CS455 Computer Vision: Practice Final Examination

University of Washington
2025

Solution:

SOLUTIONS

Name: ____________________

Student ID: ____________________

Question	Total	**Points**
True/False	11	
Multiple Choice	24	
Short answers 1: Camera Projection	16	
Short answers 2: Seam Carving	8	
Short answers 3: Dimensionality reduction	15	
Short answers 4: Videos	11	
Short answers 5: Segmentation	15	
Extra credits in short answers	10	
Total	100	

Instructions:

1. This examination contains 17 pages, including this page. You have **one hour fifty minutes** to complete the examination. As a courtesy to your classmates, we ask that you not leave during the last fifteen minutes.
2. For **True/False** and **Multiple-Choice** sections, correct answers will get the points listed next to the question. You will receive 0 for all unanswered questions and -0.5 for all incorrect answer. You can choose to not answer questions that you are unsure of to avoid receiving a negative point.
3. For **Short Answers**, you will NOT get negative points for incorrect answers. Partial credit will be assigned for showing your work and reasoning.
4. You may use one double-sided 8.5" $\times$ 11" **hand-written** sheet with notes that you have prepared. You may not use any other resources, including lecture notes, books, other students or other engineers. These notes must be submitted along with your booklet.
5. Please sign the below Honor Code statement.

In recognition of and in the spirit of the University of Washington Honor Code, I certify that I will neither give nor receive unpermitted aid on this examination.

Signature: ____________________

True/False (11 points)

Fill in the circle next to True or False, or fill in neither. Fill it in completely.
No explanations are required.

1. Using the forward seam algorithm allows us to determine the seam that minimizes the amount of energy inserted to the image.

   √ **True**
   ◯ False

2. Corners generally have high gradients in a single direction.

   ◯ True
   √ **False**

3. Harris corners are scale-invariant by default

   ◯ True
   √ **False**

4. The output size of an image after convolution is always the same as the input image.

   ◯ True
   √ **False**

5. A projective transformation always preserves straight lines but may change the relative distances between points on those lines.

   √ **True**
   ◯ False

6. Principal Component Analysis (PCA) can be used for image compression by reducing the dimensionality of image data.

   √ **True**
   ◯ False

7. When using Hough Transform to detect **lines**, we voted within **2D** cells to find the most likely lines in an image. Recall that for detecting **lines** we represented a line as $x \cdot cos(\theta) + y \cdot sin(\theta) = \rho$. If we were to use Hough Transform to detect **circles**, we would also vote in 2D cells.

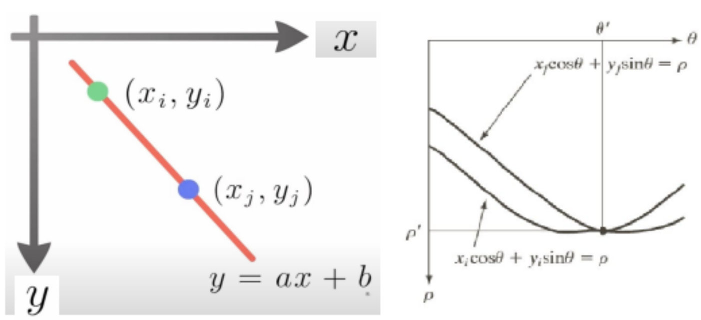


Figure 1: Hough Transform on Lines

   ◯ True
   √ **False**

8. The optical flow algorithm works better in videos with lower frame rates.

   ◯ True

   √ **False**

9. When we use linear classifiers, we will get the probability of an image input belonging to a class.

   √ **True**

   ◯ False

10. Increasing the number of clusters in k-means generally decreases the sum of squared distances within clusters.

    √ **True**

    ◯ False

11. The centroid of a cluster in k-means is always one of the data points in the dataset.

    ◯ True

    √ **False**

Multiple Choice (11points)

1. **SIFT (2 points).** SIFT descriptors are always invariant to which of the following? (*Choose all that apply*):

   √ **Translation**

   ☐ Affine transformations

   √ **Rotation**

   ☐ Occlusion

   √ **Scale**

2. **Cameras (2 points).** What is the role of the intrinsic matrix in a camera model? (*Choose the correct answer*):

   ◯ It maps 2D image coordinates to 3D world coordinates

   ◯ It describes the camera's position and orientation in the world.

   √ **It maps 3D camera coordinates to 2D image coordinates.**

   ◯ It provides the camera's field of view and resolution.

3. **Cameras (2 points).** Given a camera with intrinsic matrix K (3×3) and extrinsic parameters $[R \mid t]$ where R is 3×3 and t is 3×1, how would you obtain the 2D image coordinates P_{2D} of a 3D world point P_w (3×1)? (Hint: Verify that the matrix dimensions are consistent.)

   (*Choose the correct answer*):

   ◯ $P_{2D} = K \cdot P_w$

   ◯ $P_{2D} = R \cdot P_w + t$

   √ $P_{2D} = K \cdot (R \cdot P_w + t)$

   ◯ $P_{2D} = (R \cdot P_w + t) \cdot K$

4. **SIFT (2 points).** In the SIFT descriptor, what is the size of the descriptor vector generated for each keypoint? (*Choose the correct answer*):

   ◯ 32

   ◯ 64

   √ **128**

   ◯ 256

   Explanation: The descriptor is computed using 16 histograms with 8 orientation bins each, resulting in $16 \times 8 = 128$ values.

5. **SIFT (2 points).** If two keypoints have similar orientation histograms but one has significantly higher gradient magnitudes, how will this affect their SIFT descriptors? (*Choose the correct answer*):

   ◯ The keypoint with higher gradient magnitudes will dominate the descriptor comparison, making it more likely to match.

   ◯ The descriptors will be identical since orientation, not magnitude, defines the descriptor.

   √ **The descriptors will remain comparable because SIFT normalizes the gradient magnitudes.**

   ◯ The keypoint with lower gradient magnitudes will be discarded during descriptor computation.

6. **Convolutions (2 points).** Which of the following affects the output size of convolution operations? (*Choose all that apply*):

   √ **Kernel Size**

   √ **Padding**

   √ **Stride**

   ☐ Activation function

7. **Filters (2 points).** A moving average filter is an example of (*Choose the correct answer*):

   ◯ A system that amplifies high-frequency components

   √ **A linear, shift-invariant system used for smoothing**

   ◯ A filter that preserves sharp edges in an image

   ◯ A non-causal system that depends on future inputs

8. **Deep Learning (2 points).** Which of the following statements about Convolutional Neural Networks (CNNs) is TRUE? (*Choose the correct answer*):

   ◯ Fully connected layers are applied to entire images without any spatial structure.

   ◯ Pooling layers increase the spatial resolution of feature maps.

   √ **CNNs use weight sharing to reduce the number of parameters.**

   ◯ CNNs do not use non-linearity functions such as ReLU.

9. **Edges (2 points).** What is the primary goal of non-maximum suppression in edge detection algorithms like Canny? (*Choose the correct answer*):

   ◯ To blur the edges in an image and avoid thin edges.

   √ **To produce precise edges by removing weaker edge pixels.**

   ◯ To highlight edges belonging to object boundaries.

   ◯ To noisy edges from the corners of the image.

10. **Panoramas (2 points).** In image stitching (panorama creation), which method is commonly used to remove outliers while estimating the affine transformation matrix? (*Choose the correct answer*):
    - ◯ Principal Component Analysis (PCA).
    - ◯ Singular Value Decomposition (SVD).
    - √ **Random Sample Consensus (RANSAC).**
    - ◯ K-Means Clustering.

11. **Optical Flow (2 points).** Which assumption is fundamental to most optical flow algorithms? (*Choose the correct answer*):
    - ◯ Scale invariance
    - √ **Brightness constancy**
    - ◯ Rotation invariance
    - ◯ Temporal discontinuity

12. **Detection (2 points).** Which of the following statements about Intersection over Union (IoU) are true? (*Choose all that apply*):
    - ☐ A higher IoU threshold will result in a higher true positive rate.
    - √ **IoU is used to determine if a detected object matches the ground truth.**
    - ☐ A higher IoU threshold always improves object detection performance.
    - √ **The IoU threshold is a hyperparameter that affects the precision-recall trade-off.**

Short Answers 1: Camera Projection and Transformations (16points)

1. **(16 points)** A calibrated camera with the intrinsic matrix:

$$K = \begin{bmatrix} 1000 & 0 & 320 \\ 0 & 1000 & 240 \\ 0 & 0 & 1 \end{bmatrix}$$

is positioned at $(3, 2, 10)$ in world coordinates with its principal axis aligned with the Z-axis.

(a) **(4 points)** A 3D point $(X_w, Y_w, Z_w) = (6, 4, 15)$ is observed. Compute its camera coordinates (X_c, Y_c, Z_c).

Solution: $X_c = X_w - C = (6, 5, 15) - (3, 2, 10) = (3, 2, 5)$

(b) **(4 points)** Using the intrinsic matrix K, compute the image coordinates (u, v) of the projected 2D point corresponding to (X_c, Y_c, Z_c).

Solution: $K \cdot X_c = \begin{bmatrix} 100 & 0 & 320 \\ 0 & 1000 & 240 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 4600 \\ 3200 \\ 5 \end{bmatrix}$

$$u = \frac{4600}{5} = 920$$

$$v = \frac{3200}{5} = 640$$

(c) **(4 points)** What effect does increasing the focal length f by a factor of 2 have on the image projection? Explain mathematically.

Solution: Recall from part b how we obtained u and v.

$$u = f \cdot \frac{x_c}{z_c} + c_x \rightarrow u - c_x = f \cdot \frac{x_c}{z_c}$$

$$v = f \cdot \frac{y_c}{z_c} + c_y \rightarrow v - c_y = f \cdot \frac{y_c}{z_c}$$

Now if we doubled the focal length, we would get that

$$u' - c_x = \frac{2fx_c}{z_c}$$

$$v' - c_y = \frac{2fy_c}{z_c}$$

We can see from this that the distance of our projected point (u', v') from (c_x, c_y) has doubled, effectively zooming into the image.

(d) **(4 points)** If the camera is tilted downward by 30 degrees along the X-axis, derive the new rotation matrix and compute the transformed camera coordinates.

Solution:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

We're given that the camera is tilted downwards along the X-axis. Thus this rotation along the X-axis is $\theta = -30$ so we plug it in to get:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

We can now use this to compute the transformed camera coordinates $X_c = R_x \cdot \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ \sqrt{3} + 2.5 \\ -1 + \frac{5\sqrt{3}}{2} \end{bmatrix}$

Short Answers 2: Seam Carving (8points)

2. **(8 points)**

(a) **(4 points)** What is the cost matrix equation for cell $M(i,j)$, assuming we're doing seam carving on vertical seams and $i \neq 0$, in the following cases. You can denote the energy function as $E(i,j)$.

(i) **(2 points)** Seam carving:

Solution: $$M(i,j) = E(i,j) + min \begin{cases} M(i-1,j-1) \\ M(i-1,j) \\ M(i-1,j+1) \end{cases}$$

(ii) **(2 points)** Forward seam carving:

Solution: $$M(i,j) = E(i,j) + min \begin{cases} M(i-1,j-1) + C_{left} \\ M(i-1,j) + C_{middle} \\ M(i-1,j+1) + C_{right} \end{cases}$$
where $C_{left} = |I(i,j+1) - I(i,j-1) + |I(i-1,j) - I(i,j-1)|$
$C_{middle} = |I(i,j+1) - I(i,j-1)|$
$C_{right} = |I(i,j+1) - I(i,j-1)| + |I(i-1,j) - I(i,j+1)|$

(b) **(4 points)** Let's say we have a picture, but it has an unwanted object in the background. Describe the steps you would take to remove the object from the background, but still preserve the size of the image, using seam carving.

Solution: We want to compute the energy function $E(j, i)$ while manually setting very negative values for the pixels in the unwanted object. Perform seam carving or forward seam carving to iteratively remove seams. Once the unwanted object is removed, we can restore the width of the image by inserting low energy seams.

Short Answers 3: PCA and LDA (15points + 5 extra credits.)

3. **PCA (8 Points)**

(a) **(3 points.)** Explain in words what projecting points onto the first principal component aims to do.

Solution: Projecting points onto the first principal component aims to lower the dimensionality of the points to 1 while maximizing the variance of the points in the resulting projection

(b) **(5 points.)** Suppose your image is represented by a 2D point $X = (5, 2)$ and you've computed the first principal component of your dataset is the vector (2, 2). Project the point in X onto the first principal component. The formula for projecting a point p onto a vector v is $\frac{p \cdot v}{||v||^2} v$

Solution: First the dot product is 14 and the norm squared of v is 8. Dividing gives $\frac{7}{4}$. Multiply that by v and we get $[\frac{7}{2}, \frac{7}{2}]$

4. **LDA (7 Points)**

(a) **(3 points.)** Explain the main difference between PCA and LDA in terms of their objectives.

Solution: The key difference is supervised vs unsupervised decomposition. PCA focuses on the variance withing the data itself, irrespective of any class labels. LDA focuses on maximizing the sepration between multiple known classes.

(b) **(4 points.)** Which is generally better when dealing with a classification problems in computer vision?

Solution: LDA. The main advantage of using Linear Discriminant Analysis (LDA) over Principal Component Analysis (PCA) when dealing with classification problems in computer vision—or indeed any classification problem—stems from LDA's focus on maximizing class separability. LDA ensures that classes are better separated on the most relevant features for classification.

(c) **(Extra credit 5 points)** Suppose you have a set of 4 images, where each is represented by two values $X_t \in \mathbb{R}^2$

$$X = \begin{bmatrix} 5 & 3 & 6 & 2 \\ 2 & 4 & 5 & 1 \end{bmatrix}$$

Compute the unit length principal components, and say which one would be used if we are using the first principal component. Hint: Find the eigenvalues/eigenvectors of the covariance matrix $C = (X - \mu)(X - \mu)^T$.

Solution: First we calculate the mean $\mu = [4, 3]$. Then we center X giving $\begin{bmatrix} 1 & -1 & 2 & -2 \\ -1 & 1 & 2 & -2 \end{bmatrix}$

Then $X^T X = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$

It's eigenvectors are $[1/\sqrt{2}, 1/\sqrt{2}]$ with eigenvalue 16 and $[1/\sqrt{2}, -1/\sqrt{2}]$ with eigenvalue 4. The first eigenvector is chosen as the eigenvalue is the largest.

Short Answers 4: Videos (11points).

5. **Optical Flow (11 points)** When calculating optical flow, we made the assumption that the brightness of two consecutive frames does not differ by much. This brightness constancy equation is written as:

$$I_x u + I_y v + I_t = 0$$

(a) **(5 points.)** Explain briefly what each of the five variables in the brightness constancy equation are.

> **Solution:** $\frac{dI}{dx}$ is the x component of the image gradient.
> $\frac{dI}{dy}$ is the y component of the image gradient.
> $u = \frac{dx}{dt}$ is the x component of the motion field.
> $v = \frac{dy}{dt}$ is the y component of the motion field.
> $\frac{dI}{dt}$ is the change in image brightness over time.

(b) **(3 points.)** According to this equation, explain what is the direction in the image along which optical flow cannot be reliably estimated?

> **Solution:** The optical flow cannot be reliable estimated along the edges (perpendicular to the gradient direction). This is because the brightness constancy constraint involves the dot product between the image gradient ∇I and the optical flow u, $\nabla I \cdot u + It = 0$, which gives no equation on u when u and ∇I are orthogonal.

(c) **(3 points.)** List three key assumptions when estimating optical flow.

Solution: Brightness constancy; spatial coherence, small movement

Short Answers 5: Segmentation (15points + 5 extra credits.)

6. **(7 points.)** You are given a grayscale image of size 256×256 pixels. Your feature space is the grayscale values without any position information. The pixel intensities range from 0 to 255. You are tasked with segmenting this image using the K-means clustering algorithm.

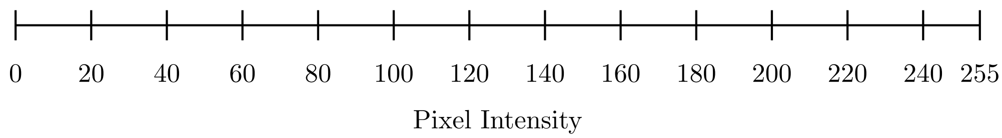
(a) **(2 points)** How does k-means algorithm work in the context of image segmentation? (2 sentences max)

> **Solution:** In image segmentation, K-means clusters pixel intensities (or features), grouping similar intensity values (or features) together to form distinct regions in the image.

(b) **(3 points)** Suppose you are using k-means with $K = 3$. Describe the steps you would take to segment the image. List out the steps of the k-means algorithm.

> **Solution:** 1) Initialize 3 random cluster centroids from the pixel internsities (features). 2) Assign each pixel to the nearest cluster centroid based on its intensity value (feature vector). 3) Update the cluster centroids by computing the mean intensity value of the pixels in each cluster. 4) Repeat steps 2 and 3 until the centroids do not change or reach maximum number of iterations.

(c) **(2 points)** After running K-means clustering on the image, you obtained three clusters with centroids at 40, 120, and 200. How would you assign the pixel values in the segmented image based on these centroids? Show the range of numbers within each cluster on the number line below. Draw two vertical lines to separate the number line into 3 clusters.



Solution: 40: 0 - 80, 120: 80 - 160, 200: 160 - 255

7. **(8 points.)** Assume $K = 3$ and and the initial centroids are 40, 120, and 200. Given the following five pixel intensities: 30, 58, 100, 180, and 220, perform the assignment and update steps of the K-means algorithm for two iterations and show the results step by step.

> **Solution:** First Iteration:
>
> Assignment: cluster 1: 30, 58 cluster 2: 100 cluster 3: 180, 220
>
> Update: New centroid for Cluster 1: $(30 + 58)/2 = 44$ New centroid for Cluster 2: 100 New centroid for Cluster 3: $(180 + 220)/2 = 200$
>
> Second Iteration:
>
> Assignment: cluster 1: 30, 58 cluster 2: 100 cluster 3: 180, 220
>
> Update: New centroid for Cluster 1: $(30 + 58)/2 = 44$ New centroid for Cluster 2: 100 New centroid for Cluster 3: $(180 + 220)/2 = 200$

8. **(Extra Credit: 5 points.)** The LAB color space is a color-opponent space with three axes: L for lightness, a for the green-red component, and b for the blue-yellow component. Unlike RGB, which is based on the primary colors of light, LAB is designed to approximate human vision and is more perceptually uniform. This means that the same amount of numerical change in these values corresponds to about the same amount of visually perceived change.
Compare the RGB and LAB color spaces in the context of image segmentation using K-means clustering. Which color space would you recommend and why? (4 sentence max)

Solution: RGB directly represents colors using red, green, and blue channels, which is simple and intuitive but may not accurately reflect perceptual differences between colors. LAB separates luminance from color information, making it more perceptually uniform and better suited for image segmentation. LAB is recommended because it aligns more closely with human vision, providing better segmentation results by separating intensity from color information.