CIS 663 Biometrics

Assignment 2

This assignment is due by the week 5 live session. If you make any assumptions, clearly state them in your answer.

1. The following represents a 10 x 10-pixel grayscale. 0 represents black and 255 represents white.

0	0	0	0	0	0	0	1	1	1
0	0	3	3	3	3	3	2	1	1
0	0	3	3	4	4	4	4	4	4
0	0	3	3	3	3	4	4	4	4
0	0	0	1	1	3	4	4	4	4
0	0	0	0	0	4	4	4	1	0
5	5	0	0	0	4	4	4	0	0
5	5	0	0	0	4	4	4	0	0
5	5	0	0	0	0	5	5	0	0
5	5	0	0	0	0	5	5	0	0

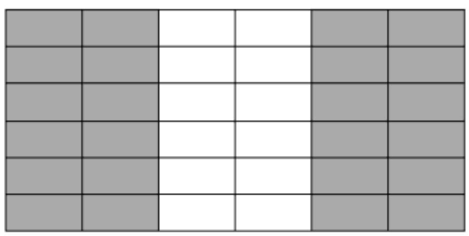
a. Convert the image to an integral image. (10pt)

0	0	0	0	0	0	0	1	2	3
0	0	3	6	9	12	15	18	20	22
0	0	**6**	**12**	**19**	**26**	**33**	**40**	46	52
0	0	**9**	**18**	**28**	**38**	**49**	**60**	70	80
0	0	**9**	**19**	**30**	**43**	**58**	**73**	87	101
0	0	**9**	**19**	**30**	**47**	**66**	**85**	100	114
5	10	19	29	40	61	84	107	122	136
10	20	29	39	50	75	102	129	144	158
15	30	39	49	60	85	117	149	164	178
20	40	49	59	70	95	132	169	184	198

b. Using the integral image, compute the sum of area from (2,2) to (5,7), shaded red above. Show your steps. (10pt)

Following the pattern of Area = C + A - B - D,
we get 85 + 0 - 18 - 0 = **67**

2. Using the grayscale image from Question 1, apply the following Haar filter to all positions that are feasible. (20pts)



0+28-19 =9	9+36-21 =24	19+38-28 =29	21+34-36 =19	28+29-38 =19
10+32-19 =23	14+44-21 =37	19+45-32 =32	21+36-44 =13	32+27-45 =14
20+30-13 =37	16+46-15 =47	13+48-30 =31	15+37-46 =6	30+25-48 =7
30+22-7 =45	18+43-8 =53	7+50-22 =35	8+34-43 =-1=1	22+17-50 =-11=11
40+16-1 =55	20+41-2 =59	1+52-16 =37	2+31-41 =-8=8	16+9-52 =-27=27

3. In Viola-Jones face detection algorithm, explain what cascading is and why it is important. (20pt)

Cascading in the Viola-Jones face detection algorithm is the strategy to use weak classifiers in groups to build confidence with a more efficient use of computing power. This means that you have more and more weak classifiers per stage as the sample passes more and more stages. This is important because it prevents the algorithm from doing too much work on non-face or easily classified images, and saves that compute time for cases that actually require it.

4. (20pts) Consider the following labeled data $(x, y) \in R^2$ (i is the example index):

i	x	y	Label
1	11	3	-
2	10	1	-
3	4	4	-
4	12	10	+
5	2	4	-
6	10	5	+
7	8	8	-
8	6	5	+
9	7	7	+
10	7	8	+

In this problem, you will use Adaboost to learn a hidden function from this set of training examples. We will use two rounds of AdaBoost to learn a hypothesis for this data set. In round number t, AdaBoost chooses a weak learner that minimizes the weighted error(t). As weak learners, you will use axis parallel lines of the form
(a) Label + if x > a, else - or
(b) Label + if y > b, else -, for some integers a, b (either one of the two forms, not a disjunction of the two).

a) The first step of AdaBoost is to create an initial data training data weight distribution D1. What are the initial weights given to data points with index 4 and 7 by the AdaBoost algorithm, respectively?

b) Which is the hypothesis h1 that minimizes the weighted error in the first round of AdaBoost, using the distribution D1 computed in the above question?

the first round of AdaBoost, using the distribution D1 computed in the above question?

c) What is the weighted error of h1 computed above?

d) After computing h1 in the previous questions, we proceed to round 2 of AdaBoost. We begin by recomputing data weights depending on the error of h1 and whether a point was (mis)classified by h1. What are the weights given to data points with index 4 and 7 according to the distribution after round 1, D2, respectively?

e) Which is the hypothesis h2 that minimizes the weighted error in the second round of AdaBoost, using the distribution D2 computed in the above question?

f) What is the weight assigned to the hypothesis of round 2, h2

g) Now that we have completed two rounds of AdaBoost, it is time to create the final output hypothesis. What is the final weighted hypothesis after two rounds of AdaBoost?

Formulas:

$$\epsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right|$$

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

Where $e_i = 0$ if input *i* is classified correctly and 1 if classified incorrectly.

$$\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$$

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

$1-\epsilon_t$

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha_t = \log \frac{1}{\beta_t}$$

5. What is Principle Component Analysis and how does it relate to face recognition? (20pts)

Principle Component Analysis (PCA) is a strategy that allows for the reduction of dimensionality of a set of vectors. This is done by using a covariance matrix and relating the data to eigenvectors of that matrix by picking subsets of the data that make up those vectors. This concept is used in facial recognition with the concept of eigenfaces, which are basically masks that can be compared to an input to see if the input is reducible to that eigenface. A given classification can be described as a percentage combination of some subset of the eigenfaces, and by measuring the combination of faces seen in a sample, it can perform the classification.