Study on Three Dimensions Body Reconstruction and Measurement by Using Kinect

Qi Luo

College of Sports Engineering and Information Technology,
Wuhan Sports University,430079, China
emeinstitute@126.com

Abstract. A point cloud is a set of data points in some coordinate system. Point clouds may be created by Three Dimensions scanners. These devices measure in an automatic way a large number of points on the surface of an object, and often output a point cloud as a data file. The point cloud represents the set of points that the device has measured. Using Microsoft Kinect to obtain the depth body data and get the depth image. In this paper, the function and the depth scanning principle of the Microsoft Kinect has been researched. The concept of Point cloud has been also introduced. Point cloud data processing has been proposed in the paper. First, the depth data obtained by Kinect are transformed into the form of Three Dimensions point cloud to store and visualize. And then, make rejections, filtering, and simplification for point cloud. Finally In the process of simplification, we take the advantage of the minimum distance method and the angular deviation method, an improved self-adapting method of simplification was introduced in the paper.

1 Introduction

The appearance of the non-contact Three Dimensions body scanner has made it to be true. Presently, non-contact Three Dimensions body scanning technology is not very advanced in our country. Non-contact Three Dimensions body scanner from foreign country is so large and expensive that many costume manufacturers cannot afford it. Meanwhile, using traditional Three Dimensions body scanner to obtain the characteristic data of the body surface is very complicated. Therefore, exploring the small volume, high precision and low cost Three Dimensions body scanner has comprehensive value in research and prospect for application. In this thesis, a scheme of Three Dimensions body scanning system based on Microsoft Kinect has been proposed.

Using Microsoft Kinect to obtain the depth body data and get the depth image. In this paper, the function and the depth scanning principle of the Microsoft Kinect has been researched. the concept of Point cloud has been also introduced.

Point cloud data processing has been proposed in the paper. First, the depth data obtained by Kinect are transformed into the form of Three Dimensions point cloud to store and visualize. And then, make rejections, filtering, and simplification for point cloud. Finally In the process of simplification, we take the advantage of the minimum

V. G. Duffy (Ed.): DHM 2014, LNCS 8529, pp. 35–42, 2014.

distance method and the angular deviation method, an improved self-adapting method of simplification was introduced in the paper.

2 Kinect

Kinect is a line of motion sensing input devices by Microsoft for Xbox 360 and Xbox One video game consoles and Windows PCs. Based around a webcam-style add-on peripheral, it enables users to control and interact with their console/computer without the need for a game controller, through a natural user interface using gestures and spoken commands. The first-generation Kinect was first introduced in November 2010 in an attempt to broaden Xbox 360's audience beyond its typical gamer base. A version for Windows was released on February 1, 2012[1]. Kinect competes with several motion controllers on other home consoles, such as Wii Remote Plus for Wii, PlayStation Move/PlayStation Eye for PlayStation 3, and PlayStation Camera for PlayStation 4. Kinect for Xbox 360 see Figure 1

Fig. 1. Kinect for Xbox 360

Kinect builds on software technology developed internally by Rare, a subsidiary of Microsoft Game Studios owned by Microsoft, and on range camera technology by Israeli developer PrimeSense, which developed a system that can interpret specific gestures, making completely hands-free control of electronic devices possible by using an infrared projector and camera and a special microchip to track the movement of objects and individuals in three dimensions. This Three Dimensions scanner system called Light Coding employs a variant of image-based Three Dimensions reconstruction.

Kinect sensor is a horizontal bar connected to a small base with a motorized pivot and is designed to be positioned lengthwise above or below the video display. The device features an "RGB camera, depth sensor and multi-array microphone running proprietary software", which provide full-body Three Dimensions motion capture, facial recognition and voice recognition capabilities. At launch, voice recognition was only made available in Japan, United Kingdom, Canada and United States. Mainland Europe received the feature later in spring 2011. Currently voice recognition is supported in Australia, Canada, France, Germany, Ireland, Italy, Japan, Mexico, New Zealand, United Kingdom and United States. Kinect sensor's microphone array enables Xbox 360 to conduct acoustic source localization and ambient noise suppression, allowing for things such as headset-free party chat over Xbox Live.

The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in Three Dimensions under any ambient light conditions. The sensing range of the depth sensor is adjustable, and Kinect software is capable of automatically calibrating the sensor based on game play and the player's physical environment, accommodating for the presence of furniture or other obstacles.

Described by Microsoft personnel as the primary innovation of Kinect, the software technology enables advanced gesture recognition, facial recognition and voice recognition. According to information supplied to retailers, Kinect is capable of simultaneously tracking up to six people, including two active players for motion analysis with a feature extraction of 20 joints per player. However, Prime Sense has stated that the number of people the device can "see" (but not process as players) is only limited by how many will fit in the field-of-view of the camera.

This infrared image shows the laser grid Kinect uses to calculate depth .The depth map is visualized here using color gradients from white (near) to blue (far)Reverse engineering has determined that the Kinect's various sensors output video at a frame rate of ~9 Hz to 30 Hz depending on resolution. The default RGB video stream uses 8-bit VGA resolution (640 × 480 pixels) with a Bayer color filter, but the hardware is capable of resolutions up to 1280x1024 (at a lower frame rate) and other colour formats such as UYVY. The monochrome depth sensing video stream is in VGA resolution (640 × 480 pixels) with 11-bit depth, which provides 2,048 levels of sensitivity. The Kinect can also stream the view from its IR camera directly (i.e.: before it has been converting into a depth map) as 640x480 video, or 1280x1024 at a lower frame rate. The Kinect sensor has a practical ranging limit of 1.2–3.5 m (3.9–11.5 ft) distance when used with the Xbox software. The area required to play Kinect is roughly 6 m2, although the sensor can maintain tracking through an extended range of approximately 0.7–6 m (2.3–19.7 ft). The sensor has an angular field of view of 57° horizontally and 43° vertically, while the motorized pivot is capable of tilting the sensor up to 27° either up or down. The horizontal field of the Kinect sensor at the minimum viewing distance of ~0.8 m (2.6 ft) is therefore ~87 cm (34 in), and the vertical field is ~63 cm (25 in), resulting in a resolution of just over 1.3 mm (0.051 in) per pixel. The microphone array features four microphone capsules and operates with each channel processing 16-bit audio at a sampling rate of 16 kHz.

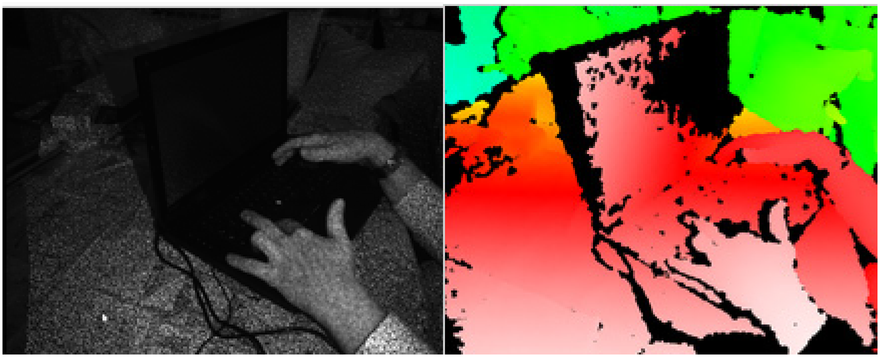

Fig. 2. The laser grid Kinect uses to calculate depth and The depth map is visualized here using color gradients from white (near) to blue (far)

Because the Kinect sensor's motorized tilt mechanism requires more power than the Xbox 360's USB ports can supply, the device makes use of a proprietary connector combining USB communication with additional power. Redesigned Xbox 360 S models include a special AUX port for accommodating the connector, while older models require a special power supply cable (included with the sensor) that splits the connection into separate USB and power connections; power is supplied from the mains by way of an AC adapter.

3 Point Cloud

A point cloud is a set of data points in some coordinate system. In a three-dimensional coordinate system, these points are usually defined by X, Y, and Z coordinates, and often are intended to represent the external surface of an object.

Point clouds may be created by Three Dimensions scanners. These devices measure in an automatic way a large number of points on the surface of an object, and often output a point cloud as a data file. The point cloud represents the set of points that the device has measured.

As the result of a Three Dimensions scanning process point clouds are used for many purposes, including to create Three Dimensions CAD models for manufactured parts, metrology/quality inspection, and a multitude of visualization, animation, rendering and mass customization applications.

While point clouds can be directly rendered and inspected, usually point clouds themselves are generally not directly usable in most Three Dimensions applications, and therefore are usually converted to polygon mesh or triangle mesh models, NURBS surface models, or CAD models through a process commonly referred to as surface reconstruction.

There are many techniques for converting a point cloud to a Three Dimensions surface. Some approaches, like Delaunay triangulation, alpha shapes, and ball pivoting, build a network of triangles over the existing vertices of the point cloud, while other approaches convert the point cloud into a volumetric distance field and reconstruct the implicit surface so defined through a marching cubes algorithm.

One application in which point clouds are directly usable is industrial metrology or inspection. The point cloud of a manufactured part can be aligned to a CAD model (or even another point cloud), and compared to check for differences. These differences can be displayed as color maps that give a visual indicator of the deviation between the manufactured part and the CAD model. Geometric dimensions and tolerances can also be extracted directly from the point cloud. Please see the figure

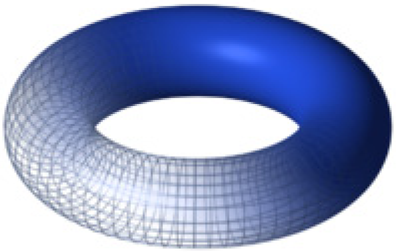

Fig. 3. A point cloud image of a torus

Point clouds can also be used to represent volumetric data used for example in medical imaging. Using point clouds multi-sampling and data compression are achieved.

In geographic information system, point clouds are one of the sources to make digital elevation model of the terrain. The point clouds are also employed in order to generate Three Dimensions model of urban environment, e.g.

4 Depth Image Processing by Kinect

Kinect is different from all other input devices, because it provides a third dimension. It does this using an infrared emitter and camera. Unlike other Kinect SDKs such as OpenNI, or libfreenect, the Microsoft SDK does not provide raw access to the IR stream. Instead, the Kinect SDK processes the IR data returned by the infrared camera to produce a depth image

The IR or depth camera has a field of view just like any other camera. The field of view of Kinect islimited, as illustrated in Figure 4. The original purpose of Kinect was to play video games within the confines of game room or living room space. Kinect's normal depth vision ranges from around two and a half feet (800mm) to just over 13 feet (4000mm). However, a recommended usage range is 3 feet to 12 feet as the reliability of the depth values degrade at the edges of the field of view.

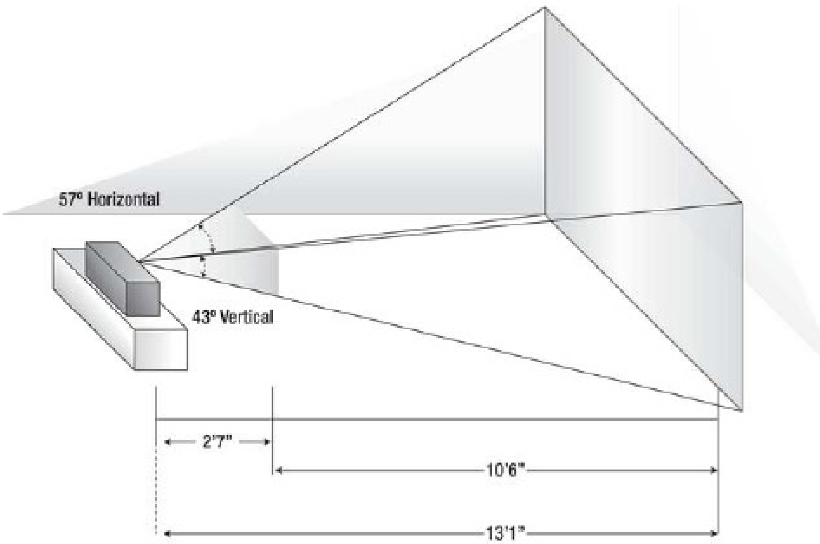

Fig. 4. Kinect field of view

Like any camera, the field of view of the depth camera is pyramid shaped. Objects farther away from the camera have a greater lateral range than objects nearer to Kinect. This means that height and width pixel dimensions, such as 640×480, do not correspond with a physical location in the camera's field of view. The depth value of each pixel, however, does map to a physical distance in the field of view. Each pixel represented in a depth frame is 16 bits, making the BytesPerPixel property of each frame a value of two. The depth value of each pixel is only 13 of the 16 bits, as shown in Figure 5.

Fig. 5. Layout of the depth bits

5 Point Cloud Data Processing

When obtaining the point cloud data, because of the large data, it is prone to errors. Hope is also known as distortion point, most due to the changes in the surrounding environment parameters and produces the equipment. First, identifying and removing the "jump point" is the first step performing point cloud data processing. In the same row or column of data, if the distance of a point from its neighboring point is big, this point can be considered as "hops".

In the process of getting the point cloud data, in addition to outliers' error caused by device parameters and environmental impact, the measurement process will also inevitably produce random errors. Therefore, the smoothing work of point cloud data is necessary. There are three main methods for data smoothing filter: median filter, mean filter and Gaussian filter.

Adaptive maximum allowable deviation streamlined method, obtaining features of the acquired point cloud on the purpose of Kinect, usually streamlines in the unit of a direction. There are currently more commonly used method--minimum distance method and angle deviation method. This section, aiming at the complexity of body surface, introduces the curvature of parameters and proposes an adaptive maximum allowable deviation streamlined method, taking the advantages of minimum distance method and the angular deviation, an improved self-adapting method of simplification as follows:

The method first set a threshold curvature K and then selected an x to calculate curvature k. When $k < K$, this section curvature curve changes slowly, then streamlines by the use of minimum distance; When $k > K$, this section curvature of curve rapidly changes, instead of directly using the minimum distance. So two parameters-- angle threshold $\theta_{\min}$ and chord high threshold $h_{\min}$ --should be set to streamline data threshold. The smaller the value of $\theta_{\min}$, the higher the precision of the reconstruction model, generally ranging between 0 -15 . It can also select good

$\theta_{\min}$ value according to the actual situation to test. $h_{\min}$ is determined by the formula :

$$h_{\min} = \mu \frac{N_b}{N_a} \sin \theta_{\min} \tag{1}$$

where u is the normal value of the distance between adjacent points, N_a the number of former streamline point, and N_b the number of latter streamline point.

Algorithm steps are as follows:

1) Given a threshold curvature K and step length x.

2) The curvature k data points per segment is calculated in accordance with sub-step x .

3) According to the curvature segment k of data points, if $k < K$, then set a minimum distance threshold $d_{\min}$.

4) From the first point of data points in one direction in order to find the distance between two adjacent d.

5) If $d < d_{\min}$, then the latter point has been deleted, whereas retention, till the last point of data points , determining whether this step is the last step length. If so, end ; if not, read the next step curvature.

6) If $k > K$, then given a threshold angle $\theta_{\min}$ and calculate the chord high threshold $h_{\min}$.

7) From the start point of data points select three adjacent points p_1, p_2, p_3, If the last point of the segment is p_3 , go to step 9). Otherwise, calculate the angle between $p_1 p_2$ and $p_2 p_3$, chord height $h = p_1 p_2 \sin \theta$.

8) If $h < h_{\min}$, then delete the point p_2 , and shift p_2 back forward. Let $p_2 = p_3 p_3 = p_4$, go to step 7) , or delete the point p_2 if $\theta < \theta_{\min}$, shifting the poin p_2 back forward. Let $p_2 = p_3 p_3 = p_4$, go to step 7), otherwise keep the point p_2, the point p_1 shift back forward, $p_1 = p_2, p_2 = p_3, p_3 = p_4$ go to step 7).

9) Determine whether this step is the last step length. If so, end ; if not, read the next step curvature.

6 Conclusion

Using Microsoft Kinect to obtain the depth body data and get the depth image. In this paper, the function and the depth scanning principle of the Microsoft Kinect has been researched. the concept of Point cloud has been also introduced. Point cloud data processing has been proposed in the paper. First, the depth data obtained by Kinect are transformed into the form of Three Dimensions point cloud to store and visualize.

And then, make rejections, filtering, and simplification for point cloud. Finally In the process of simplification, we take the advantage of the minimum distance method and the angular deviation method, an improved self-adapting method of simplification was introduced in the paper.

References

1. http://latimesblogs.latimes.com/technology/2009/06/microsofte3.html
2. Fujimoto, M., Kariya, K.: An improved method for digitized data reduction using an angle parameter. Measurement 12(2), 113–122 (1993)
3. Martin, R.R., Stroud, I.A., Marshall, A.D.: Data reduction for reverse engineering. RECCAD, Deliverable Document 1 COPERNICUS project, No 1068. Computer and Automation Institute of Hungarian Academy of Science (January 1996)
4. Veron, P., Léon, J.C.: Static polyhedron simplification using error measurements. Computer- Aided Design 29(4), 287–298 (1997)
5. Song, S.: Research on 3D body reconstruction and measurement techniques based on Kinect. Donghua University Master Dissertation 30, 31 (2013)