Indian Institute of Technology Jodhpur
CSL7360: Computer Vision, Major Exam
Date: May 10, 2024, Max Marks: 60 Max Time: 120 minutes

1. (10 points) A camera is rigidly mounted so that it views a planar table top. A projector is also rigidly mounted above the table and projects a narrow beam of light onto the table, which is visible as a point in the image of the table top. The height of the table top is precisely controllable but otherwise the positions of the camera, projector, and table are unknown. For each of the following table top heights, the point of light on the table is detected at the following image pixel coordinates:

Table Height	Image coordinates of beam of light
50mm	(100,250)
100mm	(140,340)

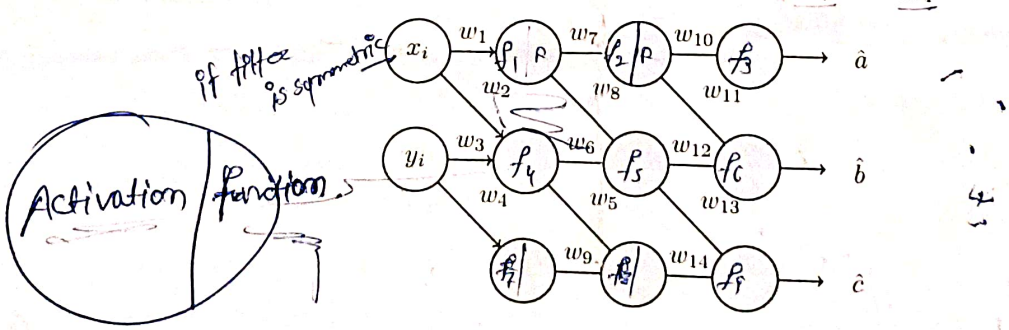
   (a) Using a projective camera model (described at the end of the question) specialized for this particular scenario, write a general formula that describes the relationship between world coordinates (x), specifying the height of the table top, and image coordinates (u, v), specifying the pixel coordinates where the point of light is detected. Give your answer using homogeneous coordinates and a projection matrix containing variables.

   (b) Once the camera is calibrated, given a new unknown height of the table and an associated image, can the height of the table be uniquely solved for? If so, give the equation(s) that is/are used. If not, describe briefly why not.

A generalized projective camera model is described as follows. Let $\mathbf{X} \in \mathbb{R}^n$ be a scene point in and $\mathbf{x} \in \mathbb{R}^m$ be the projected point in the image plane. Then, the projective camera model in the homogeneous coordinate system is described by the equation $\begin{bmatrix} \lambda \mathbf{x} \\ \lambda \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix}$. Here, λ is the depth of the scene points from the camera center and $\mathbf{M} \in \mathbb{R}^{(m+1)\times(n+1)}$ represents the projective camera matrix.

2. (10 points) Consider a set if 10 images of a static scene captured by an orthographic camera from 10 different viewpoints. Consider 20 3D points $\{\mathbf{P}_i\}_{i=1}^{20}$ in the scene are being projected in the 10 images. Let $(r_{a,b}, s_{a,b})$ be the projection of the b-th 3D point $\mathbf{P}_b$ in the a-th image. Assume that the $\{(r_{a,b}, s_{a,b})\}$ are given to you. where, $b = 1, \ldots 20$, and $a = 1, \ldots, 10$. Now, design an algorithm to find the pose of the each camera and the locations of all the 3D points $\{\mathbf{P}_i\}_{i=1}^{20}$ with respect to a world coordinate system whose origin is centered at the 3D point $2\mathbf{P}_3$.

3. (10 points) Consider a set of keypoints $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ detected in an image around a linear edge represented as a line $ax + by + c = 0$. Now in order to find the optimal parameters (a, b, c) of this line, a student has designed a neural network shown as below. Each neuron in the first and the second layers has sigmoid as the activation layer and the last layer neurons do not have any activation function. Design a loss function (call it ℓ) to train this neural network. Also, find $\frac{d\ell}{dw_7}$ and $\frac{d\ell}{dw_4}$.



4. Answer the following questions.

(a) (3 points) What is the condition in the spectral clustering of n pixels for which,

$$\text{Trace}(\mathbf{H}^T\mathbf{H}\mathbf{L}) = \text{RatioCut}(C_1, \ldots, C_k)$$

and prove it mathematically.

(b) (7 points) Perform only a single maximization (not expectation) step and find the updated cluster parameters, given the following status in a GMM clustering: number of clusters $k = 2$, the points being clustered are $p_1, p_2, p_3 = \{(1,2),(4,2),(5,1)\}$, current cluster centers are $c_1, c_2 = \{(2,2),(6,2)\}$ and

	p_1	p_2	p_3
z_{i1}	0.8	0.8	0.4
z_{i2}	0.2	0.2	0.6

, where z_{nk} denotes the probability of the n^{th} point belonging to the k^{th} cluster.

5. Assume there is a convolutional neural network containing 3 convolutional layers c_1, c_2, c_3 containing 1 filter each of size 3×3, 3×3, and 3×3, respectively, with stride 1 and no padding. The output o_1 produced by c_1 is fed to c_2 and the output o_2 produced by c_2 is fed to c_3 which produces the final output o_3.

(a) (2 points) If the input is an image of width 7 pixels and height 7 pixels, what is the final output size in pixels

(b) (2 points) Take any point in the intermediate output o_2 and find the size (width and height) of the exact region in the original 7×7 image that influences the value of that point.

(c) (2 points) What is the total number of learnable weights and biases in this convolutional neural network?

(d) (2 points) Design a fully connected layer to take the original 7×7 image as input and produce an output o_{fc} of the same size as o_3. Fully connected layer does not directly support a 2 dimensional input. So you will have to modify the input for passing it to a fully connected layer. How many neurons does this layer have?

(e) (2 points) What is the number of learnable weights and biases in the fully connected layer designed in part (d)?

6. Answer the following questions

(a) (2 points) Create a separable 2D square filter that can convert an integral image into a normal image through the correlation operation.

(b) (2 points) What are the constituent 1D filters for the above filter?

(c) (3 points) Apply this filter to the following integral image $I_{\text{integral}} = \begin{bmatrix} 1 & 6 & 13 & 16 \\ 3 & 13 & 24 & 29 \\ 6 & 22 & 35 & 41 \end{bmatrix}$ and find the original image which has the same size. Also, mention the amount of padding, if needed.

(d) (3 points) Arrange 3 Gaussian filters with standard deviations σ, 2σ, and 3σ in the decreasing order of the range of high-frequency components that get filtered out by them. Explain why?

CSL7360: Computer Vision

Quiz 3 (Questions 1 and 3) and Quiz 4 (Questions 2 and 4)

Date: April 22, 2024, Max Marks: 20+20, Max Time: 60 minutes

1. (10 points) Consider a gray-scale image I of size 2×2 where the pixel values are given as $I = \begin{array}{|c|c|}\hline x_1 & x_2 \\ \hline x_3 & x_4 \\ \hline \end{array} = \begin{array}{|c|c|}\hline 0.9 & 0.9 \\ \hline 0.9 & 0.1 \\ \hline \end{array}$. Construct a graph representing this image where each pixel denotes a vertex and weight of an edge between two vertices is determined by the similarity between the corresponding pixels and defined as $w_{ij} = \begin{cases} 1 - |x_i - x_j| & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$. Here, x_i and x_j are the intensities of the i-th and the j-th pixels, respectively. Find the Laplacian matrix of the resulting graph. Find the optimal clustering for two cluster case using the ratio-cut approach. One of the cluster should have three vertices and other cluster should have only one vertex.

2. (10 points) Consider the problem of clustering the points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ into k clusters $C_1, \ldots, C_k$ using the k-means clustering algorithm. In k-means clustering algorithm, we minimize the below cost function to find the optimal cluster centers.

$$\sum_{i=1}^{k} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \mu_i\|_2^2$$

Show that for the optimal cluster centers would be

$$\mu_i = \frac{1}{|C_i|} \sum_{\mathbf{x} \in C_i} \mathbf{x}.$$

Here, $|C_i|$ represents the number of data points in cluster C_i.

3. (10 points) On performing SVD on an observation matrix you get the following U and V matrices and the diagonal matrix obtained in SVD contains the singular values: 1,2,3,4.

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}, V = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

a) Find the observation matrix ($\mathbf{W}$) that satisfies the properties of the observation matrix of the orthographic structure from motion problem.
b) How many points are being tracked, and how many camera frames are there?

4. (10 points) Note that for the following questions, $a \times b - c$ constitutes 2 multiplication/addition/subtraction operations. Do not make any additional assumptions
a) Count the exact number of multiplication/addition/subtraction operations in computing an integral image from an image of size 10×10 using the raster scan method.
b) Given a Haar filter of size 4×10, consisting of 2 rectangles (pos & neg) of size 2×10 and 2×10, respectively, count the exact number of multiplication/addition/subtraction operations needed to apply it to the original image of size 10×10.
c) Given a Haar filter of size 4×10, consisting of 2 rectangles (pos & neg) of size 2×10 and 2×10, respectively, count the exact number of multiplication/addition/subtraction operations needed to apply it to the integral image computed above.

Indian Institute of Technology Jodhpur

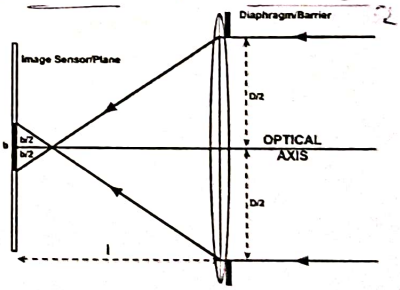
CSL7360: Computer Vision, Minor 2 Exam

Date: March 22, 2024, Max Marks: 30 Max Time: 60 minutes

1. Consider an image of size 1×3. Assume that each pixel can move only in the x-direction. Therefore, the optical flow at each point has only one component (u). The other component is zero (v) as there is no motion in vertical direction. Assume that that image gradients (I_x) and temporal gradients (I_t) at each pixel are given to you. Define the cost function for the Horn and Schunck algorithm for optical flow estimation for this image. Find the differentiation of the cost function with respect to the optical flow variables. Show that the vector $\mathsf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ containing the optimal flow at all three pixels satisfies a system of linear equations $\mathsf{Au} = \mathsf{b}$ for some matrix $\mathsf{A} \in \mathbb{R}^{3\times3}$ and a vector matrix $\mathsf{b} \in \mathbb{R}^{3\times1}$. Find the matrix A and the vector b. [8 Marks]

2. Consider a stereo camera with the given baseline $b = 5$ where both the cameras having the same intrinsic matrices $\mathsf{K}_\ell = \mathsf{K}_r = \mathsf{K} = \begin{bmatrix} \alpha & 0 & 200 \\ 0 & \alpha & 200 \\ 0 & 0 & 1 \end{bmatrix}$. Assume that the world coordinate system is aligned with the left camera coordinate system and the right camera coordinate system is a simple translation of the left camera system along x-axis by b units. Consider a pixel $\mathsf{p}_\ell = \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}$ in the left camera and its corresponding pixel $\mathsf{p}_r = \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix}$ in the right camera. Prove either $\mathsf{p}_r^\top \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\frac{5}{\alpha} \\ 0 & \frac{5}{\alpha} & 0 \end{bmatrix} \mathsf{p}_\ell = 0$. or $\mathsf{p}_\ell^\top \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\frac{5}{\alpha} \\ 0 & \frac{5}{\alpha} & 0 \end{bmatrix} \mathsf{p}_r = 0$. [7 Marks]

3. Consider a lens-based camera shown in the Figure, with aperture size D, $f-$number as N, and focal length of the lens as f. The image sensor is placed at a distance i behind the lens perpendicular to the optical axis but parallel to the lens, such that i>f. Assume that all the light rays from a scene point P reach the lens parallel to the optical axis. The optical axis passes through the center of the lens, and the lens is perpendicular to the optical axis. We observe that the lens is not able to focus these light rays from P onto a single point on the image plane leading to a blur circle of diameter b on the image sensor. Assume a thin lens.



Ques: Derive the equation for the blur circle diameter b **for this setup** showing all the steps. However, **the final expression for b should only contain i, f, and N.** [7 Marks]

4. Without using any matrix inverses, find the intrinsic matrix K and the values a, b, c in the following projection matrix P for a pin-hole camera, given the following rotation matrix R and translation vector t obtained while calibrating the camera. [8 Marks]

$$\mathsf{P} = \begin{bmatrix} 1 & -1 & 0 & a \\ 1 & 1 & 0 & b \\ 0 & 0 & 1 & c \end{bmatrix}, \mathsf{R} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathsf{t} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}.$$