

Disparity-Based Performance Analysis of the Microsoft XBox Kinect

Chris Tralie

May 2, 2014

0.1 Depth from Disparity by Triangulation

The relation between disparity and depth is derived in [1] and [2] but is repeated here for completeness. Figure 1 shows the geometry of measurements that are taken with the Kinect. To derive the relationship between disparity d and depth Z , pick a part of the IR dot pattern that is projected onto the reference plane at depth Z_{ref} (green point in the figure), which is known, and then examine the same part of the pattern that is projected onto a plane further away at depth Z onto the object point (black point in the figure), which is the point that is being imaged. The reference point gets imaged to the pixel u_{ref} on the image plane, and the object point gets imaged to the pixel u . Then, by definition, the disparity d is pixel shift between the pattern imaged on the reference plane and the pattern projected onto the object plane, or

$$d = u - u_{\text{ref}}$$

Now note that by similar triangles in a process typical for pinhole cameras

$$u_{\text{ref}} = f\left(\frac{b + X_{\text{ref}}}{Z_{\text{ref}}}\right) = f\frac{b}{Z_{\text{ref}}} + f\frac{X_{\text{ref}}}{Z_{\text{ref}}}$$
$$u = f\left(\frac{b + X}{Z}\right) = f\frac{b}{Z} + f\frac{X}{Z}$$

Note also that by a different set of similar triangles

$$\frac{X}{X_{\text{ref}}} = \frac{Z}{Z_{\text{ref}}}$$

Thus

$$u = f\frac{b}{Z} + f\frac{X_{\text{ref}}\frac{Z}{Z_{\text{ref}}}}{Z} = f\frac{b}{Z} + f\frac{X_{\text{ref}}}{Z_{\text{ref}}}$$

and the disparity d is

$$d = u - u_{\text{ref}} = \frac{bf}{Z} - \frac{bf}{Z_{\text{ref}}}$$

Thus, the disparity depends only on the depth of the point in front of the camera, not its position on the plane; the x coordinate factors out. Re-arranging and solving for the depth Z .

$$Z = \frac{bf}{\frac{bf}{Z_{\text{ref}}} - d}$$

The quantity $\frac{bf}{Z_{\text{ref}}}$ is defined as d_{off} in [2] and is referred to as the “disparity offset.” Thus, the equation in its final form is

$$Z = \frac{bf}{d_{\text{off}} - d}$$

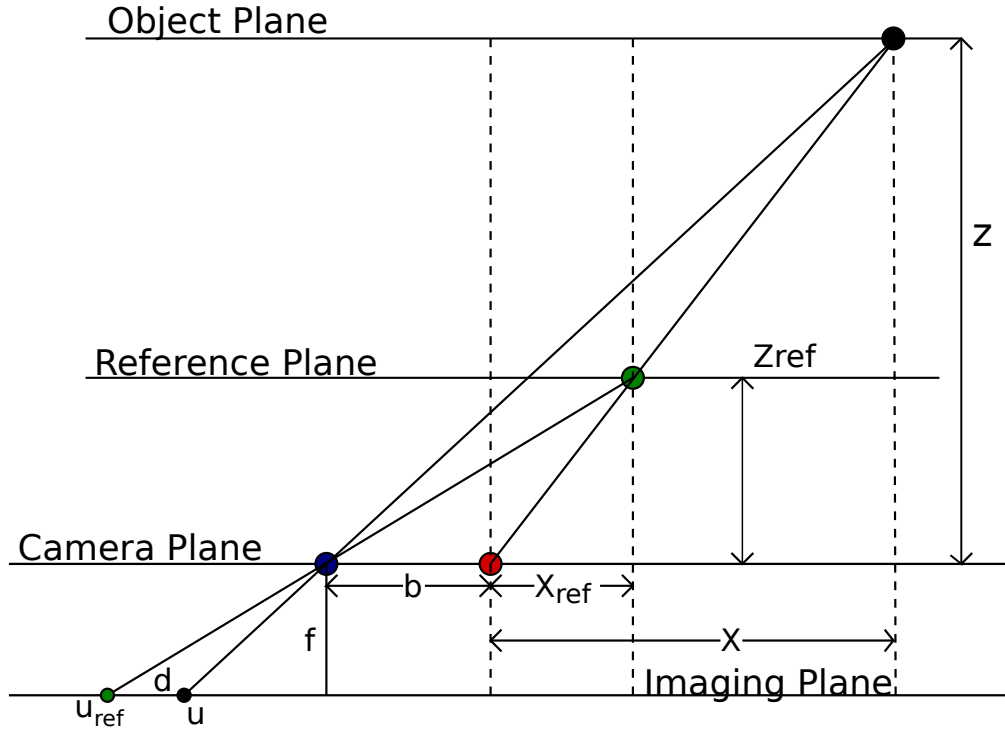


Figure 1: The geometry of converting disparity to depth. The blue dot shows the location of the IR camera and the red dot shows the location of the IR projector. The focal length is f and the baseline (separation between IR camera and IR projector) is b . Images are formed through the pinhole at the IR camera (blue dot) on the imaging plane, and the actual scene is above the camera plane. The green dot is the location of a part of the dot pattern on the reference plane, while the black dot is the same part of that pattern on another plane further away. X_{ref} is the horizontal distance of the reference point from the IR projector, and Z_{ref} is the depth of the reference point. Z is the depth of the actual point that is being measured, and X is the horizontal distance of that point from the IR projector

References

[1] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.

[2] Patrick Mihelich Kurt Konolige. Technical description of kinect calibration, April 2013.