COMP 6223 18-19 Test

1 Question 1.

Figure 1: The image of book cover

A library system is required to match an image of a book cover to a reference image (or template) stored in the library database. In the library, a user places the book face-up on a surface beneath a camera, and so the cover appears in the manner of that shown for a well-known book in Figure 1.

A computer vision system is required to match these images. Describe how this can be achieved, using a selection of computer vision approaches. Address whether colour could be used within the system, to take advantage of the delightful green and yellow colours that some books like this enjoy. Describe any likely limitations to the systems you describe. [33 marks] The solution of this problem requires two steps:
(1)

- First step: Find book:
  -Template match. The threshold can be used to speed up.
  -HT lines
- Second step:
  1. If find text:
  -Use character recognize (Templates matching, but slow)
  -Moments (Fourier descriptors)
  2. If don't use text:
  -interest point (SIFT or SURF)
  -matching

(2) If we simply use color contrast, we need to perform calibration that need lots of time for calculation. Therefore, in the tool we obtain the largest color (green) by comparing the color histogram with the target image. If the largest color in the target image is also green, proceed to the next step. Otherwise, it searches for the next book.

2 Question 2.

2.1 In terms of core principles, describe the advantages and disadvantages of edge detection as a stage in a computer vision system. [10 marks]

Property	First-order-operator	Second-order-operator
Method of edge detection	Thresholding	Zero crossing detection
advantage	Simple, only need to calculate first order differentiation.	1.Connected edge 2.Can be filtered by Gaussian filter
disadvantage	Only get partial disconnected edges(need to use advantaged operator such as Canny)	Complexity, the speed is low because of lots of calculation

Figure 2: Advantage and Disadvantage

2.2 Design an operator that can detect circular edges in images.[17 marks]

In this part, there are two main steps to get the circular shape from a raw image. The first step is using Canny operator to get all edge in image. The second is Hough transform. The details are shown as follow:

1. Gaussian smoothing: Using Gaussian filter to get the low frequency image to remove the noises.

2. Sober operator: Using Sober operator to get the edge,but the edge extracted from Sober is not accurate enough.

3. Non-Maximal suppression(Figure 2): the purpose of this step is to get the accurate edge information and give up the useless pixel value. To begin with, using the direction of the vector from Sober and get the normal.neighber pixel, and then calculating the value of M1 and M2 from their neighber.

$M_1 = \frac{\mathrm{My}}{\mathrm{Mx}} M(x+1, y-1) + \frac{\mathrm{Mx}-\mathrm{My}}{\mathrm{Mx}} M(x, y-1)$

$M_2 = \frac{\mathrm{My}}{\mathrm{Mx}} M(x-1, y+1) + \frac{\mathrm{Mx}-\mathrm{My}}{\mathrm{Mx}} M(x, y+1)$

In addition, compared with M1, M2 and Px,y, if Px,y is not the maximum, its value will be set to 0.

4. threshold with hysteresis to connect edge points: it can get a continuous edge line. Firstly give two threshold: upper switching threshold(a) and lower switching treshold(b). If a pixel value is bigger than a, it will be a edge point. And then continue to find its neighberhoods, if their values are bigger than b, they are edge points, if their values are smaller than b, they will set to 0.

5. Hough Transform: The Hough transform is a technique that locates shapes in images. In this part, Hough transform for circles can be used. The equation of a circle contains 3 unknown variables x, y, and r. The equation is $x_0 = x - r\cos(\theta) \quad y_0 = y - r\sin(\theta)$. The Hough transform is to Traverse the boundary points in the original image and replace x and y with x0 and y0 to get the equation: $x_0 = x - r\cos(\theta) \quad y_0 = y - r\sin(\theta)$ According to the x and y in the original image, take different r values to draw a circle in a three-dimensional space. The position of the most intersection of all the obtained circles is the position of the center of the circle, and its height is r. Record the center and r of all the most voted circles to locate the circle in the image.

(You can also use templates or design an operator or look for curved line with direction to answer this)

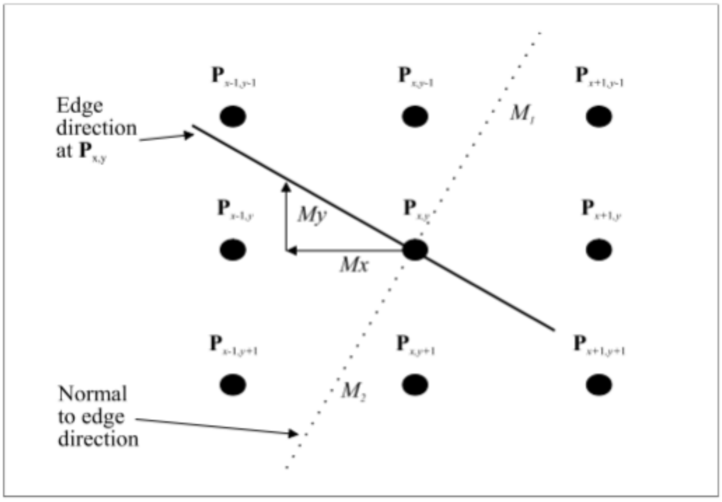


Figure 3: Non-maximal Suppression

2.3 Discuss the advantages and limitations of your approach to circular edge detection.[6 marks]

Operator uses Canny operator and Hough transform, so it inherits the advantages and disadvantages of these two calculation methods.

Advantage:

1.Due to the Gaussian convolution and the processing of the Sober operator, the operator is very resistant to noise.

2.Due to Non-maximal suppression and threshold, operator can get more accurate boundaries.

3.Due to the voting system, the Hough transform has a certain ability to resist noise and occlusion.

4.Hough transform can get the same results as template matching, but it is faster.

Disadvantages:

1.Because the Hough transform for circle needs to limit the radius range, if the value is not good, it will cause inaccurate detection.

2.But Huff transform still requires a lot of storage space and high computing requirements.

(From Speed vs complexity, Limitations vs Generality to talk)

3 Question 3.

3.1 Describe why the Fourier transform is important in image analysis[10 marks]

The Fourier transform can decomposition into frequencies. In image analysis, it is often necessary to analyze the main content, details, boundaries and other information of the image, while the high frequency part of the image represents noise and detail, and the low frequency part represents the main content. In many image analysis processes, noise needs to be removed. At this time, a low-pass filter can be used to remove the noise to make the image smoother. Sometimes you need to preserve the edges of the image, then you can use high-frequency waves. At the same time, convolution is an important tool for image analysis and processing. The Fourier transform can be used to implement the convolution process faster. First, the original image and template are Fourier transformed. In addition, dot-multiply the transformed image and template. Then perform inverse Fourier transform to achieve convolution.

3.2 [17 marks]

Given a shift T of an image along the x axis, then the Fourier transform of an image F is:$\mathcal{F}[P(x-\tau,y)] = e^{-j\omega_x\tau}P(\omega_x,\omega_y)$

(i) Explain all the symbols used and show that the magnitude of the Fourier transform is shift invariant; and

ω is the angular frequency; measured in radians/s (where the frequency f is the reciprocal of time t, f = 1/t). p(t) is a continuous signal. j is the complex variable. $\mathrm{e}^{-\mathrm{j}\omega t} = \cos(\omega t) - j\sin(\omega t)$gives the frequency components in p(t).

If we shift all the features by a fixed amount, or acquire the image from a different position, the magnitude of its Fourier transform does not change. This property is known as shift invariance.
$\Im[P(x-\tau,y] = \mathrm{e}^{-\mathrm{j}\omega\tau}P(\omega_x,\omega_y)$
$|\Im[P(x-\tau,y]| = \left|e^{-\mathrm{j}w\tau}P(\omega_x,\omega_y)\right| = \left|e^{-\mathrm{j}\omega\tau}\right||P(\omega_x,\omega_y)|$
Because $|\mathrm{e}^{-\mathrm{j}\omega t}| = |\cos(\omega t) - j\sin(\omega t)|$, $|\cos(\omega t) - j\sin(\omega t)| = \cos(\omega t)^2 + \sin(\omega t)^2 = 1$, $\left|e^{-\mathrm{j}\omega\tau}\right||P(\omega_x,\omega_y)| = |P(\omega_x,\omega_y)|$

(ii) Show by way of example how the Fourier transform of an image will rotate when the image itself is rotated.

Suppose:

$$\begin{gathered}\text{image } P = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}, \text{ image } P^T = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \\ F(P) = \sum_i \sum_j P_{ij} e^{-j\left(\frac{2\pi}{N}\right)\cdot(ui+vj)}\end{gathered} \tag{1}$$

$$\begin{aligned}&F(P(0,0)) = 1+2+2+1 = 6\cdot \\ &F(P(0,1)) = 1+2e^{-j\pi}+1+2e^{-j\pi} = 2+4e^{-j\pi} \\ &F(P(1,0)) = 1+2+e^{-j\pi}+2e^{-j\pi} = 3+3e^{-j\pi}. \\ &F(P(1,1)) = 1+2e^{-j\pi}+e^{-j\pi}+2e^{-j\pi\cdot 2} = 2e^{-j\pi\cdot 2}+3e^{-j\pi}+1\end{aligned} \tag{2}$$

$$\begin{aligned}F(P) &= \begin{bmatrix} 6, & 2+4e^{-j\pi} \\ 3+3e^{-j\pi}, & e^{-j\pi\cdot 2}+3e^{-j\pi}+1 \end{bmatrix} \\ F(P^T) &= \begin{bmatrix} 6, & 3+3e^{-j\pi} \\ 2+4e^{-j\pi}, & e^{-j\pi\cdot 2}+3e^{-j\pi}+1 \end{bmatrix} \\ F(P)^T &= \begin{bmatrix} 6, & 3+3e^{-j\pi} \\ 2+4e^{-j\pi}, & e^{-j\pi\cdot 2}+3e^{-j\pi}+1 \end{bmatrix} \\ &F\left(p^\top\right) = F(p)^\top\end{aligned} \tag{3}$$

Therefore, the Fourier transform of an image will rotate when the image itself is rotated.

3.3 Describe the importance of the invariance properties of the Fourier transform, in the automated analysis of cloth and fingerprints.[6 marks]

The differing phase implies that, in application, the magnitude of the Fourier transform of a cloth, say, will be the same irrespective of the position of the cloth in the image (i.e., the camera or the subject can move up and down), assuming that the cloth is much larger than its image version. This implies that if the Fourier transform is used to analyze an image of one of cloth, to describe it by its spatial frequency, we do not need to control the position of the camera, or the object, precisely.

4 Question 4

4.1 Consider the connected component depicted by the solid black pixels below:

Ensuring you show all working, compute the central moments µ11, µ20 and µ02.[11 marks]

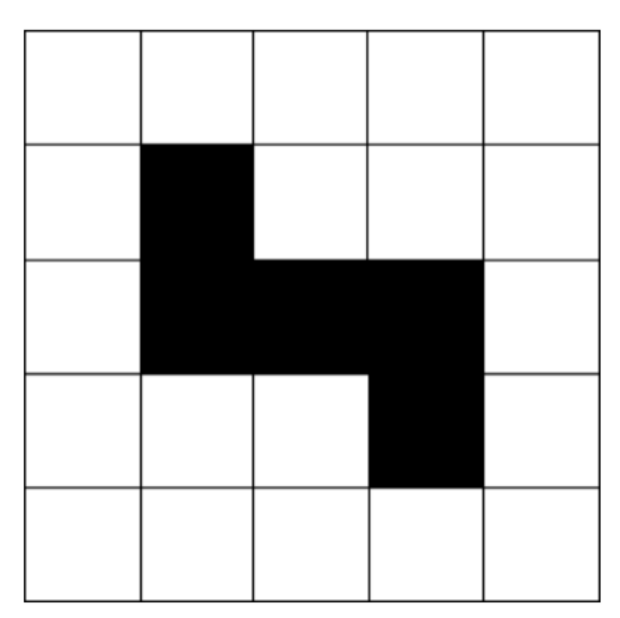


Figure 4: rotated image

$\mu_{11} = \sum_i (x_i - \bar{x})^1 (y_i - \bar{y})^1$
$\bar{x} = (1 + 1 + 2 + 3 + 3)/5 = 2$
$\bar{y} = (1 + 2 + 2 + 2 + 3)/5 = 2$
$\mu_{11} = \sum_i (x_i - \bar{x})^1 (y_i - \bar{y})^1 = 1 + 0 + 0 + 0 + 1 = 2$
$\mu_{20} = \sum_i (x_i - \bar{x})^2 = 1 + 1 + 0 + 1 + 1 = 4$
$\mu_{02} = \sum_i (y_i - \bar{y})^2 = 1 + 0 + 0 + 0 + 1 = 2$

4.2 Show why the central moments µ01 and µ10 are not useful as descriptors.[4 marks]

$\mu_{10} = \sum_i (x_i - \bar{x}) = x_1 + x_2 + x_3 + x_4 ... x_i - i\bar{x} = 0$
$\mu_{01} = \sum_i (y_i - \bar{y}) = y_1 + y_2 + y_3 + y_4 ... y_i - i\bar{y} = 0$
Because $\bar{x}$ and $\bar{y}$ is the means of x_i and y_i, µ01 and µ10 are always 0. Therefore µ01 and µ10 are not useful as descriptors.

4.3 Describe the process of creating a Point Distribution Model to describe the shape of a human face.[12 marks]

Point Distribution Models

- A few lectures ago, we saw how PCA could be applied to face images in the classic Eigenfaces algorithm. A Point Distribution Model (PDM) applies a similar process to a set of points representing a shape.
  - Taking the example of faces:
    - Sets of corresponding 2D points (i.e. points corresponding to the same physical feature [e.g. the tip of the nose]) are manually created from a set of *N* training face images (typically containing different facial expressions and poses).
      - The number of points is fixed at *M*; all *M* points must exist on all images!
    - An iterative process called Generalized Procrustes Analysis is used to align all the point sets from each image (so they all have the same size and rotation and are centred about the origin).
    - The mean shape is created by considering the average position of all of the aligned points that correspond to each other.
    - An *Nx2M* shape matrix is formed such that each row corresponds to one of the training images, and each pair of columns corresponds to the x and y coordinates of an aligned point. Corresponding points across the images are stored in the same pair of columns.
      - A good way of thinking about this, is that each column stores information on how the x or y ordinate of a point on a face can change.
    - PCA is applied to this matrix to generate a low dimensional basis that fundamentally can describe different facial expressions and poses.
      - Any shape vector can be projected against this matrix to generate a low-dimensional description (which encodes pose and expression).
      - A low dimensional description vector can be reconstructed back into a set of (x, y) coordinate pairs representing a face.

Figure 5: Image I. Numbers indicate pixel values.

4.4 Describe how you would turn a Point Distribution Model into an Active Appearance Model or Constrained Local Model.[6 marks]

Active shape models/Active appearance models/Constrained Local Models

- Active shape models (ASMs) and Constrained Local Models (CLMs) all take a point distribution model a step further, and incorporate models of what the image should look like around each point in the PDM.
    - In the simplest case, this is just a small patch of pixels (e.g. based on the average across the training images) about each point – this is a *template*.
    - The addition of this data allows the model to be fitted to an unseen image using an algorithm that iteratively updates the points to their best position (e.g. using template matching in a small window around the point), and updating the points to positions considered to be plausible by the PDM, by projecting them into the low dimensional space and then reconstructing them in the original space.
        - Essentially each point tries to move to its local optimum, whilst the PDM constrains all the points to still looking like a face!
- Active Appearance Models (AAMs) do the same thing as ASMs and CLMs, but instead of local features around each point, they try to jointly optimise the global appearance of the face (based on an Eigenfaces modal) against the PDM.

Figure 6: Image I. Numbers indicate pixel values.

5 Question 5.

This question refers to the following 7 representation of a black and white image.

0.0	0.0	0.0	0.0	1.0
0.0	0.0	0.0	1.0	1.0
0.0	0.0	1.0	1.0	1.0
0.0	0.0	0.0	1.0	1.0
0.0	0.0	0.0	0.0	1.0

Figure 7: Image I. Numbers indicate pixel values.

5.1 State the two kernels used for the 3×3 Sobel operator. Compute the 3×3 horizontal and vertical gradient images over the central portion of the image shown in Figure 8 by convolving the image with the Sobel operators.[10 marks]

The horizontal and vertical Sobel operator are shown in Fig. 8.

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} [1\ 2\ 1] \implies$$

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

(a) horizontal Sobel

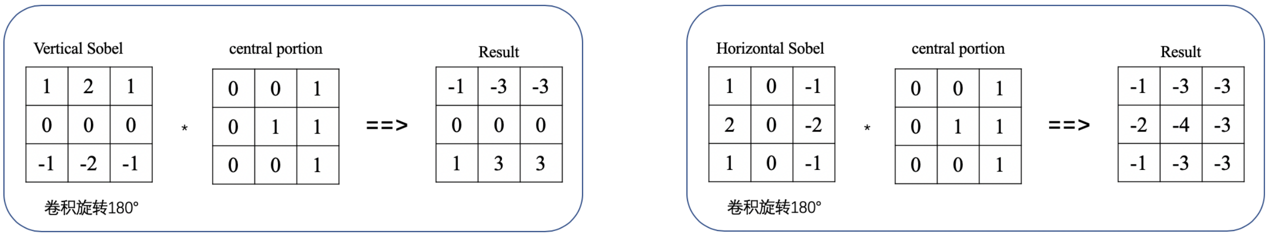
$$\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} [-1\ 0\ 1] \implies$$

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

(b) vertical Sobel

Figure 8

The results of the convolution are shown in Fig. 9.



(a) vertical Sobel (b) horizontal Sobel

Figure 9

5.2 State the formula for the Structure Tensor or Second Moment Matrix. Compute the Structure Tensor over the 3×3 window in the centre of the image in Figure 7. Show your working. [8 marks]

The solution is shown in Equation (1).

$$\mathbf{M} = \begin{bmatrix} \sum_w \left(I_x\left(x_i, y_i\right)\right)^2 & \sum_w I_x\left(x_i, y_i\right) I_y\left(x_i, y_i\right) \\ \sum_w I_x\left(x_i, y_i\right) I_y\left(x_i, y_i\right) & \sum_w \left(I_y\left(x_i, y_i\right)\right)^2 \end{bmatrix} = \begin{bmatrix} 67 & 0 \\ 0 & 38 \end{bmatrix} \tag{4}$$

5.3 State the formula for the Harris and Stephens corner response function. Assuming k = 0.04, compute the value of the response function based on the Structure Tensor previously calculated. Show your working. [8 marks]

The solution is shown in Equation (2).

$$R = \det(\mathbf{M}) - k\,\mathrm{trace}(\mathbf{M})^2 = 2105 > 0 \tag{5}$$

5.4 Briefly describe the rationale for the Harris and Stephens corner response function. Describe what different values of the corner function mean. [7 marks]

Solution: The eigenvalues and vectors tell us the rates of change and their respective directions. The formula is derived as follows.

$$\begin{aligned} E(x,y) &= \sum_W \left[I\left(x_i, y_i\right) - I\left(x_i + \Delta x, y_i + \Delta y\right)\right]^2 \\ &= \sum_W \left(I\left(x_i, y_i\right) - I\left(x_i, y_i\right) - \left[I_x\left(x_i, y_i\right) \quad I_y\left(x_i, y_i\right)\right] \begin{bmatrix} \Delta x \\ \Delta x \end{bmatrix}\right)^2 \\ &= \sum_W \left(-\left[I_x\left(x_i, y_i\right) \quad I_y\left(x_i, y_i\right)\right] \begin{bmatrix} \Delta x \\ \Delta x \end{bmatrix}\right)^2 \\ &= \sum_W \left(\left[I_x\left(x_i, y_i\right) \quad I_y\left(x_i, y_i\right)\right] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}\right)^2 \\ &= \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} \sum_w \left(I_x\left(x_i, y_i\right)\right)^2 & \sum_w I_x\left(x_i, y_i\right) I_y\left(x_i, y_i\right) \\ \sum_{w_i} I_x\left(x_i, y_i\right) I_y\left(x_i, y_i\right) & \sum_w \left(I_y\left(x_i, y_i\right)\right)^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \\ &= \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \mathbf{M} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \end{aligned} \tag{6}$$

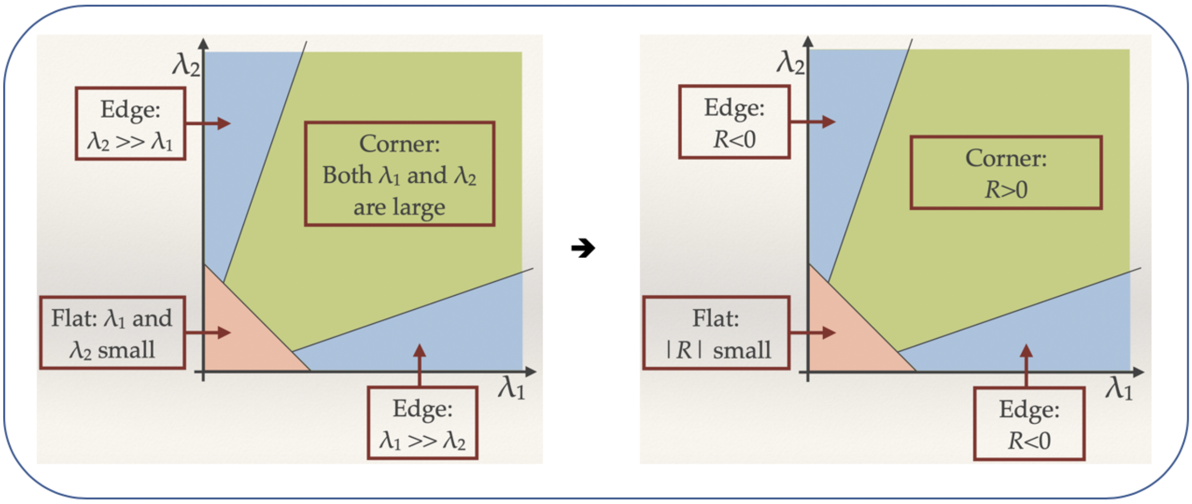


Figure 10: Principle