

MULTIMEDIA



UNIVERSITY

STUDENT ID NO

--	--	--	--	--	--	--	--	--	--

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2016/2017

TDS3651 – VISUAL INFORMATION PROCESSING
(All sections / Groups)

1 MARCH 2017
9.00 a.m. – 11.00 a.m
(2 Hours)

INSTRUCTIONS TO STUDENTS

1. This Question paper consists of 5 pages with 4 Questions only.
2. Attempt **ALL** questions. All questions carry equal marks and the distribution of the marks for each question is given.
3. Please write all your answers in the Answer Booklet provided.

Question 1

a) Define the process of **quantization** and its role in producing a digital image. [2 marks]

b) Given a large repository of 8-bit RGB images (24-bit in total) at 1600x1200 pixel dimensions, what is the ratio of storage reduction if we store all the images as 640x480 pixel grayscale images? [2 marks]

c) Alpha-trimmed mean filter is an order-statistics filter summarized by this equation,

$$\hat{f}(x, y) = \frac{1}{mn - \alpha} \sum_{(s,t) \in S_{XY}} g_r(s, t)$$

where α is the trim parameter, which dictates the $\alpha/2$ lowest and $\alpha/2$ highest values to be deleted before averaging the remaining pixel values g_r .

(i) Apply a 3x3 alpha-trimmed mean filter ($\{m, n\} = 3$; $\alpha = 4$) on the image below, applying 1-pixel zero padding along the borders. Draw the output filtered image (ensure final values are rounded to the nearest integer). [3 marks]

Input image:

180	210	112
8	177	97
70	101	195

(ii) What is the value of α for a 3x3 alpha-trimmed mean filter to function like a median filter? [1 mark]

d) Briefly describe the process of obtaining a sharpened image, through the use of a smoothed image. [2 marks]

Continued...

Question 2

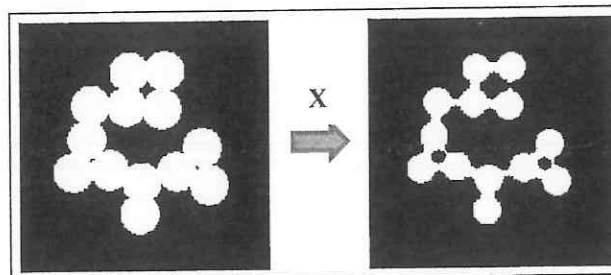
a) The Canny edge detector is one of the most widely known edge detection method that is known for good localization and thin edges. What additional steps does it have that standard gradient-based detectors lack? [2 marks]

b) What are the **three** components in the HSV color model, and what do they measure? [3 marks]

c) Label the region blobs in this binary image based on connected components. Assume pixels are 4-connected as neighbors. Use '0' for background pixels, and non-zero integers for the labeled regions. [3 marks]

0	0	0	0	0	1
0	1	1	0	1	0
1	1	1	0	1	1
0	1	1	1	0	1
1	1	1	0	0	1
1	0	1	0	0	0

d) The following binary image on the left undergoes a morphological operation **X** to produce the output binary image on the right.



(i) Identify the morphological operation **X**.

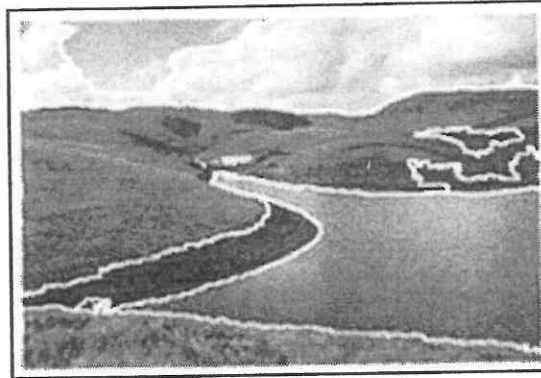
(ii) Propose another morphological operation (or a combination of operations) instead of **X** that can maintain the original size of the circular objects.

[2 marks]

Continued...

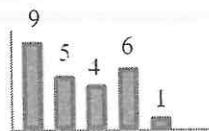
Question 3

- a) Briefly explain how Felzenszwalb's **graph-based segmentation algorithm** partitions an image into a number of distinct regions. The sample output image below shows the regions as marked by borders (Assume that the image contains colour). [3 marks]



- b) Describe the two steps involved in creating **local invariant features** – *feature detection* and *feature description*. Use SIFT descriptor as the example to elaborate your answer. [3 marks]
- c) Give an example of an application where clustering is important. Provide some details as to *how* clustering is used in your example. [2 marks]
- d) Given a vocabulary of V visual words, compute the cosine similarity score (up to 3 decimal places) between the feature vectors of the two images provided (\mathbf{x} and \mathbf{y} respectively), using the following measure:

$$\text{sim}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{\sum_{i=1}^V x_i y_i}{\sqrt{\sum_{i=1}^V x_i^2} \sqrt{\sum_{i=1}^V y_i^2}}$$



$V = 5$ visual words

[2 marks]

Continued...

Question 4

- a) Describe how **background subtraction** segments motion in videos to produce a binary foreground mask. [2 marks]

- b) The following questions are on performance metrics for image retrieval and classification:

- i) Given a query image of a bicycle (**B**), an image retrieval system returns the top 10 matches as follows:

Top 10 matches (in order of rank): **B, C, B, B, T, B, C, T, B, B**
(Abbreviations: '**B**': Bicycle, '**C**': Car, '**T**': Train)

Compute the Average Precision (AP) of the query image based on the top 10 matches (Give up to 3 decimal places).

[2 marks]

- ii) Precision and recall measures can also be used to evaluate classification performance. The confusion matrix below contains the performance of a 2-class image classification task, which predicts whether an image is labelled as a 'Bicycle' or 'Car'. Compute the **precision** and **recall** of the 'Car' category. Assume that 800 images per category were evaluated.

[2 marks]

Actual Labels	Predicted Labels	
	Bicycle	Car
Bicycle	675	125
Car	67	733

- c) A business intelligence company has tasked you to come up with a vision-based solution for **pedestrian analytics**, whereby they would like to monitor *pedestrian volume* in a public shopping street to extract *trend information* that might be useful for businesses along the street. Some outdoor video cameras have been placed along the street to monitor the scene.

Write a technical summary of the methods you can use to fulfill this task from the pedestrian detection step, leading to the extraction of the required analytics.

[4 marks]

End of Question Paper