CS 231A Computer Vision Midterm

Tuesday October 30, 2012

Full Name: ____________________

Question		Score
Multiple Choice	(22 pts)	
True or False	(10 pts)	
Detecting Patterns with Filters	(10 pts)	
Stereo Reconstruction	(10 pts)	
AdaBoost	(12 pts)	
Short Answer	(36 pts)	
Total	(100 pts)	

Welcome to the CS 231A Midterm Exam!

- The exam is 75 minutes.
- You are allowed one page of hand written notes. No calculators, cell phones, or any kind of internet connections are allowed.

I understand and agree to uphold the Stanford Honor Code during this exam.

Signature: ____________________ Date: ____________

Good luck!

1 Multiple Choice (22 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. (Pick one) 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.

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.

3. (Pick one) Which of the following processes would help avoid aliasing while downsampling 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.

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 in the above manner, reduces the number of computations.
   (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.

5. (Pick one) Which of the following is 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.

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

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 a possible change in the physical length of the lens. F is unchanged.
   (b) F can change due to a possible change in the lens characteristics. E is unchanged.
   (c) E can change due to a possible change in the lens characteristics. F is unchanged.
   (d) Both are unchanged.

8. (Pick one) Which of the following statements describes an affine camera but not a general perspective camera?
   (a) Relative sizes of visible objects 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.

9. (Circle 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's optical center.

10. (Circle all that apply) For camera calibration, we learned that since there are 11 unknown parameters, we need at least 6 correspondences to calibrate. Assuming that you couldn't find a calibration target with the minimum of 6 corners to use as correspondences, you decide to take multiple pictures from different viewpoints of a stationary pattern with N corners, where $N < 6$, which of the following statements is true?

    (a) The number of images, K must satisfy $2NK > 6$, the value of N isn't important, so long as it as $N > 0$.

    (b) The problem is unsolvable, since you do not have enough correspondences in a single image, and changing the view point changes the extrinsics.

    (c) The number of unknown parameters scales with the number of unique images taken when the calibration target is stationary.

    (d) The number of unknown parameters is fixed, but the N corners must not be co-linear.

11. (circle all that apply) Which of the following statements about correlation are true in general?

    (a) For a symmetric 1D filter, computing convolution of the filter with a signal is the same as computing correlation of the filter with the signal.

    (b) Correlation computation can be made fast through the use of Discrete Fourier Transform.

    (c) Correlation computation is not Shift Invariant.

    (d) The correlation method would be effective in solving the correspondence problem between two images of a checkerboard.

2 True or False (10 points)

True or false. Correct answers are 1 point, -1 point for each incorrect answer.

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

(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.

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

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

(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.

(g) (True/False) The Canny edge detector is a linear filter because it uses the Gaussian filter to blur the image and then uses the linear filter to compute the gradient.

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

(i) (True/False) Compared to the normalized cut algorithm, the partitions of minimum cut are always strictly smaller.

(j) (True/False) Assuming the camera coordinate system is the same as the world coordinate system, the intrinsic and extrinsic parameters of the a camera can map any point in homogenous world coordinates to a unique point in the image plane.

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 1 shows an example of a 2D Gabor Filter.

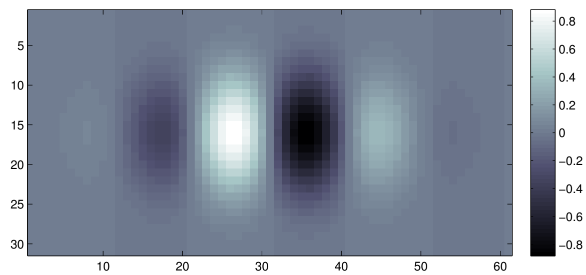


Figure 1: 2D Gabor Filter

(a) (6 points) What is the physical meaning of each of the five parameters of the Gabor filter, $\lambda, \theta, \psi, \sigma, \gamma$, 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?

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

Describe each parameter:

λ:

θ:

ψ:

σ:

γ:

(b) (4 points) Given a Gabor filter that has been tuned to maximally respond to the striped pattern in shown in Figure 2, how would these parameters, $\lambda_0, \theta_0, \psi_0, \sigma_0, \gamma_0$, have to be modified to recognize the following variations? Provide the values of the new parameters in terms of the original values.

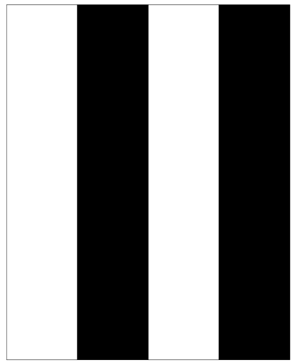
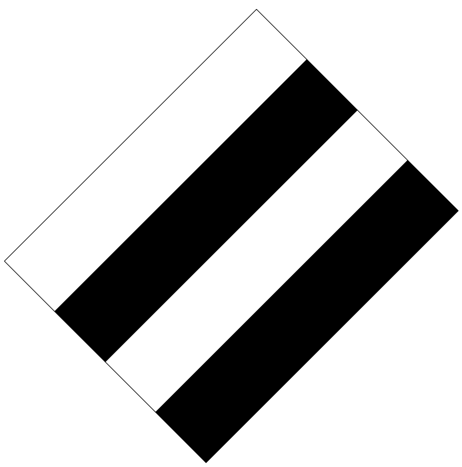
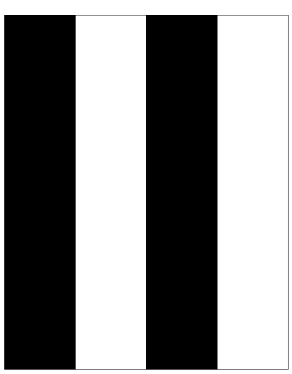


Figure 2: Reference Pattern

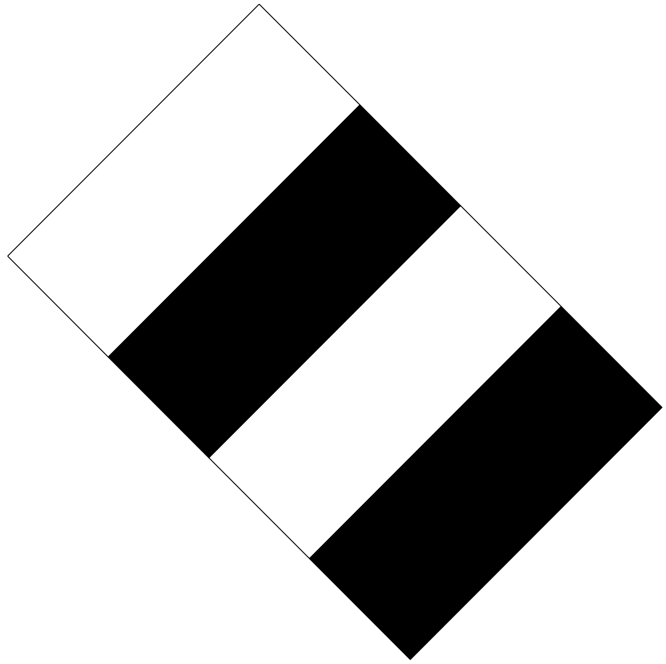
i.



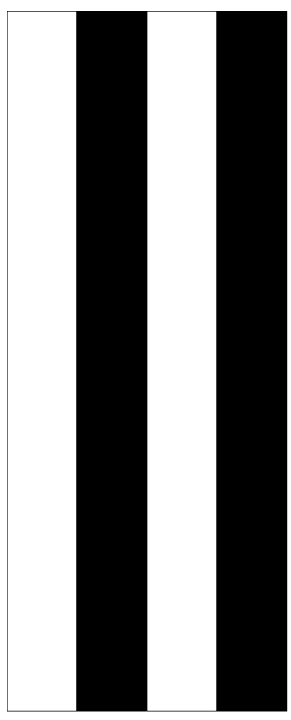
ii.



iii.



iv.



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

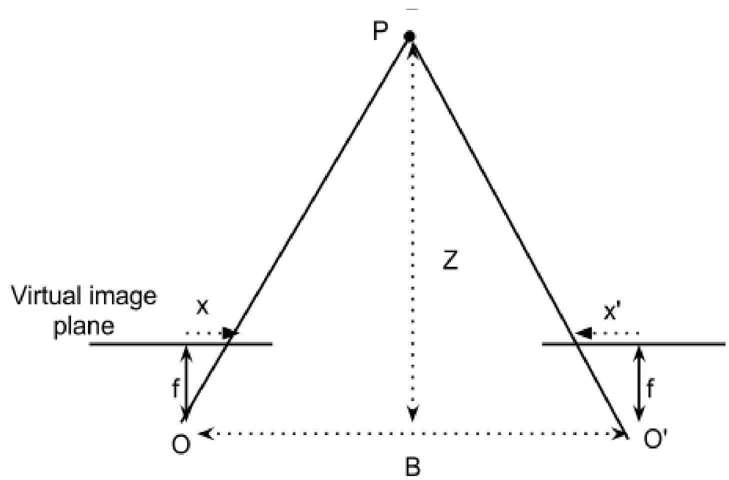


Figure 3: Rectified Stereo Rig

(a) (2 points)

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 for the rest of this question**.

(b) (4 points)

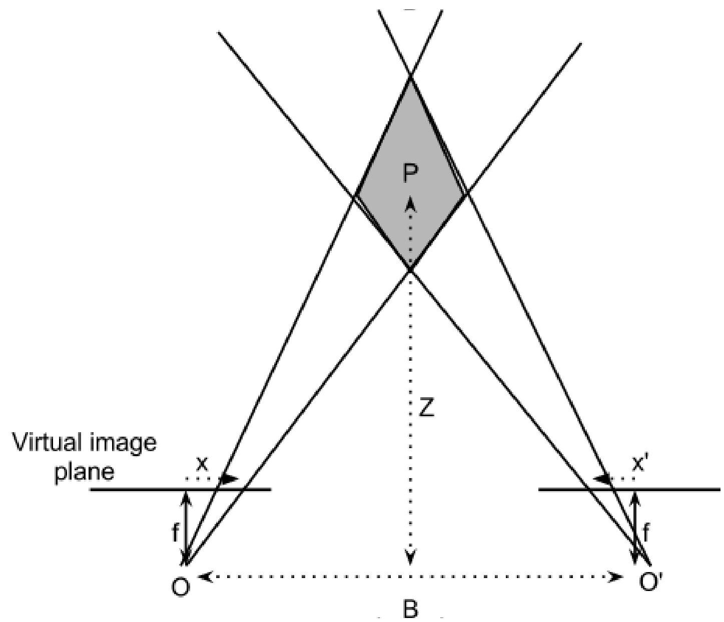


Figure 4: Rectified Stereo Rig with image plane error

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.

(c) (4 points)

Assume the X coordinate of the point P is fixed.

i. Give an expression for the uncertainty in the reconstruction of Z, in terms of the actual value of Z and the other parameters of the stereo rig.

ii. What is the depth uncertainty when Z is equal to zero?

iii. Find the depth when the uncertainty is at its maximum and give a physical interpretation and a drawing to explain.

13. **(12 points) AdaBoost algorithm for Face Detection**
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, $C_m(x)|m \in \{1,\ldots,M\}$ 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 $\beta_k, C_k \forall k \in \{1,\ldots,m-1\}$. You are also given N trainining samples $\{(x_i, y_i)\}_{i=1,\ldots,N}$, where x_i is a data point and $y_i \in \{-1,1\}$ is the corresponding output.

(a) (2 points) $(\beta_m, C_m) = \arg\min_{\beta,C} \left(\sum_{i=1}^{N} L\left[y_i, f_{m-1}(x_i) + \beta C(x_i)\right]\right)$, where $L[y,g] = \exp(-yg)$ is the loss function. Show that (β_m, C_m) can be written in the form

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

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

(b) (3 points) 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 $I[C(x) \neq y_i]$ given by

$$I[C(x_i) \neq y_i] = \begin{cases} 1, & \textit{if } C(x_i) \neq y_i \\ 0 & \textit{if } C(x_i) = y_i \end{cases}$$

(c) (3 points) 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 $w_i^{(m)} \propto w_i^{(m-1)} \cdot exp\left(2\beta_m I[y_i \neq C_m(x_i)]\right)$

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

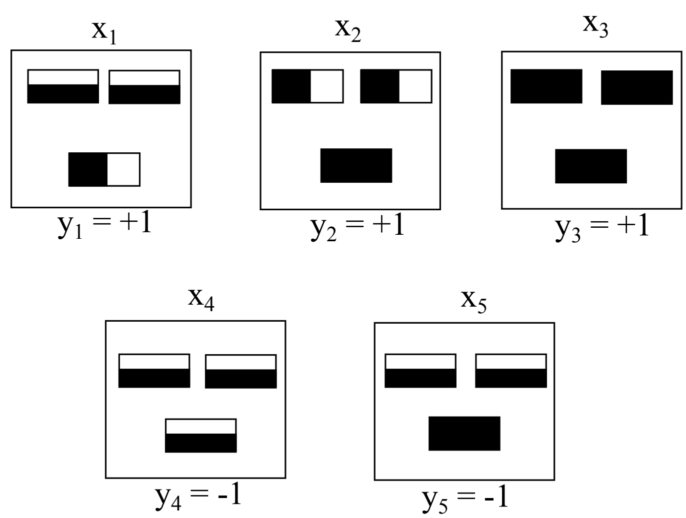


Figure 5: Training Set 'Faces'

We are also given 3 classifier patches p_1, p_2, p_3 in Fig. 6. A patch detector $I^{(+)}(x_i, p_j)$ is defined as follows:

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

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

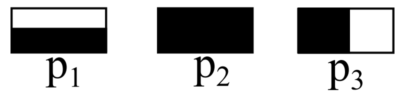


Figure 6: Classifier patches

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 $\beta_1 = 1$,

i. What is the optimal $C_2(x)$?
ii. What are the updated weights $w_i^{(1)}$?
iii. What is the final classifier $f_2(x)$ combining C_1, C_2?
iv. Does $I\left[f_2(x) > 0\right]$ correctly classify all training faces?

4 Short Answer (39 points)

14. **(6 points)** Parallel Lines under Perspective Transforms

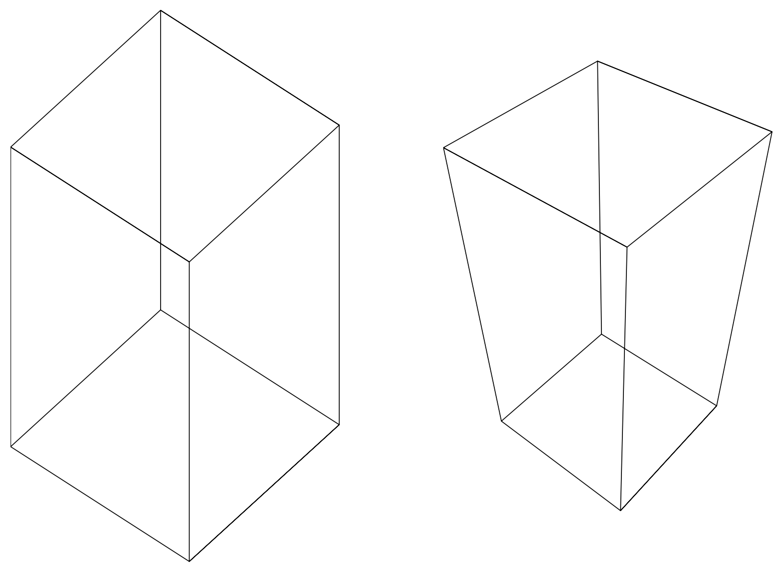


Figure 7: Boxes rendered using different projections

(a) (2 points) The two boxes in Figure 7 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.

(b) (2 points) For each projection, if the edges of the cubes were to be extended to infinity, how many intersection points would there be?

(c) (1 point) What is the maximum number of vanishing points that are possible for an arbitrary image?

(d) (1 point) How would you arrange parallel lines so that they do not appear to have a vanishing point?

15. **(6 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.

(b) (2 points) What might go wrong in solving the problem you came up with in (1) when $|I|$ is too small?

(c) (2 points) 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.

16. **(6 points)** Fast forward camera

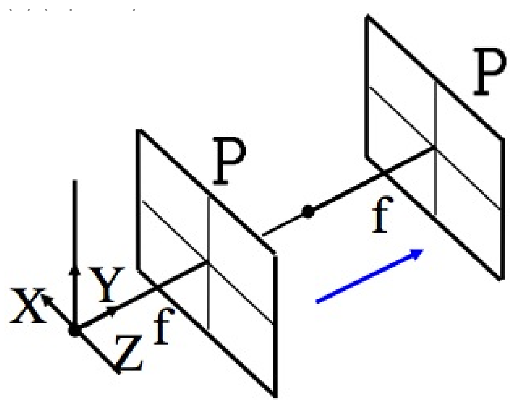


Figure 8: Camara movement

Suppose you capture two images P and P' in pure translation in the Z direction shown in Figure 8. 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?

(b) (3 points) What is the essential matrix in this case assuming the camera is calibrated?

17. **(6 points)** K-Means

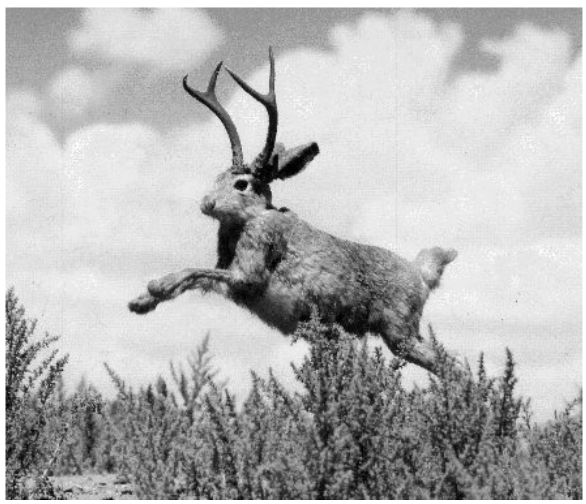


Figure 9: A wild jackalope

(a) (3 points) 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 9, and set the initial cluster centers to be the four corners of the image.

(b) (1 point) What does this tell us about using pixel locations as features?

(c) (2 points) 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?

18. **(6 points)** Canny Edge Detector

(a) (4 points) There is an edge detected using the Canny method. This detected edge is then rotated by θ as shown in Figure 10, where the relation between a point on the original edge (x, y) and a point on the rotated edge (x', y') is given by

$$x' = x\cos\theta \quad (5)$$
$$y' = x\sin\theta \quad (6)$$

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

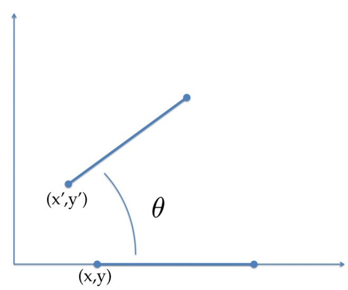


Figure 10: Edge Rotated by θ

(b) (2 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.

19. **(6 points)** Cascaded Hough transform for detecting vanishing points

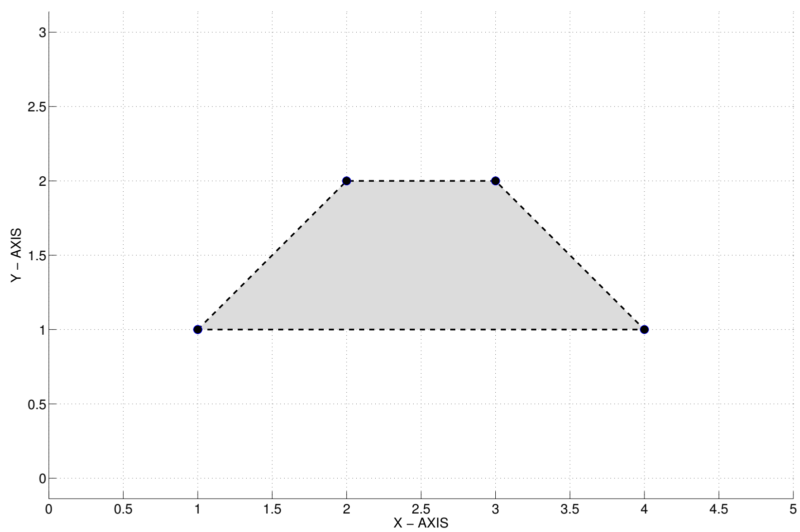


Figure 11: Hough Transform

For this problem we are going to use the slope m intercept c representation of line $y = mx + c$. The attached Figure 11 shows the vertices of a rectangular patch under perspective transformation. We wish to find the vanishing point in the image through Hough Transform.

(a) (2 points) Plot the Hough transform 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.

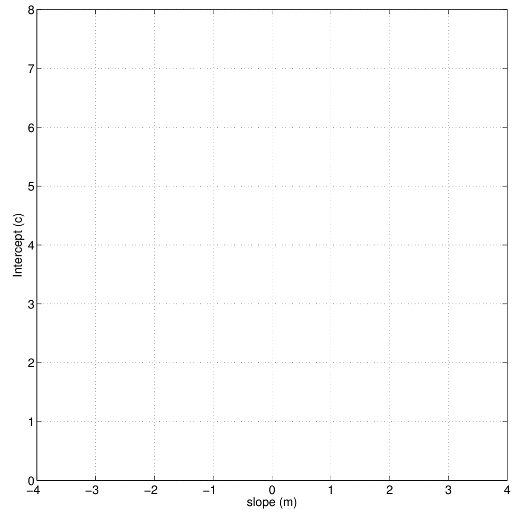


Figure 12: Make your plot here

(b) (2 points) 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.

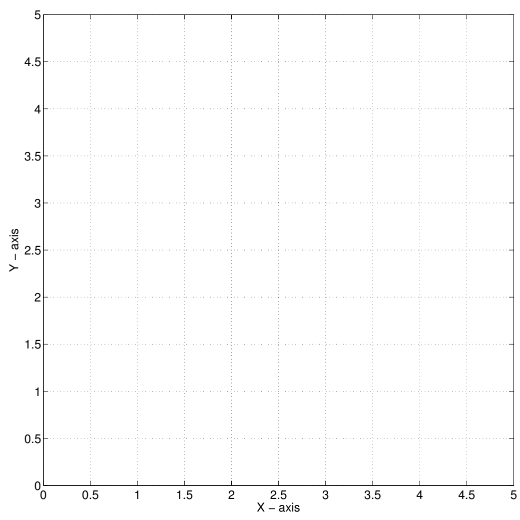


Figure 13: Make your plot here

(c) (2 points)Find the vanishing point from the representation in part (b)