RUTGERS UNIVERSITY

CAPSTONE DESIGN

ELECTRICAL AND COMPUTER ENGINEERING

Computer Vision-Based 3D Reconstruction for Object Replication

Authors:
Ryan CULLINANE
Cady MOTYKA
Elie ROSEN

Supervisor:
Kristen DANA

March 10, 2013

1 Introduction

The Computer Vision-Based 3D Reconstruction for Object Replication is accomplished by using a Kinect for windows. Originally, the Kinect was created for entertainment, but recently it has been introduced to the field of robotics and computer vision. The Kinect is a quick, reliable and affordable tool that uses a near-infrared laser pattern projector and an IR camera, along with the sensor and software development kit, to calculate 3D measurements.

The 3D printer is another part of the robotics field that is beginning to find an increasing number of uses. The most innovative aspect of the 3D printer is the ability to print an object, regardless of interconnecting internal components, and have it function as intended. This means that any connecting gears that are printed with the 3D printer will in fact turn as they are supposed to.

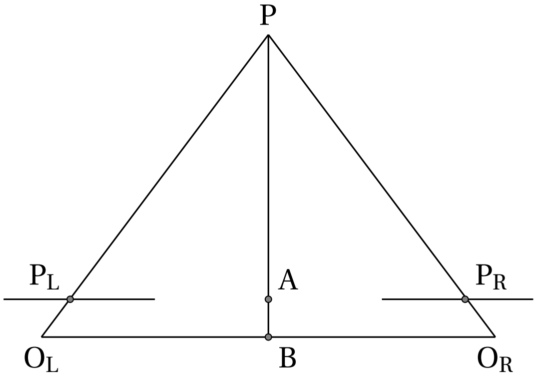
2 Methods

Calibration

The Kinect can be calibrated in a way similar to other cameras for computer vision, the only difference is that changes in the depth have to be present with the pattern in order to calibrate the depth camera. The Kinect needs to take an image of a checkerboard pattern.

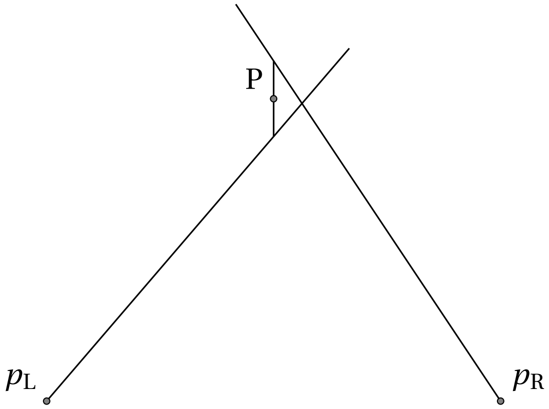
Stereo Reconstruction

Once the Kinect has been calibrated, all that is needed now for the stereo reconstruction is a triangulation of viewing rays.



P is the location of the object in the world, O_L and O_R are the left and right camera centers, P_R and P_L are the appearance of the point P in the two image planes where $P_L = \begin{bmatrix} x_L \\ y_L \end{bmatrix}$. The distance between O_L and O_R is T, or the distance between the left and right camera. The distance between A and B is the focal length of the cameras. If we define the distance between P and B as distance Z, the following equation can be used to represent the ratio between T and Z, using the theorem of like triangles: $\frac{T}{Z} = \frac{T+x_L-x_R}{Z-f} or \frac{T-x_R-x_L}{Z-f}$ Cross multiplying these equations results in: $\frac{Z(T-x_R-x_L)}{Z-f} = \frac{T(Z-f)}{Z}$ These calculations show that depth, or Z, is inversely proportional to disparity. This means that $P_L = \frac{f^L P}{Z_L}$. and $P_R = \frac{f^R P}{Z_R}$.

Once there is a corresponding point pair for P from the two images, an algorithm would undo the scale and shift of the pixel points in order to obtain the 2 dimensional camera coordinates. The midpoint algorithm is then used to find the real three-dimensional world coordinate that corresponds to that point pair.



Above are the rays $\overrightarrow{O_R p_R}$ and $\overrightarrow{O_L p_L}$ are drawn. The line connecting the two vectors, which is also perpendicular to both, is obtained by taking the cross product of these two vectors. The vector from p_L is equal to $a\vec{p}_L$, since point p_R is distance T away from p_L, the vector from p_R is equal to $b^L R_R \vec{p}_R + T$. The segment connecting these two vectors can be represented as $c\vec{p}_L x^L R_R \vec{p}_R$, where a, b and c are unknown constants that can be solved using the three equations explained above.
The point P lies on the center of this line and be found by $^L P = a\vec{p}_L + \frac{c}{2}\vec{p}_L x^L R_R \vec{p}_R$ In order to get the world point M, this point would just be divided by the Intrinsic and extrinsic matrices.

The Kinect accomplishes this triangulation by using the known information about the sensor, the data obtained from the infrared projection and the image received from the camera. The sensor will project invisible light onto an object, the light bounces back and the infrared sensor reads back the data. These clusters of light that are read back can be matched to the hard-coded images the Kinect has of the normal projected pattern and allows for a search for correlations, or the matching points. While looking through the camera's focal point, the point of interest will fall on a specific pixel, depending on how close or far away it is, this means that we know along which trajectory this point is from the camera. The relative line of trajectory from the projector and from the camera, along with the known information about the distance between the cameras on the Kinect sensor, are used in the above described triangulation process to find the three-dimensional coordinates of the point.

In order to make this data more manageable, a bilateral filter is used to remove the erroneous measurements. This filter will just take every point, and recalculate the value of that point based on the waited average of the surrounding pixels. The process takes away some of the sharpness of the depth map, but it removes the noise that will skew the results of the three dimensional reconstruction. Next, in order to represent the depth map in the true three dimensional points, a vertex must be created at each point where x, and y are the pixel values and the depth is the z coordinate. These points are then multiplied by a calibrated matrix to convert them into a vector map, or point cloud. The normal of each vertex is calculated by the cross product of the neighboring pixels. This process is combined with the process of computing a pyramid of the data, multiple copied of the depth map are made with smaller resolutions, at leach layer the vertices and their normals are calculated and stored.

The next step is to take a previously calculated vertex and normal map and run an Iterative Closest Point Algorithm on the four maps. This step creates a rotation and translation that minimizes distant errors between the point clouds. This algorithm is useful because it will find the best way to position the point cloud before the reconstruction so that every part of the object is represented correctly. Once the best fit is found, the existing depth data can be combined with the existing model to get a more refined result. The filtering got rid of the noise, and adding the raw depth data back to this will add the important details back into the final model. A Truncated Surface Distance Function is used to fit the depth data and the model back together. Finally, a weighted average of the existing model and the latest depth measurements is taken and used to generate the surface using a ray-casting algorithm, to express the reconstructed object to the user. The model can be exported as an STL file to the Gcode converter, but some work still needs to be done in order to represent to the user what he ore she is trying to print. The ray-casting algorithm will tell a virtual camera looking at the virtual model what to display on the GUI. A ray is cast from every pixel in the image through the focal point of the digital camera, and the first surface that the ray intersects with is excited, displaying it to the user. [8]

Gcode Conversion

The RepRap firmware uses G-code to communicate to the 3D printer, specifically to define the print head movements. This code has commands that tell the print head to move to a certain point with rapid or controlled movement, turn on a cooling fan, or selecting a different tool. Since this 3D printer does not have as many features, the G-code generator does not have to add much complicated code, but rather just instructions to the printer head. Since the printer continuously dispenses plastic, it is necessary to find a path for it to take that will build up the reconstructed object layer by layer without placing too much plastic in any specific area. This requires cutting up the reconstructed object into layers and then finding the best path to traverse that layer without overlapping any part of that path. The G-code converter takes in the STL file, cuts it up into horizontal layers and then calculates the about of material that is needed to fill each slice.

3 Experimental Results

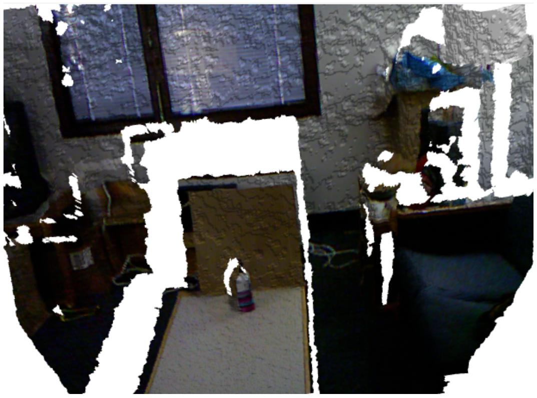


Figure 1: The Kinect depth field without bilateral filter

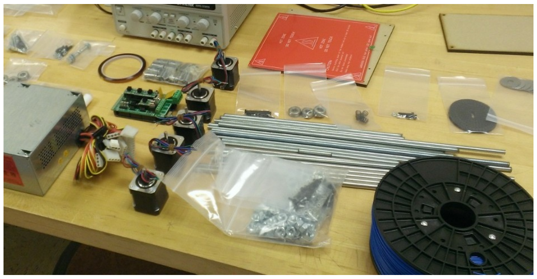


Figure 2: Parts of the 3D Printer

4 Discussion

5 Cost Analysis

Printbot LC- §549.00

3D Printer for developing reconstructed objects

Microsoft Kinect- §109.99

The most powerful sensor for its price, contains color filtering and depth filtering in one package

ikg 3mm ABS Spool-§46.00

Material that the 2D Printer uses for creating objects, spool is necessary for feeding into the printer

2x 1lb 3mm ABS- §36.00

Extra printing material, no extra spool required

Arduino Leonardo- §24.95

Microcontroller to plan orientation of object during reconstruction process

EasyDriver Stepper Motor Driver- §14.95

Allows the ability to control a stepper motion at lower voltages such as from an Arduino microcontroller

Stepper Motor with Cable- §14.95

Motor that can be controlled in "steps" this allows a more precise method for orienting objects being reconstructed

Total:

§795.84

6 Current Trends in Robotics and Computer Vision

One of the reasons that the Kinect has become so popular for computer vision projects is that it is a cheap, quick, and highly reliable for 3D measurements. Many researchers are beginning to look into the possibility of using this device to achieve everything from a 3D reconstruction of a scene to aiding in a SLAM algorithm. The fact that this device is so affordable, and so many new resources are available, makes the Kinect a viable device for conducting research in the field of robotics and computer vision.

The KinectFusion Project is slightly different than other projects that were using the Kinect; instead of using both the RGB cameras and the sensor, this project tracks the 3D sensor pose and preforms a reconstruction in real time using exclusively the depth data. This paper points out that depth cameras aren't exactly new, but the Kinect is a low-cost, real-time, depth camera that is much more accessible. The accuracy of the Kinect is called into questions, the point cloud that the depth data creates does usually contain noise and sometimes has holes where no readings were obtained. This project also considered the Kinect's low X/Y resolution and depth accuracy and fixes the quality of the images using depth super resolution. KinectFusion also looks into using multiple Kinects to preform a 3D body scan; this raises more issues because the quality of the overlapping sections of the images is compromised.

Another KinectFusion Project is the Real-time 3D Reconstruction and Interaction, this project is impressive because the entire process is done using a moving depth camera. With this software, the user can hold a Kinect camera up to a scene, and a 3D construction would be made. Not only would the user be able to see the 3D Reconstruction, but they would be able to interact with it; for instance, if they were to throw a handful of spheres onto the scene, they would land on the top of appropriate surfaces, and fall under appropriate objects following the rules of physics. To accomplish this, the depth camera is used to track the 3D pose and the sensor is used to reconstruct the scene. Different views of the scene are taken and fussed together into a singe representation, the pipe line segments the objects in the scene and uses them to create a global surface based reconstruction. This project shows the real-time capabilities of then Kinect and why that makes it an innovative tool for computer vision.

A study shown in the Asia Simulation Conference in 2011 demonstrated that a calibrated Kinect can be combined with Structure from Motion to find the 3D data of a scene and reconstruct the surface by Multi-view Stereo. This study proved that the Kinect was more accurate for this procedure than a Swiss-Ranger SR-4000 3D-TOF camera and close to a medium resolution SLR Stereo rigs. The Kinect works by using a near-infrared laser pattern projector and an IR camera as a stereo pair to triangulate points in 3D space, then the RGB camera is used to reconstruct the correct texture to the 3D points. This RGB camera, which outputs medium quality images, can also be used for recognition. One issue this study found was that the resulting IR and Depth images were shifted. To figure out what the shift was, the Kinect recorded pictures of a circle from different distances. The shift was found to be around 4 pixels in the u direction and three pixels in the v direction. Even after the camera has been fully calibrated, there are a few remaining residual errors in the close range 3D measurements. An easy fix for this error was to we form a z-correction image of z values constructed as the pixel-wise mean of all residual images and then subtract that correction image from the z coordinates of the 3D image. [1] Though the SLR Stereo was the most accurate, the error e (or the Euclidean distance between the points returned by the sensors and points reconstructed in the process of calibration) of the SR-400 was much higher than the Kinect and the SLR. This study shows that the Kinect is possible cheaper and simpler alternative to previously used cameras and rigs in the computer vision field.

Another subject of research that is looking into using the Kinect is the simultaneous localization and mapping algorithm, used to create a 3D map of the world so that the robot can avoid collision with obstacles or walls. The

SLAM problem could be solved using GPS if the robot is outside, but inside one needs to use wheel or visual odometry. Visual odometry determines the position and the orientation of the robot using the associated camera images, algorithms like Scale Invariant Feature Transformation (SIFT), used to find the interest points, and laser sensors, used to collect depth data. Since the Kinect has both the RGB camera and a laser sensor, this piece of technology is a good piece of hardware to use for robots computing the SLAM Algorithm. In the study conducted by the students in the Graduates School of Science and Technology, at Meiji University, they found that the Kinect worked well for this process for horizontal and straight movement, but they had errors when they tried to recreate an earlier experiment, this means that their algorithm successfully solves the initial problem, but accuracy fell over time. [2] They found that the issue was not with the Kinect, and that it could be solved using the Speed-Up Robust Feature algorithm (SURF) and Smirnov-Grubbs test to further improve the accuracy of their SLAM Algorithm. This study proved that the Kinect was a reasonable, inexpensive and non-special piece of equipment that is capable of preforming well in computer vision applications.

It seems as though the Kinect is a popular choice in current robotics and computer vision. This device is affordable, easily obtainable, and capable of a lot more than is expected from a video game add on. The Kinect combines a near-infrared laser pattern projector and an IR camera in one tool, and when combined with this eliminates the set up of some other configuration. The Kinect is also surprisingly accurate, requiring only some optimization software to make the results comparable to the results from a medium resolution SLR Stereo rig.

One of the most innovated uses for the 3D Printer is its applications in the medical field. Since 2010, people have been using 3D printers to print out prosthetic limbs. One company in California has been printing the totally customizable prosthetics, which cost about one tenth of traditional prothetic limbs. Another company is looking at the possibility of using a 3D printer to print a house. Right now the design fits on the back of a tractor trailer and the 3D printer prints out custom concrete parts the are then assembled to complete the house. Some 3D printers have the ability to change the printing head, so it can begin printing with one material and then switch to a different material, all based on the code it receives, this means that a 3D printer could theoretically print the concrete part of the house and switch to printing the plastic siding or the glass windows all on the same path around the outside of the house. The most importunes aspect of these 3D printer applications is that it drastically cuts down on production costs, allowing the consumer to pay a lower price and get a completely customized product. Rather than paying a person to design the object, and then have a bother construct it, with a 3D printer all that needs to be done is the design and the 3D printer automates the entire construction process. For example, the 3D printed prosthetics cost 5,000 dollars to print and customize by covering the 3D printed material in a shoe or sleeve while a normal generic prothetic would cost about

60,000 dollars. [5]

[6]

[7]

7

References

[1] Helmut Grabner Xiaofeng Ren-Kurt Konolige Andrea Fossati, Juergen Gall, editor. *Consumer Depth Cameras for Computer Vision Research Topics and Applications*. Springer-Verlag London, 2013.

[2] Satoshi Tanaka Soo-Hyun Park Jong-Hyun Kim, Kangsun Lee. *Advanced Methods, Techniques, and Applications in Modeling and Simulation*. Springer Japan, 2012.

[3] Jing Tong, Jin Zhou, Ligang Liu, Zhigeng Pan, and Hao Yan *Scanning 3D Full Human Bodies Using Kinects*. IEEE Transactions on Visualization and Computer Graphics, 2012

[4] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbo *KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera*. ACM Symposium on User Interface Software and Technology, 2011

[5] Ashlee Vance *3-D Printing Spurs a Manufacturing Revolution*. New York TImes, 2010

[6] Richard A. D'Aveni *3-D Printing Will Change the World*. Harvard Business Review, 2013

[7] Lesliei Gordon *The changing face of 3D printing*. Machine Design, 2013

[8] "How Kinect and Kinect Fusion (Kinfu) Works" Internet: http://razorvision.tumblr.com/post/15039827747/how-kinect-and-kinect-fusion-kinfu-work, December 2011 [2/10/2013]

8 Appendix