AVD624 - Computer Vision

Assignment

Quiz I Solutions

Submitted by

Vatsalya Gupta
SC19B098
B.Tech. ECE VI Semester

Department of Avionics
Indian Institute of Space Science and Technology
Thiruvananthapuram - 695 547
April 2022

Question - 1

Explain what are the key ideas that are used in edge detection. What all things can cause an edge in an image. Name a few methods to detect edges in the image. How Sobel edge detection method works.

The basic idea of edge detection is that there would be pixel intensity variation very suddenly in an image in spatial domain. High frequency implies we can apply high pass filtering to detect the edges in the image. Hence, we take derivative.

Edges can be caused by intensity variation due to object boundary, variation in light falling at a certain area of some object, etc.

But HPF might take in high frequency noise when trying to detect edges. So we need to apply some kind of smoothing or blurring first before taking derivative. One such method is Sobel edge detection, which applies blurring before taking the derivative and convolving with image.

Other methods include convolving the image with derivative of Gaussian or Laplacian (second order derivative) of Gaussian, since Gaussian is inherently applying blur and is followed by derivative. DoG would give a peak at the edges as output when convolved with image, whereas LoG would give a zero crossing at the edges.

Question - 2

The image below is an image of a 3 pixel thick vertical line.

(a) Show the resulting image obtained after convolution of the original with the following approximation of the derivative filter [-1; 0; 1] in the horizontal direction. How many local maxima of the filter response do you obtain?

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(b) Suggest a filter which when convolved with the same image would yield a single maximum in the middle of the line. Demonstrate the result of the convolution on the original image.

(c) What is a difference between Gaussian smoothing and median filtering? How would you decide to use one vs. another?

(a) The response of each row after convolution with the above filter will be

0 0 1 1 0 -1 -1 0 0

There are two local extrema at 1 and -1.

(b) Convolving the image with the following filter will yield single maximum in the middle of the line $g = [1, 2, 1]^T$. The response of each row after convolution with this filter will be

0 0 1 3 4 3 1 0 0

(c) Gaussian smoothing is a filtering technique in which a 2D uniform Gaussian distribution is arranged in the form of a square matrix and convolved with the image matrix. It acts as low pass filtering. Median filtering is the technique in which the iage pixel value is replaced with the median of its neighbouring pixels.

Gaussian is used for smoothing and it is better when the noise is not spiky. Median filter is used for noise removal, and is more suitable for salt and pepper noise.

Question - 3

Given two lines in the image denoted by their projective coordinates l_1 and l_2, how would you compute an intersection point of these two lines?

$$l_1 = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \quad , \quad l_2 = \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$$

We can get their intersection point ($\mathbf{x}$) by their cross product.

$$l_1 \times l_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = \mathbf{i}(b_1c_2 - b_2c_1) + \mathbf{j}(a_2c_1 - a_1c_2) + \mathbf{k}(a_1b_2 - a_2b_1) = \begin{bmatrix} b_1c_2 - b_2c_1 \\ a_2c_1 - a_1c_2 \\ a_1b_2 - a_2b_1 \end{bmatrix} = \mathbf{x}$$

Question - 4

Show that the line through two 2D points x and x' is $l = x \times x'$.

Let $\mathbf{x} = [a_1 \ b_1 \ 1]^T$ and $\mathbf{x'} = [a_2 \ b_2 \ 1]^T$ in homogeneous coordinates. Then their cross product can be written as

$$\mathbf{x} \times \mathbf{x'} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & b_1 & 1 \\ a_2 & b_2 & 1 \end{vmatrix} = \mathbf{i}(b_1 - b_2) + \mathbf{j}(a_2 - a_1) + \mathbf{k}(a_1b_2 - a_2b_1) = \begin{bmatrix} b_1 - b_2 \\ a_2 - a_1 \\ a_1b_2 - a_2b_1 \end{bmatrix} = \mathbf{l}$$

Hence, we obtain the equation of the line passing through the two points as $ax + by + c = 0$, where $a = b_1 - b_2$, $b = a_2 - a_1$ and $c = a_1b_2 - a_2b_1$.

Question - 5

Under homography, how can we write the transformation of points in 3D from camera 1 to camera 2? Explain (mention all the steps needed) how do we estimate the homography.

We can estimate homography using the following steps:

1. Select 4 correspondence points on the image.
2. Form 8 equations using the equation x = HX, where H is the homography matrix. Each point (x,y) gives 2 equations.
3. Let $h_{33} = 1$. Solve the 8 equations, given 8 variables (h_{ij}, $i, j \in [1, 3]$).
4. We get the homography matrix using the SVD for solution.
5. If some points lie on same plane and lead to infeasible, then take more than 4 points and least squares method would give the solution.

Question - 6

What does it mean for the 2D convolution kernel (filter) to be separable?

An $N \times N$ convolution kernel is called separable if it can be written as a product of 1D kernels $N \times 1$ and $1 \times N$.

$$[K]_{N\times N} = [K_1]_{N\times 1} \times [K_2]_{1\times N}$$

The necessary condition is that the original kernel $[K]_{N\times N}$ should be of rank 1 for it to be separable.

Question - 7

Which of the following statements are true of a pinhole camera? Which are false?

(a) Images in a pinhole camera are upside down.

(b) A pinhole camera has a fixed focal length, f.

(c) Images in a pinhole camera are a perspective projection.

The statements (a) and (c) are true, whereas (b) is false.

Question - 8

(Write all that apply) Which of the following could affect the intrinsic parameters of a camera?

(a) A crooked lens system.

(b) Diamond/Rhombus shaped pixels with non right angles.

(c) The aperture configuration and construction.

(d) Any offset of the image sensor from the lens' optical center.

The statements (a), (b) and (d) apply, whereas (c) does not apply.

Question - 9

State true or false (with a proper reasoning).

(a) Pincushion distortion modifies the image only along the vertical direction and barrel distortion modifies the image only along the horizontal direction.

(b) The vanishing point associated with a line in 3D space (when viewed through a pinhole camera) can never be a point at infinity.

(c) The result of applying a scale, rotation, and translation (in that order) to a vector $\mathbf{v}$ is equal to the result of applying the same rotation, translation, and scale (in that order) to the same vector $\mathbf{v}$.

(d) It is impossible to estimate the intrinsic parameters of the camera given a single image even with prior information about the scene.

(a) False. Both of them modify the image along both directions. In case of barrel distortion, all axes bend outwards, whereas in pincushion distortion the axes bend inwards, be it vertical or horizontal. Only the central axes remain straight in both cases.

(b) False. The vanishing point can be at infinity if the line in 3D space is in a plane parallel to image plane.

(c) False. Both the sequences of operations would give different results as they would be matrix multiplications, and matrix multiplication is not commutative, i.e. $[A][B] \neq [B][A]$.

(d) False. If the scene information is given, then we can get the correspondence points and obtain the projection matrix using DLT and SVD. Then, using QR decomposition we can get the intrinsic matrix K.

Question - 10

Which of the following always hold(s) under affine transformations?

(a) Parallel lines remain parallel.

(b) Ratio of lengths of parallel line segments remain the same.

(c) Ratio of areas remain the same.

(d) Perpendicular lines remain perpendicular.

(e) Angles between two line segments remain the same.

The statements (a), (b) and (c) always hold under affine transformation.

Question - 11

The figure (1) below shows the outputs of applying one of transformations (projective, affine, similarity, and isometric) to a square with vertices at (1,1), (1,-1), (-1,-1), (-1,1) tell which is the most specific transformation used to generate each output. if possible write the transformation matrix and write the degrees of freedom.

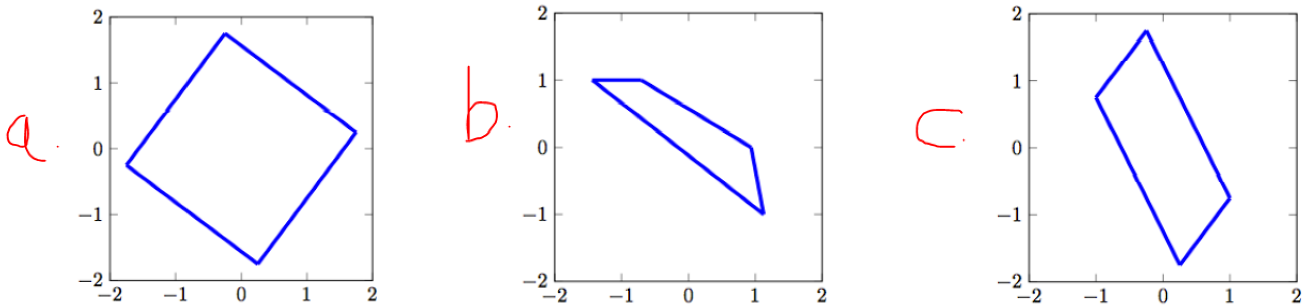

Figure 1: Figure for understanding the various transforms.

(a) Similarity transformation. It preserves angles and has 4 degrees of freedom. Its transformation matrix is $[s\mathbf{R}\ \mathbf{t}]_{2\times3}$.

(b) Projective transformation. It preserves straight lines and has 8 degrees of freedom. Its transformation matrix is $[\tilde{\mathbf{H}}]_{3\times3}$.

(c) Affine transformation. It preserves parallelism and has 6 degrees of freedom. Its transformation matrix is $[\mathbf{A}]_{2\times3}$.

Question - 12

Suppose we want to solve for the camera matrix K and that we know that the world coordinate system is the same as the camera coordinate system. Assume the matrix K has the structure outlined below. Note that K_{33} is an unknown. Assume that we are given n correspondences. Each correspondence consists of a world point (x_i, y_i, z_i) and its projection (u_i, v_i) for $i = 1, ..., n$.

$$\begin{pmatrix} K_{11} & K_{12} & K_{13} \\ 0 & K_{22} & K_{23} \\ 0 & 0 & K_{33} \end{pmatrix}$$

(a) What is the minimum number of correspondences needed to solve for the unknowns in the matrix K?

(b) Set up an equation of the form $Ax = 0$ to solve for the unknowns in K (where A is a matrix, and x and 0 are vectors). Be specific about what the matrix A and vector x are.

(c) Explain how to solve for the unknowns in the camera matrix K. Make sure $K_{33} = 1$.

(a) We have six unknowns in the camera matrix. Each correspondence gives two equations. Therefore, we need at least 3 correspondences.

(b) Each correspondence point will give us an equation of the form

$$K_{33} z_i \begin{bmatrix} u_i \\ v_i \end{bmatrix} - \begin{bmatrix} K_{11} x_i + K_{12} y_i + K_{13} z_i \\ K_{22} y_i + K_{23} z_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

On simplifying, we get

$$\begin{bmatrix} x_i & y_i & z_i & 0 & 0 & -z_i u_i \\ 0 & 0 & 0 & y_i & z_i & z_i v_i \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{22} & K_{23} & K_{33} \end{bmatrix}^T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Therefore, we get linear system of the form $Ax = 0$, where

$$A = \begin{bmatrix} x_1 & y_1 & z_1 & 0 & 0 & -z_1 u_1 \\ 0 & 0 & 0 & y_1 & z_1 & z_1 v_1 \\ x_2 & y_2 & z_2 & 0 & 0 & -z_2 u_2 \\ 0 & 0 & 0 & y_2 & z_2 & z_2 v_2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_n & y_n & z_n & 0 & 0 & -z_n u_n \\ 0 & 0 & 0 & y_n & z_n & z_n v_n \end{bmatrix} \quad , \quad x = \begin{bmatrix} K_{11} \\ K_{12} \\ K_{13} \\ K_{22} \\ K_{23} \\ K_{33} \end{bmatrix}$$

(c) We want to minimize $||Ax||$ subject to $|x| = 1$. So, we apply SVD on A to get the eigenvector corresponding to the smallest eigenvalue. We divide each element in the eigenvector by the element corresponding to K_{33}, so that $K_{33} = 1$. The result is a vector of the form $(K_{11}, K_{12}, K_{13}, K_{22}, K_{23}, 1)$.

■ ■ ■