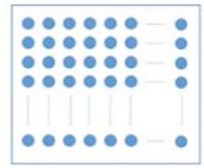
HW1 Carson Stevens

1. (25 pts) We can treat the human fovea as a square sensor array of size 1.5 mm x 1.5 mm, containing about 337,000 cones (sensor elements) (See the figure below). Assume that the space between cones is equal to width of a cone, and that the focal length of the eye is 17 mm.

1.1. What is the field of view (in degrees) of the human fovea?

$$FOV = 2\tan^{-1}\left(\frac{w/2}{f}\right) = 2\tan^{-1}\left(\frac{1.5/2}{17}\right) \approx \boxed{5.0522^\circ}$$



1.2. Estimate the distance from Brown Hall to the top of South Table Mountain (you can find this using a map, or a webtool such as Microsoft Bing Maps, or Google Earth). What is the minimum size object that you can see with the naked eye on top of the mountain? Can you see a person on top of the mountain? Assume for simplicity that size of the image of the object must cover at least two receptors (cones).

- To find cones (N x N): $\sqrt{\sim 337{,}000} = 580\text{ cones} \times 580\text{ cones}$
- To find cone diameter: $C_d = \frac{1.5m}{580\underbrace{(2)}_{\text{space}}} = 0.0013\text{ mm}$
- To find min size of cones to see: $d_s = 2(C_d)\cdot\underbrace{2}_{\text{space}} = 4(0.0013) = 0.0052\text{ mm}$
- To find min size of object to be seen *(Distance from Brown to South Table ~ 1080m)* :

$$\frac{e_d}{d} = \frac{d_s}{f} \Rightarrow e_d = \frac{d(d_s)}{f} = \frac{1080(0.0000052)}{0.0017} \approx \boxed{0.330353m}$$

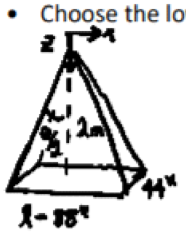
Since the minimum size we can see is ~ 0.330353m and a human is bigger, we should be able to see a human

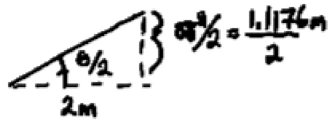
2. (25 pts) A pool-playing robot uses an overhead camera to determine the positions of the balls on the pool table. Assume that:

We are using a standard billiard table of size 44" x 88"
We are using standard 57mm Billiards balls
We need at least 100 square pixels per ball to reliably determine the identity of each ball
The center of the ball can be located to a precision of ± one pixel in the image
We need to locate the ball on the table to an accuracy of ± one cm
We are going to mount the camera on the ceiling, looking straight down. The distance from the camera to the table is 2 m.

Determine a configuration of the camera resolution and lens FOV that will meet these requirements. Assume that you can choose from the following parts:

- Lenses with field of view 30, 60, 90 degrees
- Cameras with resolutions of 256x256, 512x512, or 1024x1024 pixels
- Choose the lowest resolution that will meet the requirements.



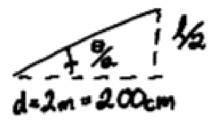


- To find FOV of camera:

$$\theta = 2\tan^{-1}\left(\frac{\frac{1.1176}{2}}{2}\right) \approx 58.3929^\circ$$

*To cover the whole table, a FOV of **60°***

- To find NxN camera resolution:
  - To find length of image for camera:



$$\frac{l}{2} = d\tan\left(\frac{\theta}{2}\right)$$
$$l = 2\cdot 200\tan(30^\circ)$$
$$l \approx 230.940\text{cm}$$

  - To find diameter of ball in pixels *(57mm => 5.7 cm)* :

$$l\cdot d_B = N\cdot 5.7 \qquad d_B = \frac{N\cdot 5.7}{l} = \frac{N\cdot 5.7}{230.940} \approx 0.024681\cdot N_{px}$$

  - To fine area of ball in image:

$$A = \pi r^2 = \pi\left(\frac{0.024681}{2}\cdot N\right)^2 \approx 4.7845\times 10^{-4}\cdot N^2\ px^2$$

  - To account for ball constraint *(100px / Area ball)* :

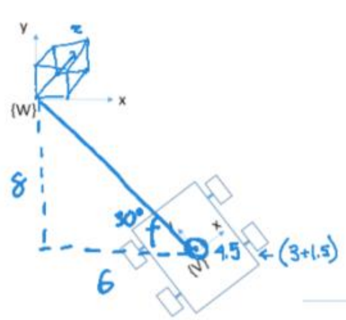
$$\frac{100px}{4.7845\times 10^{-4}\cdot N^2} \Rightarrow N = \sqrt{\frac{100}{4.7845\times 10^{-4}}} = 457.1719_{px}$$

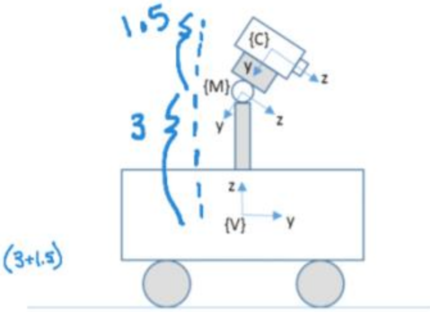
*To encompass the ~457.1719px, must use camera resolution of **512px x 512px***

3. A vehicle {V} is positioned at (6,-8,1) with respect to the world {W}. It is rotated by 30 degrees about the world Z axis, which points up. The figure below shows a top down view of the scene.

   A camera {C} is mounted on a rotational mount {M} on the vehicle, as shown in the figure below. The mount {M} is positioned directly above the vehicle origin at a distance = 3. It is tilted down by 30 degrees. The camera is rigidly attached to the mount. It is positioned directly above the mount origin at a distance = 1.5.

   A pyramid has vertices in world coordinates: (-1,-1,0), (1,-1,0), (1,1,0), (-1,1,0), (0,0,3). Using Python, generate an image of a wireframe model of the pyramid as if were seen by the camera, similar to the figure below, and a 3D plot showing the poses of the camera, mount, vehicle, and pyramid . Assume a pinhole camera model, with focal length = 600 pixels, where the image size is 640 pixels wide by 480 pixels high.





X	Y
185	34
273	45
329	26
247	17
266	172

Method: I first tried to visualize it and then I created Rotation Matrices and translation matrices for each coordinate system. I then found the Homogeneous Matrix for each of the rotation/translation pairs. I then multiplied them together to get the Homogeneous Camera to world and world to camera matrices. I then plotted these to make sure that my coordinate system was the same as the problem statement. I used the world to camera to homogeneous to find the Mext. I then multiplied the world points by Mext and K (K @ Mext @ P) to get the camera coordinates. I then plotted the points and drew lines to connect the points. Finally I save and show the image