# Question 1

Taking the partial derivative of D w.r.t to disparity d:

Conclusion:  
The dependence of the error in-depth estimation of a 3D point as a function of:

* Baseline length increase means a decrease in the error (inverse relation)
* Focal length increase means decrease in error (inverse relation)
* Stereo matching error increase means increase in error (proportional relationship)
* 3D point depth increase means increase in error (proportional relationship)

# Question 2

A picture containing graphical user interface

Description automatically generated

No, simply rotating about the optical center yields no depth information, it captures images of a 3D scene as if the scene were painted on a plane infinitely far away from the camera. We get no depth information since we don’t have a translation between two views (i.e. we can’t triangulate the multiple frames), so we don’t have an 3D information. Depth cues (parallax) can only be recovered when T is nonzero.

Yes, since we’re translating along the optical axis’s (i.e. along the Baseline from to ). We have (translation depth Z):

Chart, radar chart

Description automatically generated

Then with both rotation and translation (R, T) we net 3D information among the two image planes in reference to the point P.

# Question 3

Human:

1. Parallax and stereo motion in our eyes. For example, if we observe the same object in two different scenarios, our brains know to calibrate a weighted combination of information. The object in question would appear “large” In a small room and “small” a large room.
2. We can then relate object size to the size of nearby reference objects, like a brick wall, ignoring perceived sizes SL and SR when making their match.
3. In a static environment, performance depends on the distance of comparison cubes alone.
4. We can infer a lot of information of a blurry image sequence based on motion alone.
5. Triangulation of the human eyes, a view of profundity arises from a unique combination of a given 3D point in our left and right retinal pictures (for example when we try to directly look at our nose and “cross” our eyes).

Machine:

1. Video Coding and Compression & Video View interpolation
2. Robot navigation (robot vision)
3. Laser Scanning models
4. Recovering 3D models (Computer Vision) & Image-based Rendering
5. Anaglyph from Mars Rover

# Question 4 (attempt)

Fundamental Matrix:  
Text

Description automatically generated

Fig 1. Code Eight\_Point

Map

Description automatically generated

Fig 2. Epipole attempt 1 (using in-built)

I tried implementing the mouse click feature but I could only manage the red star output from two clicks.

I resorted to using in-built functions to net the epipoles in this attempt as yellow dashed lines. However, they don’t quite line up.

Map

Description automatically generated

Fig 3. Attempting to correct the outliers.

I tried to align the lines again to no avail.

Graphical user interface, text

Description automatically generated

Fundamental Matrix output

Graphical user interface, application

Description automatically generated

Fig 4. Two mouse clicks to generate a red point

Text

Description automatically generated

Fig 5. SURF features attempt 2

A picture containing graphical user interface

Description automatically generated

Fig 6. Epipoles matching the correct features but not interactive

Diagram, map

Description automatically generated

Fig 7. Surf features

SURF features highlighted the most prominent features of the images (green circles) so I could set up for triangulation in 3D.

Map

Description automatically generated with low confidence

Fig 8. 3D construction

I tried aligned the feature matches together to net a 3D estimation. Not interactive though.

Discussion: The images above have corners, edges, smooth regions (roof and parking lot), textured regions (grass), and occluded regions (behind the stadium).

My vision system was partially successful in finding point matches (mainly in the second surf features attempt) yet fails to correctly highlight 8 correct prominent matches. It would also fail to fully detect occluded features due to the nature of 2D imagery and the computer not knowing the 3D structure like we humans would. We can “fill in” occluded space in our mind through a fairly accurate estimation. Machines don’t have that luxury.