Computer Vision Questions Set

1. Derive the rotation projection matrix in clockwise direction at angle θ along the origin in 2D using shear projection matrix with shearing factors shx and shy. Include any other projections in the process of derivation if required.

2. You are given a 3D point P (-1, 1, 2) and the corresponding projected point P1 (-1, 3, 5) in the same 3D coordinate system.

a) Compute the projection Matrix.

b) Estimate the possible sequence of atomic projections.

3. Consider a real pinhole camera with an ideal diameter. Consider changing the diameter. As the diameter is increased, the image degrades. Likewise, when the diameter is shrunk, the image also degrades. Why is this?

4. Given two ideal pinhole cameras where:

- The baseline of the two cameras is parallel to their scanlines,
- The optical axes of the two cameras intersect to form an angle of 90 degrees,
- The two centers of projection are at equal distances from the intersection of the optical axis, and
- The field of view of each camera is 90 degrees.

Draw the epipoles and a few epipolar lines.

5. A camera calibration software gave the following parameters for a particular camera:

$$\alpha_x = 650,\ \alpha_y = 650,\ \beta_x = 303,\ \beta_y = 242$$

The camera rotation matrix and the translation matrix with respect to the world coordinate system are

$$R = \begin{matrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \text{and} \qquad \overline{C} = [0 \quad 0 \quad 2]^T$$

i. Computer the camera matric P

ii. Compare the Cartesian coordinate of the image of the point $X = [2\ \ 3\ \ 10\ \ 1]^T$

iii. What will happen to the size of the image of an object when $\alpha = \alpha_x = \alpha_y$ increase?

6. Consider a 3D world scene where the cameraman is walking and recording the scene. Suppose you get his camcorder and recorded video after one year. Give a step-wise algorithm/procedure on how you will estimate the 3D model including the depth of the real-world scene using the computer vision geometry concepts you have studied. Support your answers with a few sketches to explain your approach. Make appropriate assumptions.

7. Compute the descriptor of center pixel in the following 3 × 3 × 1 grayscale image (Table 1).

0	0	1
0	1	0
1	0	0

Table 1: Grayscale Image

8. (2 points) Explain the applicability of Sobel filter over Prewit filter for edge detection with an appropriate example.

9. To compute the Laplacian of the given function

$$f(x,y) = \frac{1}{\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x^2+y^2)}$$

and compute the equivalent mask of size 3 × 3 (Table 2), where, x and y represents the index of a pixel (x, y) in the mask.

(0,0)	(0,1)	(0,2)
(1,0)	(1,1)	(1,2)
(2,0)	(2,1)	(2,2)

Table 2: Filter

10. Compute the Fourier transform for the following grayscale image (Table 1).

11. Explain the Laplacian of Gaussian (LoG) and difference of Gaussians (DoG) with an example. Why DoG is preferred over LoG in SIFT based feature extraction in the image data.

12. You are given an image database containing images of animals and human faces. You are required to extract discriminative features for each of the image in the database and design an inference system to automatically detect between animals and human faces. Describe all the steps with an example.

13. Discuss the following part with an appropriate example.

a) Orthogonal Projections and Oblique Projections.

b) 1/2/3 - Vanishing Points in Perspective Projections.

c) Homogeneous Coordinate System and Homogeneous Division

d) Cavalier and Cabinet Projection.

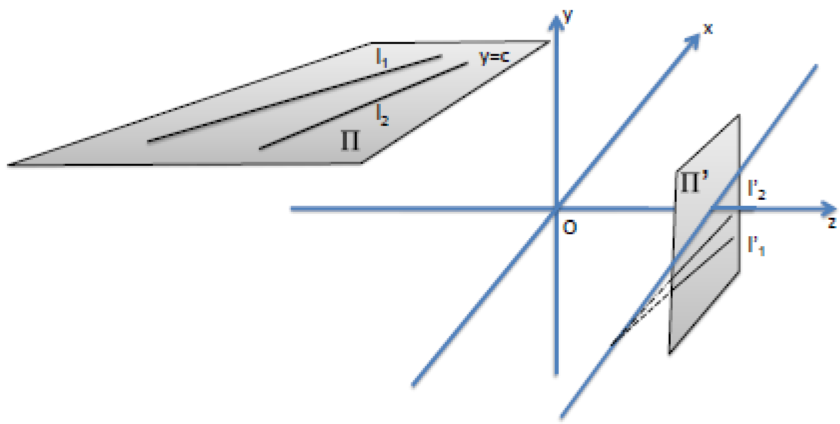
14. What is a vanishing point? Can there be more than one vanishing point in the image?

15. Homogeneous coordinates of point P in the image are $P = [40\ 164\ 2]^T$. Which row and column contains P if homogeneous coordinates are measured from the top left corner of the image?

16. State the camera calibration problem.

17. What is the motion between two images of a stereo pair whose epipoles are located: (a) in the center of the images, (b) in the infinity;

18. (a) In Figure 1, there are two parallel lines l_1 and l_2 lying on the same plane II. l'_1 and l'_2 are their projections through the optical center O on the image plane II'. Let's define plane II by y = c, line l_1 by equation $ax + bz = d_1$, and line l_2 by equation $ax + bz = d_2$. For any point P = (x; y) on l_1 or l_2, use the perspective projection equation below to find the projected point P' = (x', y') on the image plane. f' is the focal length of the camera.



(b) It turns out l'_1 and l'_2 appear to converge on the intersection of the image plane II' given by $z = f'$ and the plane y = 0. Explain why.

19. A scene point at coordinates (400, 600, 1200) is perspectively projected into an image at coordinates (24, 36), where both coordinates are given in millimeters in the camera coordinate frame and the camera's principal point is at coordinates (0,0,f). Assuming the aspect ratio of the pixels in the camera is 1, what is the focal length of the camera?

20. We want a method for *corner detection* for use with 3D images, i.e., there is an intensity value for each (x,y,z) voxel. Describe a generalization of either the Harris corner detector or the Tomasi-Kanade corner detector by giving the main steps of an algorithm, including a test to decide when a voxel is a corner point.

21. The SIFT descriptor is a popular method for describing selected feature points based on local neighborhood properties so that they can be matched reliably across images. Assuming feature points have been previously detected using the SIFT feature detector, (i) briefly describe the main steps of creating the SIFT feature descriptor at a given feature point, and (ii) name three (3) scene or image changes that the SIFT descriptor is invariant to (i.e., relatively insensitive to).

22. After running your favorite stereo algorithm assume you have produced a dense depth map such that for each pixel in the input image you have its associated scene point's (X, Y, Z) coordinates in the camera coordinate frame. Assume the image is of a scene that contains a single dominant plane (e.g., the front wall of a building) at unknown orientation, plus smaller numbers of other scene points (e.g., from trees, poles and a street) that are not part of this plane. As you know, the plane equation is given by $ax + by + cz + d = 0$. Define a *Hough transform* based algorithm for detecting the orientation of the plane in the scene. That is, define the dimensions of your Hough space, a procedure for mapping the scene points (i.e., the (X, Y, Z) coordinates for each pixel) into this space, and how the plane's orientation is determined.

23. Under perspective projection, what does a circle in the scene project to on the image plane?

24. For each of the following properties of edge detectors, indicate whether the **Canny edge detector or the Marr-Hildreth** detector is better with respect to this property, and explain briefly why.

- **a)** Fewer number of parameters that must be set.
- **b)** Closed chains of edges detected.
- **c)** Computes the edge orientation.
- **d)** Better localization of the true edge position.
- **e)** Can be implemented more efficiently (i.e., faster execution time).
- **f)** Requires non-maximum suppression to thin edges.
- **g)** Is an isotropic operator.
- **h)** Less likely to round corners where the boundary curvature is high.
- **i)** Less sensitive to noise.

25. (a) Perform 2D linear Convolution between two matrices

$$x[n] = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad h[n] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

(b) Consider the two image subsets S1 and S2, shown in the following figure. For V=[1], determine whether these two are (a) 4-adjacent, (b) 8-adjacent or (c) m-adjacent

		S_1				S_2			
0	0	0	0	0	0	0	1	1	0
1	0	0	1	0	0	1	0	0	1
1	0	0	1	0	1	1	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	1	1	1

26. Show that convolving an image with a discrete, separable 2D filter kernel is equivalent to convolving with two 1D filter kernels. Estimate the number of operations saved for an NxN image and a $2k + 1 \times 2k + 1$ kernel.

27. Consider a grayscale image represented by a 4x4 matrix of pixel values and represent its intensity value in 8-bit binary form.

100	50	150	200
75	25	125	175
225	175	75	25
125	200	100	150

i. Explain the concept of bit-plane slicing in the context of image processing.

ii. Perform bit-plane slicing, breaking down its binary representation into individual bit-planes (MSB, Bits 6-4, Bits 3-1, LSB).

28. Define the process of bit-plane slicing use the image 5x5

(a)

3	6	5	2	0
1	4	7	2	3
4	2	0	6	5
3	1	5	7	4
6	4	2	1	0

(b) Apply the bit plane slicing of the following image 3x3,

7	6	5
4	3	2
1	1	0

29. (a) Analyze the digital negative of the following 3x3 grey scale image

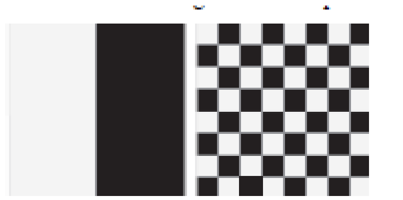
121	205	217
139	127	157
252	117	236

(b) Perform thresholding on image segment f(x, y) with t =128.

0	10	50	100
5	95	150	200
110	150	190	210
175	210	255	100

30. The two images shown in the following figure are quite different, but their histograms are the same. Suppose that each image is blurred using a 3 × 3 box kernel.

(a) Would the histograms of the blurred images still be equal? Explain.



(b) If your answer is no, either sketch the two histograms or give two tables detailing the histogram components.