

1. Consider the stereo pair camera model given by $K_l=K_r=I$, $R_l=R_r=I$ (I is the 3×3

identity matrix), $t_l = \begin{bmatrix} -1/2 \\ 0 \\ 0 \end{bmatrix}$, $t_r = \begin{bmatrix} +1/2 \\ 0 \\ 0 \end{bmatrix}$.

(a) Find the point X_w in world coordinates which is imaged in the left camera at

$\begin{bmatrix} u_l \\ v_l \end{bmatrix} = \begin{bmatrix} -1/8 \\ 1/4 \end{bmatrix}$ and in the right camera at $\begin{bmatrix} u_r \\ v_r \end{bmatrix} = \begin{bmatrix} -3/8 \\ 1/4 \end{bmatrix}$.

REMINDER: the expression for a ray as derived in class is $\alpha R^{-1} K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + t$.

The rays intersect at the point (in camera coordinates)

$$\alpha_l \begin{bmatrix} -1/8 \\ 1/4 \\ 1 \end{bmatrix} + \begin{bmatrix} -1/2 \\ 0 \\ 0 \end{bmatrix} = \alpha_r \begin{bmatrix} -3/8 \\ 1/4 \\ 1 \end{bmatrix} + \begin{bmatrix} +1/2 \\ 0 \\ 0 \end{bmatrix}$$

From third eqn $\alpha_l = \alpha_r$, and plugging into first eqn get $\alpha_l = \alpha_r = 4$. So plugging into ray

expression for right camera, get $\begin{bmatrix} -3/2 \\ 1 \\ 4 \end{bmatrix}$. But that is the point in right camera coordinates.

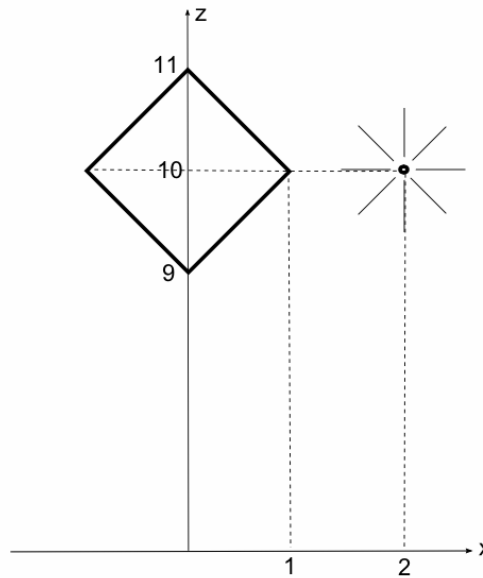
Must convert using $X_w = R^{-1} X_c + t$, which yields $X_w = \begin{bmatrix} -1 \\ 1 \\ 4 \end{bmatrix}$.

(b) If the focal lengths of both cameras were doubled but the camera model otherwise unchanged, where would the point X_w be imaged in the two cameras?

By similar triangles, when f is doubled the size of the image doubles. So $\begin{bmatrix} u_l \\ v_l \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1/2 \end{bmatrix}$,

$$\begin{bmatrix} u_r \\ v_r \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/2 \end{bmatrix}$$

2. The diagram below shows the crosssection in the xz plane of a three dimensional scene. The Lambertian object in the center has a square crosssection which is constant for all y , like a very tall square building. This object is illuminated by a line source of light (like a thin florescent tube) oriented parallel to the y axis, so in this crosssection it appears as a single point at $x=2, z=10$.



(a) Using orthographic projection, what point or points $\begin{bmatrix} x \\ y \end{bmatrix}$ in the image plane will be illuminated the brightest by the light reflected from the object? Explain your reasoning.

The reflected irradiance on the image plane will depend on two factors: the cosine of the angle between the source and the object surface normal, and the distance from the line source. On both counts, maxima occur at $x=1$, if dx/dz is defined as -1 at that point, else just to the left of that value of x , and for all y .

(b) Compute the ratio of irradiance at the point $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ in the image plane to the irradiance at the point $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \end{bmatrix}$.

The irradiance of the line source at the object falls off linearly with distance, and the reflected radiance from the object is proportional to the irradiance and to the cosine of the angle between the surface normal, which is $[+1, -1]^T$ for both points, and the line connecting the point with the light source.

For $(x,z)=(1/2,19/2)$ the distance to the source at $(2,10)$ is $\sqrt{5/2}$ and the vector to the source is $[3/2 \ 1/2]^T$. The cosine is the inner product of $[+1,-1]^T$ and $[3/2 \ 1/2]^T$ divided by the product of their norms, which is $1/((\sqrt{2})*\sqrt{5/2})=1/\sqrt{5}$.

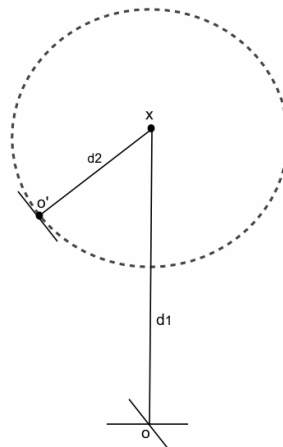
For $(x,z)=(1/4,37/4)$ the distance to the source at $(2,10)$ is $3*\sqrt{6}/4$ and the vector to the source is $[7/4 \ 3/4]^T$. The cosine is the inner product of $[+1,-1]^T$ and $[7/4 \ 3/4]^T$ divided by the product of their norms, which is $1/((\sqrt{2})*3*\sqrt{6}/4)=2/(3\sqrt{3})$. So the desired ratio is

$$\frac{\text{irradiance at } (x,y)=(1/2,1/2)}{\text{irradiance at } (x,y)=(1/4,1/4)} = \frac{3*\sqrt{5}}{2*\sqrt{5}} \frac{3*\sqrt{6}/4}{\sqrt{5/2}} = \frac{9*\sqrt{12/5}}{8} = 1.74$$

3 Suppose you have a robot equipped with a stereo camera pair where you could control the robot's translational motion, rotation of the stereo pair baseline ("neck" rotation) and rotation of each camera relative to the baseline ("eyeball" rotation).

(a) The Barth-Tsuji algorithm egomotion algorithm requires only one camera. Devise an algorithm for egomotion estimation for your robot which uses a stereo camera pair. For maximum credit, your algorithm should work even when there is just a single visible scene feature with no other ones near it (Barth-Tsuji fails in that case). You need not specify all the mathematical details, just show that egomotion can be estimated using your algorithm.

There are many possible answers to this question. Here is one! Fixate both eyes on a stationary point x in the world. As you move, maintain fixation on x by rotating the neck (baseline) while not rotating the eyes. Use stereopsis to get your distance at $d1$ and then $d2$ a small time later. Now there is only a single point o' that satisfies both the $d2$ distance constraint and the rotation of the neck angle. Once o' is known the angle between $o'o$ and ox is known and the neck can be rotated that amount to bring the optical axis in line with the velocity vector $o'o$.



(b) Comment on the degrees to which your algorithm adheres to these four paradigms: Marr, purposive vision, animate vision, active vision.

Marr: not at all, no recovery being done. Purposive: high degree of match with the motion heirarchy approach and minimal representations idea but uses quantitative measurements. Animate: stereo yes, fovea would be useful but is not necessary, indexical reference very little, gaze control, yes very important.