculate the distance of each point from the sensor.

To understand how the Kinect does this, Person can perform a simple experiment involving a darkened room, a piece of paper, a flashlight, and a helpful friend. Person need to adjust the flashlight beam so it's tightly focused and makes a small spot. Now, get Person friend to stand about 5 feet (1.5 meters) away from Person, slightly to Person right. Ask Person friend to hold the paper to the front of Person, holding the torch in Person left hand, shine the torch dot onto the piece of paper. Now ask Person friend to move forward toward Person. As the person comes closer, Person will see that the dot on the paper moves a little to the left because it now hits the paper before it has traveled quite as far to the right.
Figure 8 shows how this works. If Person know the place Person are aiming the dot, Person can work out how far away Person friend is by the position of the dot on the paper. The impressive thing about the Kinect sensor is that it performs that calculation for thousands of dots, many times a second. The infrared camera in the Kinect allows it to "see" where the dot appears in the image. Because the software knows the pattern that the infrared transmitter is drawing, the hardware inside the Kinect does all the calculations that are required to produce the "depth image" of the scene that is sent to the computer or Xbox.

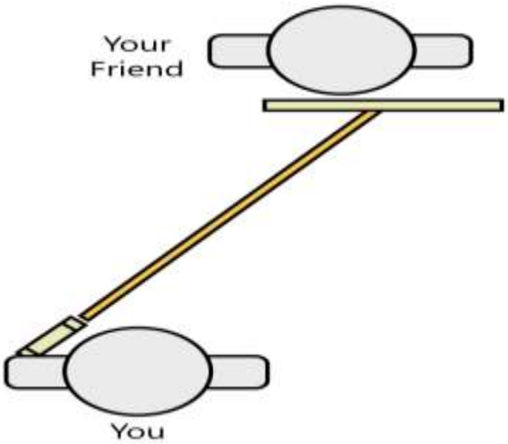

Figure 8 Showing how the Kinect distance sensor works.

This technique is interesting because it is completely different from the way that humans see distance. Each human eye gets a slightly different view of a scene, which means that the closer an object is to a human, the greater the difference between the images seen by each eye. The brain identifies the objects in the scene, determines how much difference there is between the image from each eye, and then assigns a distance value to each object.
In contrast, the Kinect sensor shines a tightly focused spot of light on points in the scene and then works out how far away that point is from the sensor by analyzing the spot's reflection. The Kinect itself doesn't identify any objects in a scene; that task is performed by software in an Xbox or computer, as Person'll see later.

D) The Kinect Microphones

The Kinect sensor also contains four microphones arranged along the bottom of the bar. Person can see them in Figure 2: two on the left and right ends, and two more on the right side of the unit. The Kinect uses these microphones to help determine from where in a room a particular voice is coming. This works because sound takes time to travel through air. Sound travels much more slowly than light, which is why Person often hear a thunderclap long after seeing the corresponding bolt of lightning.
When Person speak to the Kinect sensor, Person voice will arrive at each microphone at different times, because each microphone is a slightly different distance away from the sound source. Software can then extract Person voice waveform from the sound signal produced by each microphone and—using the timing information—calculate where the sound source is in the room. If several people are in a room with the Kinect, it can even work out which person is talking by calculating the direction from which their voice is coming, and can then "direct" the microphone array to listen to that area of the room. It can then remove "unwanted" sounds from that signal to make it easier to understand the speech content.
From a control point of view, when a program knows where the speech is coming from (perhaps by using the distance sensor), it can direct the microphone array in that direction, essentially creating a software version of the directional microphones that are physically pointed at actors to record their voices when filming motion pictures.

II. Recognizing People with Kinect

One very popular use for the Kinect sensor is recognizing and tracking people standing in front of it. The Kinect sensor itself does not recognize people; it simply sends the depth image to the host device, such as an Xbox or computer. Software running on the host device contains logic to decode the information and recognize elements in the image with characteristic human shapes. The software has been "trained" with a wide variety of body shapes. It uses the alignment of the various body parts, along with the way that they move, to identify and track them.
Figure 9 shows the output produced by the body-tracking software as a "stick figure" with lines joining the various elements.

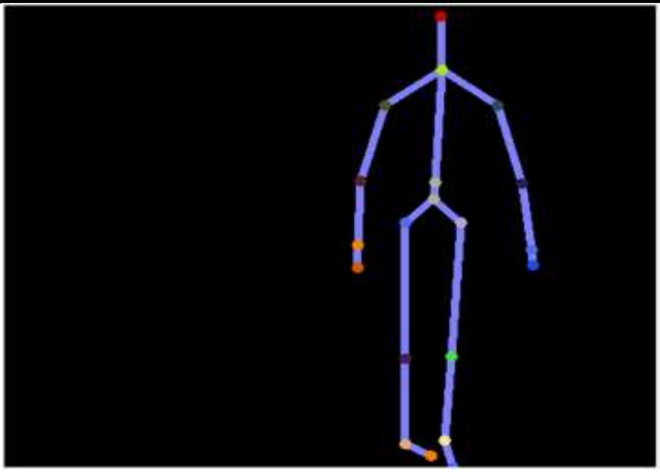

Figure 9 Skeleton information retrieved using the Kinect software.

The Kinect software can also recognize the height and proportions of a particular person. For example, this feature lets Xbox Live users “train” their Xbox so it recognizes them when they walk into a room.

III. Programming the Kinect

The software described in the previous sections, and which Person’ll see more of in this book, is called the Kinect for Windows Software Development Kit (SDK). Installing the SDK lets Person write programs that use the power of the Kinect at different levels. Person can obtain direct access to the low-level video and depth signals and create applications that use that low-level data, or Person can make use of the powerful library features built into the SDK that make it easy for a program to identify and track users.

Person can download the Kinect for Windows SDK for free. The SDK provides a set of libraries that Person can add to Person own programs and games so they can use the sensor. The SDK also contains all the drivers that Person need to link a Kinect to Person computer.

Person can use the Kinect SDK from a managed code programming language (such as C# or Visual Basic.NET) or from unmanaged C++. The SDK provides a set of objects that expose properties and methods Person can use in Person programs. The following chapters explore how Person can write programs that use these objects to create some novel and fun programs that support completely new ways of interacting with a computer.

The next chapter describes how to install the SDK on Person computer and get it connected and talking to the Kinect.

IV. Kinect for Xbox and Kinect for Windows

Person can write programs that use either the Kinect for Xbox sensor or the Kinect for Windows sensor. The Kinect for Xbox sensor has been set up to allow it to be most effective when tracking the figures of game players. This means that it can track objects that are up to 12 feet (4.0 meters) away from the sensor but cannot track any objects that are closer than 24 inches (80 cm). The Kinect for Windows sensor has been set up to allow it to track a single user of a computer, and it has much better short-range performance as it is able to track objects as close to the sensor as 12 inches (40 cm).

The Kinect for Windows SDK was, as the name implies, primarily created for use with the Kinect for Windows sensor, but it will also work with an Xbox 360 Kinect sensor. Microsoft engineers will provide support into the future for Xbox Kinect from this SDK, but for best results, particularly if Person want to track objects very close to the sensor bar, Person should invest in a Kinect for Windows sensor device. The Kinect for Windows device can even track individual finger movements and gestures of the computer user.

The bottom line is that if Person have an Xbox 360 with a Kinect device attached to it, Person can use that sensor to have some fun learning how to create programs that can see, measure distance, and hear users. However, if Person want to get serious about providing a product of Person own that is based on the Kinect sensor, Person should target the Kinect for Windows device. If Person want complete details of how this all works, read the detailed End User License here:

http://www.microsoft.com/en-us/kinectforwindows/develop/sdk-eula.aspx

V. Software

The software is what makes the Kinect a breakthrough device. Developers for the Kinect gathered an incredible amount of data regarding motion-capture of actual moving things in real-life scenarios. Processing all of this data using a special artificial intelligence machine-learning algorithm allows the Kinect to map the visual data it collects to models representing people of different backgrounds (age, height, gender, body type, clothing and more). This is just one of the ways that developers were able to help the Kinect "learn" about its surroundings and what it is actually seeing.

The Kinect's "brain" is really the secret. Stored in the system is enough intelligence to analyze what it sees and align that with stored collection of skeletal structures to interpret Person movements. Once the brain has enough data on Person body parts, it outputs this reference data into a simplified 3D avatar shape. Beyond gauging player movements, the Kinect must also judge the distances of different points on Person body throughout the entire game. To do this it uses a host of **sensors** and analyzes all this data 30 times a second.

Figure - 10 *Microsoft's clusters of computers are the "learning brain" that feeds all Kinects*

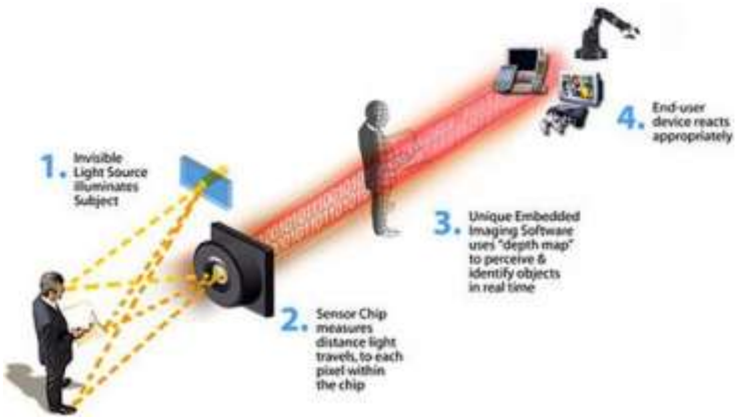

Figure - 11 *Depth perception using the infrared camera*

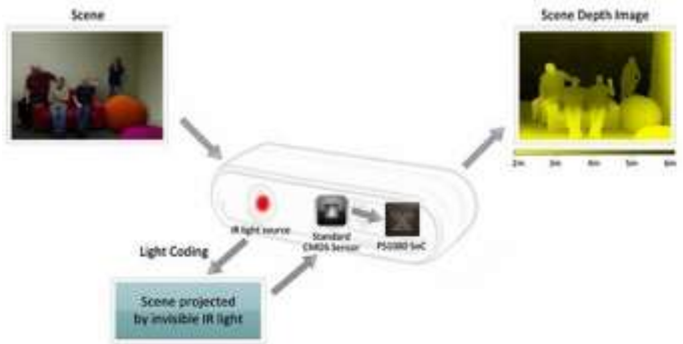

Figure - 12 *Transfer of information from the camera to the TV screen Person see*

VI. Hardware

The Kinect contains three vital pieces that work together to detect Person motion and create Person physical image on the screen: an RGB color VGA video camera, a depth sensor, and a multi-array microphone.

The camera detects the red, green, and blue color components as well as body-type and facial features. It has a pixel resolution of 640x480 and a frame rate of 30 fps. This helps in facial recognition and body recognition.

The depth sensor contains a monochrome CMOS sensor and infrared projector that help create the 3D imagery throughout the room. It also measures the distance of each point of the player's body by transmitting invisible near-infrared light and measuring its "time of flight" after it reflects off the objects.\The microphone is actually an array of four microphones that can isolate the voices of the player from other background noises allowing players to use their voices as an added control feature. These components come together to detect and track 48 different points on each player's body and repeats 30 times every second.

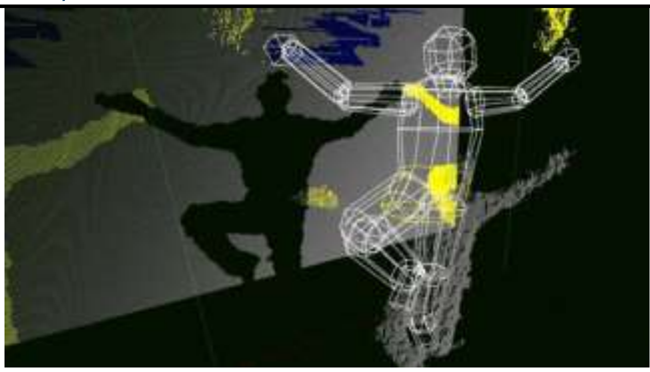

Figure - 13 *Graphical image of an Xbox player*

Putting both hardware and software together give the Kinect the ability to generate 3D images and recognize human beings within its field of vision. It can analyze the person in front of it and go through multiple "filters" to try and determine which type of body structure matches with the correct type programmed in its system. Data is constantly being transferred back and forth between the Kinect and the objects in its field of vision while Person simply enjoy the fun of being a character in a game, without holding anything in Person hands.

Figure - 14 *Game: Kinect Sports Hurdles (Jump over hurdles without leaving the living room)*

As great as it sounds to be able to play a game without the controller, it doesn't stop at just playing video games. There are tons of possible applications with the Kinect far from the gaming world. Read about some of the truly remarkable applications below.

VII. Robotic Applications

Because the Kinect has an infrared projector, infrared camera and color camera, it's a great imaging tool, even for **robots**. In order to enhance the range and autonomous nature of robots, they need to be able to see the environment around them. One way they do this is through simultaneous localization and mapping, or SLAM.

Traditionally, these kinds of sensors are either expensive and cumbersome or cheap and unreliable. Laser arrays are expensive and heavy and can only map in two dimensions. Stereo cameras are light and can map in 3D, but require colossal computing power. Ken Conley of Willow Garage can now sell his Kinect-equipped TurtleBot for $500. A gigantic savings from the previous non-Kinect version that cost over $250,000!

The TurtleBot is a customizable mobile robotic platform that rides on an iRobot Create platform and uses the open-source ROS (Robot Operating System) platform. The TurtleBot uses the Kinect to see the world in 3D and for detecting and tracking people. Right out of the box, Person can program TurtleBot to build maps of Person home and navigate from Person kitchen to Person favorite seat in the living room. It also has the capability to take pictures from around Person house and stitch them together to create a 360-degree panorama.

With heavier-duty and more robust platforms, a user can also give gesture commands to control the robot, like the PR2 Robot, and even remotely control the limbs as if they were Person own. Perhaps one day we can use them to lift heavy objects for us and do other chores.

VIII. Scientific Applications

Last year, Ph.D student, Ken Mankoff squeezed his way into a small cavern underneath the Rieperbreen Glacier in Svalbard, Norway with a backpack containing a laptop, a battery pack and a Kinect. Once inside, he used the Kinect sensor to scan the cave floor in 3D to map its size and the irregularities on the surface. This helps the scientists better understand how the ice above flows toward the sea. The Kinect is quickly becoming a vital tool because of the 3D data it captures in visible and infrared wavelengths with very high accuracy.

The Kinect is in a league of its own effectively capturing 9 million data points per second. Traditional scanning tools can be bulky and use LIDAR (Light Detection And Ranging) to send laser pulses to accurately measure surfaces over many miles, but these systems cost between $10,000 and $200,000 and have be ordered from special manufacturers and operated by trained professionals. On the other hand, the Kinect costs around $120 and takes measurements in the three to 16 foot range, and it can fit in Person pocket. The Kinect is an inspiring device because of its low cost and most students are already familiar with it.

Small and cheap hardly means incapable. The Kinect even has relevance in space. Naor Movshovitz, a planetary Ph.D. student at UC Santa Cruz, said the data would be useful for future missions where we may have to deflect medium to
large asteroids that threaten to impact Earth.

We have pretty good data on how objects impact the Earth surface, but how do impacts differ when there is extremely low gravity? The idea is to use one of NASA's gravity-reduced airplanes to study how a small projectile would impact a dirt pile, while the Kinect would be used to measure the three-dimensional position of objects to get data about how debris is ejected after the projectile's impact.

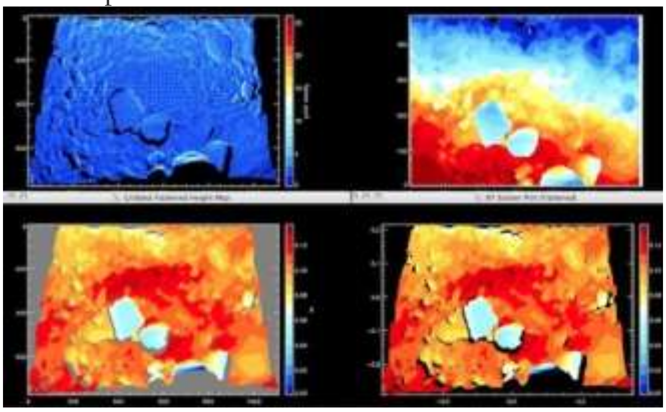

Figure - 15 *Images: A sample of the 3-D data from the Kinect's scan of the glacier cave. Ken Mankoff.*

IX. Summary

This study gave Person a look inside the Kinect sensor so Person could see (without having to take Person own Kinect apart) how complex it is. Person saw that the Kinect contains two cameras (one infrared camera and one video camera) and a special infrared transmitter that produces a grid of dots that measure the distance of objects from the Kinect and to compose a “depth map” of the image. Person also learned that the Kinect sensor contains four microphones that can be used to remove background noise from an audio signal and to listen to sound from particular parts of a room.

Person also saw that the Kinect sensor sends this data to a host device (Xbox or computer), which then processes the data in various ways, including recognizing the position, movement, and even the identity of people in front of the Kinect.

Person also found out that two Kinect sensor bars are available, both of which can be used with the Kinect for Windows Software Development Kit (SDK). The Kinect for Xbox device has a good long-range performance for tracking game players, and the Kinect for Windows device has been optimized for shorter-range tracking so that a single computer user can use it to interact with a system that is nearby.

REFERENCES

1. Rob Miles, “Start Here! Learn the Kinect API”,eBook 978-0-7356-6394-7, Published 7/15/2012
2. Robert Cong & Ryan Winters, “How Does The Xbox Kinect Work”, JEMECO electronics, 2016
3. Wenjun Zeng, “Microsoft Kinect Sensor and Its Effect”, 1070-986X/12/$31.00 c 2012 IEEE Chai Joon Lip, Kamarulzaman Kamarudin, Syed Muhammad Mamduh, “Human 3d Reconstruction and Identification Using Kinect Sensor”, 978-1-5386-8369-9/18/$31.00 ©2018 IEEE
4. Dan Song 1, Ruofeng Tong1, Jiang Du1, Yun Zhang2, And Yao Jin3, “Data-Driven 3-D Human Body Customization With a Mobile Device”, 2169-3536 2018 IEEE.
5. S.Priyadharsun1 , S.Lakshigan2 , S.S Baheerathan3, S.Rajasooriyar4, U.U.S.K. Rajapaksha5, S.M.Buddika Harshanath6, “Parade in the Virtual Dressing Room”, 978-1-5386-5495-8/18/$31.00 ©2018 IEEE
6. Bhalekar Sourabh, Chitte Darshan, Dhamal Hemant, Ganeshwade Priyanka, Rankhambe J.P., “Implementation of Virtual Dressing Room using Newton’s Mechanics”, 10.23956/ijarcsse/SV7I5/0107
7. Adjeisah Michael, Zhao Chen, Guohua Liu, Yang Yi, “A New Approach For Tracking Human Body Movements by Kinect Sensor”, 2017
8. Hanwen Li, Zilong Liu, Wenting Wang, “A More Practical Automatic Dressing Method for Clothes Animation”, 978-1-5090-0768-4/16 $31.00 © 2016 IEEE DOI 10.1109/IHMSC.2016.220

9. Nikita Deshmukh, Ishani Patil, Sudehi Patwari, Aarati Deshmukh, Pradnya Mehta, “Real Time Virtual Dressing Room”, 2277-5420, 2016
10. Dardan Maraj, Arianit Maraj, Adhurim Hajzeraj, “Application Interface for Gesture Recognition with Kinect Sensor”, 978-1-5090-3471-0/16/$31.00 ©20 16 IEEE
11. Muhammed Kotan1 and Cemil Oz. “Virtual Dressing Room Application with Virtual Human Using Kinect Sensor”,Journal of Mechanics Engineering and Automation 5, (2015) 322-326 doi: 10.17265/2159-5275/2015.05.008,
12. Mingliang Chen, Weiyao Lin, Bing Zhou, “A Real-time Virtual Dressing System with RGB-D Camera”, 978-988-14768-0-7©2015 APSIPA
13. Hesham Alabbasi, Alex Gradinaru, Florica Moldoveanu, Alin Moldoveanu, “Human Motion Tracking & Evaluation using Kinect V2 Sensor”, 978-1-4673-7545-0/15/$31.00 ©2015 IEEE
14. Jungsu Shin, Kyeong-Ri Ko, and Sung Bum Pan, “Automation of Human Body M1 o del Data Measurement Using Kinect in Motion Capture System”, 978-1-4799-7543-3/15/$31.00 ©2015 IEEE
15. Jaychand Upadhyay, Divya Shukla, Nidhi Patel, Sheetal Nangare, “Virtual Makeover and Virtual Trial Dressing”, 10.15680/ijircce.2015.0303090, 2015
16. Flavio Minos Pineda-L´opez., Marco J. Flores C., Gustavo J. Ortiz V, Willian A. Mosquera Y., “Prototype for the Analysis of Human Body Movement with Kinect Sensor”, 978-1-4673-9461-1/15/$31.00 c 2015 IEEE
17. Mr. Rohan Tirmale, Miss. Pushpanjali Rout, Miss. Ritu Joshi, Mr. Brian Pereira, Prof. S. N. Shelke,”A SURVEY ON VIRTUAL REALITY FOR DRESSING ROOM”,OAIJSE, ISO 3297:2007 Certified ISSN (Online) 2456-3293, Volume 2, Special Issue, December 2017.
18. Aditi Nalawade, Ankita Kotgire, Niti Pawar, Pragati Sarap, Prof. Rupali Mahajan (2020), “Novel Approach for Virtual Dressing Room using Augmented Reality”, e-ISSN: 2319-8753, p-ISSN: 2320-6710, IJIRSET.Volume 9, Issue 9, September 2020.