McGill

October 2024
Midterm Examination

INTRODUCTION TO COMPUTER VISION

ECSE-415

(There are 10 questions in total, each is worth 10 marks)

This take-home exam is to be done online, on mycourses.

The exam will be available on mycourses for a 48-hour period. You can submit your answers at any time during this period. You can do multiple submissions, but only the last one will be graded.

The exam is open-book, but you must do the exam by yourself, with no outside assistance.

1. Cameras

Suppose you are at the camera store and are aiming to buy a new lens for your digital camera. The choices available at the store are:

a. 135mm focal length, f-stop of f/2.0
b. 24mm focal length, f-stop of f/2.8

a) Which lens would you choose if you wanted the widest field of view?
b) Which lens would you choose if you wanted the shallowest depth of field?
c) Which lens would be the least expensive?
d) Which lens provides the most light at the camera's image sensor (when viewing a scene with uniform brightness)?
e) Which lens would give more perspective distortion?

2. Filtering

Given the five convolution kernels shown in a) through e) apply each convolution mask to the central pixel of the image patch depicted in diagram f) below (i.e. get a single number for each).

a)

1	1	1
1	1	1
1	1	1

b)

-1	-2	3
-1	2	-1
3	2	1

c)

-1	-2	-1
0	0	0
1	2	1

d)

-1	0	1
-2	0	2
-1	0	1

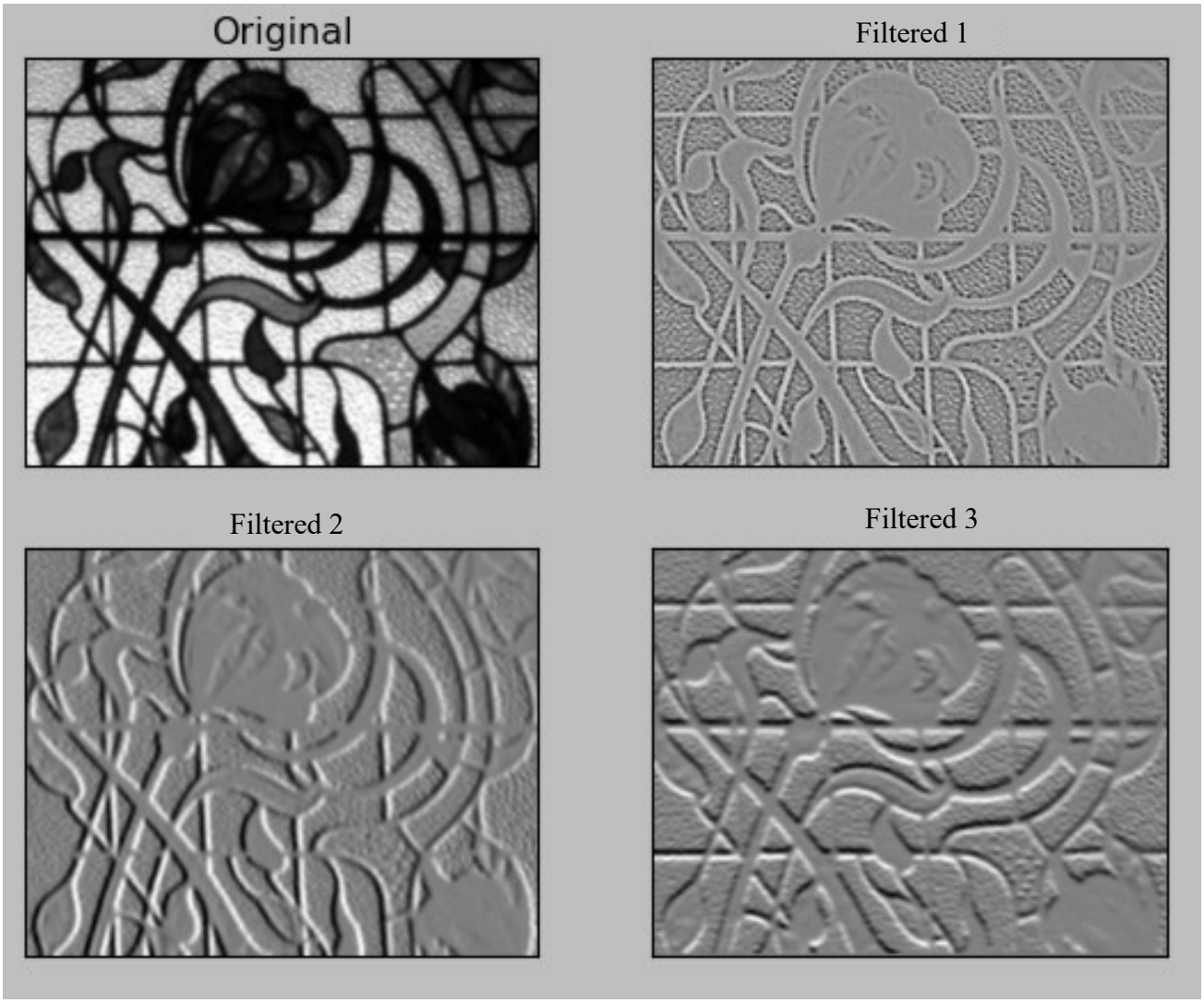
e)

0	1	0
1	-4	1
0	1	0

f)

7	**3**	**1**
7	**5**	**0**
8	**4**	**1**

Value for a): ……
Value for b): ……
Value for c): ……
Value for d): ……
Value for e): ……

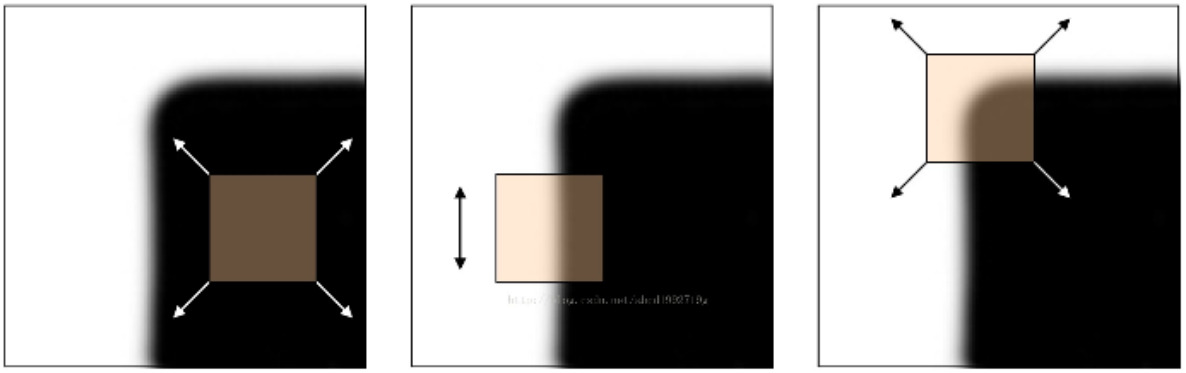


Match each filtered image with the convolution kernel that was used to generate it (from the previous page)

Filtered 1: convolution kernel ……. (out of choices a, b, c, d, e)
Filtered 2: convolution kernel ……. (out of choices a, b, c, d, e)
Filtered 3: convolution kernel ……. (out of choices a, b, c, d, e)

3. Harris Corner Features

a) Which of the following image patch locations would produce the highest response for a Harris operator? Answer = …… (a,b or c)



(a) (b) (c)

b) The Harris operator is given by:

$$f = \frac{\lambda_- \lambda_+}{\lambda_- + \lambda_+}$$

Suppose we have the following 4x4 pixel image patch:

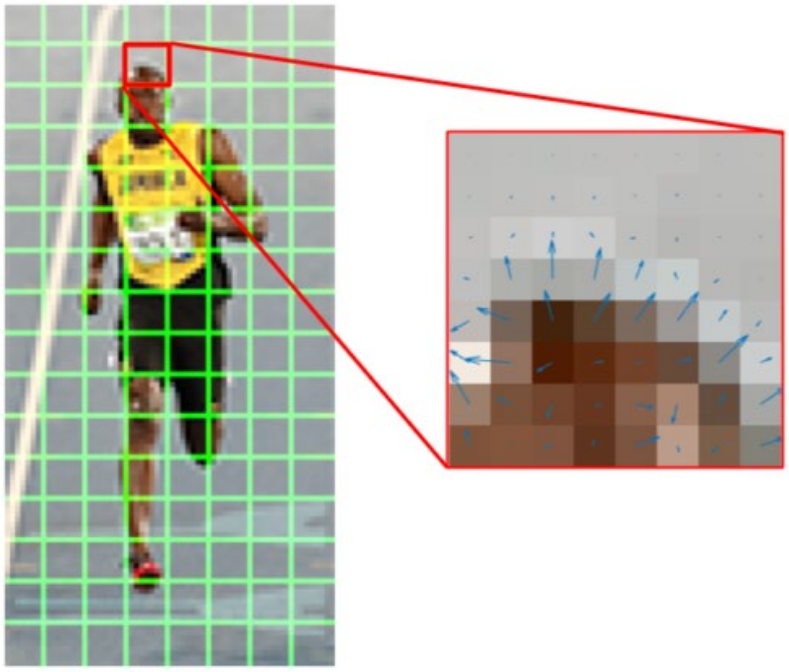
3	**4**	**3**	**1**
4	**6**	**5**	**0**
6	**8**	**4**	**1**
6	**9**	**7**	**2**

Compute the value of *f* for the 2x2 pixel window in the middle of this patch.
Answer = ……..

4. SIFT/HoG Descriptors

a) What image transformations is SIFT invariant to? (select all that apply)
   a. Rotation in the image plane
   b. Translation
   c. Non-rigid object deformation
   d. Affine intensity change
   e. Logarithmic image intensity compression
   f. Image intensity inversion (e.g. 255-I(x,y))
   g. Mirror reflections
   h. Scaling (changes in size of objects)
   i. Similarity transforms
   j. Perspective projections

b) In the image block shown highlighted below, the gradient magnitude and gradient directions have been computed. Assuming the orientation histogram has 9 bins covering the range of angles from 0 to 180 degrees:
(0-19, 20-39, 40-59, 60-79, 80-99, 100-119, 120-139, 140-159, 160-179)
Compute the 9 entries of the HOG histogram for this block, weighted by the gradient magnitudes (i.e. increment the corresponding bin by an amount equal to the gradient magnitude).



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

Answer = [……, ……., ……., …....., ……, ……., ……., ……., …….]

5. Homography and Image matching

a) Match the following homography matrices to the type of transformation that they produce.

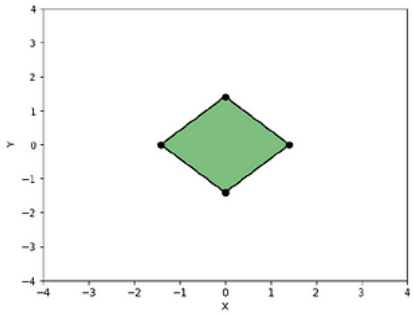
1	0	1
0	1	2
0	0	1

0.866	-0.5	0
0.5	0.866	0
0	0	1

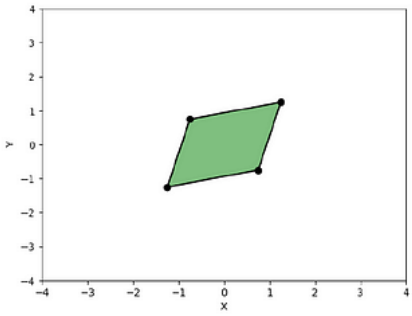
1	0.25	0
0.25	1	0
0	0	1

2	0	0
0	2	0
0	0	1

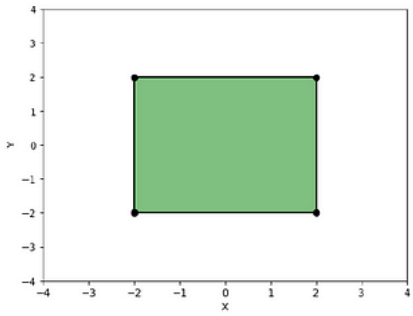
0	-1	0
1	0	0
0	0	1



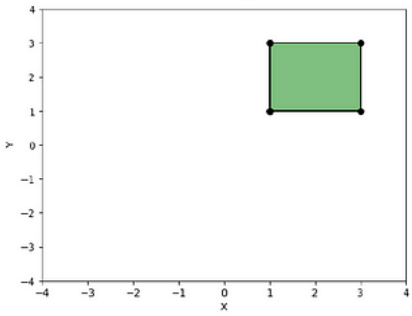
Rotation



Skewing



Scaling



Translation

b) RANSAC: ***for each of the 10 pairs of points*** in the dataset of 5 different 3D features shown below, ***find the number of inliers*** to the line between those points. Define an inlier as a point that has a distance of 1.0 units or less to the line (hint: look back at your math class notes to remember how to compute the distance between a point and a line).

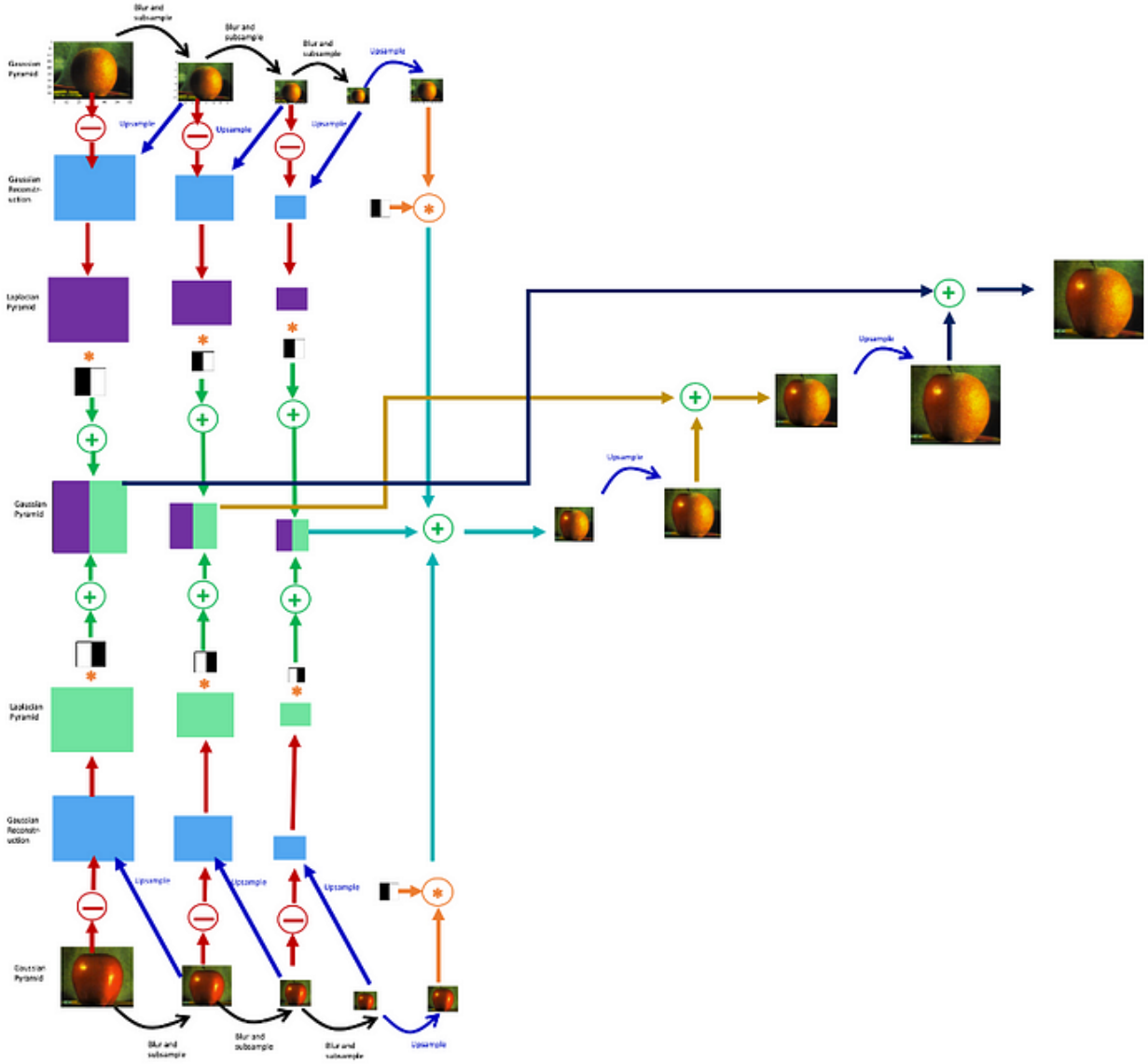
Point 1: (1.0, 1.0, -2.2)
Point 2: (0.0, 1.0, -1.4)
Point 3: (1.0, -1.0, 2.0)
Point 4: (2.0, 0.2, -2.0)
Point 5: (-1.0, 0.8, 0.6)

Pair [Point1, Point2] # inliers = ……
Pair [Point1, Point3] # inliers = ……
Pair [Point1, Point4] # inliers = ……
Pair [Point1, Point5] # inliers = ……
Pair [Point2, Point3] # inliers = ……
Pair [Point2, Point4] # inliers = ……
Pair [Point2, Point5] # inliers = ……
Pair [Point3, Point4] # inliers = ……
Pair [Point3, Point5] # inliers = ……
Pair [Point4, Point5] # inliers = ……

6. Scale Space

a) Explain what a Laplacian or DoG pyramid is and how it can be constructed.

b) With the aid of the diagram below explain how multi-scale image blending works, such as used in making panoramic images.



7. Faces

a) What are two characteristics of a good image representation for face detection?

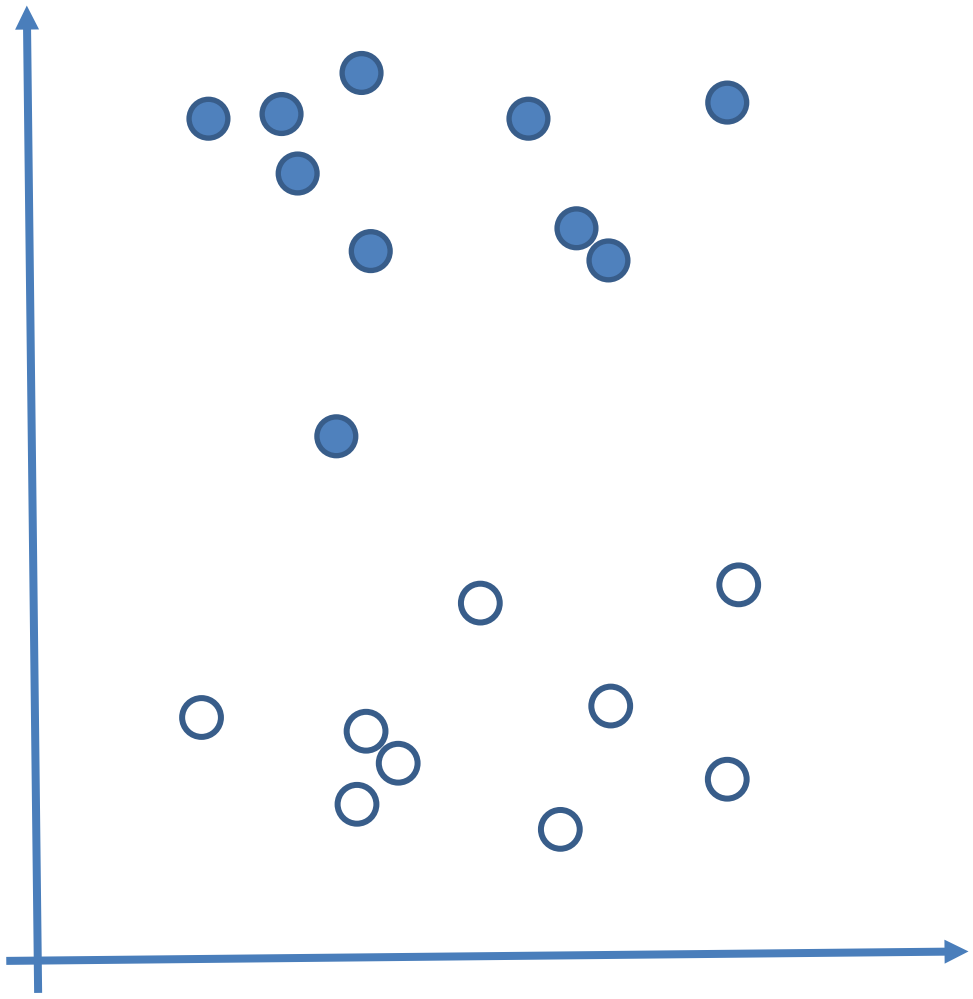
b) Explain the difference between face detection and face recognition.

c) Suppose you want to represent a face image patch with three principal components. How do you select the principal components that would give the best reconstruction of the original face image patch?

8. Support Vector Machines

Consider a 2-dimensional 2-class classification problem, with samples (training data) shown in the diagram below.

a) Draw the ***maximum margin classifier's*** decision boundary and show the ***margin*** (or at least where you think they should be, approximately)
b) Circle at least one ***support vector*** on the diagram



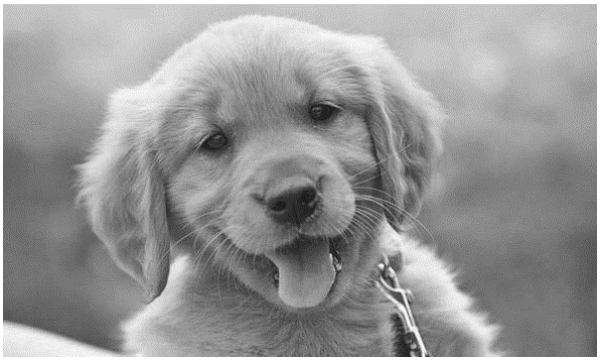
c) Suppose that for a given 2-D feature vector (x1,x2) the training data can be nonlinearly separated by the curve defined by $x^3 + y^3 + xy = 0$.

Define a new set of features (z1,z2) for which the data can be linearly separated by the hyperplane $z1+z2 = 0$.

z1 =

z2 =

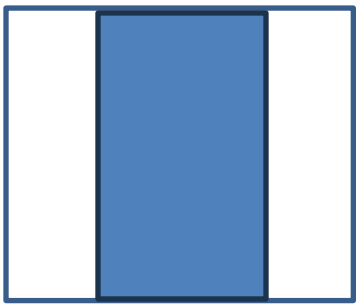
d) Which one of these images is the most likely to be a support vector if they were used as training examples for an SVM classifier? Explain your answer.



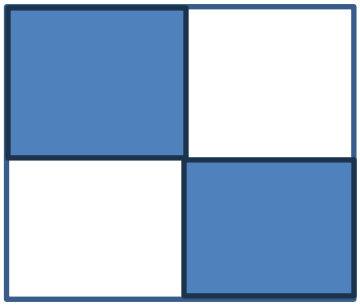
9. Haar features

Shown below are 3 different Haar features:
(on a 4x4 grid of pixels, blue = -1, white = +1)

Haar Feature A



Haar Feature B



Haar Feature C

a. In the box on the right, fill in the values of the INTEGRAL IMAGE corresponding to the small image whose values are shown on the left.

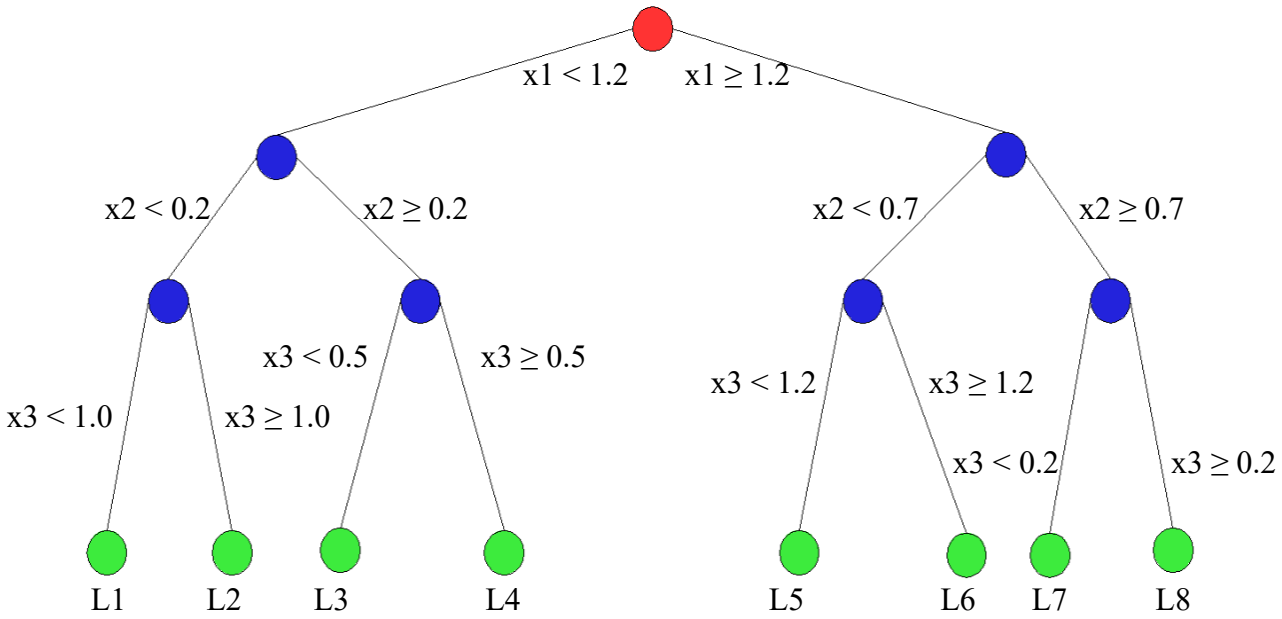
1	3	0	7
1	1	2	3
4	1	2	2
4	3	0	1

b. Compute the value for each of the Haar features for the image above.

Answer: Haar Feature A Value = ……
Answer: Haar Feature B Value = ……
Answer: Haar Feature C Value = ……

10. Decision Tree Classification:

Consider the decision tree shown below which uses three features x1, x2, x3 to classify into one of three classes C1, C2, C3.



Suppose we are given the following training data:

x1	x2	x3	Class	Leaf Node
0.0	0.0	0.0	C1	
1.0	1.0	1.5	C3	
1.5	0.1	0.9	C1	
1.8	0.1	1.1	C2	
1.3	0.8	2.0	C1	
0.6	0.6	1.7	C3	
0.4	1.2	0.7	C1	
2.2	0.0	1.0	C2	
0.0	0.8	1.3	C2	
1.2	0.5	1.1	C3	

a) Fill in the last column of the table (indicate which leaf node is selected).

b) Given new feature values x1=2.0, x2=0.5, x3=0.9, what is the probability of each class?