CS 231A Computer Vision Midterm

Tuesday October 30, 2012

Solution Set

1 Multiple Choice (20 points)

Each question is worth **2 points**. To discourage random guessing, **1 point will be deducted** for a wrong answer on multiple choice questions! **For answers with multiple answers, 2 points will only be awarded if all correct choices are selected, otherwise, it is wrong and will incur a 1 point penalty**. Please draw a circle around the option(s) to indicate your answer. No credit will be awarded for unclear/ambiguous answers.

1. (Circle all that apply) If all of our data points are in $\mathbb{R}^2$, which of the following clustering algorithms can handle clusters of arbitrary shape?

   (a) k-means

   (b) k-means++

   (c) EM with a gaussian mixture model

   (d) mean-shift

   Solution Only d

2. (Pick one) Suppose we are using a Hough transform to do line fitting, but we notice that our system is detecting two lines where there is actually one in some example image. Which of the following is most likely to alleviate this problem?

   (a) Increase the size of the bins in the Hough transform

   (b) Decrease the size of the bins in the Hough transform

   (c) Sharpen the image

   (d) Make the image larger

   Solution a

3. (Pick one) Which of the following processes would help avoid aliasing while down sampling an image?

(a) Image sharpening
(b) Image blurring
(c) Median filtering where you replace every pixel by the median of pixels in a window around the pixel
(d) Histogram equalization

Solution b

4. (Circle all that apply) A Sobel filter can be written as

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \tag{1}$$

Which of the following statements are true

(a) Separating the filter, reduces the number of computations by 30%
(b) It is similar to applying a gaussian filter followed by a derivative.
(c) Separation leads to spurious edge artifacts
(d) Separation approximates the first derivative of gaussian.

Solution a and b

5. (Pick one) Which of the following are true for Eigenfaces (PCA)

(a) Can be used to effectively detect deformable objects
(b) Invariant to Affine Transforms
(c) Can be used for lossy image compression
(d) Is invariant to shadows

Solution c

6. (Pick one) Downsampling can lead to aliasing because

(a) Sampling leads to additions of low frequency noise
(b) Sampled high frequency components result in apparent low frequency components.
(c) Sampling increases the frequency components in an image.
(d) Sampling leads to spurious high frequency noise

Solution b

7. (Pick one) If we replace one lens on a calibrated stereo rig with a bigger one, what can we say about the essential matrix, E, and the fundamental matrix, F?

(a) E can change due to the possible change in physical length of the lens. F is unchanged.

(b) F can change due to the possible change in lens characteristics. E is unchanged.

(c) E can change due to the possible change in lens characteristics. F is unchanged.

(d) Both are unchanged.

Solution b

8. (circle all that apply) Which of the following describes an affine camera but not a general perspective camera?

   (a) Relative sizes of visible features in a scene can be determined without prior knowledge

   (b) Can be used to determine the distance from a object of a known height

   (c) Approximates the human visual system

   (d) An infinitely long plane can be viewed as a line from the right angle.

   Solution a, d

The following questions refer to this picture in Figure 1

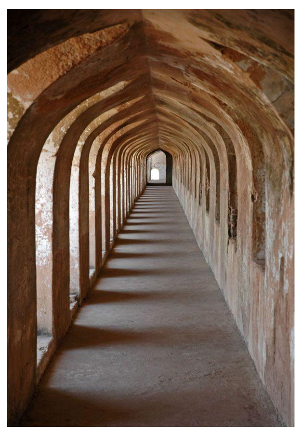


Figure 1: Long hallway

9. (circle all that apply) Assuming the hallway is very long and very straight, if an affine camera were directed down the hallway, which conditions must be met to see anything at the end of the hallway?

   (a) The camera is in the exact middle of the hallway.

   (b) The camera is on the floor

   (c) The camera's image plane must be very very large.

   (d) The vector representing the camera's viewing direction is exactly parallel to the floor and the walls.

   Solution D

10. (circle all that apply) Which of the following characteristics of the hallway shown can be determined using an affine camera pointed straight down the hallway?

(a) The color/texture of the walls
(b) The number of openings in the hallway
(c) If the hallway is getting narrower
(d) If the hallway is getting wider

Solution C

2 True or False (20 points)

True or false, briefly justify your answer in one sentence or less. Correct answers are 2 points, -1 point for each incorrect answer.

(a) (True/False) Fisherfaces works better at discrimination than Eigenfaces because Eigenfaces assumes that the faces are aligned.

Solution (False), both assume the faces are aligned

(b) (True/False) If you don't normalize your data to have zero mean, then the first principal component found via PCA is likely to be uninformative.

Solution (True)

(c) (True/False) Given sufficiently many weak classifiers, boosting is guaranteed to get perfect accuracy on the training set no matter what the training data looks like.

Solution (False), one could have two datapoints which are identical except for their label

(d) (True/False) Boosting always makes your algorithm generalize better.

Solution (False), you can overfit

(e) (True/False) It is possible to blur an image using a linear filter.

Solution (True)

(f) (True/False) When extracting the eigenvectors of the similarity matrix for an image to do clustering, the first eigenvector to use should be the one corresponding to the second largest eigenvalue, not the largest.

Solution (False), it's the largest one

(g) (True/False) The Canny edge detector is a linear filter.

Solution (False), It has non-linear operations, thresholding, hysteresis, non-maximal supression

(h) (True/False) A zero skew intrinsic matrix is not full rank because it has one less DOF.

Solution (True), It is still full rank, even if it has one less DOF.

(i) (True/False) A picture of an outdoor field will never have a horizon with an affine camera.

Solution (False), The field is only of finite length, and still on earth, which is approximately spherical, so it will curve "down" relative to the camera and occlude itself, which yields a "horizon".

(j) (True/False) Any two cameras with overlapping fields of view can be used in a useful stereo rig when using appearance-based features for correspondence.

Solution (False) If the cameras are facing each other, they can't see features on the surface, except on the edges which could be hard to extract since the backgrounds are different from the two cameras, even if the object is the same.

3 Long Answer (40 points)

11. **(10 points) Detecting Patterns with Filters** A Gabor filter is a linear filter that is used in image processing to detect patterns of various orientations and frequencies. A Gabor filter is composed of a Gaussian kernel function that has been modulated by a sinusoidal plane wave. The real value version of the filter is shown below.

$g\left(x, y; \lambda, \theta, \psi, \sigma, \gamma,\right) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right)$

Where

$x' = x\ \cos\left(\theta\right) + y\ \sin\left(\theta\right)$

$y' = -x\ \sin\left(\theta\right) + y\ \cos\left(\theta\right)$

Figure 2 shows an example of a 2D Gabor Filter.

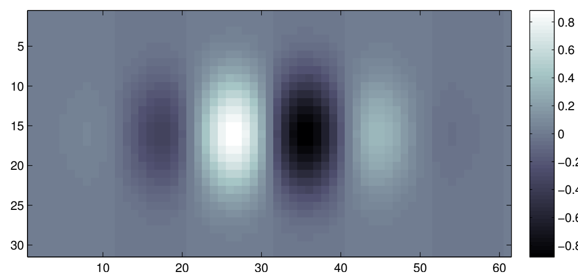


Figure 2: 2D Gabor Filter

(a) (6 point) What is the physical meaning of each of the five parameters of the Gabor filter, $\lambda, \theta, \phi, \sigma, \gamma$, and how do they affect the impulse response?

Hint: The impulse response of a gaussian filter is shown in Equation 2, it is normally radially symmetric, how would you make this filter elliptical? How would you make this filter steerable? What does the 2D cosine modulation do to this filter?

- λ:
- θ:
- ϕ:
- σ:
- γ:

$$gaussian\left(x, y\right) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

Solution

λ: represents the wavelength of the sinusoidal factor

θ: represents the orientation of the normal to the parallel stripes of a Gabor function

ϕ: phase offest

σ: Sigma of the gaussian envelope

γ: Spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function

(b) (4 point) Given a Gabor filter that has been tuned to maximally respond to the striped pattern in shown in Figure 3, how would these parameters, $\lambda_0, \theta_0, \phi_0, \sigma_0, \gamma_0$, have to be modified to recognize the following variations? Provide the values of the new parameters in terms of the original values.

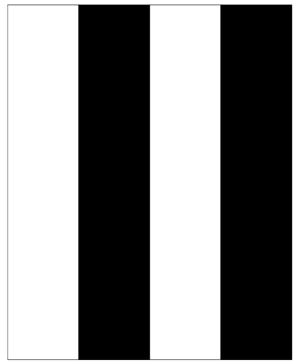
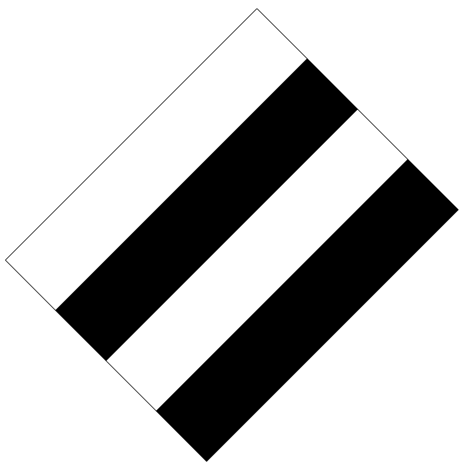
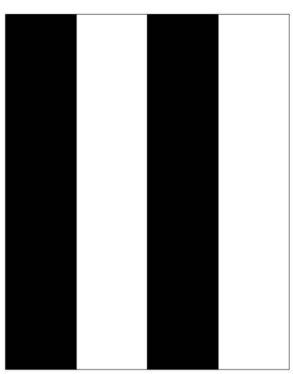


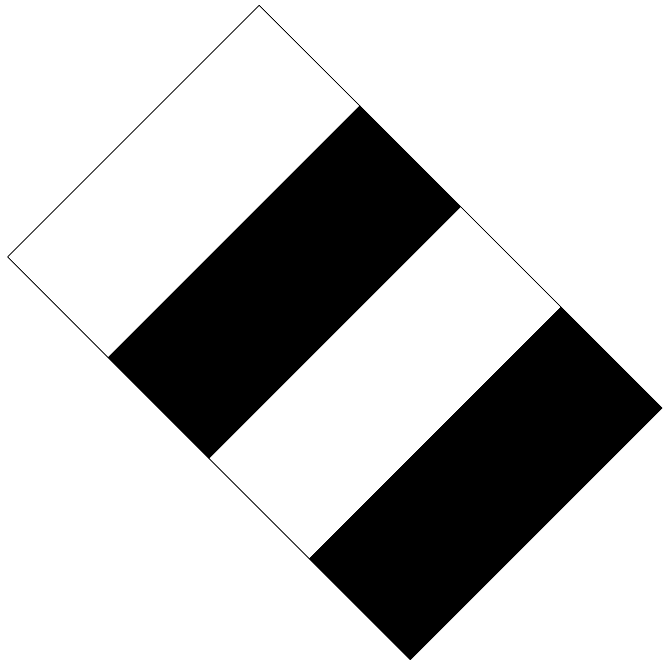
Figure 3: Reference Pattern

i. 

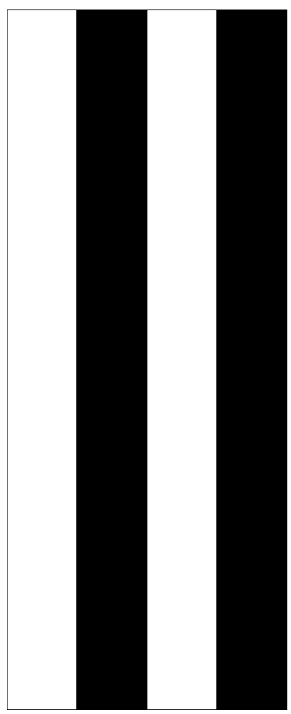
Solution
$\theta = \theta_0 + \frac{\pi}{4}$

ii. 

Solution
$\psi = \psi_0 + \frac{\pi}{2}$

iii. 

Solution
$\theta = \theta_0 + \frac{\pi}{4}, \sigma = 2\sigma_0, \lambda = 2\lambda_0$

iv. 

Solution
$\gamma = \frac{1}{2}\gamma_0$

12. **(10 points) Stereo Reconstruction**

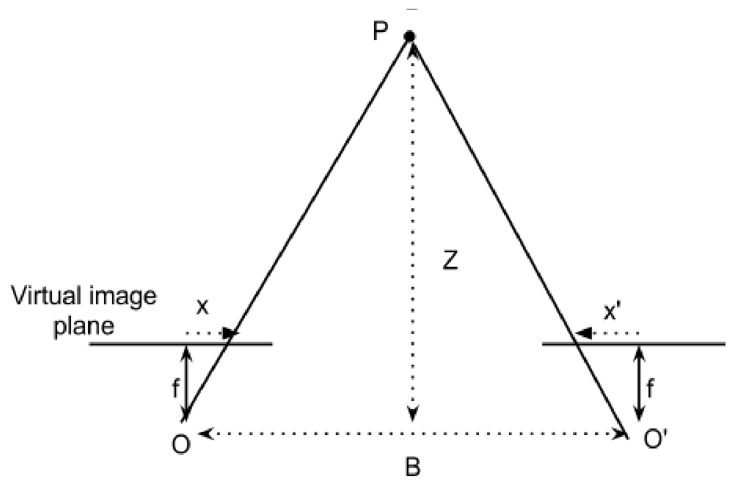


Figure 4: Rectified Stereo Rig

(a) (4 point)

The figure above shows a rectified stereo rig with camera centers O and O', focal length f and baseline B. x and x' are the projected point locations on the virtual image planes by the point P; note that since x' is to the left of O', it is negative. Give an expression for the depth of the point P, shown in the diagram as Z. Also give an expression for the X coordinate of the point P in world coordinates, assuming an origin of O. You can assume that the two are pinhole cameras.

Solution

$Z = \frac{fB}{x-x'}$

$X = \frac{xZ}{f}$

(b) (6 point)

Now assume that the camera system can't perfectly capture the projected points location on the image planes, so there is now some uncertainty about the point's location since a real digital camera's image plane is discretized. Assume that the original x and x' positions now have an uncertainty of $\pm e$, which is related to discretization of the image plane. Give an expression of the X, Z locations of the 4 intersection points resulting from the virtual image plane uncertainty. Give an expression for the maximum uncertainty in the X and Z directions of the point P's location in world coordinates. All expressions should be in terms of image coordinates only, you can assume that x is always positive and x' is always negative.

Solution

$Z_{min} = \frac{fB}{(x+e)-(x'-e)}$

$Z_{max} = \frac{fB}{(x-e)-(x'+e)}$

$d = x - x'$

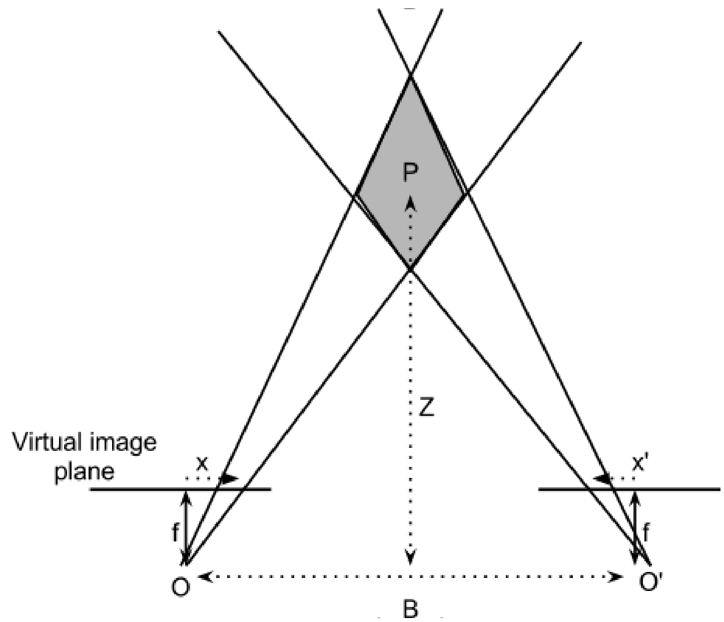


Figure 5: Rectified Stereo Rig with image plane error

$Z_{diff} = Z_{max} - Z_{min} = fB\frac{4e}{(d-2e)(d+2e)}$
$Z_{mid} = \frac{fB}{(x-e)-(x'-e)} = \frac{fB}{x-x'}$
$X_{min} = \frac{(x-e)Z_{mid}}{f}$
$X_{max} = \frac{(x+e)Z_{mid}}{f}$
$X_{max} - X_{min} = \frac{Z_{mid}(2e)}{f}$

item (? point)
Assuming the X coordinate of the point P is fixed, give an expression of of the uncertainty of the reconstruction of Z with in terms of the actual value of Z and the other parameters of the stereo rig. What is the depth uncertainty when Z is equal to zero? Find the depth when the uncertainty is at its maximum and give a physical interpretation and a drawing to explain.

Solution

$Z_{min} = \frac{fB}{(x+e)-(x'-e)}$
$Z_{max} = \frac{fB}{(x-e)-(x'+e)}$
$d = x - x'$
$Z_{diff} = Z_{max} - Z_{min} = fB\frac{4x_o}{(d-2e)(d+2e)}$
$Z_{mid} = \frac{fB}{(x-e)-(x'-e)} = \frac{fB}{x-x'}$
$X_{min} = \frac{(x-e)Z_{mid}}{f}$
$X_{max} = \frac{(x+e)Z_{mid}}{f}$
$X_{max} - X_{min} = \frac{Z_{mid}(2e)}{f}$

13. **(13 points) AdaBoost algorithm (greedy training)** Let $f_m(x)$ be the classifying function learnt after the m^{th} iteration.

$$f_M(x) = \sum_{m=1}^{M} \beta_m C_m(x) \tag{3}$$

Where, $C_m(x)|m \in \{1, \dots, M\}$ is a bunch of weak classifiers learned in M iterations. $C_m : \mathbb{R} \rightarrow \{-1, 1\}$. We will now look at a derivation for the optimal β, C at the m^{th} iteration given $\beta_i, C_i \forall i \in \{1, \dots, m-1\}$

(a) (1 point) $(\beta_m, C_m) = \underset{\beta, C}{\arg\min} \left(\sum_{i=1}^{N} L\left[y_i, f_{m-1}(x_i) + \beta C(x_i)\right] \right)$, where $L[y, g] = exp(-yg)$ is the loss function. Show that (β_m, C_m) can be written in the form

$$(\beta_m, C_m) = \underset{\beta, C}{\arg\min} \sum_{i=1}^{N} w_i^{(m-1)} exp\{-\beta y_i C(x_i)\} \tag{4}$$

and define $w_i^{(m-1)}$. Note that $w_i^{(m-1)}$ is the weight associated with the i^{th} data point after $m-1$ iterations.

Solution
$\beta_m, C_m = \underset{\beta, C}{\arg\min} \sum_{i=1}^{N} exp\{-y_i f_{m-1}(x_i) - \beta y_i C(x_i)\}$
$= \underset{\beta, C}{\arg\min} \sum_{i=1}^{N} w_i^{(m-1)} exp\{-\beta y_i C(x_i)\}$
$w_i^{(m-1)} = exp\{-y_i f_{m-1}(x_i)\}$

(b) (3 point) Express the optimal C_m in the form $argmin_C Err(C)$, where $Err(C)$ is an error function and is independent of β. $Err(C)$ should be defined in terms of the indicator function $I[C(x) \neq y_i]$ defined by
$I[C(x) \neq y_i] = \begin{cases} 1, & \textit{if } C(x) \neq y_i \\ 0 & \textit{if } C(x) = y_i \end{cases}$

Solution
$C_m = \underset{C}{\arg\min}\, e^{-\beta} \cdot \sum_{y_i = C(x_i)} w_i^{(m-1)} + e^{\beta} \cdot \sum_{y_i \neq C(x_i)} w_i^{(m-1)}$
$C_m = \underset{C}{\arg\min}\, (e^{\beta} - e^{-\beta}) \cdot \sum_{i=1}^{N} w_i^{(m-1)} I[y_i \neq C(x_i)] + e^{-\beta} \cdot \sum_{i=1}^{N} w_i^{(m-1)}$
$C_m = \underset{C}{\arg\min} \sum_{i=1}^{N} w_i^{(m-1)} I[y_i \neq C(x_i)]$

(c) (2 point) Using the C_m from part (a), the optimal expression for β_m can be obtained as

$\beta_m = \frac{1}{2} log \left(\frac{1-err_m}{err_m} \right)$
where $err_m = \frac{\sum_{i=1}^{N} w_i^{(m)} I[y_i \neq C_m(x_i)]}{\sum_{i=1}^{N} w_i^{(m)}}$.

Now, find the update equation for $w_i^{(m)}$ and show that $w_i^{(m)} \propto w_i^{(m-1)} \cdot exp\left(2\beta_m I[y \neq C_m(x_i)]\right)$

Solution
From part(a), we have
$w_i^{(m)} = w_i^{(m-1)} \cdot e^{-\beta_m y_i C_m(x_i)}$
Since, $y_i C_m(x_i) = 1 - 2I[y_i \neq C_m(x_i)]$
$w_i^{(m+1)} = w_i^{(m)} \cdot e^{2\beta_m I[y_i \neq C_m(x_i)]} \cdot e^{-\beta_m}$

(d) (4 point) We will use this algorithm to classify some simple faces. The set of training images is given in Fig. 6. x_i is the face and y_i is the corresponding face label, $\forall i \in \{1, 2, 3, 4, 5\}$.

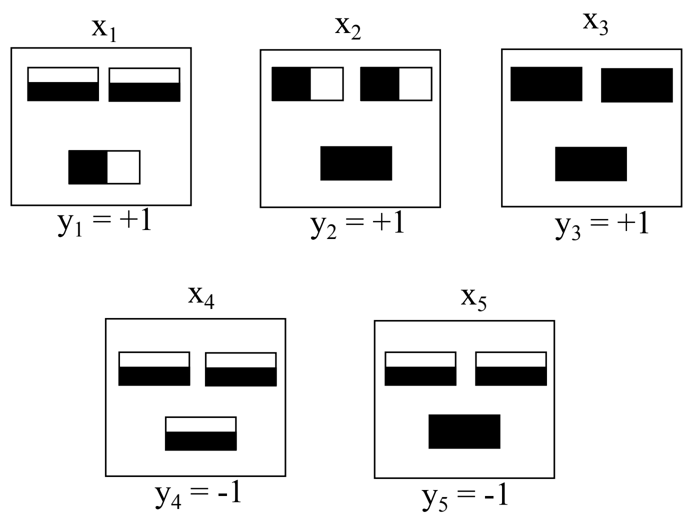


Figure 6: Training Set 'Faces'

We are also given 3 classifier patches p_1, p_2, p_3 in Fig. 7. A patch detector $I^{(+)}(x_i, p_j)$ is defined follows:

$$I^{(+)}(x_i, p_j) = \begin{cases} 1, & \text{if image } x_i \text{ contains patch } p_j \\ -1, & \text{otherwise} \end{cases}$$

$I^{(-)}(x_i, p_j) = -I^{(+)}(x_i, p_j)$

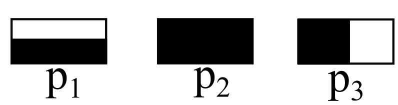


Figure 7: Classifier patches

All classifiers C_m are restricted to belong to one of the 6 patch detectors, i.e. $C_m(x) \in \{I^{(\pm)}(x, p_1), I^{(\pm)}(x, p_2), I^{(\pm)}(x, p_3)\}$.

If $C_1(x) = I^{(+)}(x, p_2)$, $w_i^{(0)} = 1$, $\forall i \in \{1, 2, 3, 4, 5\}$ and $\beta_1 = 1$,
what is the optimal $C_2(x)$? What are the updated weights $w_i^{(1)}$? What is the final classifier $f_2(x)$ combining C_1, C_2? Does $I[f_2(x) > 0]$ correctly classify all training faces?

Solution

$C_2(x) = I^{(+)}(x, p_3)$
$w_i^{(1)} \propto exp(2\beta_2)$ for $i = 1, 5$
$w_i^{(1)} \propto 1$ for $i = 2, 3, 4$
where, $\beta_2 = 0.5log(1.5)$
$f_2(x) = C_1(x) + 0.5log(1.5) \cdot C_2(x)$
Yes, it correctly classifies all the training images.

14. **(10 points) Implicit Shape Model (Application of Generalized Hough Transform for joint image classification and segmentation)**

You are given a 'dictionary' D of 1000 patches extracted from an image database. $M(e, D)$ is a retrieval operation which returns the closest dictionary patches from D for a query patch f. You are also provided training images with K descriptive image patches f_k extracted at different locations d_k in each image, i.e you are given $\{(f_k, d_k)\}_{k=1,\ldots,K}$.

(a) (2 point) Training: Let N_L be the total number of locations in an object measured from the object center. We wish to produce a $1000 \times N_L$ probability matrix $\mathbb{P}$, with $\mathbb{P}(D_i, L_j)$ recording the probability of finding a patch $D_i \in D$ at a location L_j with respect to the object center x. All training images are annotated with the object center. Write simple pseudocode to find $\mathbb{P}(I_i, L_j)$ from the training images in a manner similar to Hough Transform.

Solution

i. Repeat following steps for every patch (e, L) on a training image, with object center x
   Find the closest disctionary patched using $G = M(e, D)$
   Increment the bins in $\mathbb{P}$ corresponding to $(G_i, L - x)$ $\forall G_i \in G$
ii. Normalize $\mathbb{P}$ to obtain the probability matrix

(b) (3 point) Testing (Classification): Let $\{(e_k, d_k)\}_{k=1,\ldots,K}$ be the set of patches extracted at distinctive points from a Test image at locations d_k.

$$P(o, x|e_k, d_k) = \sum_i P(o, x|I_i, d_k).P(G_i|e_k) \tag{5}$$

is the probability of finding the object o at center x given the patch (e_k, d_k). $G_i \in \{M(e_k, D)\}$ is the set of all matching patches in D returned by $M(e_k, D)$. Simplify $P(o, x|G_i, d_k)$ and show how the final terms are related to $\mathbb{P}$ computed in part (a).

Solution

$P(o, x|G_i, d_k) = P(x|o, G_i, d_k)P(o|G_i, d_k)$ (1 point)
$P(o|G_i, d_k) = P(o|G_i) \propto \sum_d \mathbb{P}(G_i, d)$ (1 point)
$P(x|o, G_i, d_k) \propto \mathbb{P}(G_i, d_k - x)$ (1 point)

(c) (2 point) Testing (Classification): Given $P(o, x|e_k, d_k)$ for all patches in a test image, provide an expression for $score(x, o)$ denoting the probability of finding the object centered at x. Assume $P(e_k, d_k)$ is a constant corresponding to a uniform prior.

Solution

$score(x, o) \propto \sum_k P(o, x|e_k, d_k)$

(d) (3 points) Now you are provided a segmentation mask for objects during training. Provide a method to obtain $\mathbb{P}$ ('pixel' on object$|o, I, L)$, denoting the probability of the "pixel" of a dictionary patch I located at location L (measured from the object center) being on the object o.

Solution

Maintain a segmentation codebook for every patch I and location L. Each codebook entry corresponds to a pixel on the patch. During training for every patch I on the image at a location L with respect to center, find the pixels under the segmentation mask and increment the corresponding pixel codebook entries. Finally normalize all codebooks to obtain the probability.

4 Short Answer (39 points)

15. **(6 points)** Parallel Lines under Perspective Transforms

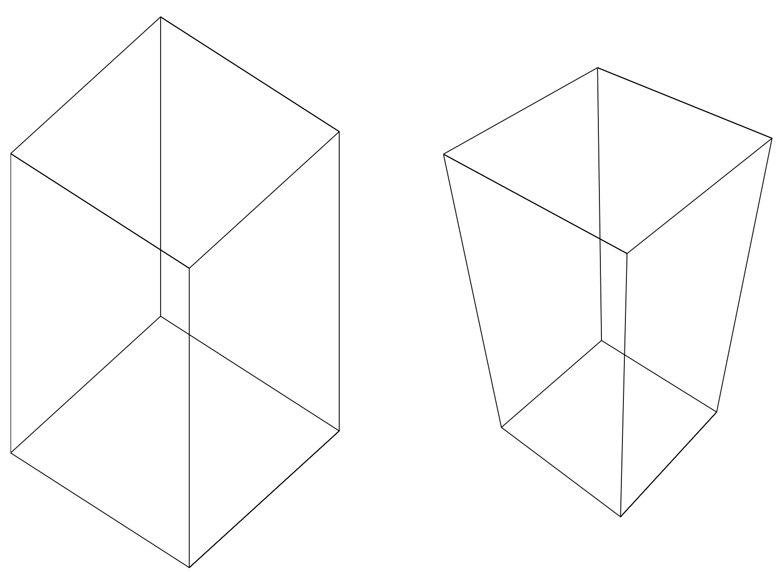


Figure 8: Boxes rendered using different projections

(a) (2 point) The two boxes in Figure 8 represent the same 3D shape rendered using two projective techniques, explain their different appearance and the types of projections used to map the objects to the image plane.
Solution Figure on the right is an orthographic projection, parallel lines are parallel, figure on the left is perspective, parallel lines at an angle to the camera plane have a vanishing point

(b) (2 point) For each projection, if the edges of the cubes were to be extended to infinity, how many intersection points would there be?
Solution Left, none, right, 3 vanishing points.

(c) (1 point) What is the maximum number of vanishing points that are possible for an arbitrary image?
Solution There is no limit, any set of parallel lines at an angle to the camera will converge at a vanishing point

(d) (1 point) How would you arrange parallel lines so that they do not appear to have a vanishing point?
Solution Place lines that are parallel to the camera plane, they will converge at the point at infinity

16. **(6 points)** Cascaded Hough transform for detecting vanishing points

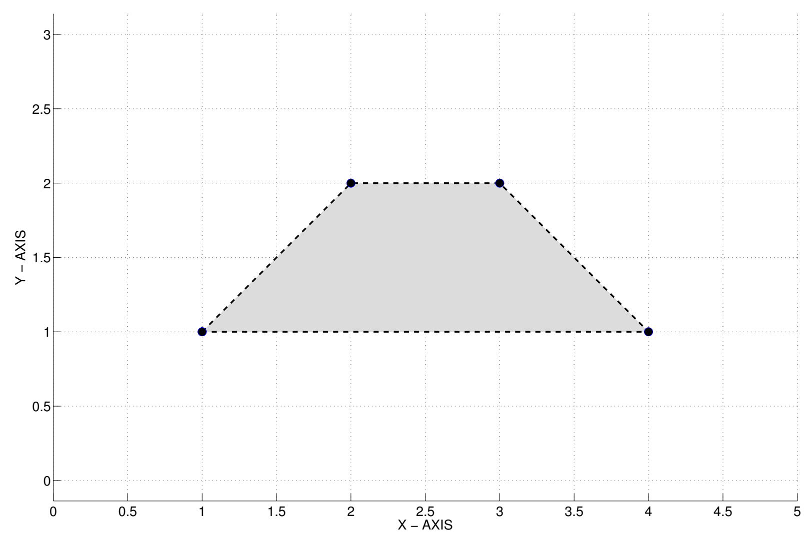


Figure 9: Hough Transform

For this problem we are going to use the slope m intercept c representation of line: $y = mx + c$. The attached Figure 9 shows the vertices of a rectangular patch under a perspective transformation. We wish to find the vanishing point in the image using a Hough Transform.

(a) (2 point) Plot the Hough Transform (for line fitting) representation of the image. Assume no binning and make plots in a continuous (m, c) space. Just show the points with two or more votes.
Solution See figure 10

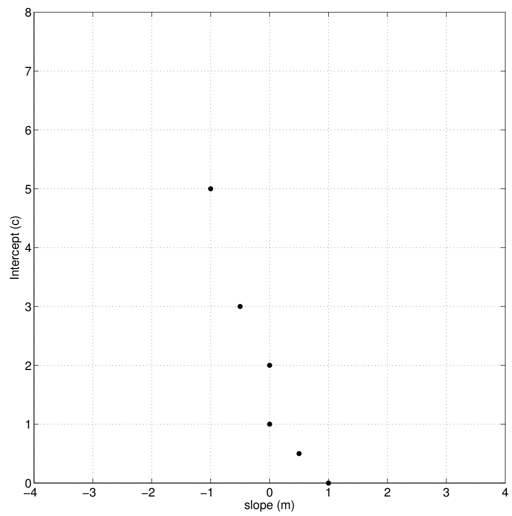


Figure 10: Make your plot here

(b) (2 point) Now using $y = mx + c$ representation, run Hough transform on the results from part (a) (after using a threshold of 2 votes) to get a representation in the (x, y) space again.
Solution
The same plot as the plot in the problem statement, with an intersection at (2.5, 1.75)

(c) (2 point) Explain how to extract the vanishing point from the representation in part (b).
Solution The vanishing point is (2.5, 1.75)

17. **(5 points)** Using RANSAC to find circles

Suppose we would like to use RANSAC to find circles in $\mathbb{R}^2$. Let $D = \{(x_i, y_i)\}_{i=1}^n$ be our data, and let I be the random seed group of points used in RANSAC.

(a) (2 points) The next step of RANSAC is to fit a circle to the points in I. Formulate this as an optimization problem. That is, represent fitting a circle to the points as a problem of the form

$$\text{minimize} \quad \sum_{i \in I} L(x_i, y_i, c_x, c_y, r)$$

where L is a function for you to determine which gives the distance from (x_i, y_i) to the circle with center (c_x, c_y) and radius r.

Solution

$$L(x_i, y_i, c_x, c_y, r) = |sqrt((x_i - c_x)^2 + (y_i - c_y)^2) - r|$$

(b) (2 points) What might go wrong in solving the problem you came up with in (1) when $|I|$ is too small?
Solution
The problem is underdetermined. With e.g. 2 points there are infinitely many circles one can fit perfectly.

(c) (1 point) The next step in our RANSAC procedure is to determine what the inliers are, given the circle (c_x, c_y, r). Using these inliers we refit the circle and determine new inliers in an iterative fashion. Define mathematically what an inlier is for this problem. Mention any free variables.
Solution
An inlier is a point (x, y) such that $|sqrt((x_i - c_x)^2 + (y_i - c_y)^2) - r| \leq T$ for some threshold T.

18. **(6 points)** Fast forward camera

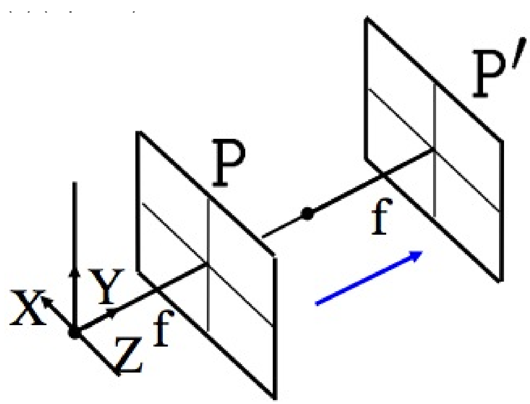


Figure 11: Camara movement

Suppose you capture two images P and P' in pure translation in the Z direction shown in Figure 11. Image planes are parallel to the XY plane.

(a) (3 points) Suppose the center of an image is (0,0). For a point (a, b) on image P', what is the corresponding epipolar line on image P?
Solution the line goes through (0, 0) and (a, b) on image P.

(b) (3 points) What is the fundamental matrix in this case?
Solution
0 -1 0
1 0 0
0 0 0

19. **(6 points)** K-Means

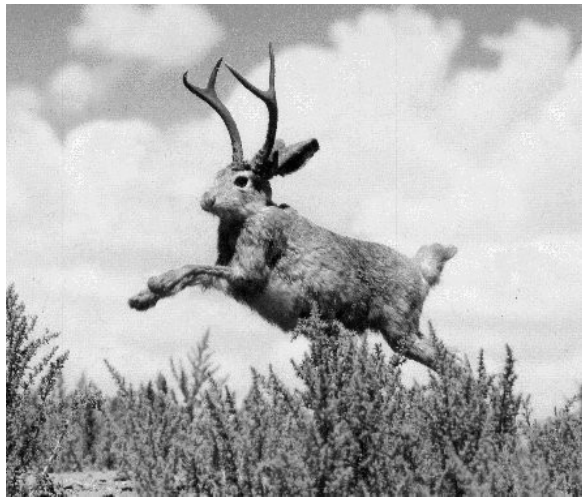


Figure 12: A wild jackalope

(a) (3 point) What is likely to happen if we run k-means to cluster pixels when we only represent pixels by their location? With k=4, draw the boundary around each cluster and mark each cluster center with a point for some clusters that might result when running k-means to convergence. Draw on Figure 12, and set the initial cluster centers to be the four corners of the image.
Solution Just a bunch of (roughly) equal-sized clusters tiling the plane, depending on the starting locations. For the image, we expect each quadrant to be its own cluster.

(b) (1 point) What does this tell us about using pixel locations as features?
Solution It's not sufficient, we need richer features.

(c) (2 point) We replace the *sum of squared distances of all points to the nearest cluster center* criterion in k-means with *sum of absolute distances of all points to the nearest cluster center*, i.e. our distance is now given by $d(x_1, x_2) = \|x_1 - x_2\|_1$. How would the update step change for finding the cluster center?
Solution At every iteration
1. Assign points to the closest cluster center
2. Update the cluster centers with the median of points belonging to a cluster

20. **(6 points)** Canny Edge Detector

(a) (3 points) There is an edge detected using the Canny method. This detected edge is then rotated by θ as shown in Figure 13, where the relation between a point on the original edge (x, y) and a point on the rotated edge (x', y') is given by

$$x' = x\cos\theta \tag{6}$$
$$y' = x\sin\theta \tag{7}$$

Will the rotated edge be detected using the Canny method? Provide either a mathematical proof or a counter example.

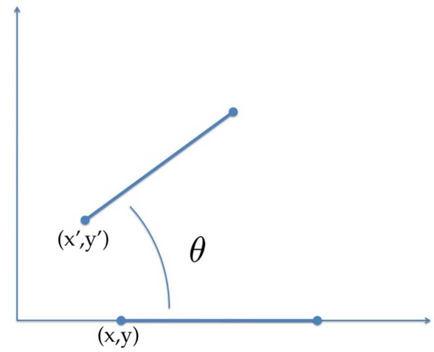


Figure 13: Edge Rotated by θ

Solution
Our rotation is given by

$$x' = x\cos\theta$$
$$y' = x\sin\theta$$

Our canny edge depends on the magnitude of the derivative which is the only part of the algorithm which could have really changed. This is given by

$$\begin{aligned}\sqrt{D^2_{x'x'} + D^2_{y'y'}} &= \sqrt{\cos^2\theta D^2_{xx} + \sin^2\theta D^2_{yy}} \\ &= \sqrt{D^2_{xx}}\end{aligned}$$

which is the same rule for the original edge thus we have shown that the Canny method is rotationally invariant.

(b) (3 points) After running the Canny edge detector on an image, you notice that long edges are broken into short segments separated by gaps. In addition, some spurious edges appear. For each of the two thresholds (low and high) used in hysteresis thresholding, state how you would adjust the threshold (up or down) to address both problems. Assume that a setting exists for the two thresholds that produces the desired result. Explain your answer very briefly.

Solution
The gaps in the long edges require a lower low threshold: parts of the long edge are detected, so the high threshold is low enough for these edges, but the edges are disconnected because the low threshold is too high. Lowering the low threshold will include more pixels of the long edges. Eliminating the spurious edges requires a higher high threshold. The high threshold should be increased only slightly, so as not to make the long edges disappear. The assumption in the problem statement ensures that this is possible.

21. **(4 points)** Image Segmentation

Both the minimum cut and the minimum normalized cut algorithms for image segmentation have problems relating to the size of the partitions that they produce.

(a) (2 point) What is the problem with the minimum cut formulation, and how is this dealt with by the minimum normalized cut formulation?
Solution
Min cut prefers very small and very large partitions

(b) (2 point) Although using normalized cuts is much better than using minimum cuts, it still has a problem concerning the size of the partitions it produces. In what way is it biased?
Solution
Prefers partitions of roughly equal size