day / date:

- It is the use of digital computer to process digital image
- enhancement, medicine, GIS, Industrial inspection, Human computer interaction
- sign language, signature, robotics, OCR

high level : computer vision → input is attribute/image, output is understanding.

mid level : input is image, output is attribute (image segmentation)

low level : image processing → input and output are both images.

Pattern Recognition

- Patterns are structures in data that can be categorized.
- Recognition includes
  → learns from data
  → identifying patterns even if partially visible.
  → recognize patterns which are from different angles and shapes
- Classification (supervised) → use labeled training data to categorize images.
- Clustering (unsupervised) → groups similar images without predefined labels.
- Machine learning constructs and studies systems that learn from data.
- Pattern recognition focuses on recognition patterns in data.

Feature Extraction

class → a collection of similar objects.

interclass ~~variability~~ similarity → similarities between objects of different classes.

e.g. number "0" and alphabet "O" in handwritten text recognition.

intraclass variability → objects belonging to the same class may have difference in shape, size, color etc.

e.g. different images of cats may look different due to breed variations.

features → measurable properties of an image used for learning

feature vector → set of features used for learning/recognition

dimensionality → number of features used.

day / date:

Image Acquisition

- capturing image using different sensors.

Image Representation

- discrete representation of images
- each pixel has an intensity value (0 for black, 1 for white)

Image Source and Electromagnetic Spectrum

sources:

→ electromagnetic band imaging (gama, x-ray, UV, infra-red, radio, micro)

→ non em imaging (ultrasonic, electron microscopy, synthetic images)

electromagnetic waves

→ a stream of mass less particles each travelling in a wave like pattern, moving at speed of light

- visible light can be sensed by human eye.

Image Processing Operation

1. Noise Removal → • unwanted variations in image pixel values, which degrades image quality.
   • replace each pixel value with the average of its neighbouring pixel values to reduce noise, but blurs edge
2. Noise Reduction → minimizes it while preserving image details
3. compression → reduces file size of an image while maintaining as much visual quality.
4. enhancement → increasing brightness, enhancing sharpness, reduce noise
5. contrast adjustment → difference between darkest and brightest areas in image.
6. edge detection → object boundaries by detecting changes in intensity.
7. restoration → corrects image distortions e.g low lighting, sensor defects.
8. inpainting → reconstructs missing or damaged parts of an image.
9. special effects → applying artistic transformation.
10. image mosaic → composition of multiple images stitched together to form a larger image.

day / date:

Image Formation

- projection of 3D → 2D
- loss of 1D, but understand geometry and maths.

Visual Perspective

- use of geometric construction line
- illusion → stimulus is generated on the retina of our eye is the same as that will be generated by looking at a 3D scene.

Geometry of perspective

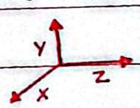
- specular reflection
  - ↳ angle of incidence = angle of reflection. e.g. mirror.
- lambertian reflection
  - ↳ light scattered equally in all directions

$$I = I_0 \cos(\theta)$$

I_0 incident light intensity

- pinhole camera
  - ↳ lensless camera
  - ↳ produces dim and inverted images, can be improved by allowing more light
- camera obscura
  - ↳ rays of light pass through a small hole into a dark space form an image where they strike a surface, resulting in inverted and reversed image

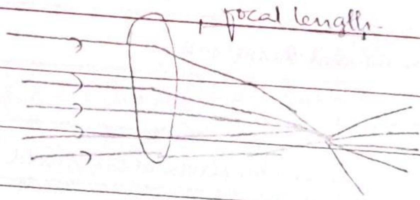
$$\frac{y}{z} = \frac{Y}{f}$$



$$\frac{x}{z} = \frac{X}{f}$$

- consequence of projection
  - → parallel lines appear to converge, shapes like circle may appear as ellipse.
- perspective projection
  - → representing 3D → 2D
  - → angles are not preserved
  - → conics are transformed.

Lenses



- focal length affects brightness

  F number = f / lens diameter (focal length) → $F = \frac{f}{\varnothing}$

- a smaller focal number, means a larger aperture, allowing more light to reach the camera sensor.
- semi-conductor imaging chip
  - → large array of light sensitive elements
- for focus adjustment
  - → move imaging plane, change distance between lens and image plane.
- light field camera
  - → captures intensity and direction of light rays.
  - → can focus after image capture.

Digital Images

- found on facebook, instagram
- cameras are everywhere → phone, laptop
- storage is cheap, making digital photography more accesible.
- transition from chemical imaging to digital
- a digital image is a grid of squares, each contains a value in a pixel.
- a pixel is also called Picture elements.
- color images have 3 values per pixel (Red, Green, Blue)
- monochrome have 1 value for pixel.

Image File Format

- 3 components (header, metadata, image pixels)
- JPEG image uses EXIF as a container.
- uses JPEG compression to store the image data efficiently.
- unsigned 8-bit integer (uint8) most common pixel representation.
  - ↳ (0-255) where 0 is black
  - ↳ n-bit integer, max value is $2^n - 1$
- uint 8 is efficient for integer arithmetic, and minimizes memory while keeping sufficient image detail

day / date:

Sampling and Quantization

1. sampling

→ how many pixels are captured (spatial resolution), higher pixel count means better detail.

→ picture is smooth, if sampling is maximum.

2. quantization

→ how many shades are used (digitizing color / intensity levels)

→ continuous color mapped to a finite number of colors.

→ more quantization = better image quality.

→ 8 bit quantization = 256 intensity level → 2^n

→ 1 bit quantization = binary image (black/white)

sampling → no. of sensors on a CCD array = no. of pixels = no. of samples taken

quantization → no. of partition = no. of gray levels

Image Representation

• image intensity levels depend on the number of bits per pixel

$L = 2^k$ where k = no. of bits $[0, L-1]$

• number of bits required to store an image

$b = M \times N \times k$ M, N = dimensions (width, height) k = bits per pixel

Spatial and Intensity Resolution

• spatial resolution determined by how sampling was carried out

• 3 measures for image size / resolution

↳ pixel count, physical size, resolution

• more intensity levels, finer level of detail in an image.

Gray level Reduction

• high bit to low bit image

• use of thresholding or logical shift operators.

• 8 bit to 5 bit → 1) 8-5 = 3

2^8 to 2^5 bit → 2) divide each element with 2^3

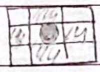
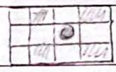
DPI → no. of printed dots per inch

PPI → pixels per inch

day / date:

- In histograms, dark images representate concentration on left
- bright images represent concentration on right

Neighbourhood of Pixels.

- adjacent pixels of a particular pixel
  - $N_4(p)$ → left, right, up, down 
  - $N_D(p)$ → diagonal position 
  - $N_8(p)$ → all surrounding pixels 

Connected Component labelling

- establishing boundaries of objects and components.
- two pixels p and q are connected if
  - → they've adjacent
  - → gray level satisfy a specified criteria
- 4 connectivity → connected through N_4 neighbours.
- 8 connectivity → m-connectivity, remove common co-ordinates that are co
- a path is a sequence of pixels connecting two points.
- a group of pixels that share the same intensity.

Distance metrics

- euclidean distance $D_e(p,q) = \sqrt{(x-s)^2 + (y-t)^2}$ radius r, centre (x, y)
  straight line
- manhattan distance $D_4(p,q) = |x-s| + |y-t|$
  grid-based
- chessboard distance $D_8(p,q) = \max(|x-s|, |y-t|)$ square centre p(x, y)
  diagonal movement

day / date:

Image Resizing or Interpolation

- to zoom in/out, enhance resolution
- enlarging an image

  4 pixels → 8 pixels

  to zoom 4x → fill in every 4th pixel in every 4th row with original pixel value.
- nearest neighbour interpolation

  → copies nearest pixel value   duplicates rows and columns.
- bilinear interpolation

  → uses weighted average of 4 nearest pixel
  - take adjacent 2 values. for rows side by side,
  - for columns take up and below.
- shrinking

  → row - column deletion.
- bicubic interpolation

  → uses 16 neighbouring pixels.

Steganography

- writing hidden messages by modifying pixel values
- store image data in LSB of another image.
- for 6-bit → 8-bit

$$\boxed{b\,|\,b\,|\,b\,|\,b\,|\,b\,|\,b} \rightarrow \boxed{\underbrace{b\,|\,b\,|\,b\,|\,b\,|\,b\,|\,b}_{b = 0 \text{ or } 1}\,|\,\underbrace{0\,|\,0}_{\text{always } 0}}$$

steps:

image 1 → right shift 2 → left shift 2

image 2 → right shift 6 ← +

- to recover second image, simply <u>left shift the combined image by 6 bits.</u>

day / date:

Monadic Image Processing

- each output pixel is a function of corresponding input pixel.
- example: if we apply a brightness increase function, each pixel modified independently.
- operations are performed sequentially. (one pixel at a time)
- functions like sqrt(x) or log(x) applied to each pixel to transform the image.
- example: sqrt(x) will make darker regions brighter
- changing data types, brightness, thresholding,
- brightness: adding or subtracting a constant value from pixel intensities
- contrast: stretching the range of pixel intensities → $f(x) = 2x$
- saturation: adjusting the intensity of colors in an image
- negative: inverting pixel values → for 8 bit → 255 levels so 255 - pixel value.

Histogram Normalization

- enhances fine details in an image.

example: if a grayscale image has more dark pixels, normalization can redistribute the intensity values to improve visibility.

Gamma Correction

- adjusts brightness non-linearly to match human vision perception. $f(x) = x^{\gamma}$
- applied twice (2):
  - → when capturing an image (camera sensor)
  - → when displaying an image (monitor)

$L' = V^{\gamma}$

luminance — voltage

- gamma encoding (from camera)
  - → captures a realworld luminance and applies a gamma correction

$V = L^{1/\gamma}$

- gamma decoding

$L' = V^{\gamma}$

- since camera already encoded gamma correction using $V = L^{1/\gamma}$, when the display applies it becomes

$L' = (L^{1/\gamma})^{\gamma} = L$

- output luminance matches original scene

day / date:

Image Thresholding

- converts an image to binary form
- example. threshold = 128, pixels above become white (255) and below become black (0)
- achieved through logical operators

Histogram

- shows pixel intensity distribution
- gray-scale histogram : 256 bins (intensity 0-255)
- color histogram : separate for red, green, blue.

example. a dark image will have a histogram concentrated on the left

- $\frac{R+G+B}{3}$ and 1 for grayscale.

example. for low contrast image, histogram is narrow and centered toward the middle of gray scale.

Histogram Equalization

- reassigns the intensity values of pixels in the input image
- uses CDF for normalization

Image Enhancement

- modify pixels in an image to enhance or correct features
- spatial domain, transform domain

Types of image enhancement

1. point/pixel operation

   → output depends only on input pixel at (x,y), monadic and diadic

2. local operation

   → output depends on neighbourhood.

3. global operations

   → output depends on all pixels in the image, histogram equalization.

day / date:

Point processing

- each pixel value is transformed based on its value alone, independent of its neighbours.
- linear, logarithmic, power law.

Power law transformations

- $f(x) = x^{\gamma}$
- $\gamma < 1 \rightarrow$ expands dark pixel values and compresses bright ones.
- $\gamma > 1 \rightarrow$ compresses dark pixel values and expands bright ones.
- $\gamma = 1 \rightarrow$ no transformation.
- for $\gamma = 0.5$, dark regions become brighter.
- if γ becomes too low, the image may lose contrast and appear washed out.
- if $\gamma = 5.0$, high contrast, some regions become too dark. (4.0 is suitable)

Thresholding

- segment an object from background based on a threshold value.

example: threshold = 128, all pixels with values > 128 are set to 1 (white)

Logarithmic Transformation

$$s = c \times \log(1 + r)$$

- maps a narrow range of low input values to a wider range of output values.

Contrast stretching

- Increases dynamic range of gray levels for low contrast images.
- increasing pixel values between 50–150, to 0–255, improving visibility.

$$s = T(r) = (r - r_{min}) \left[\frac{L-1}{r_{max} - r_{min}} \right]$$

- provides a more natural enhancement compared to histogram equalization

day / date:

Diadic Image Processing

- each output pixel is a function of two corresponding input pixels.
- example: adding 2 images pixel by pixel
- adding, subtracting, multiplying, sqrt

combine image, enhance brightness. | detect changes | masking operations.

- both operands must have same size.
- result is unsigned, cannot be negative.
- in saturation if a sum exceeds the range, then assign max value of range.

example. sum = 300, make it 255

difference = -10, make it 0

- scale an image to a range

$$f_m = f - \min(f) \quad \text{then,} \quad f_s = K\left[f_m / \max(f_m)\right]$$

current pixel - min

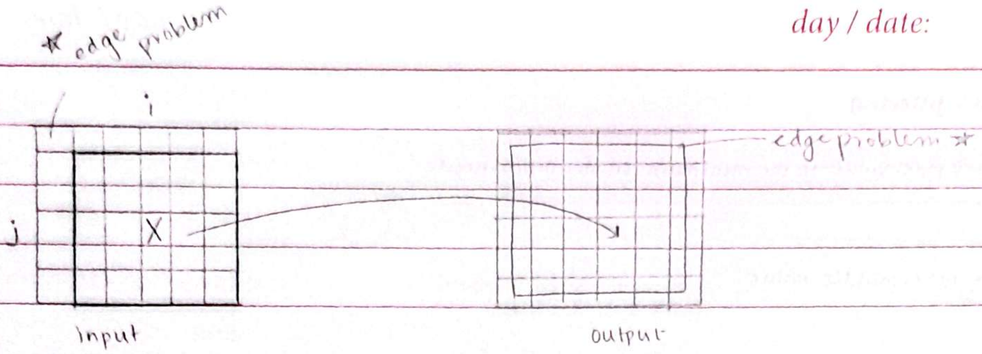
an image whose minimum is 0

scaled image f_s with values in range [0 K]

Green screen effect

- green pixels in one image are replaced with pixels from another image
- creating a binary mask to identify green pixels.
- green screen $\times$ mask matrix → } → add both
- background $\times$ binary matrix → }
- addition, subtraction, multiplication, division
- logical operation on binary images.

day / date:



Spatial filtering

- a small kernel is applied over an image to modify pixel values
- each output pixel is a function of the corresponding input pixel and its neighbors.

Why is the window always odd?

- to ensure there is a clear center pixel for transformation.
- window width is $2h+1$

    where h : pixels from the center to the edge

- larger windows (21 x 21) produce more significant smoothing but may lose more detail.
- smaller windows preserve details but are more sensitive to noise.

* Edge problem

- either discard edges i.e. don't calculate
- zero padding : surround with 0 pixels.
- average padding : average of surrounding pixels.
- pixel replication : surround with same pixels.
- when applying a filter

r	s	t
u	v	w
x	y	z

X original image :

a	b	c
d	e	f
g	h	i

to find centre → e

→ $ar + bs + ct + du + ev + fw + gx + hy + iz$

Properties of convolution

- commutative, associative, distributive, linear

$I = G \otimes (G \otimes I)$

Smoothing Spatial filters

- in box filters, assign equal weight to all pixels

$\frac{1}{9} \times$

1	1	1
1	1	1
1	1	1

- weighted average filters give more weight to pixels closer to centre.

$\frac{1}{16} \times$

1	2	1
2	4	2
1	2	1

1	1	2	1	1
1	2	2	2	1
2	3	4	3	2
1	2	3	2	1
1	1	2	1	1

* spatial masks are used and convolved over the entire image

Order statistic filtering

min → set pixel value to the minimum in neighbourhood.

max

median → takes middle value.

130	130	255
124	200	123
121	126	0

→ median = 0, 120, 121, 123, <u>124</u>, 126, 130, 200, 255

↳ replace with center pixel

- effective for salt and pepper noise, preserves edges when median filter applied.
- a noisy image retains more detail.
- applying median filter multiple times removes more noise but may also blur the image.

Convolution vs Correlation

- convolution involves flipping of kernel
- for symmetric kernel, both produce the same result

$e_{processed} = az + by + cx + dw + ev + eu + hr + gs + tf$

- convolution filter →

r	s	t
u	v	w
x	y	z

a	b	c
d	e	e
f	g	h

- flipping kernel both horizontally and vertically.

Gaussian filtering.

- used to blur image and remove noise and detail.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}} \qquad \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-u^2-v^2}{2\sigma^2}}$$

- assumed to have mean = 0, where σ = standard deviation.
- larger σ values produces wider gaussian kernels, resulting in more blurring.

Designing Gausian filter

$$G(u,v) = \frac{1}{2\pi\sigma^2} e^{\frac{-(u^2+v^2)}{2\sigma^2}}, \quad \sigma = 0.6, \quad size = 3\times3$$

step1 : solve for σ → $\dfrac{1}{2\times3.14\times0.16\times0.16} = \dfrac{625}{1413}$

value decrease ← → value increase

X =

-1	0	1
-1	0	1
-1	0	1

Y =

-1	-1	-1
0	0	0
1	1	1

value decrease ↑ value increase ↓

→

-1, -1	0, -1	1, -1
-1, 0	0, 0	1, 0
-1, 1	0, 1	1, 1

day / date:

step2. solve for $\dfrac{-(x^2+y^2)}{2\sigma^2}$

$$\begin{matrix} -2.7778 & -1.3829 & -2.7778 \\ -1.3829 & 0 & -1.3889 \\ -2.788 & -1.3829 & -2.788 \end{matrix}$$

step3: gaussian kernel → $\dfrac{625}{1413} e^{\square}$ → matrix of 3x3

step4. convolve on an img

$$\begin{matrix} 72 & 68 & 88 & 159 \\ 69 & 66 & 87 & 162 \\ 70 & 66 & 83 & 161 \\ 70 & 66 & 78 & 154 \end{matrix} \times \begin{matrix} 0.0275 & 0.1102 & 0.0275 \\ 0.1102 & 0.4421 & 0.1102 \\ 0.0275 & 0.1102 & 0.0275 \end{matrix}$$

step5: after multiplying above 2 matrix, add up the values in vector = 94.9269 (centre pixel)

centre has highest intensity i.e. 0.4421 and 38.4624

- a noisy image becomes smoother, blurred
- blurring effect is to make something less clear or distinct.
- low pass filtering is also known as blurring or smoothing
  - ↳ removes high-frequency noise from an image.
  - ↳ gaussian filter.
  - ↳ blurring often used before thresholding or edge detection.

Rank filtering

- assigns specific order based value from a pixel neighbourhood.

$$\begin{matrix} 120 & 123 & 125 \\ 121 & 125 & 0 \\ 121 & 124 & 132 \end{matrix}$$

mean = 110

rank-5

median = 0, 120, 121, 121, <u>123</u>, 124, 125, 125, 132

↳ rank 9 (minimum) ↳ rank 1 - max

after convolving.

a) $\frac{1}{4}\begin{bmatrix} 0 & 2 & 0 \\ 1 & -4 & 1 \\ 1 & 6 & 3 \end{bmatrix}$

b) $K3 = \begin{bmatrix} 1 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 1 \end{bmatrix}$

day / date:

Sharpening filters

- highlights fine detail
- remove blurring from image
- highlight edges

→ spatial differentiation

- an image shows a section of grayscale (marked from A to B)
- a dashed orange line represents a profile line across the image where gray levels are analyzed.
- x-axis represents image strip, while y-axis shows the corresponding gray levels of the pixels.

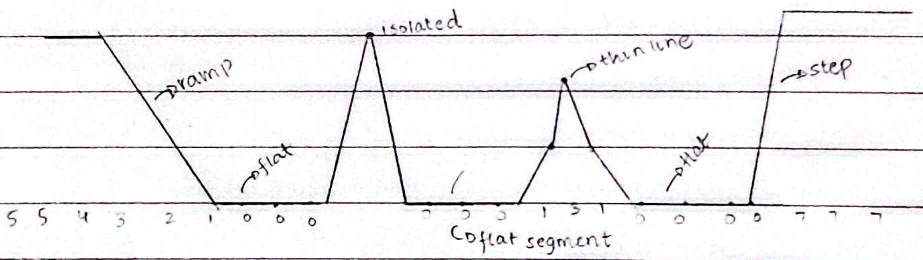
Ramp: gradual increase or decrease in pixel intensity, representing transition between two regions. An edge with a smooth transition

Isolated Point: sharp change in intensity at a single pixel; small change

Thin line: intensity changes abruptly over a small number of pixels, can be interpreted as thin edge or thin line.

Flat segment: intensity remains constant

Step: sudden change in intensity, representing a clear edge between 2 regions with contrasting intensities.



1st derivative

- measures immediate change in intensity $\frac{\partial f}{\partial x} = f(x+1) - f(x)$ +ve → dark to light

Co after Co current

2nd derivative

- measures curvature or rate of change of the 1st derivative. $\frac{\partial^2 f}{\partial^2 x} = f(x+1) + f(x-1) - 2f(x)$

Co after Co before Co current

- stronger response to fine detail.

SF: laplacian filter

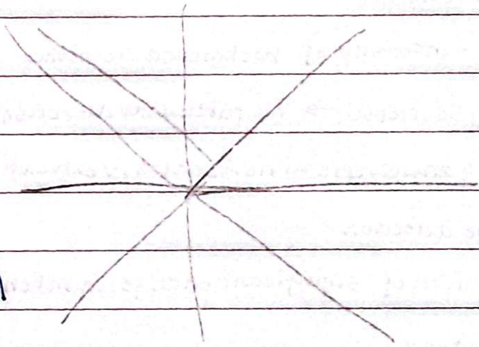
$$\nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \qquad \frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y)$$

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y)$$

- useful in edge detection for finding sharp changes
- for image enhancement subtract laplacian from original image.

$$g(x,y) = f(x,y) - \nabla^2 f$$

$$f(x,y) + \nabla^2 f$$



1^{st} derivative and gradient : SIMPLEST OPERATOR

$$G_x = \frac{\partial f}{\partial x}, \quad G_y = \frac{\partial f}{\partial y}$$

$$\text{Magnitude } |\nabla f| = \sqrt{G_x^2 + G_y^2} \simeq |G_x| + |G_y|$$

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

$$\frac{\partial f}{\partial y} = z_8 - z_5 \qquad \frac{\partial f}{\partial x} = z_6 - z_5$$

$$|\nabla f| = \sqrt{(z_6 - z_5)^2 + (z_8 - z_5)^2}$$

put it in this formula.

Prewitt kernel : PREWITT OPERATOR

- horizontal and vertical edge detection.

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

$$\frac{\partial f}{\partial y} = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3)$$

$$\frac{\partial f}{\partial x} = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)$$

- extract horizontal and vertical edges (∂x)

$$\frac{\partial f}{\partial y} = \begin{matrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{matrix} \qquad \frac{\partial f}{\partial x} = \begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix}$$

day / date:

Sobel kernel : SOBEL OPERATOR

$$\frac{\partial f}{\partial y} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad \frac{\partial f}{\partial x} = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & 1 \end{bmatrix}$$

extract horizontal   extract vertical

$$\nabla f \approx |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)|$$

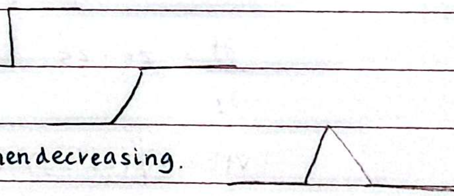
$$\frac{\partial f}{\partial y} = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \qquad \frac{\partial f}{\partial x} = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

Line detection

- intensity of background ie either much higher or lower than intensity of line pixels.
- isotropic → no particular direction
- anisotropic → horizontal, vertical, $-45°$, $+45°$

Edge detection

- area of significant change in intensity.
- types
  - → step : abrupt intensity change
  - → ramp : gradual intensity change
  - → roof : intensity change increasing then decreasing.



- image smoothing, edge localization.
- $|G| = \sqrt{G_x^2 + G_y^2}$
- $\alpha = \tan^{-1}\left(\frac{G_y}{G_x}\right)$
- for 2D

$$G_x = \frac{\partial f}{\partial x} = f(x+1, y) - f(x, y) \qquad G_y = f(x, y+1) - f(x, y)$$

canny edge

1. gaussian.
2. derivative.
3. gradient
4. non-max
5. threshold.

vertical edge

horizontal edge.

dark to bright → negative.

bright to dark → positive

day / date:

Template Matching

- find a particular pattern in an image
- similarity measure
  - → sum of absolute difference (SAD)
    - ↳ lower the value (0 indicates exact match)

$$s = \sum |I_1[u,v] - I_2[u,v]|$$

    - ↳ 0 to 1
  - → sum of squared differences (SSD)

$$s = \sum (I_1[u,v] - I_2[u,v])^2$$

  - → zero mean normalized cross correlation (ZNCC)
    - ↳ values range from -1 to +1, with +1 indicating perfect similarity
      0 means not correlated.
      -1 means they're opposite.

KNN

- classification is supervised i.e. known categories.
- clustering is unsupervised i.e. learning categories
- KNN used as a classifier for pattern recognition
- finds nearest neighbour of a given test pattern.
- algorithm
  - → select the number k of neighbours.
  - → euclidean distance from test sample to all training sample.
  - → take k-nearest neighbour as per the calculated euclidean distance.
  - → count number of data points belonging each category for voting.
  - → assign new data points to that category for which number of neighbour is maximum.
- always choose odd value of k
- small value of k can be noisy and sensitive to outliers.
- large values of k make model very robust, can be too smooth and less sensitive to small differences
- consider feature vectors of the templates to look for closest templates instead of exact matches

day / date:

Image Segmentation

- groups similar pixels into regions based on specific criteria.
- used for detecting medical features, identifying objects in aerial
- 2 basic properties of grayscale.
  - → Discontinuity.
    - based on abrupt changes in gray-scale values
    - edges, lines, isolated point
  - → Similarity
    - thresholding, region growing, splitting based on pixel similarity.

Thresholding

- divides an image into 2 groups, foreground and background.
- assigns pixels to these groups based on a set threshold value.

1. Global thresholding.
   - single threshold value applied across entire image
   - useful when image has uniform illumination and the foreground and background have distinct intensity differences. turns a grayscale into black and white.
   - STEPS
     - → estimate an initial threshold (T)
     - → segment the image in 2 groups — add all pixel values in G1 and divide by # of G1 name it m1.
     - → compute average intensity values of 2 regions and adjust threshold. $T = \frac{1}{2}(m1 + m2)$
     - → repeat until convergence
   - multilevel thresholding
   - global thresholding fails at images having varying lighting conditions
2. Adaptive thresholding
   - calculates threshold for each pixel based on local neighbourhood around it.
   - handles varying illumination easily.
   - calculating mean or median within a neighborhood of pixels.
   - $T = \text{mean}$

     $T = \text{median}$

     $T = \frac{\max + \min}{2}$

KAGHAZ www.kaghaz.pk

CamScanner

5. create bins and draw numbers randomly between 0 and

day / date:

· Niblack algorithm

→ threshold at each pixel is determined by the mean and standard deviation of the pixel values in a local window (WxW)

$T = m + k \times s$

↓ mean ↓ niblack constant ↳ standard dev.

→ ideal for binarizing text in document

→ mostly $k = -0.2$ for text

→ document binarization converts text images to binary form.

→ especially when written in different fonts or intensities

· Sauvola's algorithm

→ adjusts the threshold using a dynamic range factor R

→ $T = m \times (1 - k(1 - \frac{s}{R}))$

↳ dynamic range of std

↓ control factor

→ for 8-bit grayscale, R is set to 128

→ k ranges from 0.2 to 0.5

Region-based segmentation

· divides image into regions that are homogenous based on a predicate (such as having similar intensity values)

· region growing is a technique where the segmentation starts from seed pixels and grows by adding neighbouring pixels that have similar properties.

STEPS

→ choose seed pixels

→ check neighbouring pixel for similarity and add them to the region.

→ repeat until no more pixels can be added.

· region splitting starts with the whole image as a single region. It then splits into smaller regions if condition is not satisfied.

· region merging follows to merge adjacent regions if they satisfy condition together.

Deep learning

- uses multi-layered neural networks to model complex data patterns.

Semantic Gap and challenges

- raw pixel data does not easily map to meaningful high-level concepts, such as objects or scenes.

$$300 \times 100 \times \underbrace{3}_{\text{RGB channels}}$$

- Viewpoint Variation → objects may appear differently depending on the angle.
- Illumination → variations in lighting can make objects hard to recognize.
- Deformation → objects may change shape, complicating recognition.
- Occlusion → Part of an object might be hidden, affecting classification.
- Background clutter → complex backgrounds can interfere with recognizing the object.
- Intra-class variation → differences within same class (e.g. breeds of dogs)

Convolutional Neural network

- CNN's by using convolutional layers are more efficient in image processing.
- MLP/ANN don't capture spatial relationships between pixels.

→ Image Reduction

- reduce images to smaller forms while preserving important features.

→ Automated Feature Engineering

- also learns to extract features, eliminating need for manual feature engineering.

Architecture

$$\underbrace{\text{Convolution} \rightarrow \text{Pooling} \rightarrow \text{Convolution} \rightarrow \text{Pooling}}_{\text{feature learning}} \rightarrow \underbrace{\text{Flatten} \rightarrow \text{Fully Connected} \rightarrow \text{Softmax}}_{\text{classification}}$$

Convolution

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 19 & 25 \\ 37 & 43 \end{bmatrix}$$

- the first convlayer is responsible for capturing the low level features such as edges, color, gradient
- with added layers, high level features are captured.

5. create bins and draw numbers randomly between

day / date:

Pooling

- reduces dimensionality of the input
- prevents overfitting, as we are removing some features from image
- max pooling also performs as a noise suppressant.

Flattening

- transforms multi-dimensional data from previous layers into a one dimensional vector

Fully Connected Layer

- convolutional and pooling layer, forms i-th layer of a CNN.
- these layers may be extended for capturing low level details if image is complex

Spatial Dimensions

output size $= \frac{N-F}{\text{stride}} + 1$ ∴ F = filter, N = rows x columns.

∴ to check if a stride is valid for a certain image size.

- zero padding involves adding zeros around the border of the input image so if we are padding then output size increases by 2 (rows and columns)

zero padding $= \frac{F-1}{2}$ $2 \times 2 + 1 = F$

- parameters : $w^T x + b \rightarrow$ filter height x filter width x input channels ~~Ebias~~ $\frac{(2 \times 2 + 1)}{2}$
- output size : $\frac{\text{input size} + 2 \times \text{Padding} - \text{Filter}}{\text{stride} + \cancel{1}} + 1$ $= \frac{32 - (5)}{2} = \frac{27}{2}$
- pool doesn't have parameters

KAGHAZ www.kagha

day / date:

DIGITAL IMAGE PROCESSING

IMAGE SEGMENTATION

- segmentation is based on
  - → discontinuity : detects edges, lines or points where image intensity changes abruptly.
  - → similarity : groups pixels with similar values. e.g. thresholding, region growing.

K-MEANS CLUSTERING

1. choose number of k clusters and randomly select the centroids of each cluster.
2. For each point:
   - → calculate each distance to centroid.
   - → assign it to nearest cluster.
3. recalculate centroids.
4. repeat until convergence

- how increasing k refines segmentation (more detailed clusters)

MORPHOLOGY

- refers to the structure and form of organisms.
- in image processing, it refers to techniques for shape analysis using set theory.

EROSION

- output is true (white) if all pixels in S are true (white)
- this operation shrinks foreground objects by removing boundary pixels.

DILATION

- output is true if any pixels in S are true (white)
- expands foreground objects

day / date:

OPENING AND CLOSING

- opening → erode then dilate, removes small objects and noise
- closing → dilation then erode, restores shape of larger objects

COMPLEX STRUCTURING ELEMENT

- using diagonal lines.
- defining them as a subset of pixels in a window where 0s indicate "don't care" and 1's are part of element
- only lines that are compatible are retained in the output.

DC	DC	S
DC	S	DC
S	DC	DC

HIT AND MISS TRANSFORM

Hit → All 1s in the SE must align with foreground pixels in the image, and all 0s in the SE must align with background pixels (0s).

Miss → if any 1 or 0 in the SE does not match the corresponding pixel in the image, it's a miss and no output is marked.

Fit → SE perfectly matches the image's pixel pattern at a given position.

SET THEORY

$a \in A$ a is an element of A.

$A = \emptyset$ null

$A \cap B = \emptyset$ disjoint if no common element.

$A \subseteq B$ = subset

A^c complement.

$A - B = A \cap B^c$ difference.

$\hat{B}$ reflection of set B.

$(A)_z$ translation of set A by point z.

day / date:

How many colors are there?

- Rainbow → 7
- crayola crayons → 120
- HTML named colors → 121
- X11/rgb.txt → 881
- english corpus → 4000

Red (G=0, B=0, R varies) : Red to yellow to white

Green (R=0, B=0, G varies) Green to yellow to white.

Blue (R=0, G=0, B varies) Blue to magenta to white.

Black, white, red, yellow, green, blue, brown.

↳ this order remains same for all languages.

COLOR PLANES / CHANNELS.

Red, Green, Blue.

Red + Green = Yellow

Red + Green + Blue = white.

Red + Blue = purple.

Blue + Green = cyan

WHERE THE COLOR COMES FROM?

- light reflected from objects enter our eyes and hit rods and cones ; chemical reaction.
- it reaches visual cortex in brain leading to different colors.
- our eyes have 3 types of cones cell, sensitive to RGB.
- each generate a different signal called tristimulus which we perceive as a particular color.
- some animals hav are dichromats, tetrachromats
- some have defective cones that cause color blindness.

LIGHT IS AN ELECTROMAGNETIC WAVE

- temperature and wavelength are inversely proportional.

PLANCK'S LAW = $E(\lambda) = \dfrac{2hc^2}{\lambda^5 \left(e^{hc/k\lambda T} - 1\right)}$

where h = plancks constant 6.626×10^{-34} J

c = speed of light 3×10^8 m/s

λ = wavelength.

k = boltzmann constant.

T = temperature.

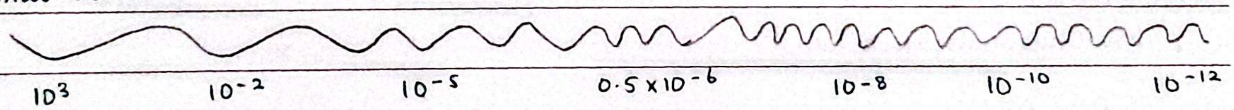
WIEN DISPLACEMENT LAW = $\lambda_{max} = \dfrac{2.8978 \times 10^{-3}}{T}$ m

- when wavelength is less, temperature increases
- color changes underwater

EM SPECTRUM

Radio $\rightarrow$ Microwave $\rightarrow$ Infrared $\rightarrow$ Visible $\rightarrow$ Ultraviolet $\rightarrow$ X-ray $\rightarrow$ Gamma.

- thermal radiation



- thermal imagery $\rightarrow$ detects heat
- infrared sensing.

BAYER FILTER PATTERN*

- camera sensors act like an 'artificial retina' to capture light
- each pixel gets one color filter; but doesn't know which one.
- 2 green on diagonals as because our eyes ~~see~~ see it best.
- for every pixel, we can measure one color and estimate 2 other.
- raw format $\rightarrow$ $n \times m$, later we convert to RGB

SEPARATING INTENSITY AND COLOR :

$r = \frac{R}{R+G+B}$, $g = \frac{G}{R+G+B}$, $b = \frac{B}{R+G+B}$

- $r+g+b = 1$
- chromaticity

→ upper triangle : value of blue is negative. which is not possible

steps :

1. R = load first channel .
2. $Y = R+G+B$
3. $r = R/Y$
4. $g = G/Y$
5. set threshold.

FINDING A COLOURED OBJECT

colorspaces . RGB, YCbCr, HSV

- HSV hue, saturation, value .

Color	Angle .
Red	0-60
Yellow	60-120
Green	120-180
cyan	180-240
blue	240-300
magenta.	300-360

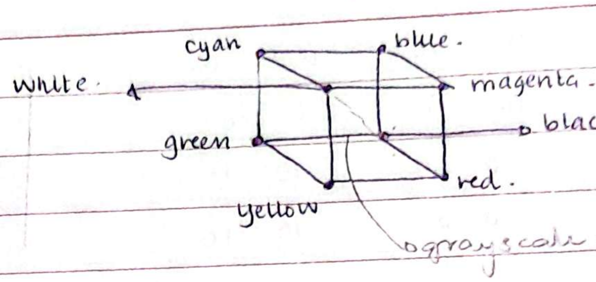
- saturation : amount of gray in color (0-100)

can be (0-1) where 0 is gray and 1 primary colour .

- value : brightness or intensity of the color .

day / date:

COLOR MODELS



- RGB and HSI
- cyan, magenta yellow are 3 corners
- black is origin.

- # of bits used to represent each pixel is referred to as the color depth. 2^{bit}

HSI COLOR MODEL

- Hue → color attribute that describes a pure color. what color?
- Saturation → how much a pure color is diluted with white light. how strong?
- Intensity → • achromatic notion that we have seen in grey level images.
  - extracted from RGB images.
  - can be determined by passing a plane perpendicular to the intensity axis and containing the color point.

- after passing intensity axis, if its near the top (white), color is bright.

★ CONVERTING FROM RGB TO HSI

$$h = \theta = \cos^{-1}\left[\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{1/2}}\right]$$

$$H = \begin{cases} \theta & \text{if } B \le G. \\ 360-\theta & \text{if } B > G \end{cases}$$

$$I = \frac{1}{3}(R+G+B)$$

$$S = 1 - \frac{3}{(R+G+B)}\left[\min(R,G,B)\right]$$

- assigning colors to grey values based on a specific criteria.