

# Image Formation

## 1. What determines where a point in the 3D world appears on a 2D image?

*Camera position and orientation relative to 3D point.*

## 2. What determines how bright the image of that point is?

*Illumination (intensity, colour-spectrum, number of sources, locations of sources, etc.)*

*Reflectance (material properties, amount of light absorbed, reflected, etc.)*

*Sensor (sensitivity to different frequencies of light)*

## 3. Given the RGB values of a pixel, how can we tell the colour of the surface that is shown at that point in the image?

*We can't. The RGB values of the image will depend both on the properties of the surface (including its colour) and the properties of the light it is reflecting. Without knowledge of the latter, we can't know the former.*

## 4. Briefly define what is meant by the terms “focus” and “exposure”.

*Focus means that all rays coming from a scene point converge into a single image point.*

*Exposure is the time needed to allow enough light through to form an image*

## 5. Write down the thin lens equation, which relates the focal length of a lens to the depths of the image and object.

$$\frac{1}{f} = \frac{1}{\|z\|} + \frac{1}{\|z'\|}$$

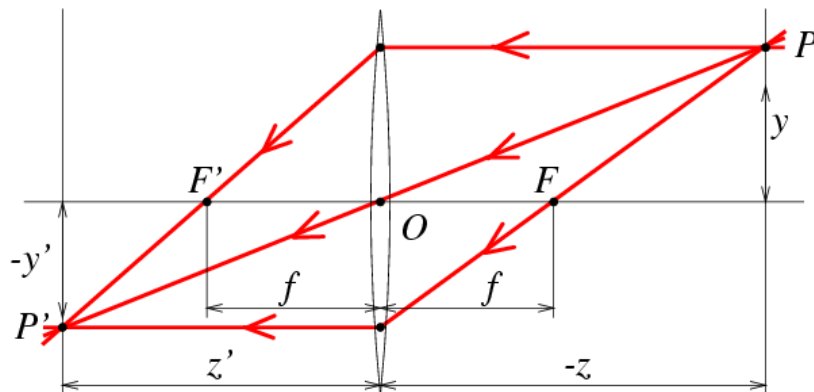
*where:*

*f = focal length of lens*

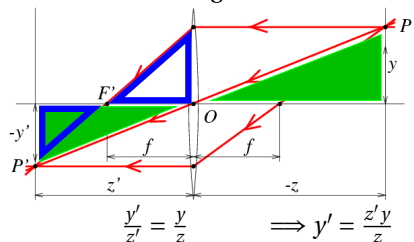
*z = distance of object from the lens*

*z' = distance of image from the lens*

## 6. Derive the thin lens equation.



*From similar triangles:*



*and*

$$\frac{y'}{z'} = \frac{y}{z} \quad \Rightarrow \quad y' = \frac{z'y}{z}$$

$$\frac{y'}{z'-f} = \frac{y}{f} \quad \Rightarrow \quad y' = \frac{(z'-f)y}{f}$$

$$\begin{aligned}
\text{equating for } y': \quad & \frac{z'y}{z} = \frac{(z'-f)y}{f} \\
y \text{ cancels, hence:} \quad & \frac{z'}{z} = \frac{(z'-f)}{f} = \frac{z'}{f} - 1 \\
\text{dividing both sides by } z': \quad & \frac{1}{z} = \frac{1}{f} - \frac{1}{z'} \quad \implies \frac{1}{z} + \frac{1}{z'} = \frac{1}{f}
\end{aligned}$$

**7. If a lens has a focal length of 35mm at what depth should the image plane be placed to bring an object 3m from the camera into focus? What if the object is at 0.5m?**

*For object at 3m:*

$$\begin{aligned}
\frac{1}{f} &= \frac{1}{\|z\|} + \frac{1}{\|z'\|} \\
\frac{1}{35} &= \frac{1}{\|3000\|} + \frac{1}{\|z'\|} \\
\frac{1}{\|z'\|} &= \frac{1}{35} - \frac{1}{\|3000\|} \\
z' &= 35.4 \text{ mm}
\end{aligned}$$

*For object at 0.5m:*

$$\begin{aligned}
\frac{1}{\|z'\|} &= \frac{1}{35} - \frac{1}{\|500\|} \\
z' &= 37.6 \text{ mm}
\end{aligned}$$

**8. If an object is at a distance of 3m from a *pinhole* camera, where should the image plane be placed to get a focused image?**

*A pinhole camera has an infinite focal range, meaning that the image will be in focus no matter what distance the image plane is from the pinhole.*

**9. Briefly compare the mechanisms used for focusing a camera and an eye.**

*A camera lens has a fixed shape, and hence, a fixed focal length. Focusing is achieved by moving the lens to change the distance to the image plane.*

*An eye lens has an adjustable shape, and hence, a variable focal length. Whereas the distance between the lens and the image plane (the retina) is fixed. Focusing is achieved by changing the focal length of the lens.*

**10. Briefly compare the mechanisms used for sampling the image in a camera and in an eye.**

*Camera: Has sensing elements sensitive to three wavelengths (RGB). These elements occur in a fixed ratio across the whole image plane. The sampling density is uniform across the whole image plane.*

*Eye: Has sensing elements sensitive to four wavelengths (RGBW). These elements occur in a variable ratios across the image plane (cone density highest at fovea, rod density highest outside fovea). The sampling density is non-uniform across the image plane (density is highest at the fovea).*

**11. A point in 3D space has coordinates [10,10,500] mm relative to the camera reference frame. If the image principal point is at coordinates [244,180] pixels, and the magnification factors in the x and y directions are 925 and 740, then determine the location of the point in the image. Assume that the camera does not suffer from skew or any other defect.**

**How do the coordinates of the image change if (a) the world coordinates are [20,20,500] (i.e. object is twice as large), (b) the world coordinates are [10,10,250] (i.e. camera is half as far from the object), (c) the world coordinates are [10,10,500] but the CCD array doubles its resolution from 20 pixels/mm to 40 pixels/mm**

$$\begin{aligned}
\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \frac{1}{z} \begin{pmatrix} \alpha & 0 & o_x \\ 0 & \beta & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \\
\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= \frac{1}{500} \begin{pmatrix} 925 & 0 & 244 \\ 0 & 740 & 180 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 10 \\ 500 \\ 1 \end{pmatrix} = \begin{pmatrix} 262.5 \\ 194.8 \\ 1 \end{pmatrix} \\
&\text{(a) the world coordinates are [20,20,500].}
\end{aligned}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{500} \begin{pmatrix} 925 & 0 & 244 \\ 0 & 740 & 180 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 20 \\ 20 \\ 500 \\ 1 \end{pmatrix} = \begin{pmatrix} 281 \\ 209.6 \\ 1 \end{pmatrix}$$

Distance from image centre to image of point originally:

$$\begin{pmatrix} 262.5 \\ 194.8 \end{pmatrix} - \begin{pmatrix} 244 \\ 180 \end{pmatrix} = \begin{pmatrix} 18.5 \\ 14.8 \end{pmatrix}$$

$$\text{Now: } \begin{pmatrix} 281 \\ 209.6 \end{pmatrix} - \begin{pmatrix} 244 \\ 180 \end{pmatrix} = \begin{pmatrix} 37 \\ 29.6 \end{pmatrix}$$

i.e., image doubles in size when the object doubles in size.

(b) the world coordinates are [10,10,250].

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{250} \begin{pmatrix} 925 & 0 & 244 \\ 0 & 740 & 180 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 10 \\ 250 \\ 1 \end{pmatrix} = \begin{pmatrix} 281 \\ 209.6 \\ 1 \end{pmatrix}$$

i.e., image doubles in size when distance to object is halved.

(c) the world coordinates are [10,10,500] but the CCD array doubles its resolution from 20 pixels/mm to 40 pixels/mm.

$$\alpha = \frac{-f'}{s_x}; \quad \beta = \frac{-f'}{s_y}$$

where  $s_x$  and  $s_y$  are the pixel sizes in the x and y directions (in mm/pixel)

If the resolution of the camera doubles, we can assume that the sizes of the pixels halves, hence the magnification factors double.

$$\text{Hence, } \alpha = 2 * 925 = 1850; \quad \beta = 2 * 740 = 1480$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{500} \begin{pmatrix} 1850 & 0 & 244 \\ 0 & 1480 & 180 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 10 \\ 500 \\ 1 \end{pmatrix} = \begin{pmatrix} 281 \\ 209.6 \\ 1 \end{pmatrix}$$

i.e., doubling the camera resolution doubles the image size (in pixels).

**12. The RGB channels for a small patch of image are shown below. Convert this image patch to a greyscale representation using an equal weighting for each channel and (a) 8 bits per pixel, (b) 2 bits per pixel.**

$$I_R = \begin{bmatrix} 205 & 195 \\ 238 & 203 \end{bmatrix}, \quad I_G = \begin{bmatrix} 143 & 138 \\ 166 & 143 \end{bmatrix}, \quad I_B = \begin{bmatrix} 154 & 145 \\ 174 & 151 \end{bmatrix}$$

(a) using 8 bits per pixel gives 256 grey levels, represented by the integers 0 to 255.

$$(205+143+154)/3 = 167.3333$$

$$(195+138+145)/3 = 159.3333$$

$$(238+166+174)/3 = 192.6667$$

$$(203+143+151)/3 = 165.6667$$

Therefore, rounding to nearest integer:

$$I_{\text{Grey}} = \begin{bmatrix} 167 & 159 \\ 193 & 166 \end{bmatrix}$$

(b) using 2 bits per pixel gives 4 grey levels, represented by the integers 0 to 3. Hence, 8-bit values in range [0, 63] map to 0, [64, 127] map to 1, [128, 191] map to 2, [192, 255] map to 3.

$$I_{\text{Grey}} = \begin{bmatrix} 2 & 2 \\ 3 & 2 \end{bmatrix}$$

**13. The image below shows the values of the red, green and blue pixels in a Bayer masked image. Calculate the RGB values for each pixel at the four central locations (i.e., locations 33, 34, 43, and 44), using: a) bilinear interpolation; b) smooth hue transition interpolation; c) edge-directed interpolation.**

R11	G12	R13	G14	R15	G16
G21	B22	G23	B24	G25	B26
R31	G32	R33	G34	R35	G36
G41	B42	G43	B44	G45	B46
R51	G52	R53	G54	R55	G56
G61	B62	G63	B64	G65	B66

*a) bilinear interpolation*

At Pixel 33:

$$\text{Red} = R33$$

$$\text{Green} = (G23+G34+G32+G43) / 4$$

$$\text{Blue} = (B22+B24+B42+B44) / 4$$

At Pixel 34:

$$\text{Red} = (R33+R35) / 2$$

$$\text{Green} = G34$$

$$\text{Blue} = (B24+B44) / 2$$

At Pixel 43:

$$\text{Red} = (R33+R53) / 2$$

$$\text{Green} = G43$$

$$\text{Blue} = (B42+B44) / 2$$

At Pixel 44:

$$\text{Red} = (R33+R35+R53+R55) / 4$$

$$\text{Green} = (R33+R35+R53+R55) / 4$$

$$\text{Blue} = B44$$

*b) smooth hue transition interpolation*

Interpolation of green pixels as above, call these values G33, G34, G43, and G44.

At Pixel 33:

$$\text{Red} = R33$$

$$\text{Blue} = G33 * (B22 / G22 + B24 / G24 + B42 / G42 + B44 / G44) / 4$$

At Pixel 34:

$$\text{Red} = G34 * (R33 / G33 + R35 / G35) / 2$$

$$\text{Blue} = G34 * (B24 / G24 + B44 / G44) / 2$$

At Pixel 43:

$$\text{Red} = G43 * (R33 / G33 + R53 / G53) / 2$$

$$\text{Blue} = G43 * (B42 / G42 + B44 / G44) / 2$$

At Pixel 44:

$$\text{Red} = G44 * (R33 / G33 + R35 / G35 + R53 / G53 + R55 / G55) / 4$$

$$\text{Blue} = B44$$

*c) edge-directed interpolation*

At Pixel 33:

$$\Delta H = \|G32 - G34\| \quad \Delta V = \|G23 - G43\|$$

$$\text{Green} = (G32+G34)/2 \text{ if } \Delta H < \Delta V$$

$$\text{Green} = (G23+G43)/2 \text{ if } \Delta H > \Delta V$$

$$\text{Green} = (G32+G34+G23+G43)/4 \text{ if } \Delta H = \Delta V$$

At Pixel 34:

$$\text{Green} = G34$$

At Pixel 43:

$$\text{Green} = G43$$

At Pixel 44:

$$\Delta H = \|G43 - G45\| \quad \Delta V = \|G34 - G54\|$$

$$\text{Green} = (G43+G45)/2 \text{ if } \Delta H < \Delta V$$

$$\text{Green} = (G34+G54)/2 \text{ if } \Delta H > \Delta V$$

$$\text{Green} = (G43+G45+G34+G54)/4 \text{ if } \Delta H = \Delta V$$

Interpolation of Red and Blue pixels as in (b).

**14. Briefly describe the different characteristics of the fovea and periphery of the retina.**

*Fovea:*

- *high resolution (acuity) - due to high density of photoreceptors*
- *colour - due to photoreceptors being cones*
- *low sensitivity - due to response characteristics of cones*

*Periphery:*

- *low resolution (acuity) - due to low density of photoreceptors and large convergence of photoreceptors to ganglion cells*
- *monochrome - due to photoreceptors being rods*
- *high sensitivity - due to response characteristics of rods*

**15. Describe the general receptive field structure of a retinal ganglion cell. Describe the full range of ganglion cell RFs found in the normally functioning human retina.**

*Ganglion cells have centre-surround RFs. There are two classes On-centre, off-surround (active if central stimulus is brighter than background) and Off-centre, on-surround (active if central stimulus is darker than background). The drive to the centre and surround can come from rods or cones, giving rise to various combinations of RF properties:*

<i>Centre</i>	<i>Surround</i>
W+	W-
W-	W+
R+	G-
R-	G+
G+	R-
G-	R+
B+	Y-
B-	Y+
Y+	B-
Y-	B+

**16. Briefly describe how Ganglion cell RFs give rise to: (a) efficient image coding, (b) invariance to illumination, (c) edge enhancement.**

*A centre-surround cell only responds strongly when there is a difference between the intensity of the stimulus in the centre compared to that in the surround (e.g. centre is bright, surround is dark). When the intensity across the RF is uniform, the cell responds weakly.*

*(a) efficient image coding.*

*In a typical image, across a large proportion of the image, intensity changes slowly. This means that only relatively few ganglion cells respond strongly. This minimises the number of active neurons. Also, each active neuron is encoding the difference between the centre and surround, this difference in intensity is less than the absolute intensity and hence can be encoded to the same precision using a lower bandwidth.*

*(b) invariance to illumination.*

*The absolute intensity of different parts of the image will change with the intensity of the illumination, however, the relative intensities of different parts of the image will remain constant. Each active neuron is encoding the difference between the centre and surround, and hence the response is relatively unaffected by illumination strength.*

*(c) edge enhancement.*

*Edges cause changes in intensity, changes in intensity cause centre-surround cells to become active. Hence, edges cause strong responses from ganglion cells.*