Computer Vision
Spring 2024

Quiz-1 Solutions

Question 1

Assigned TA: Sreenya Chitluri
Solution:
Parallel lines in the real world appear to converge to a point on the image plane due to perspective projection. Perspective projection is a phenomenon where objects appear smaller as they move further away from the observer. This effect causes parallel lines, which are actually equidistant from each other in physical space, to appear to converge towards a vanishing point in the distance when projected onto a two-dimensional surface, such as a photograph or a canvas.
Example - Parallel railway tracks intersect in an image.

Other logical or mathematical reasonings have also been accepted if they are relevant and complete.

Question 2

Assigned TA: Sreenya Chitluri
Solution:

1. Limited Semantic Understanding
2. Loss of Spatial Information
3. Vocabulary Size and Generalization

Stating the reason and explaining it briefly is expected. Other reasons which have not been stated above and are correct are also accepted.

Question 3

Assigned TA: Mohd Hozaifa Khan
Solution:
We can exploit a stereo imaging setup and use the concept of *disparity* to calculate depth. However, with a textureless wall, *correspondence finding methods using* ***wall features*** *are ineffective.* Instead, we can use *our main camera as a source of unique features.*
We position two additional cameras facing the main camera on the wall, creating a parallel stereo camera setup, as shown in 1. Using the equation

$$z = \frac{f \cdot b}{d}$$

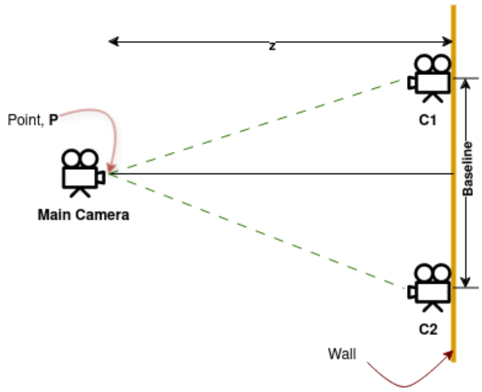

Figure 1: Q3-Stereo Setup

, we can compute depth, where f is the focal length, b is the baseline (the distance between the optical centres of the two cameras), and d is disparity.
Disparity computation involves solving a correspondence matching problem to find corresponding points in different images. For this, we can use a template or feature-matching method, with the search space limited to a line due to our parallel setup. We can calculate disparity using the equation $d = (x_l - x_r)$, where x_l and x_r are the x-coordinates of the corresponding points in the left and right images, respectively.
By using our main camera as the source of distinct features, we can accurately compute depth in our stereo imaging setup, even with a textureless wall.

In case of alternate solutions, the following criteria will be used:

- Clarity and completeness of solution
- Violation of constraints given in the question
- Correctness of solution given your assumptions

Question 4

Assigned TA: Sanyam
Solution:
Definition of Epipolar Constraint with Fundamental Matrix equation (0.5): "We have epipolar plane P created using baseline B and ray R1. e1 and e2 are epipoles, and L2 is the epipolar line. Based on the epipolar geometry, search space for pixel in image i2 corresponding to pixel x1 is constrained to a single 2D line which is the epipolar line l2. This is called the epipolar constraint."
Explanation for Fundamental Matrix Equation: $x_1^T F x_2 = 0$
Examples (0.5): Reduced Search Space, Stereo Disparity, Parallel Images (Any one)

Question 5

Assigned TA: Sanyam
Solution:
Part 1 (0.5): Rectangular Pixels/Individual Sensors

Part 2 (0.5): Calibration Matrix with explicit explanation of f_x and f_y (and how it is related to magnification) The focal length f_x (for example) is actually the product of the physical focal length of the lens and the size s_x of the individual imager elements (this should make sense because s_x has units of pixels per millimeter† while F has units of millimeters, which means that f_x is in the required units of pixels). Of course, similar statements hold for f_y and s_y.
$f_x = F * s_x$ and $f_y = F * s_y$

Question 6

Assigned TA: Brunda
Solution:
0.5 - For writing the assumptions where we can write 3*4 projection matrix as 3*3 0.5 - For writing example in the real world.
Approach 1: Assuming the object is planar (say xy plane and Z=0), then we can write the 3*4 projection matrix as 3*3 by eliminating the third column of the matrix Example: Capturing a planar wall (any relevant example)
Approach2: If the translation component is zero in the projection or when the camera coordinates and world coordinates are same or the camera is assumed to be at the world origin.(Any Relevant Example)

Question 7

Assigned TA: Shreya
Solution:
For a patch P in an image, let λ_1 and λ_2 be the largest and smallest eigen values of H, and x_1 and x_2 be the corresponding eigen vectors.
Here x_1 and x_2 are the directions of the largest and smallest intensity gradients respectively. And λ_1 and λ_2 are the corresponding magnitudes of gradients.
If the patch P contains an edge, the gradient of intensity along the edge will be zero or close to zero, and the gradient perpendicular to the edge will be high. Hence, λ_2 will be small (close to zero), and λ_1 will be large.
A corner consists of two non-parallel edges. Hence, a significant directional gradient will exist in all directions. i.e. both λ_1 and λ_2 will be large and comparable.
Hence, to check if a P contains an edge, or a corner, we can check the following:

- if $\lambda_1 \sim \lambda_2$ and $\lambda_1, \lambda_2 \gg 0$ then P contains a corner.
- else if $\lambda_1 \gg \lambda_2$ and $\lambda_2 \sim 0$ then P contains an edge.
- else if $\lambda_1, \lambda_2 \sim 0$ then P contains a flat surface.