Report: Mapping between world and image coordinates

by: Ankita Dhar, Student Number: 1154197, University of Melbourne

Abstract:
This report provides an analysis and result of an experiment to understand depth of images using "pinhole projection model" with application of basic geometry. In this experiment, we will work on an image captured by pinhole projection model. Our objective is to calculate co-ordinates of objects in real-world.

Theory:
Let's understand a simple camera system – a system that can record an image of an object or scene in the 3D world. This camera system can be designed by placing a barrier with a small aperture between the 3D object and a photographic film or sensor.

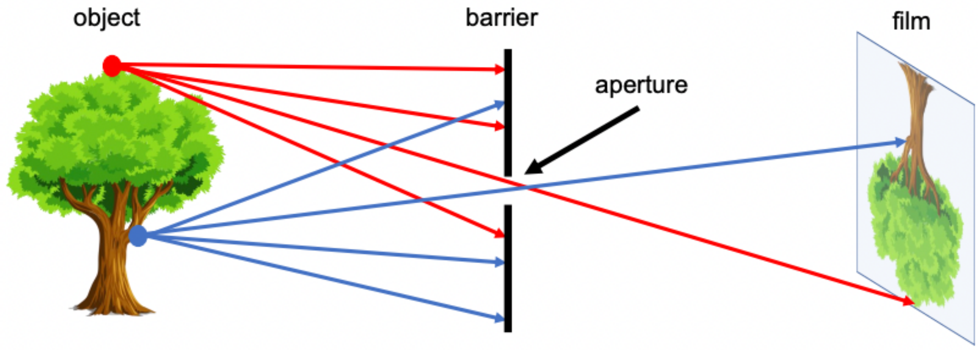

Figure 1.1: The pinhole projection model.

As Figure 1.1 shows, each point on the 3D object emits multiple rays of light outwards. Without a barrier in place, every point on the film will be influenced by light rays emitted from every point on the 3D object. Due to the barrier, only one (or a few) of these rays of light passes through the aperture and hits the film. Therefore, we can establish a one-to-one mapping between points on the 3D object and the film. The result is that the film gets exposed by an "image" of the 3D object by means of this mapping. This simple model is known as the pinhole projection model [1].

Assumptions:

- The aperture is assumed to be aligned on real-world z-axis with camera aperture at (0,0,0).
- Table surface plane as y=0.
- From the given image, the base of the coffee mug (point D in diagram 1.1) is found to be at 517 pixels from the top of the image. Therefore, for this experiment we are assuming that the (x, y) coordinates of the real-world origin, i.e. (0, 0), is corresponding to $(750\ pixel, 517\ pixel)$ in the 2D image.
- camera is oriented along z-axis.
- vertical axis of the image is exactly aligned with the vertical (y) axis in the world
- horizontal axis of the image is exactly aligned with the horizontal (x) axis in the world

Experiment:
For this experiment we have considered the image 1.1 with 1500 x 1000 pixels dimensions.

Image 1.1: Experimental image

Following facts are given:

- coffee pot height: 21.5 cm
- mug height: 9 cm
- camera's focal length: 32 mm
- imaging sensor height: 14.8 mm

Now, we need to calculate the (x, y, z) coordinates of point A and point B in the real-world. Let's say point A has coordinates (x_a, y_a, z_a) and point B has coordinates (x_b, y_b, z_b).

To begin with, let's understand the structure of this image capturing setup with the representation in diagram 1.1.

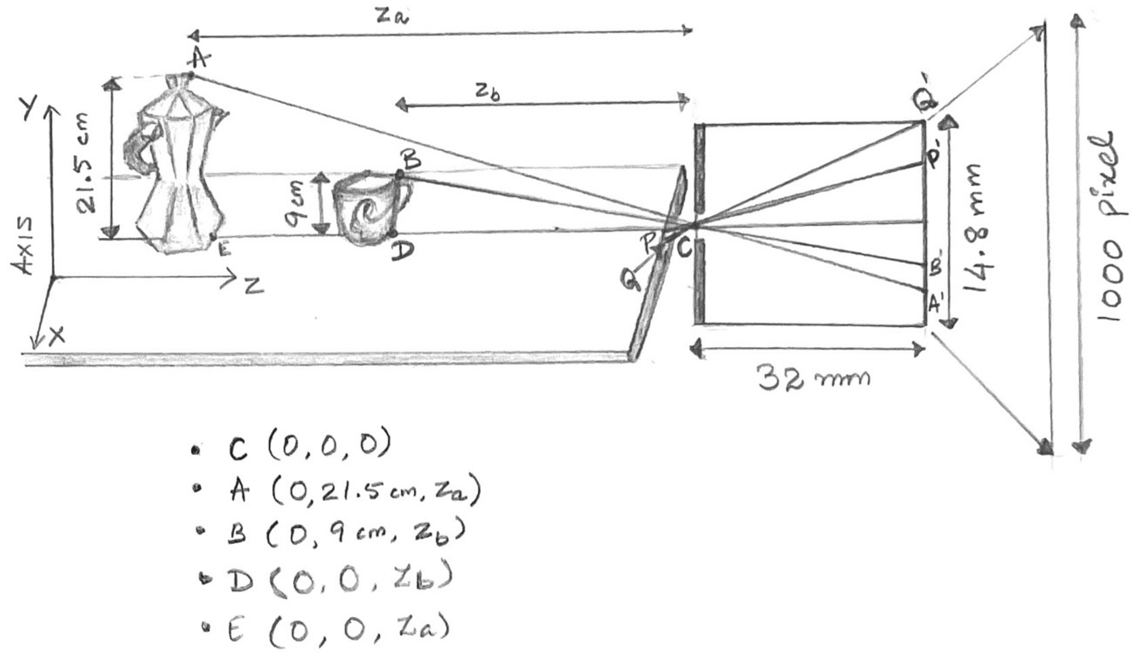

Diagram 1.1: Illustration of dimensions of experimental setup on x, y, z plane

From the diagram, we understand that $x_a = x_b = 0$, as centre of the image corresponds to $(0,0,z)$ and point A and point B are horizontally at the centre of the image; as both the points lie at 750 pixel distance from the either sides of the image. [Refer screenshots in Appendix] Point P' and Q' on senser plane, corresponding to real world points P and Q, are merely illustrating the darkness at the bottom of the image.

Here, we have taken into consideration a point D, which is at the base of the coffee mug, in real world, whose y-axis value (0) also corresponds to that of the base of coffee pot at point E. Both points D and E are also aligned with the x-axis value of points B and A which is also 0. Therefore, the coordinates of point D evaluates to be $(0,0,z_b)$ and that of point E are $(0,0,z_a)$). From the given image, the base of the coffee mug (point D) is found to be at the distance of 517 pixels from the top of the image.

As demonstrated in the diagram 1.1, the height of the sensor plane can be mapped to height of the image.

$$\Rightarrow 1000\ pixels = 14.8\ mm\ on\ sensor\ plane$$

Using this information, we can find out the point of coffee mug base (point D) on sensor plane, corresponding to 517 pixels in the image.

$$let's\ say, \qquad 517\ pixels = D'(y)$$

$$\Rightarrow D'(y) = E'(y) = \frac{517\ pixels \times 14.8\ mm}{1000 pixels}$$

$$\Rightarrow D'(y) = E'(y) = 7.6516\ mm.$$

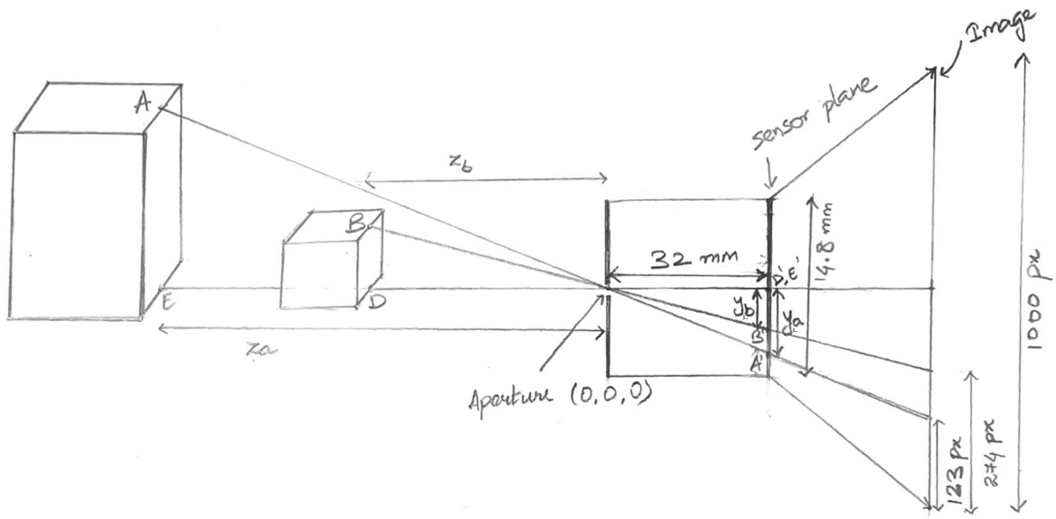

Diagram 1.2: Illustration of dimensions of experimental setup

Now, let's consider the diagram 1.2 for rest of the calculations.

Since the ratios of dimensions of the image on sensor plane will be maintained on image, we can say that,

$$123\ pixels = \frac{123\ pixels \times 14.8\ mm}{1000 pixels} = 1.8204\ mm\ on\ sensor\ plane$$

$$\Rightarrow y_{ia} = (7.6516 - 1.8204) mm = 5.8312\ mm\ on\ sensor\ plane$$

We know $y_a = 21.5\ cm = 215\ mm$, now using the similar triangle's formula, we can say:

$$z_a = f\ \frac{y_a}{y_{ia}}$$

$$\Rightarrow z_a = 32\ mm\ \frac{215\ mm}{5.8312\ mm} = 1179.86\ mm \approx 118\ cm$$

So, point A has (0, 21.5 cm, 118 cm) as coordinates in the real-world.

Similarly,

$$274\ pixels = \frac{274\ pixels \times 14.8\ mm}{1000 pixels} = 4.0552\ mm\ on\ sensor\ plane$$

$$\Rightarrow y_{ib} = (7.6516 - 4.0552) mm = 3.5964\ mm\ on\ sensor\ plane$$

We know $y_b = 9\ cm = 90\ mm$, now using the similar triangle's formula, we can say:

$$z_b = f\ \frac{y_b}{y_{ib}}$$

$$\Rightarrow z_b = 32\ mm\ \frac{90\ mm}{3.5964\ mm} = 800.801\ mm \approx 80\ cm$$

So, point B has (0, 9 cm, 80 cm) as coordinates in the real-world.

Findings:

Points	*Coordinates*
Point A	(0, 21.5 cm, 118 cm)
Point B	(0, 9 cm, 80 cm)

Table 1.1: Real-world co-ordinates of points A and B

Analysis:

As per our findings we see that Point B is much nearer to the camera aperture than Point A. This is in consistent with the fact that coffee pot is more than double the height of coffee mug and yet the difference of their heights is not much evident from the image, since nearer objects seem bigger and farther objects seem smaller.

Conclusion:

The experiment was successfully conducted to find the real-world coordinates of point A and point B.

References:

[1] Hata, K. and Savarese, S., 2021. *Camera Models*. [online] Web.stanford.edu. Available at: <https://web.stanford.edu/class/cs231a/course_notes/01-camera-models.pdf> [Accessed 18 August 2021].

Appendix

Images showing the co-ordinates of points A, B and D.

A
B
D
750, 517px
1500 × 1000px
Size: 452.8KB