Chapter I Introduction to Digital Image Processing

Introduction

The pictorial representation of an object is known as formation of an image. A digital image is a 2D light intensity function $f(x,y)$ where 'x' and 'y' denotes spatial co-ordinates and value of function 'f' at any point (x,y) is propotional to the brightness (gray level) of an image at that point (x,y).

Digital image processing refers to the processing of digital image by using digital computer on the basis of digitization (sampling & quantization).

Digital Image Representation

A digital image is an image $f(x,y)$ that has been discretized both in spatial co-ordinates and in brightness.

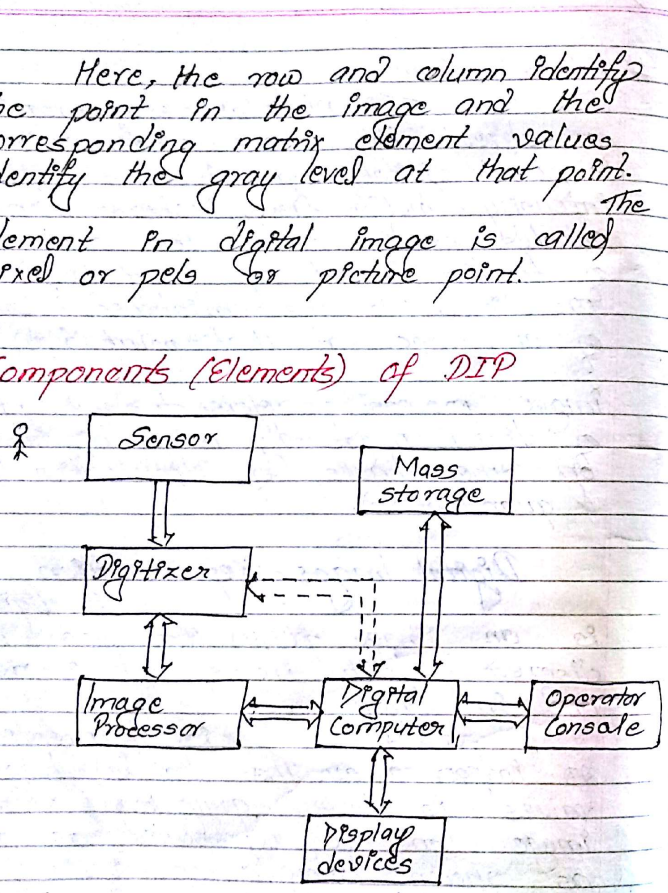
It is represented by 2D integer array. The digitized brightness values is called gray level values. Digital image can be represented as a matrix as shown below:

$$f(x,y) = \begin{bmatrix} f(0,0), f(1,0) \ldots \ldots \\ f(0,1), \ldots \ldots \\ \ldots \ldots \\ \ldots \ldots \quad f(x_{max}, y_{max}) \end{bmatrix}$$

Scanned by CamScanner

Here, the row and column identify the point in the image and the corresponding matrix element values identify the gray level at that point. The element in digital image is called pixel or pels or picture point.

Components (Elements) of DIP



(fig: Elements of digital Image Processing)

(i) Sensor:

It intercepts the energy propagating from the scene and transform it to produce an intensity image. Camera and Scanner are the example of photo sensor.

(ii) Digitizer:

The digitizer produce an image by the digitization technique. Normally, a digitizer converts an intensity image into numerical representation which is suitable for input into a digital computer

(iii) Image Processors:

It interface with digital computer to provide easy of programming. For this purpose we can use micro computer or mainframe computer. Specially, it does functions like image acquisition, low level processing, scaling & rotating

(iv) Digital Computer:

It always digitize the required image by the help of image pre-processors.

Scanned by CamScanner

(v) Mass Storage:
It is used to store digital image line by line. The storage devices may be magnetic tape, hard disk, etc.

(vi) Display devices:
It produce and shows a visual of numerical values stored in a computer. The example of display devices is CRT monitor, printer, etc.

(vii) Operator Console:
It control the full mechanism of image processing system.

Application of Image Processing

(i) Office Automation
Example: OCR (Optical Character Reader) and document Processing

(ii) Industrial automation
Example: Robotics, ~~Autom~~ automatic assembly, etc.

(iii) Biomedical automation
Example: ECG, EEG, X-Ray, etc.

(iv) Information Technology (IT)
Example: Videophone, fax, Video Conferencing

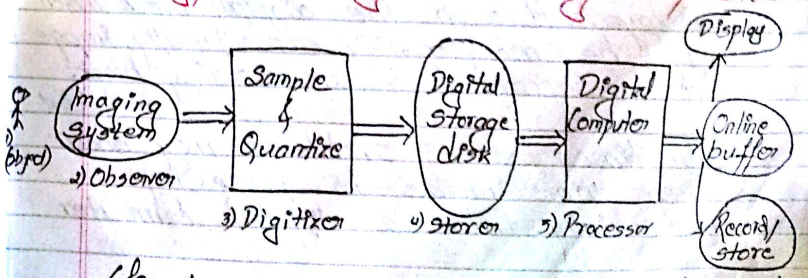
(v) Entertainment
Example: Videogame

(vi) Military applications
Example: target identification, missile guidance & detection

(vii) Criminology
Example: Biometric detection such as fingerprint detection, face detection, retina detection, etc.

Imp

A typical Image Processing Sequence



(fig: Image Processing sequence)

Scanned by CamScanner

In typical image processing sequence involves various components. The image of an object travels along the different blocks.

Firstly, the object is sensed by the system and image is formed. This image is digitized by the process of sampling and quantization.

Now, image is in the digital form. A digital image is stored in digital storage system. The digital computer is connected to digital storage disk where all the manipulation of digital image is done.

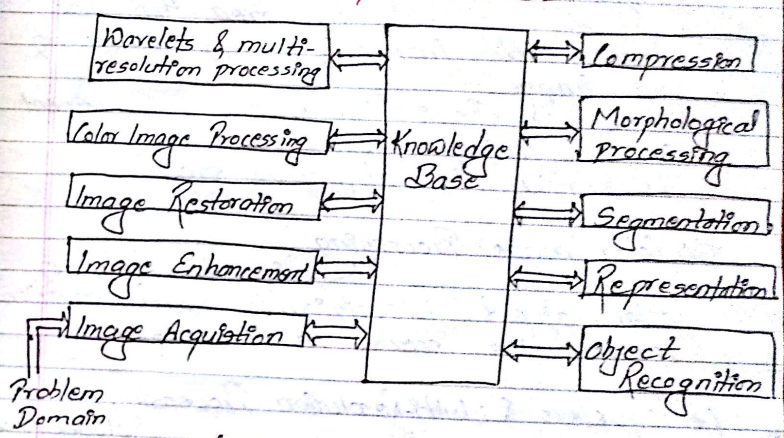
Here, also online buffer mechanism is available where the information are stored as a memory.

The image is displayed through the display block which may be monitor, TV, etc. Finally, image can be recorded in a recording block.

But sometimes some problems may occur when the image is processing sequentially into a digital computer.

V. Imp

Fundamental steps in DIP



(fig: steps in DIP)

(i) Image Acquisition

It is the first process in which pre-processing activities like scaling and rotating are done.

(ii) Image Enhancement

In this stage, the original image is processed in such a manner that the resultant image will be more suitable with compared to original image.

Scanned by CamScanner

(iii) Image Restoration

It is area that also deals with improving appearance of an image.

Basically, image enhancement is subjective process but image restoration is objective process.

(iv) Color Image Processing

With the help of color, object identification and extraction from the scene is processed.

(v) Wavelets & Multi-Resolution Processing

Foundation for the representing image in various degree of resolution deals to the wavelets and multi-resolution.

Sometimes, it is also used for image compression.

(vi) Compression

It deals with the technique for reducing the storage required to save an image to transmit it.

(vii) Morphological Processing

It is the combination of structure. So, it deals with tools for extraction of an image component that are useful in representation and description of shape.

(viii) Segmentation.

It deals with the partition of an image into different parts and objectives

(ix) Representation

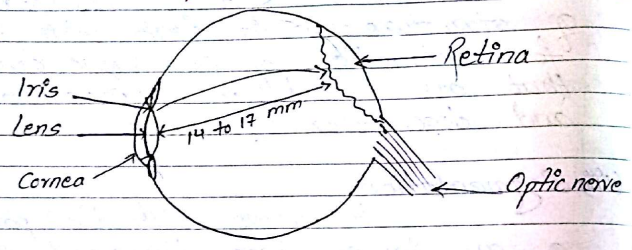
After an image has been segmented into sub-regions then the resultant segmented pixels is represented and described in a suitable form for further analysis.

(x) Object Recognition

It process that the level to an object based on its descriptors.

Scanned by CamScanner

Elements of Visual Perception



(fig: Cross section of human eye)

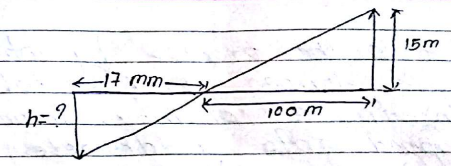
Image analysis plays central role in the image processing. Therefore the study of human visual perception is important which is represented as cross-section of human eye is as in given figure above.

The human eye nearly a sphere with an average diameter of 14 mm to 17 mm.

The lens of the eye is flexible than the ordinary optical lens.

Let the observer is looking at a tree 15 m high at distance of 100 m. Find the height of image retrieved ?

Sol



We have,

$$\frac{h}{17\,mm} = \frac{15\,m}{100\,m}$$

$$h = \frac{15 \times 17}{100}$$

$$= 2.55\,mm$$

∴ Height of retrieved image (h) = 2.55 mm

1 Imp

Sampling And Quantization
[Digitization]

A physical image is changed into digital image by performing sampling and quantization. This process is also known as digitization.

Sampling:

It is the process where measurements are taken at regular space intervals. Digitizing the spatial co-ordinate values (x, y) is called sampling.

Scanned by CamScanner

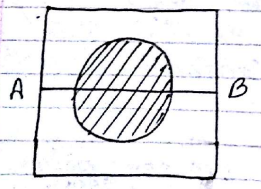
It is the process by which image is formed over a patch (domain) in a continous domain is mapped into a discrete point with integer co-ordinates.

Hence, sampling can be defined as a selection of discrete location in the continous 2D space.

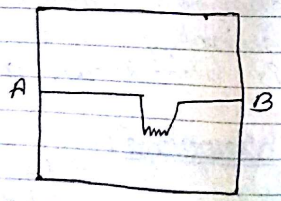
Quantization:

It is the process of mapping the measured intensity to one of the finite number of discrete levels.

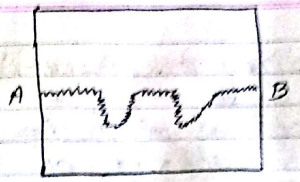
Hence, digitizing the amplitude values is called quantization or gray level quantization.



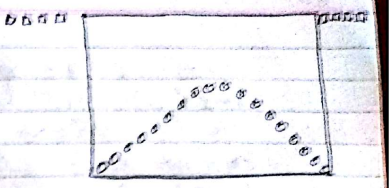
(i) Continous type



(ii) Ideal scan line A to B without noise

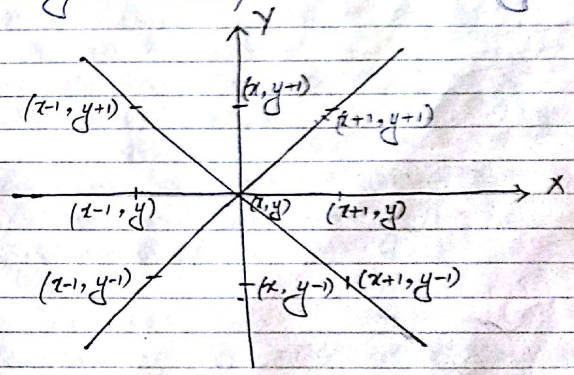


(iii) Scan line A to B in image with noise



(iv) Sampling & Quantization

Neighbours Of Pixel
[Arrangement of Pixel in Image]



(fig: neighbourhood pixel)

A pixel at co-ordinate (x, y) has a four horizontal and four vertical neighbours whose co-ordinate are given by (x-1, y+1), (x, y+1), (x+1, y+1)

Scanned by CamScanner

$(x+1, y)$, $(x+1, y-1)$, $(x, y-1)$, $(x-1, y-1)$, & $(x-1, y)$.

These set of pixel is known as four neighbours of horizontal & four neighbours of vertical.

Each pixel is at unit distance from (x, y) and called as $N_8(P)$

Chapter II Image Enhancement in Spatial Domain

[-7 hours]

~~Point Operation~~ Introduction

Image enhancement refers to the sharpening or to improve the quality of an image by enhancement of pixel or point in a spatial domain.

Some basic gray level transformation is needed to enhance the image in a spatial domain. Some basic gray level transformation is listed as below:

(i) Point Operations
(ii) Contrast Stretching
(iii) Thresholding
(iv) Digital Negative
(v) Intensity level slicing
(vi) Bit plane slicing

Point Operation:

Point operation is defined as a function that are performed on each pixel of an image.

Point operation is unary point operation if single image is modified. The operation can be binary which means that two images combine in the same manner.

Scanned by CamScanner

The operation is called as ternary if three images are used in same manner at operation.

Unary operation can be defined as,

T_{xy} = function $(P(x,y))$

where,
T_{xy} = Transformed Pixel Value
function = Numeric transform
$P(x,y)$ = Input Pixel Value

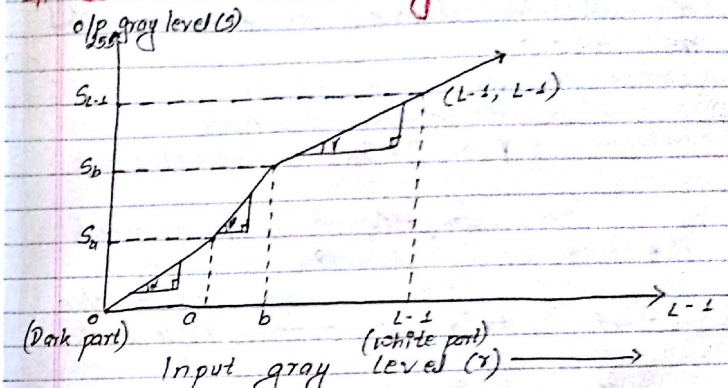
This operation can be used in many situations and different purpose such as: contrast stretching, noise clipping, window slicing and histogram modeling.

So, in point operation a given gray level, $u \in [0, L-1]$ is mapped into a gray level $v \in [0, L-1]$ according to a transformation.

ie $\boxed{v = f(u)}$

$L-1 = 255$
L = max gray level

Contrast Stretching:



(fig: Contrast stretching transformation)

where, α, β & γ are the slopes.

In many times, we obtain low contrast image due to poor illumination or due to the wrong setting of lens aperture.

The idea behind contrast stretching is to increase the dynamic range of gray level in the image. The main idea is to increase the contrast of the image by making the dark portion darker & bright portion brighter.

Scanned by CamScanner

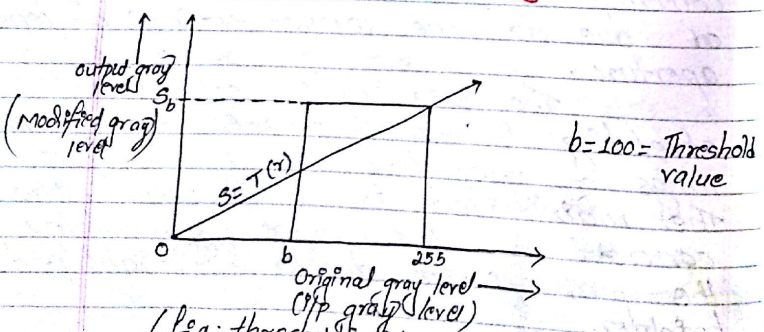
To obtain the contrast stretching, we should have some formulations as below:

$$S = \begin{cases} \alpha r, & 0 \le r \le a \\ \beta(r-a)+S_a, & a \le r < b \\ \gamma(r-b)+S_b, & b \le r < L-1 \end{cases}$$

In figure, 'r' represents the horizontal axis ie input gray level and 'S' represents the vertical axis ie output gray level.

Here, darker gray level is obtained by making a slope < 1 and brighter is obtained by making a slope > 1.

Thresholding (Binary Image)



(fig: thresholding)

It is the simplest method of the image processing that can be used to create binary image.

If we observe the contrast stretching diagram, we notice that the first and last slopes are made zero.

If $r_1 = r_2$ & $S_1 = 0$, $S_2 = L-1$

Now, the ~~corresponding~~ thresholding function is given by:

$$S = 0, \text{ if } r < b \Rightarrow 0$$
$$S = L-1, \text{ if } r \ge b \Rightarrow 1$$

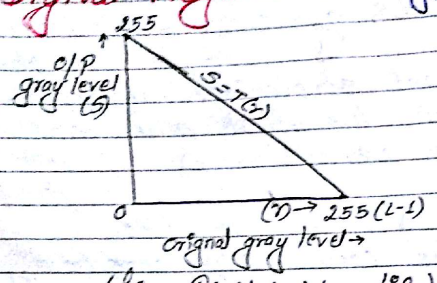
where,
b = thresholding value
L = maximum gray level

Important notice is that, threshold image has maximum contrast on it having black and white gray level only.

$$\boxed{T(r) = \{(L-1) - r\}}$$

Scanned by CamScanner

Digital Negative (X-Ray)



(fig: Digital Negative)

It is useful in a lots of application. Common example of digital negative is X-Ray image. As the name suggest, negative means simply inverting the gray level ie, black into white and vice versa. So, the negative of digital image with gray level in the range from 0 to (L-1) is obtained by using negative transformation.

i.e, $s = T(r) = [(L-1) - r]$

where,

r = input pixel value

L = No. of gray level / Max. gray

T(r) = Transformation function

s = output gray level

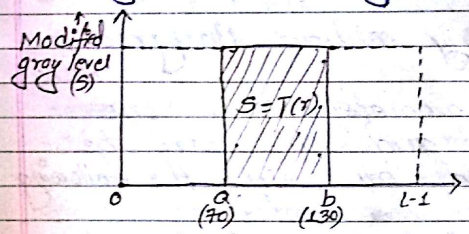
~~Hence~~,

When $r = 0$, $s = 255$ (white)

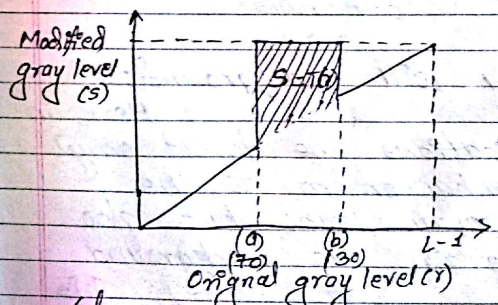
$r = 255$, $s = 0$ (black)

V. Imp

Intensity level slicing [Gray Level Slicing]



(fig: slicing without background)



(fig: slicing with background)

Scanned by CamScanner

Thresholding splits the gray level into two parts. But in many times we need highlight a specific range of gray level. In such situations, use transformation function on a gray level called as gray level ~~slit~~ slicing.

There are two types of slicing:

(i) Slicing without Background: Here we have completely lost the background and this can be implemented by using the following function:

ie,

$s = L-1$, if $a \le r \le b$
$s = 0$, otherwise

(ii) Slicing with Background: Here, in some applications we do not only need of enhancement to the band of gray level but also need to retain the background. This can be implemented by using the following function:

ie $s = L-1$, if $a \le r \le b$
$s = r$, otherwise

Bit plane Slicing (Clipping)

It is used to enhance features with an image. We must provide a threshold level to determine how the clipping occurs. The ~~ther~~ threshold level or value is a reference value where other values set changed. In general it is performed for maximum and minimum operators, And expression that contain an image with threshold levels.

For example: if ~~whe~~ we have image variable & want to slice it to ≥ 125, then the resulting image image stated as:

Clipped Image = ((Image (Original)) > 125)

Here, it means if the element values is less than

125, then it is set as equal to 125 otherwise it is unchanged.

V. Imp

Introduction to Histogram

Histogram of an image provides a global description of the appearance of an image.

The information obtained from histogram is innermost & hence, histogram modeling is important in digital image processing.

By the defination histogram of an image represents the relative frequency of occurance of various gray level in an image.

Histogram of an image can be plotted as:

(i) The X-axis has the gray level and Y-axis has the number of pixels in each gray level.

(ii) The X-axis represent gray level & Y-axis represents the probability of occurance of that gray level.

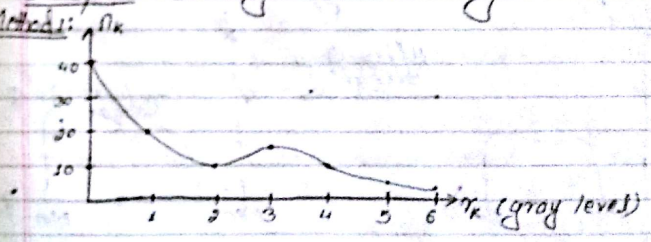
The method of histogram is shown as below:

Step 1: Original Image

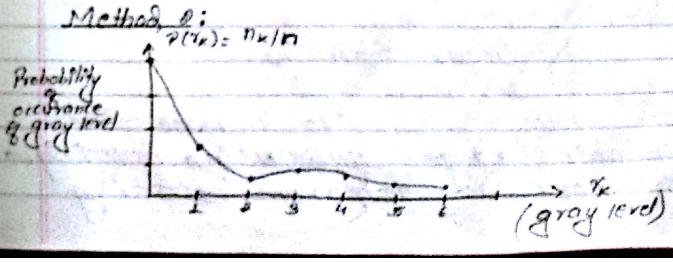
Gray level (r_k)	No. of pixel (n_k)
0	40
1	20
2	10
3	15
4	10
5	3
6	2

Step 2: Histogram Modeling

Method 1:



Method 2:



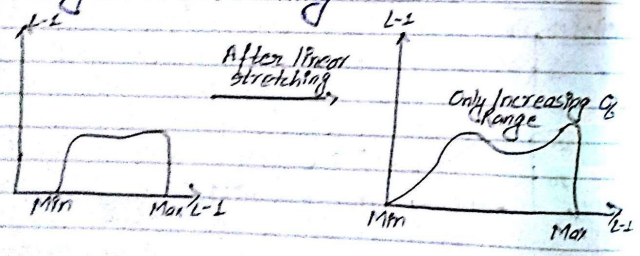
Scanned by CamScanner

Here,
r_k = k^{th} gray level
n_k = No. of pixel in k^{th} gray level
n = total no of pixel.

Histogram Techniques

(i) Histogram Stretching / Linear Stretching
(ii) Histogram Equalization / Histogram Linearization
(iii) Histogram Specification / Histogram Matching

1) Histogram Stretching



(fig: Linear Stretching)

To increase the dynamic range by using a technique is known as histogram stretching.

In this method, we don't change basic shape of histogram but spread it over the entire dynamic range. For this we use this technique on the basis of slope. It can be given as:-

$$\text{Slope } (m) = \frac{S_{max} - S_{min}}{r_{max} - r_{min}} \quad \text{---- (i)}$$

For transformation of histogram stretching then, transformation can be defined as:

$$S = T(r) = \frac{S_{max} - S_{min}}{r_{max} - r_{min}} (r - r_{min}) + S_{min}$$

or, $$S = \frac{S_{max} - S_{min}}{r_{max} - r_{min}} (r - r_{min}) + S_{min}$$

where,
S_{max} = max^m gray level in o/p image
S_{min} = min^m gray level in o/p image
r_{max} = max^m gray level in i/p image
r_{min} = min^m gray level in i/p image

This transformation function stretches or increase the dynamic range of given image.

Scanned by CamScanner

Q

Gray level	0	1	2	3	4	5	6	7
No. of pixel	0	0	50	60	50	20	~~10~~ 10	0

Now, perform histogram stretching so that the new image has a dynamic range of entire range ie [0,7]

Soln

$r_{min} = 2$ $\quad$ $S_{min} = 0$

$r_{max} = 6$ $\quad$ $S_{max} = 7$

Now,

we use transformation function on H.S as:

$$S = T(r) = \frac{S_{max} - S_{min}}{r_{max} - r_{min}} (r - r_{min}) + S_{min}$$

$$S = \frac{7-0}{6-2}(r-2) + 0$$

$$S = \frac{7}{4}(r-2)$$

When $r = 2$, $S = \frac{7}{4}(2-2) = 0$

$r = 3$, $S = \frac{7}{4}(3-2) = \frac{7}{4} \approx 2$

$r = 4$, $S = \frac{7}{4}(4-2) = \frac{7}{2} \approx 4$

$r = 5$, $S = \frac{7}{4}(5-2) = \frac{21}{4} \simeq 5$

$r = 6$, $S = \frac{7}{4}(6-2) = 7$

Hence, modified histogram or stretched range can be illustrated as:

Gray level	0	1	2	3	4	5	6	7
No. of pixels	50	0	60	0	50	20	0	~~10~~ 0

Now, output histogram is as shown below.

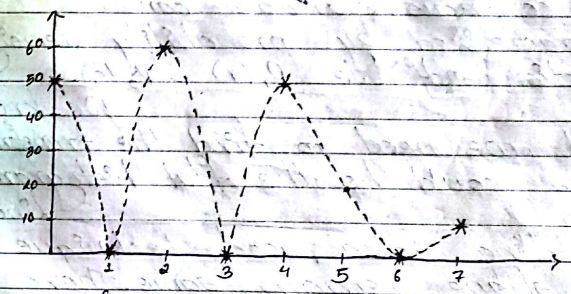
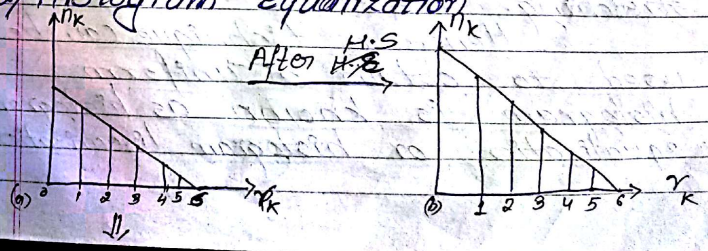


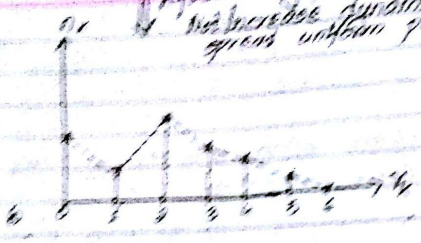
fig: stretched/modified o/p histogram

2) Histogram Equalization



Scanned by CamScanner

After Histogram Equalization
↓ the increase dynamic range but gives uniform pixel frequency



(fig: Histogram Linearization)

The gray level for continuous variable can be characterized by probability density function (pdf) as $P_r(r)$ & $P_s(s)$.

We know that, we need to find the transformation which could be the flat histogram.

Linear stretching is a good technique but shape remain same. There are many application where we need a flat histogram so that this cannot be achieved by histogram stretching.

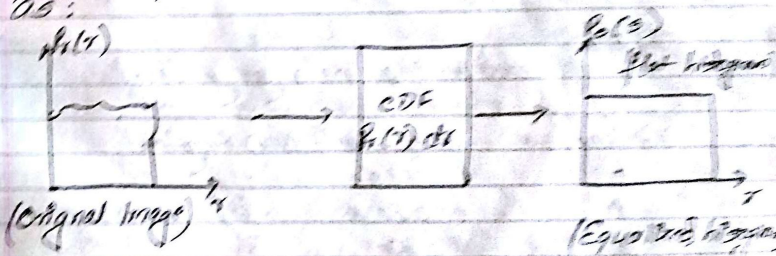
Hence, a technique can be used to obtain a uniform histogram is known as histogram equalization or histogram linearization.

A perfect image is one which has equal number of pixels in all gray level. Therefore, image is not only spread the dynamic range but also to have equal number of pixels in all gray level.

We know that,

$$s = T(r) = \int_0^r P_r(r)\,dr \quad , 0 \le r \le 1$$

Above equation shows the transformation function of cumulative pdf (cdf) and it can be shown as:



Q. Equalize the given histogram:

Gray level	0	1	2	3	4	5	6	7
No. of pixel	790	1023	850	656	329	245	122	81

Scanned by CamScanner

Solⁿ Step 1:

Here,

Gray level	n_k	P.D.F $P_r(r_k) = n_k/n$	C.D.F $S_k = \Sigma P_r(r_k)$	$7 * S_k$	Rounding off
0	790	0.19	0.19	1.33	1
1	1023	0.25	0.44	3.08	3
2	850	0.21	0.65	4.55	5
3	656	0.16	0.81	5.67	6
4	329	0.08	0.89	6.23	6
5	245	0.06	0.95	6.65	7
6	122	0.03	0.98	6.86	7
7	81	0.02	1.00	7	7
	n = 4096				

Step 2:

Representing new gray level

Old Gray level	No. of pixel	New Gray level
0	790	1
1	1023	3
2	850	5
3	656	6 (6 Add)
4	329	6
5	245	7 (7 Add)
6	122	7
7	81	7

Step 3:

Modified Histogram

Ans

Gray level	0	1	2	3	4	5	6	7
No. of pixel	0	790	0	1023	0	850	985	488

Hence, Equalized Histogram:

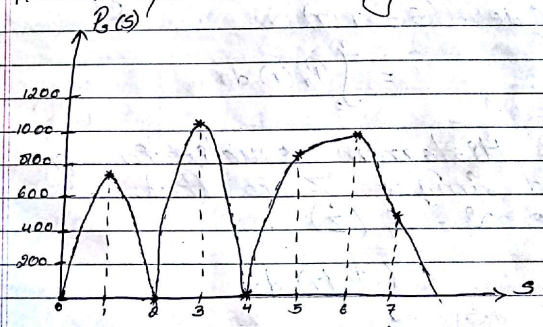


fig: Equalized Histogram

(Im+Nm)

3) Histogram Specification

Histogram matching is not interactive because it always gives one result i.e An approximation to an uniform histogram.

It is at time desirable to have an interactive methods in which certain gray levels are highlighted.

Scanned by CamScanner

Let us suppose $P_r(r)$ is the original p.d.f.

$P_z(z)$ is the desired p.d.f

Suppose that histogram equalization is first applied on the original image i.e $S = T(r)$

$$= \int_0^r P_r(r)\,dr$$

Histogram equalization of desired image 'Z' is that:

i.e $v = G(z)$

$$= \int_0^z P_z(r)\,dz$$

Now,

The inverse process $z = G^{-1}(S)$ which have the desired p.d.f that provide histogram specification.

i.e $z = G^{-1}(S) = G^{-1}(T(r))$

$$\boxed{z = G^{-1} \int_0^r P_r(r)\,dr}$$

Q. Match the given histogram:

Histogram for image (a)

Gray level	0	1	2	3	4	5	6	7
No of pixel	790	1023	850	656	329	245	122	81

Histogram for image (b)

Gray level	0	1	2	3	4	5	6	7
No of pixel	0	0	0	614	819	1230	819	614

Soln Step 1:

Equalized Histogram of image (a)

Gray level	0	1	2	3	4	5	6	7
No of pixel	0	790	0	1023	0	850	985	448

Now,

Equalize histogram of image (b)

Gray level	n_k	$P_r(r_k) = n_k/n$	$S_k = \Sigma P_r(r_k)$	$7 \times S_k$	Rounding off
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	614	0.15	0.15	1.05	1
4	819	0.20	0.35	2.45	2
5	1230	0.30	0.65	4.55	5
6	819	0.2	0.85	5.77	6
7	614	0.15	1	7	7

$n = 4096$

Scanned by CamScanner

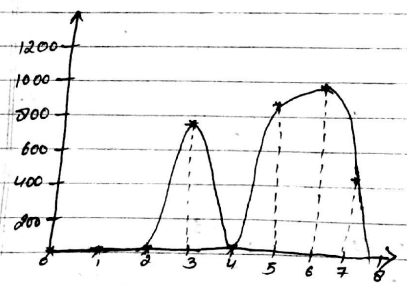
Equalized histogram of image (b)

Gray level	0	1	2	3	4	5	6	7
No. of pixels	0	614	819	0	0	1230	819	614

Step 2:
To obtain histogram specification, we apply inverse transformation comparing both equalized histogram.

Gray level	0	1	2	3	4	5	6	7
No. of pixel	0	0	0	790	0	850	985	448

Hence, the matched or resultant histogram is as shown below:



(fig: specified image)

Spatial Operations

Many image enhancement techniques are based on spatial operation. Most of the spatial operation are performed on the basis of local neighbourhood of the input pixel.

Some of the example of spatial operation for image enhancement are as follows:

(i) Neighbourhood Processing
(ii) Spatial Averaging
(iii) Smoothing Spatial filter
(iv) Zooming
(v) Low Pass, High Pass & Band pass

1) Neighbourhood Processing.

$$\begin{array}{c|ccc} & -1 & 0 & 1 \\ \hline -1 & g(x-1,y+1) & g(x,y+1) & g(x+1,y+1) \\ 0 & g(x-1,y) & g(x,y) & g(x+1,y) \\ 1 & g(x-1,y-1) & g(x,y-1) & g(x+1,y-1) \end{array}$$

(fig: 3*3 nieighbourhood Pixel)

Scanned by CamScanner

We change the value of pixel based on the values of its eight (8) neighbours as shown above.

Insted of 3*3 neighbourhood, we can also use 5*5, 7*7, etc. Neighbourhood processing deals with the point operation.

The neighbourhood pixel shown above can be masked into window as shown in figure below:

k \ l	-1	0	1
-1	w_1	w_2	w_3
0	w_4	w_5	w_6
1	w_7	w_8	w_9

(fig: 3*3 window)

The above figure is a one type of template which is also called as mask or window.

To achieve the neighbourhood processing we place this 3*3 mask on the image corresponding to each of the component and get some information at the centre.

If g( ) is the original image and f is the modified image then f is derived as,

$$f(x,y) = g(x-1,y+1)*w_1 + g(x,y+1)*w_2 + g(x+1,y+1)*w_3 + g(x-1,y)*w_4 + g(x,y)*w_5 + g(x+1,y)*w_6 + g(x-1,y-1)*w_7 + g(x,y-1)*w_8 + g(x+1,y-1)*w_9$$

Hence, the sum of the product of the mask coefficients with corresponding pixel take place.

2) Spatial Averaging [Averaging Filtering]

In spatial averaging, each pixel is replaced by the weighted average of its neighbourhood pixel ie

$$v(m,n) = \sum_{(k,l)\in w}\sum a(k,l)\, y(m-k, n-l)$$

where,

$y(m-k, n-l)$ = i/p image

$v(m,n)$ = o/p image

w = choosen window

$a(k,l)$ = filter weight

A common class of spatial average filter has all equal weight. ie

i.e.,

$$v(m,n) = \frac{1}{N_w} \sum_{(k,l)\in w} \sum y(m-k, n-l)$$

where,

$a(k,l) = \frac{1}{N_w}$, N_w is the no. of pixel in window

Now, each pixel is replaced by its average with the average of its nearest pixel value.

Spatial average is used for noise smoothing low pass filtering and sub sampling of the image.

Example of spatial averaging can be represented as:

k \ l	-1	0	1
-1	1/9	1/9	1/9
0	1/9	1/9	1/9
1	1/9	1/9	1/9

(fig: 3*3 mask window)

k=2 \ l=2		
	1/4	1/4
	1/4	1/4

(fig: 2*2 mask window)

3) Smoothing Spatial Filter [Median and Mean Filter]

Median filter:

The median filter is also a spatial filter but it replaces the central value in the window with median of all the pixel value in the window. The kernel or template or mask is usually square but can be ~~initiated~~ initiated in different window.

An example of median filtering is:

Unfiltered values in order like
0, 1, 2, 3, 4, 6, 10, 15, 19, 97

Now,

o/p value $= \left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{9+1}{2}\right)^{th}$ item

$= 5^{th}$ item

$= 6$

Hence,

6	2	0
3	97	4
19	15	10

(Before filtering)

Median filtering →

6	2	0
3	6	4
19	15	10

(After filtering)

Scanned by CamScanner

The center value 97 is replaced by the median value of all nine values ie 6.

Formula for median filtering is:

$$v(m,n) = \text{Median}\{y(m-k, n-l) \,(k,l) \in W\}$$

<u>Mean filter</u>:

Mean filter is a spatial filter that replace the center value in the window with the average of all the pixel values in the window.

Here, the window can be of any shape but can be initiated as a square mask. Therefore average or smoothing filter shows the properties of low pass filter. So, it reduces the noise from an image.

An example of mean filter is:

Unfiltered values in order like

5, 3, 6, 2, 1, 9, 8, 4, 7

Now,

o/p value or mean value $= \frac{\Sigma x}{N} = 5$

Hence,

5	3	6
2	1	9
8	4	7

(Before filtering)

Mean filtering →

5	3	6
2	5	9
8	4	7

(After filtering)

The center value '1' is replaced by the mean value '5'.

(V. Imp)

4) Zooming

[Magnification & Interpolation]

Another important application is image enhancement where spatial domain neighbourhood operation is used.

Such type of image enhancement technique is known as zooming.

Actually, it is the process of making an image larger. It involves following two steps.

(i) Creating of new pixel location

(ii) Assigning of grey level to those new location

Scanned by CamScanner

Different types of zooming techniques can be used. They are:

(i) Replication
(ii) Linear Interpolation

1) Replication

In replication, we simply replicate each pixel and then replicate row.

Let us consider an image as:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 8 & 6 & 7 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

As start from the first row, first we replicate each pixel and then replicate each row.

Now, The first row looks like:

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \end{bmatrix}$$

We will now replicate this row then it looks as:

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \end{bmatrix}$$

Now, performing replication operation on an entire image then we obtain output image as:

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ 5 & 5 & 6 & 6 & 7 & 7 & 8 & 8 \\ 5 & 5 & 6 & 6 & 7 & 7 & 8 & 8 \\ 9 & 9 & 8 & 8 & 6 & 6 & 7 & 7 \\ 9 & 9 & 8 & 8 & 6 & 6 & 7 & 7 \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 & 3 \end{bmatrix}$$

Here, 4x4 image is zoomed to 8x8 image.

This method can be repeated to get larger image but no new data is added.

Zooming by zero Interlace Technique

Step 1:

Adding zero to every after pixel of the first row, then we obtain row as:

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 & 4 & 0 \end{bmatrix}$$

Scanned by CamScanner

2) 2-D Fourier Transform

A 2D function $f(x,y)$ has 2D transform as:

$$F[f(x,y)] = F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-j2\pi(ux+vy)}\, dx\, dy$$

Inverse, fourier transform is

$$F^{-1}[F(u,v)] = f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\, e^{j2\pi(ux+vy)}\, du\, dv$$

Properties of fourier transform

(i) Addition Theorem:

$$F[f(x,y) + g(x,y)] = F(u,v) + G(u,v)$$

(ii) Shift Theorem:

$$F[f(x-a, y-b)] = F(u,v)\, e^{-j2\pi(ua+vb)}$$

(iii) Similarity Theorem:

$$F[f(ax, by)] = \frac{1}{|ab|} F\left(\frac{u}{a}, \frac{v}{b}\right)$$

(iv) Convolution Theorem:

$$F[f(x,y) * g(x,y)] = F(u,v) * G(u,v)$$

(v) Rayleigh's Theorem:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x,y)|^2\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F(u,v)|^2\, du\, dv$$

(vi) Evenness & Oddness:

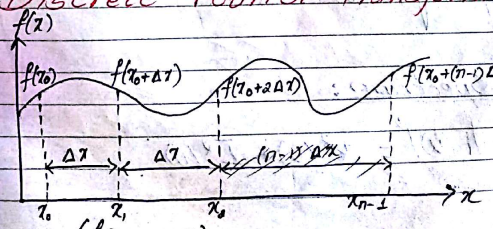
$f_e(x) \Rightarrow f(x) = f(-x)$

$f_o(x) \Rightarrow f(-x) = -f(x)$

(vii) Power Theorem:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) * g^*(x,y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v) * G^*(u,v)\, du$$

Discrete Fourier Transform (DFT)



(fig: DFT)

Scanned by CamScanner

Step 2:

Adding zero at every column.

Now,

Insert row of zero & then this is known as zero-Interlacing.

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 3:

Performing zero-interlace technique on an entire image then we obtain output image as:

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 6 & 0 & 7 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 9 & 0 & 8 & 0 & 6 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \end{bmatrix}$$

(fig: 8x8 final Pseudo image)

Hence, this image is known as zero-Interlace image.

2) Linear Interpolation

In this method, instead of replicating each pixel, average of the two adjacent pixel along the row is taken and placed between two pixels.

Accordingly same operation is then performed along the column.

Linear Interpolation is also derived from the zero-interlace technique.

Step 1:

Zero Interlace

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 6 & 0 & 7 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 9 & 0 & 8 & 0 & 6 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 2

Interpolate row

Scanned by CamScanner

1	1.5	2	2.5	3	3.5	4	2
0	0	0	0	0	0	0	0
5	5.5	6	6.5	7	7.5	8	4
0	0	0	0	0	0	0	0
8	8.5	8	7	6	6.5	7	3.5
0	0	0	0	0	0	0	0
0	0.5	1	1.5	2	2.5	3	1.5
0	0	0	0	0	0	0	0

Step 3

Interpolate column

1	1.5	2	2.5	3	3.5	4	2
3	3.5	4	4.5	5	5.5	6	3
5	5.5	6	6.5	7	7.5	8	4
7	7	7	6.25	6.5	7	7.5	3.25
9	8.5	8	7	6	6.5	7	3.5
4.5	4.5	4.5	4.5	4	4.5	5	2.5
0	0.5	1	1.5	2	2.5	3	1.5
0	0.25	0.5	0.75	1	1.25	1.5	0.7

5) Filtering

(i) Low Pass Filtering

It is employed to remove high frequency (unnecessary noise) from a digital image. This type of filtering is usually used to attenuate the image noise. Here, an image is smoothened by decreasing the disparity between the pixel value by average value.

Example of mask or kernel or window of low pass filter can be shown as below:

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

(fig: 3*3 mask)

1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16
1/16	1/16	1/16	1/16

(fig: 4*4 mask)

(ii) High Pass Filtering

A pass filtering is the basis for most sharpening methods. Sharpening is the opposite of smoothing.

The sharpening can make it possible to highlight the borders between homogenous region. This is used to return high frequency information. The kernel of the high pass filter is designed to increase the brightness of center pixel.

Scanned by CamScanner

Example of mask or kernel or window of high pass filter is as shown below:

1/9	1/9	1/9
1/9	-1/9	1/9
1/9	1/9	1/9

(fig: 3*3 mask high pass filter)

(iii) Band Pass Filter

This filter is used to remove selected frequency region between low and high frequency.

It is also used for image restoration. It passes a limited range of frequency.

Calculating spatial averaging for band pass filtering is:

$$v(m,n:\theta) = \frac{1}{N_\theta} \sum\sum_{(k,l)\in W_\theta} y(m-k, n-l)$$

where

W_θ = Neighbourhood mask along direction

N_θ = No of pixel within the neighbourhood

Chapter III Image Enhancement In Frequency Domain

[6 hours]

Fourier Transform

Fourier transform is a mathematical tool to analyze and design linear system. It is used to reduce the number of calculation to a fraction.

It also helps to quantify the effects of digitizing system.

The application of fourier transform is:

- Image enhancement
- X-Ray
- Enhancement of TV transmission & video signal.

1) 1-D Fourier Transform

Let $f(x)$ be a continous real variable of x

Then,

Fourier Transform = $F[f(t)] = F(U)$

$$F(U) = \int_{-\infty}^{\infty} f(x)\, e^{-j2\pi Ux}\, dx$$

Inverse fourier transform is

$$F^{-1}[F(u)] = f(x) = \int_{-\infty}^{\infty} F(u)\, e^{j2\pi Ux}\, dx$$

Scanned by CamScanner

Here, $f(x)$ is a continous function to be discretized by taking 'n' samples i.e $\{f(x_0), f(x_0+\Delta x), \ldots f(x_0+(n-1)\Delta x)\}$.

It is defined as

$$F(u) = \frac{1}{n}\sum_{x=0}^{n-1} f(x)\, e^{-j2\pi \frac{ux}{n}}$$

Inverse discrete fourier transform is defined as:

$$f(x) = \sum_{u=0}^{n-1} F(u)\, e^{j2\pi \frac{ux}{n}}$$

These two given equations are discrete fourier transform and it is used to calculate the following:

(i) Fourier Spectrum:

$$|F(u)| = \sqrt{R^2(u) + I^2(u)}$$

(ii) Power Spectrum:

$$|P(u)| = |F(u)|^2 = R^2(u) + I^2(u)$$

(iii) Phase angle spectrum:

$$\phi = \tan^{-1}\left(\frac{I(u)}{R(u)}\right)$$

Where,
'u' is transform frequency variable
'x' is spatial or image variable

Discrete Fourier Transform (DFT) is most widely used orthogonal transform in the field of image processing & it can diagonalized circuiland matrix. However, in this transform, it takes n^2 multiplication and $n(n-1)$ addition to calculate 1D discrete fourier transform of 'n' data points.

Discrete Cosine Transform & Its Properties

Discrete Cosine Transform is a fourier related transform similar to DFT but it is only concerned with real number only.

It can be defined as:

Scanned by CamScanner

$$DCT = C(U) = \alpha(U) \sum_{x=0}^{N-1} f(x) \cos\left(\frac{(2x+1)U\pi}{2N}\right)$$

where,

$U = 0, 1, \ldots N-1$

$$\alpha(U) = \sqrt{\frac{1}{N}} = \frac{1}{\sqrt{N}} \text{ for } U = 0$$

$$= \frac{2}{\sqrt{N}} \text{ for } U = 1, 2, \ldots N-1$$

Inverse DCT is given by:

$$IDCT = f(x)$$

$$= \sum_{U=0}^{N-1} \alpha(U) \cos\left(\frac{(2x+1)U\pi}{2N}\right)$$

where, $U = 0, 1, 2 \ldots N-1$

<u>For 2D</u>

$$DCT = C(u,v) = \alpha(u,v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

where,

$$\alpha(u,v) = \frac{1}{N} \text{ for } u,v = 0$$

$$= \sqrt{\frac{2}{N}} \text{ for } u,v = 1, \ldots N-1$$

~~Sim~~

$$IDCT = f(x,y)$$

$$= \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u,v) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2x+1)v\pi}{2N}\right)$$

where,

$x = y = 0, 1, 2, \ldots N-1$

Properties of DCT

(i) The cosine transform is real and orthogonal.

$$C = C^* \Rightarrow C^{-1} = C^T$$

where,

C^* = Conjugate of C

C^T = Transform of C

(ii) The cosine transform is not real part of unitary DFT.

ie $T^{-1} = T^{*T}$

Scanned by CamScanner

(iii) It is a fast transform. The cosine transform of 'N' elements can be calculated in $O(N \log_2 N)$ operation through 'N' point FFT (fast fourier Transform).

(iv) It has excellent energy compaction for highly correlated data.

Uses of DCT

DCT is used in image processing & signal processing especially for loose data compression. Since, it has a strong energy compaction features.

It is also used for JPEG image compression, MPEG video compression and MJPEG video compression.

Imp ~~Hand~~ Hadamard Transform

Hadamard transform is a fast transform which is real, symmetric and orthogonal.

1-D hadamard transform can be implemented in $O(N \log_2 N)$ addition & subtraction.

It has only binary values i.e 1 or -1 in its kernel matrix. It is defined as:

$$H(U) = \frac{1}{\sqrt{N}} \underbrace{\sum_{x=0}^{N-1} f(x) (-1) \sum_{x=0}^{n-1} b_i(x)\, b_i(U)}_{\text{Kernel}}$$

for $N = 2$,

$N = 2^n \Rightarrow 2 = 2^n \Rightarrow n = 1$

where,

$u = 0, 1, \ldots n-1$

Now,

Hadamard matrix can be defined as,

$$H_n = \frac{1}{\sqrt{N}} \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -(H_{n-1}) \end{bmatrix}$$

where,

$\frac{1}{\sqrt{N}}$ = Sample value

$\begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -(H_{n-1}) \end{bmatrix}$ = Indicate kernel

$H_0 = 1$

Scanned by CamScanner

Hadamard matrix at 1, $H_0 = 1$

$$H_n = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -(H_{n-1}) \end{bmatrix}$$ // N=2 i.e (0, 1)

$$H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} H_0 & H_0 \\ H_0 & -H_0 \end{bmatrix}$$

$$H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Also for $N=4$, $N=2^n \Rightarrow n=2$

$$H_2 = \frac{1}{\sqrt{4}} \begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Properties of Hadamard transform

(i) It is real, symmetric and orthogonal.
i.e $H = H^* \Rightarrow H^{-1} = H^T$

(ii) It is a fast transform and 1-D Hadamard transform can be implemented in $O(N \log_2 N)$ addition & subtraction

(iii) It has only binary values i.e 1 or -1 ~~it~~ in its kernel matrix.
No multiplications are required in the transform.

(iv) It is used for digital image processing and digital signal processing

(v) It has good energy compaction for highly ~~correlated~~ corelated image.

(vi) It is also used in digital hardware implementation of image processing algorithm.

Haar Transform

Haar transform is the orthogonal transform which is derived from the Haar matrix.
It is also a fast transform and can be implemented in $O(N \log_2 N)$ operation.
The main advantage of this transform is that

Scanned by CamScanner

It samples the input data sequence into fine resolution and takes the difference between the adjacent pairs.

The resolution increases by a power of 2 and as a result, in the transform domain differential energy is more & more localized.

It only exist for $n \cdot N = 2^n$ where, n is an integer.

Haar transform is defined as:

$H_T = H_k(x)$ where, $k = 0, 1, \ldots N-1$

k is given as:

$$k = 2^p + q - 1$$

and $0 \le p \le (n-1)$

$1 \le q \le 2^p$ for $p \ne 0$

$q = 0, 1$ for $p = 0$

Properties of Haar Transform

(i) It is symmetric, seperable, unitary transform that uses Haar function for its basis.

(ii) It is orthogonal and real i.e,

$$T^{-1} = T^T \Rightarrow H_T = H_T^*$$

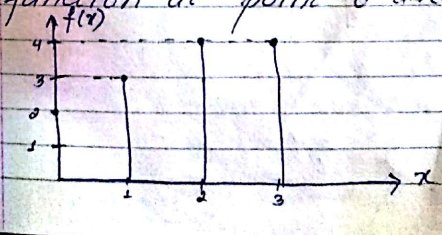
(iii) It is a fast transform and can be implemented in $O(N)$ operation where N is the number of samples.

(iv) It exists for $N = 2^n$ where n is an integer.

(v) It has poor energy compaction property.

Numericals

Q. Find the Fourier spectrum of the function at point 0 and 1 for given figure



$f(0) = 2$

$f(1) = 3$

$f(2) = 4$

$f(3) = 4$

Scanned by CamScanner

Soln

Here,

$f(0) = 2$

$f(1) = 3$

$f(2) = 4$

$f(3) = 4$

We know,

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x)\, e^{\frac{-j2\pi ux}{N}}$$

where,

$N = 4 \quad (0, 1, 2, 3)$

Now,

for $u = 0$

$$F(0) = \frac{1}{4} \sum_{x=0}^{3} f(x)\, e^{\frac{-j2\pi 0 . x}{4}}$$

$$= \frac{1}{4} \sum_{x=0}^{3} f(x)$$

$$= \frac{1}{4} \left[f(0) + f(1) + f(2) + f(3) \right]$$

$$= \frac{1}{4} \left[2 + 3 + 4 + 4 \right]$$

$$= \frac{13}{4}$$

fourier Spectrum at point 0 is:

$$|F(0)| = \sqrt{R^2(0) + I^2(0)}$$

$$= \sqrt{\left(\frac{13}{4}\right)^2 + 0}$$

$$|F(0)| = \frac{13}{4}$$

Power spectrum at point 0 is:

$$|P(0)| = |F(0)|^2$$

$$= \left(\frac{13}{4}\right)^2 = \frac{169}{16}$$

Phase angle spectrum is:

$$\varphi = \tan^{-1}\left(\frac{I(u)}{R(u)}\right)$$

$$= \tan^{-1}\left(\frac{0}{13/4}\right)$$

$$= \tan^{-1}(0)$$

$$= 0$$

Scanned by CamScanner

At $u = 1$

$$F(1) = \frac{1}{4}\sum_{x=0}^{3} f(x)\, e^{-j\frac{2\pi \cdot 1 \cdot x}{4}}$$

$$= \frac{1}{4}\sum_{x=0}^{3} f(x)\, e^{-j\frac{\pi x}{2}}$$

$$= \frac{1}{4}\left[2e^{-j\frac{\pi}{2}\cdot 0} + 3e^{-j\frac{\pi}{2}} + 4e^{-j\frac{\pi}{2}\cdot 2} + 4e^{-j\frac{3\pi}{2}}\right]$$

$$= \frac{1}{4}\left[2 + 3e^{-j\frac{\pi}{2}} + 4e^{-j\pi} + 4e^{-\frac{3j\pi}{2}}\right]$$

$$= \frac{1}{2} + \frac{3}{4}e^{-j\frac{\pi}{2}} + e^{-j\pi} + e^{-\frac{3j\pi}{2}}$$

$$= \frac{1}{2} + \frac{3}{4}\left(\cos\frac{\pi}{2} - j\sin\frac{\pi}{2}\right) + \left(\cos\pi - j\sin\pi\right) + \left(\cos\frac{3\pi}{2} - j\sin\frac{3\pi}{2}\right)$$

$$= \frac{1}{2} + \frac{3}{4}(0 - j\cdot 1) + (-1 - j\cdot 0) + (0 - j(-1))$$

$$= \frac{1}{2} + \frac{3}{4}(-j) + j - 1$$

$$= \frac{1}{2} - \frac{3}{4}j + j - 1$$

$$= \frac{2 - 3j + 4j - 4}{4}$$

$$= \frac{j - 2}{4}$$

$$= \frac{1}{4}j - \frac{1}{2}$$

$$= \left(-\frac{1}{2}\right) + \left(\frac{1}{4}\right)j$$

Now,

Fourier spectrum at 1 is

$$|F(u)| = \sqrt{(R(u))^2 + (I(u))^2}$$

$$= \sqrt{\left(-\frac{1}{2}\right)^2 + \left(\frac{1}{4}\right)^2}$$

$$= \sqrt{\frac{1}{4} + \frac{1}{16}} = \sqrt{\frac{5}{16}} = \frac{\sqrt{5}}{4}$$

Power spectrum at 1 is

$$|P(u)| = |F(u)|^2 = \frac{5}{16}$$

Phase angle spectrum is:

$$\phi = \tan^{-1}\left(\frac{I(u)}{R(u)}\right) = \tan^{-1}\left(\frac{1/4}{-1/2}\right)$$

$$= \tan^{-1}\left(-\frac{1}{2}\right)$$

Scanned by CamScanner

Chapter IV Image Restoration

(~4 hours)

Introduction

The objective of image restoration is to improve a given image in some predefined sense.

Image enhancement is a subjective process whereas image restoration is objective process.

Image restoration is to reconstruct or recover an image that has been degraded by using a prior knowledge of the degradation phenomena.

So, this technique is a process of degradation and applying the inverse process in order to recover the original image.

If there is a presence of noise and image blurring we use image restoration.

A model of Image Degradation / Restoration Process

The figure below shows a model of image degradation / restoration process:

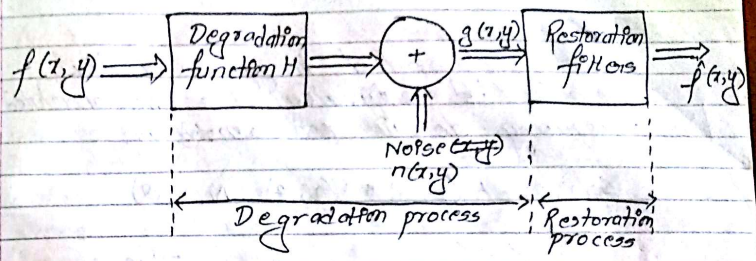


fig: A model of Image degradation / restoration process.

The degraded image $g(x,y)$ can be produced as:

$$g(x,y) = H[f(x,y)] + n(x,y)$$

where,

$f(x,y)$ = I/p image

$n(x,y)$ = Additive noise ~~form~~ term

Now, The degraded image can be defined in spatial domain as:

$$g(x,y) = h(x,y) * f(x,y) + n(x,y)$$

where,

$h(x,y)$ = Spatial representation of the degradation function

Scanned by CamScanner

Now,
the degraded image in spatial domain is equivalent to the convolution frequency domain represented as:

$$G(u,v) = H(u,v) * F(u,v) + N(u,v)$$

The degradation process is also referred to as convolution of the image with a point spread function (PSF).

Similarly, restoration process is defined or referred to as deconvolution.

Hence,
Restoration process can be obtained as:

$\hat{f}(x,y) = R[g(x,y)]$, In linear spatial invariant form

~~$\hat{f}(x,y)$~~

$\hat{f}(u,v) = R[G(u,v)]$, In deconvolution form.

Noise Model Restoration in the Presence of Noise Only Spatial Filtering

The method of choice for reducing noise is spatial filering. After filtering the noise, there should have a restoration process to have or obtain original image.

For noise reduction, following summarize and implement two types of spatial filters:

(i) Spatial Noise filters
(ii) Adaptive Spatial filters

1) Spatial Noise Filter

In spatial noise filters image independently of how image characteristics vary from one location to another location.

In this condition the spatial noise filters can be used & explained as:

(i) Arithmetic mean,

$$\hat{f}(x,y) = \frac{1}{mn} \sum_{(s,t) \in S_{xy}} g(s,t)$$

Scanned by CamScanner

(ii) Geometric mean,

$$\hat{f}(x,y) = \left[\prod_{(s,t)\in S_{xy}} g(s,t)\right]^{1/mn}$$

(iii) Harmonic mean,

$$\hat{f}(x,y) = \frac{mn}{\sum_{(s,t)\in S_{xy}} \frac{1}{g(s,t)}}$$

(iv) Median,

$$\hat{f}(x,y) = \underset{(s,t)\in S_{xy}}{\text{Median}} \{g(s,t)\}$$

(v) Max,

$$\hat{f}(x,y) = \max_{(s,t)\in S_{xy}} \{g(s,t)\}$$

(vi) Min,

$$\hat{f}(x,y) = \min_{(s,t)\in S_{xy}} \{g(s,t)\}$$

where,

m & n = variables denotes image rows & columns spanned by the filter respectively

S_{xy} denotes an m*n sub image of the input noisy image g.

$\hat{f}(x,y)$ = filter response at (x,y) condition (co-ordinates)

(x,y) = co-ordinates of sub image

2) Adaptive Spatial Filter

In some application, results can be improved by using filters capable of adapting their behaviour based on the characteristic of image in the region being filtered.

The adaptive median filtering algorithm uses two processing levels for adaptive spatial filter that can be denoted as:

Level A:
If $Z_{min} < Z_{med} < Z_{max}$ go to level B
Else, increase the window size
If window size $\leq S_{max}$, repeat level A
Else output Z_{med}

Scanned by CamScanner

Level B:
If $Z_{min} < Z_{xy} < Z_{max}$, output Z_{xy}
Else output Z_{med}

where,
Z_{min} = Minimum intensity value
Z_{max} = Maximum intensity value
Z_{med} = Median intensity value
Z_{xy} = Intensity value at co-ordinate (x, y)
S_{xy} = maximum allowed size of adaptive filter window

Periodic Noise Reduction Using Frequency Domain Filtering

Periodic noise produces impulse like bursts that often are visible in the fourier spectrum.

The principal approach for filtering these components is to use notch reject filtering.

The general expression for notch reject filtering having 'Q' notch pair can be defined as:

$$H_{NR}(u,v) = \prod_{k=1}^{Q} H_k(u,v)\, H_{-k}(u,v) \quad \text{--- (i)}$$

where,
$H_k(u,v)$ & $H_{-k}(u,v)$ are high pass filter with centre at (u_k, v_k) & $(-u_k, -v_k)$ respectively

These translated centres are specified with respect to the centre of the frequency rectangle $(M/2, N/2)$.

Therefore, the distance computation for the filters are given by following expression.

$$D_k(u,v) = \left[(u - M/2 - u_k)^2 + (v - N/2 - v_k)^2\right]^{1/2}, \text{ for } (u_k, v_k)$$

and,

$$D_k(u,v) = \left[(u - M/2 + u_k)^2 + (v - N/2 + v_k)^2\right]^{1/2}, \text{ for } (-u_k, -v$$

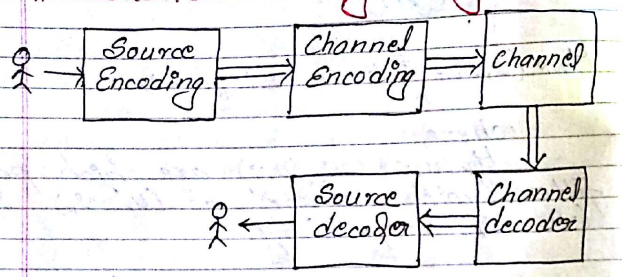
Scanned by CamScanner

Chapter V

Image Compression

(~6 hours)

Introduction to Image Coding & Compression



(fig: Block diagram of Image Compression)

Image Compression addresses the problem of reducing the amount of data required to represent a digital image. The underlying basis of the reduction process is removal of redundant data.

The compressed image is decompressed to reconstruct the original image and approximation of it.

Image (data) compression is the reduction of number of bits required to transmit the image probably without any remarkable loss of information.

The image encoding process through the channel is called image compression and image decoding process is known as image decompression.

Application of Data Compression

(i) Broadcasting TV
(ii) Remote Sensing through satellite
(iii) Military Communication
(iv) Fascimile transmission & telecommunication (Fax)

Data Redundancy: It contains data that either provide no relevent information or simply restate that which is already known. This is called data redundancy.

Image Compression Technique:

(i) Pixel Coding Technique
(ii) Predictive Coding
(iii) Transform Coding
(iv) Other methods

Scanned by CamScanner

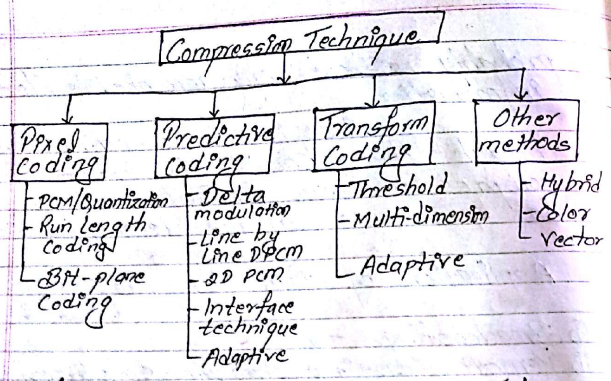


fig: Different Image Compression Technique

1) Pixel Coding Technique

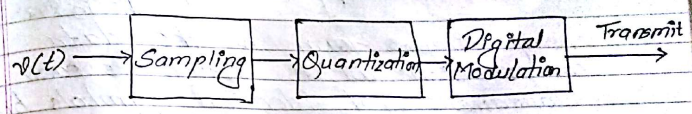
In this technique, each pixel is processed independently ignoring the inter-pixel dependencies.

This technique firstly remove the inter dependencies between the neighbourhood pixel and then removal of redundant data.

Pixel Coding Technique is further divided as:

(a) PCM /Quantization
(b) Run length Coding
(c) Bit-plane Coding

A) Pulse Code Modulation (PCM):



(fig: PCM /Quantization)

In PCM technique, the incoming video signal is sampled, quantized and coded by the suitable code word before inputting it to a digital modulator for transmission.

The quantized output is coded by a fixed length binary code word having usually 8-bits.

The formula for achieving rate of compression of PCM is given by the rate distortion formula.

i.e $R_{PCM} = \frac{1}{2} \log_2 \left(\frac{\sigma_v^2}{\sigma_q^2}\right)$, $\sigma_q^2 < \sigma_v^2$

where,
σ_v^2 = variance of quantizer i/p
σ_q^2 = quantizer mean square deviation

Scanned by CamScanner

B) Run length Coding:

It is a very simple form of the data compression which represents each subsequence of identical symbols by a ~~pa~~ pair like (L, a) where L is the subsequence and a is the recurring symbols.

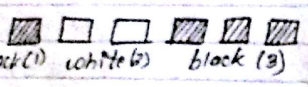
Run length coding describes each row of the image by a sequence of length that describes successive runs.

<u>Example</u>: 'aaabbbaaa' is coded as '3a3b3a'.

The most common methods are as follows:

(i) To specify the value of the first run of each row

(ii) To assume that each row begins with a white run

for example: 

from above figure, the run length code can be written as:

1 black 2 white 3 black

In this technique, the numeric run length coding can also be defined as:

00000 2222111 = 0/5 2/4 1/3

Entropy Coding:

Entropy codes encoding the given set of symbols with minimum number of bits required to represent them.

The most important entropy coding is Huffman coding. The theoritical minimum average of bits that are required to transmit a particular source string is known as entropy of source. It can be computed by using the following formula:

$$\text{Entropy}(H) = -\sum_{i=1}^{N} P_i \log_2 P_i$$

It has two major parts
(i) Construction of probability tree
(ii) Assigning code to each node of constructed tree

Scanned by CamScanner

It gives the variable length code words and highest probability assigns ~~sth~~ ~~pa~~ short path and lowest probability assign longest path ie ~~use~~ code length.

Huffman Coding Algorithm

(i) Arrange the symbol probabilities P in an decreasing order & consider them as leaf node of a tree

(ii) While there is more than one node,
   (a) Manage the 2 nodes with smallest probability to form a new node.

   (b) Assign 1 and 0 to each pair of branches merging into a node

(iii) Read sequentially from root node to leaf node.

Q. Source generates the symbol S_1, S_2, S_3, S_4 and S_5 randomly with probability $P_1 = 0.4$, $P_2 = 0.2$, $P_3 = 0.2$, $P_4 = 0.1$ & $P_5 = 0.1$ respectively.

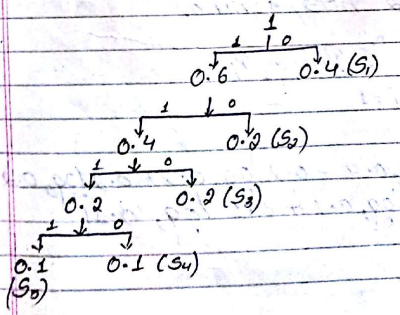
Generate the codeword for each symbol using Huffman coding. Also calculate the compression ratio & efficiency of the system.

Sol^n Step 1:

Symbol	Original Size Probability (P_i)	step 1	2	3	4
S_1	0.4	0.4	0.4	0.6	1
S_2	0.2	0.2	0.4	0.4	
S_3	0.2	0.2	0.2		
S_4	0.1	0.2			
S_5	0.1				

Step 2:

Construct Huffman tree, assume highest probability 0.6 = 1 & lowed probability 0.4 = 0 (step 3 ~~ओर~~)



Step 3
Generate codeword length

Symbol	Probability (P_i)	Code Length (L_i)
S_1	0.4	0 = 1
S_2	0.2	10 = 2
S_3	0.2	110 = 3
S_4	0.1	1110 = 4
S_1	0.1	1111 = 4

Now,
Compression ratio (L_{avg}) $= \sum_{i=1}^{N} P_i L_i$

$= 0.4\times1 + 0.2\times2 + 0.2\times3 + 0.1\times4 + 0.1\times4$

$= 0.4 + 0.4 + 0.6 + 0.4 + 0.4$

$= 2.2$ bits/symbol

Entropy (H) $= -\sum_{i=1}^{N} P_i \log_2 P_i$

$= -[0.4 \log_2 0.4 + 0.2 \log_2 0.2 + 0.2 \log_2 0.2 + 0.1 \log_2 0.1 + 0.1 \log_2 0.1]$

$= 2.12$

Efficiency $= \frac{\text{Entropy}}{L_{avg}} \times 100\%$

$= \frac{2.12}{2.2} \times 100\%$

$= 96.36\%$

Q. Calculate the compression ratio and efficiency from the image given below:

r	0	1	2	3	4	5	6	7
n_k	400	1350	659	2034	816	2560	250	1500

Soln
Step 1: Calculate the probability in each gray level.

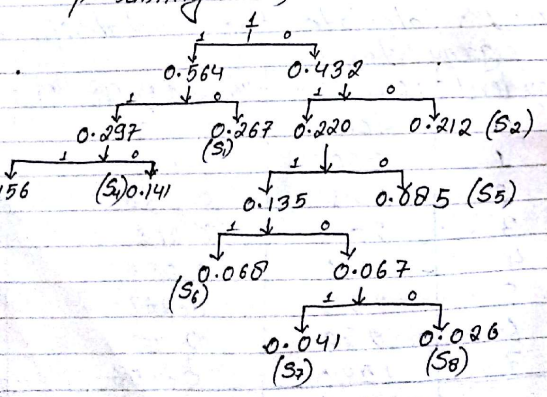
Gray level	no. of pixel	Probability (P_i) $= n_k/n$
0	400	0.041
1	1350	0.141
2	659	0.068
3	2034	0.212
4	816	0.085
5	2560	0.267
6	250	0.026
7	1500	0.156
	n = 9568	

Scanned by CamScanner

ep 2 : Assume symbol for each probablity descending order

mbol	P_i	step 1	2	3	4	5	6	7
S_1	0.267	0.267	0.267	0.267	0.297	0.432	0.564	0.996 ≈ 1
S_2	0.212	0.212	0.212	0.220	0.267	0.297	0.432	
S_3	0.156	0.156	0.156	0.212	0.220	0.267		
S_4	0.141	0.141	0.141	0.156	0.212			
S_5	0.085	0.085	0.135	0.141				
S_6	0.068	0.068	0.085					
S_7	0.041	0.067						
S_8	0.026							

p 3 : Construct Huffman tree, assume highest probability = 1 & lowest = 0



Step 4 : Generate Codeword length

Symbol	Probability (Pi)	Code length (Li)
S_1	0.267	10 = 2
S_2	0.212	00 = 2
S_3	0.156	111 = 3
S_4	0.141	110 = 3
S_5	0.085	010 = 3
S_6	0.068	01101 = 4
S_7	0.041	01111 = 5
S_8	0.026	01110 = 5

Now,

Compression ratio $(L_{avg}) = \sum_{i=1}^{N} P_i L_i$

$= 0.267\times2 + 0.212\times2 + 0.156\times3 + 0.141\times3 + 0.085\times3 + 0.068\times4 + 0.041\times5 + 0.026\times5$

$= 0.534 + 0.424 + 0.468 + 0.423 + 0.255 + 0.272 + 0.205 + 0.13$

$= 2.711$

Entropy $(H) = -\sum_{i=1}^{N} P_i \log_2 P_i$

$= -[0.267 \log_2(0.267) + 0.212 \log_2(0.212) + 0.156 \log_2(0.156) + 0.141 \log_2(0.141) + 0.085 \log_2(0.085) + 0.068 \log_2(0.068) + 0.041 \log_2(0.041) + 0.026 \log_2(0.026)]$

$= 2.687$

Scanned by CamScanner

Now,

$$\text{Efficiency}(\eta) = \frac{\text{Entropy}}{L_{avg}} \times 100\,\%$$

$$= \frac{2.687}{2.711} \times 100\,\%$$

$$= 99.11\,\%$$

c) Bit plane Coding: It breaks the image into bit planes and apply run length coding to each plane.

Let, 'I' be an image where every pixel value is n-bit long. We can express every pixel in binary using 'n-bit from out of an image and 'n' binary matrix (called bit plane) where n^{th} matrix consist of ith bit of the pixel of I.

It can be illustrated as:

$$I = \begin{bmatrix} 101 & 110 \\ 111 & 011 \end{bmatrix}$$, corresponding 3 bit plane & 2x2 matrix.

Now,

$$I' = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Bit-plane coding uses the run length coding but in case of m-bit gray level scale image at polynomial 2-base is defined as:

$$a_{m-1} 2^{m-1} + a_{m-2} 2^{m-2} + \cdots a_1 2^1 + a_0 2^0$$

2) Predictive Coding future value will be predicted on the basis of past values. If we know that, the past behaviour of a signal upto a certain point in time then it is impossible to make some prediction about its future values then such process is called prediction.

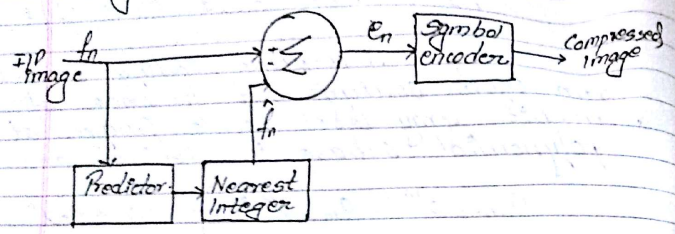
It is of two types.

They are:

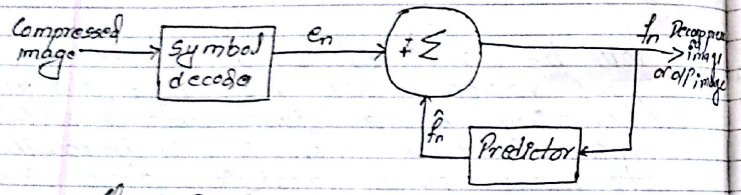
(i) Lossless Predictive Coding
(ii) Lossy Predictive Coding

1) lossless Predictive Coding There is no information loss & image can be reconstructed exactly the same as the original. So, compre

process is reversible in lossless coding.



(fig (a) Encoder)



(fig (b) Decoder)

(fig: Lossless Predictive Coding)

The system consist of an encoder and a decoder, each containing an identical predictor.

The predictor generates anticipated value of that pixel based on some number of past values. The o/p is then rounded to the nearest integer.

Then, the differences or prediction error is calculated as:

$$e_n = f_n - \hat{f}_n$$

where,

e_n = prediction error

f_n = i/p image symbol

$\hat{f}_n$ = prediction function.

In case of encoding, symbol encoder uses variable length coding technique to encode symbol.

In case of decoding, symbol decoder uses decoding function as:

$$f_n = e_n + \hat{f}_n$$

where,

f_n = decompressed image or o/p image

In most cases, the predictor is formed by a linear combination of m previous pixel.

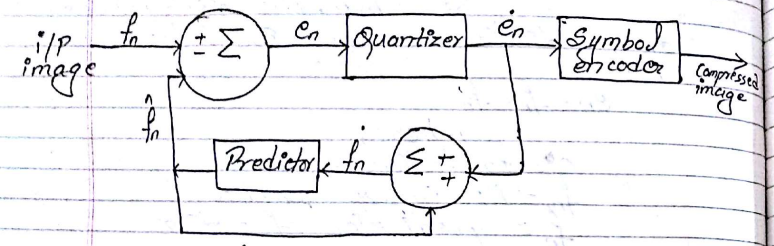
i.e $\hat{f}_n = \text{round} \sum_{i=1}^{m} \alpha_i f_{m-i}$

α = predictor coefficient

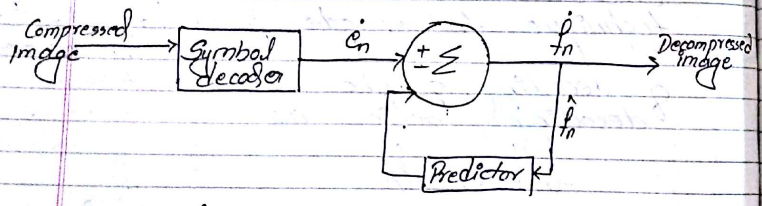
m = order of linear prediction

Scanned by CamScanner

2) Lossy Predictive Coding



(fig: Encoder)



(fig: decoder)

The major difference in this technique is that a quantizer is added between the symbol encoder and the point where the prediction error is calculated.

The quantizer absorbs the nearest value of a prediction error & maps it into a limited range of output denoted by $\dot{e}_n$ at encoding and $\dot{f}_n$ at decoding, which is defined as:

$$\dot{e}_n = f_n - \hat{f}_n$$

&, $\dot{f}_n = \dot{e}_n + \hat{f}_n$

Interframe & Intraframe Coding

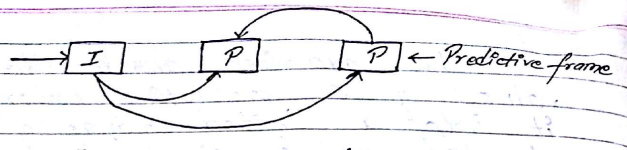
It is used for video compression. The intraframe coding is also known as I-frame. Intraframe compression uses only current frame for encoding.

Interframe coding is encoding and compression technique between the video signal.

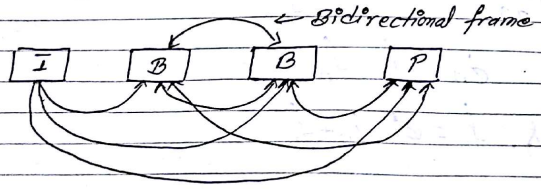
Interframe & Intraframe coding is categorized into 3 types on the basis of predicted frame.

(i) Predictive frame (P-frame)
(ii) Intraframe or Independent frame (I-fra
(iii) Bidirectional frame (B-frame).

Scanned by CamScanner



(fig (a) : I and P frame)



(fig (b): I, P and B frame)

Chapter VI Introduction to Morphological Image Processing

[4 hours]

Logic Operations Involving Binary Images

The language and theory of mathematical morphology often present a dual (but equivolent) view of binary images.

Thus, we have considered a binary image to be a bivalued function of spatial co-ordinates x and y.

Morphological theory views a binary image as a set of foreground (one valued) pixel and elements of which are in z^2 (cartesian product) i.e, pair of elements.

Set operations such as union and intersection can be applied directly to binary image sets.

<u>Example:</u>

If A and B are binary images then $C = A \cup B$ is a binary image. where, a pixel in C is a foreground pixel if either or both of the corresponding pixel in A and B are foreground pixel.

The function C is given b

two view:

(i) First view

$$C(x,y) = \begin{cases} 1 & \text{if either } A(x,y) \text{ or } B(x,y) \text{ or both is 1} \\ 0, & \text{otherwise} \end{cases}$$

(ii) Second view

$$C = \{(x,y) : (x,y) \in A \text{ or } (x,y) \in B \text{ or } (x,y) \in (A \text{ and } B)\}$$

where, A & B are binary image where the elements of A and B are 1-valued.

Thus, we see that the function point of view deals with the both foreground (1) and background (0) pixel simultaneously.

The set point of view deals only with foreground pixel & it is understood that all pixels are not foreground pixel, that constitute the background.

The set operations can be performed on binary images using following logical operations:

Set Operation	Logical Operation	Name
$A \cup B$	$A \mid B$	OR
$A \cap B$	$A \& B$	AND
A^c	$\sim A$	NOT
$A - B$	$A \& \sim B$	DIFFERENCE

V.Imp

Dilation & Erosion

Dilation:

Dilation is an operation that grows or thickens objects in an image. The specific manner & extent of this thickening is controlled by a set referred to as a structuring element.

Graphically structuring element can be represented either by a matrix of 0's and 1's or as a set of foreground (1-valued) pixels.

So, the origin of the structuring element must be clearly identified by using the both representation.

Scanned by CamScanner

Dilation operation can be done by using following steps:

(i) Representation of original image with rectangular object.

Let,

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

(ii) Structuring element with 5 pixels arranged in a diagonal line.

		1
	1	
1		

$\boxed{1}$ ← origin or center of the structuring element as shown with a dark border

(iii) Structuring element translated to several location to the image

			*	*	*	*	$\boxed{*}$	*	*	*	*
		*	*	*	*	*	*	*	*	*
	1	1	1	1	1	1	1	*	*
	*	1	1	1	1	1	1	1	*
*	*	1	1	1	1	1	1	1	
*	*	*	*	*	*	*	*	*	
*	*	*	*	*	*	*	*	*	

overlap 1-valued pixel in original image

doesnot overlap 1-valued pixel in orgnal image

(iv) Output image ie shaded region will be ~~re~~ replaced by 1's in the original image

			1	1	1	1	1	1	1	1	1
		1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1		
1	1	1	1	1	1	1	1			

Now,

the dilation of A by $\mathcal{B}$ denoted $A \oplus \mathcal{B}$ is defined as the set operation as:

$$A \oplus B = \{ z \mid (\hat{\mathcal{B}})_z \cap A \neq \phi \}$$

where,

ϕ = empty set

$\mathcal{B}$ = structuring element

The dilation of A by $\mathcal{B}$ is the set consisting of all the structuring element origin location where the reflected and translated $\mathcal{B}$ overlaps at least one element of A.

It is c convention in image processing to the

Scanned by CamScanner

first operand of $A \oplus B$ be the image and second operand be the structuring element which usually is much smaller than the image.

Dilation is of two types as:

→ Dilation is associative:

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C$$

→ Dilation is commutative:

$$A \oplus B = B \oplus A$$

Erosion :

Erosion is an operation that thins or shrinks objects in a binary image.

As in dilation, the manner and extent of erosion is controlled by a structuring element.

The erosion process can be done by using following steps:

(i) Representation of original image with rectangular object

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

(ii) Structuring element with 3-pixels arranged in vertical line

$$\begin{matrix} 1 \\ \boxed{1} \\ 1 \end{matrix}$$

← The origin of the structuring element is shown with dark border

(iii) Structuring element translated into several location in the image

1	1	1	1	1	1	$\boxed{1}$
1	1	1	$\boxed{1}$	1	1	1
1	1	1	1	1	1	1

← The result is 0 at this location in the o/p image because all or part of the structuring element overlaps the background

The result is 1 at this location in the o/p image because the structuring element fits entirely within the foreground.

(iv) Output image, the shaded region shows the location of 1's in the original image.

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

Scanned by CamScanner

Now,
the erosion of A by B is denoted as $A \ominus B$ and is defined as:

$$A \ominus B = \{ z \mid (B)_z \subseteq A \}$$

where,
$(B)_z \subseteq A$ means that $(B)_z$ is subset of A.

This equation says that the erosion of A by B is the set of all points z such that B translated by z is contained in A.
Thus, B is contained in A is equivalent to B not sharing any elements with the background of A.
The equivalent expression as the defination of erosion is defined as:

$$A \ominus B = \{ z \mid (B)_z \cap (A)^c = \phi \}$$

Opening & Closing

Opening:
The morphological opening of A by B, denoted as $A \circ B$, is defined as the erosion of A by B and followed by a dilation of the result by B.

ie $A \circ B = (A \ominus B) \oplus B$

Now, an equivalent formulation of opening is:

$$A \circ B = \cup \{ (B)_z \mid (B)_z \subseteq A \}$$

where,
$\cup \{\}$ denotes the union of all sets inside the brakes
$A \circ B$ is union of all translation of B that fits entirely within A.

Closing:
The morphological closing of A by B, denoted as $A \bullet B$ is a dilation of A by B and followed by an erosion of the result by B

ie, $A \bullet B = (A \oplus B) \ominus B$

Geometrically $A \bullet B$ is the complement of the union of all the translation of B that do not overlap A.

Scanned by CamScanner

The opening and closing can be illustrated as:

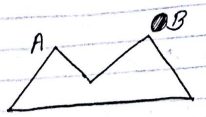


fig (a): Set A & structuring element B

fig (b): Translation of B that fit entirely within set A

fig (c): The complete opening

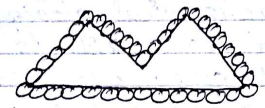


fig (d): Translation of B outside the border of A

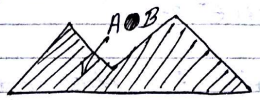


fig (e): The complete closing

(fig: Opening & Closing as Union of Translated Structuring Element)

Chapter VII Image Segmentation

[-7 hours]

Introduction

Image segmentation is a process of partitioning a digital image into multiple regions (set of pixels).

The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyse.

It is typically used to locate objects and boundaries (line and curves) in an image.

The result of an image segmentation is a set of regions that collectively cover the entire image. Each of the pixel in a region are similar with respect to the some characteristics such as ~~com~~ color intensity or texture.

Segmentation accuracy determines the success or failure of computerized analysis procedure.

Some practical applications of image segmentation are as follows:

(i) Medical imaging:
→ Measurement of tissue volumes

→ diagnosis
→ Treatment planning
→ Study of anatomical structure

(ii) Locate object in satellite image

(iii) Face recognition

(iv) Fingerprint recognition

(v) Machine vision

(vi) Iris recognition

Image segmentation algorithm are generally based on two properties:

(a) Discontinuity
(b) Similarity (Continuity)

Discontinuity

This approach is to partition an image based on abrupt changes in intensity such as edges in an image.

There are three basic types of gray level discontinuities in an image. ie, points or spot, lines and edges.

Similarity

This approach is based on partitioning an image into regions that are similar according to predefined criteria such as thresholding, region growing, region merging and splitting.

Types of Image Segmentation on the Property of Similarity

[Segmentation Techniques / Methods]

(i) Region Based Segmentation

(ii) Segmentation Using Region Growing

(iii) Using Region Splitting & Merging

(iv) Segmentation by Thresholding

(v) Basic Global Thersholding

Scanned by CamScanner

1) Region Based Segmentation

The objective of image segmentation is to partition an image into multiple regions. This technique is based on finding the regions directly. To apply this technique, following basic formulation condition must be applied.

Let, R represent the entire image region, segmentation partition R into 'n' sub regions as $R_1, R_2, R_3, \ldots R_n$ such that:

(a) $\bigcup_{i=1}^{n} R_i = R$:

It indicate that every pixel must be in a region.

(b) $R_i \cap R_j = \phi$:

It indicates that the multiple regions must be disjoint.

(c) $P(R_i) = TRUE$:

It indicates that properties must be satisfied by the pixels in a segmented regions.

(d) $P(R_i \cup R_j) = FALSE$:

It indicates that region R_i and R_j are different

2) Segmentation Using Region Growing

Region growing is a technique that groups pixels or sub region into larger regions based on predefined criteria.

Let us pick up an arbitary (seed pixel) pixel (r, c) from the domain of an image to be segmented. This pixel is called seed pixel.

The basic approach is to start with a set of seed points. Then, examine the nearest neighbour (i.e, 4 or 8 neighbour) of (r, c) one-by-one and neighbourhood pixel accepted belongs to the same region as (r, c).

Once, a new pixel is accepted due to the homogeneity property of a region as a member of the current region.

Now, the nearest neighbour of this new pixel are examined. This process goes on recursing until no more pixel is accepted.

Scanned by CamScanner

Then, same procedure is repeated for all seed pixel.

3) Using Region Splitting & Merging

This technique starts with the whole image as one region and splits into sub regions until the sub region satisfy the predefined image homogeneity property.

On the other hand, merge technique starts by taking each pixel in a region and merge small regions into a larger if it satisfy the homogeneity property.

Hence, the hybrid of two approaches i.e, splitting & merging technique is known as region splitting & merging technique.

For this type of segmentation it may take following steps:

(i) In this technique initially an image is subdivided into a set of arbitary region.

(ii) If the formula or condition is satisfied then merge all subregions.

This technique starts somewhere at the middle level.

Suppose, we start with rectangular region of size m*n pixels. To each region Homogeneity property is tested.

If the test fails, the region is splitted into four quadrants each of size $m/2 * n/2$.

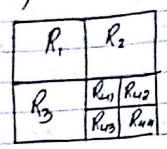
Now, if the region satisfies the homogeneity property then merging process is followed to form a size $2m * 2n$.

So, the procedure is summarized as,

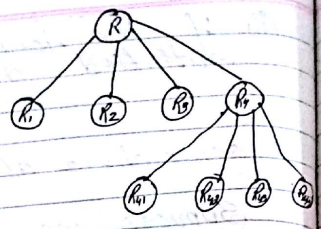
- split into 4 disjoint quadrants any region R_i for which $P(R_i) = FALSE$
- Merge any adjacent regions R_j & R_k for which $P(R_j \cup R_k) = TRUE$
- Stop when no further splitting or merging is possible.

Scanned by CamScanner

Example:



(fig (a): Partitioned Image)

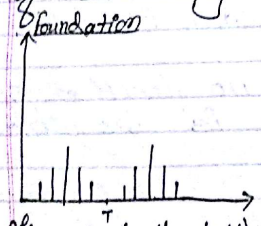


(fig (b): Corresponding Quadtree)

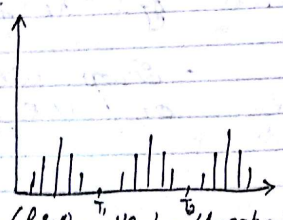
4) Segmentation By Thresholding

An image contains various regions corresponding to different object. The pixel comparing a region and receive information from point of corresponding object.

Hence, thresholding is the simplest method of image segmentation. It plays central role in application of an image segmentation.



(fig(a). single threshold)



(fig(b). multiple threshold)

(fig: Gray level histogram that can be partitioned)

Suppose that, the gray level histogram of an image $f(x,y)$ composed of light object on a dark background as shown in figure above.

In fig(a), it shows to extract the object from the background by selecting threshold 'T' that seperates the object from background then, any point (x,y) for which $f(x,y) > T$ is called an object point otherwise background point.

In fig (b), it shows a slightly more general case of this approach (multiple thresholding). There are two types of light object on a dark background.

Here, multilevel thresholding classifies a point (x,y) as belonging to one ~~of th~~ object class.

It can be expressed as:

If $T_1 < f(x,y) < T_2$ to the other object

if $f(x,y) > T_2$, to the background

if $f(x,y) \leq T_1$, to the background

Scanned by CamScanner

Here, the difference between the first & third row of an 3X3 image region approximates the derivatives in the x direction.

The difference between the first & third column approximates the derivatives in y direction.

c) Slobel operators

In slobel operators higher weights are assigned to the pixels close to the candidate pixel.

So, it is defined as:

$$d_1 = G_x = [(Z_7 + 2Z_8 + Z_9) - (Z_1 + 2Z_2 + Z_3)]$$

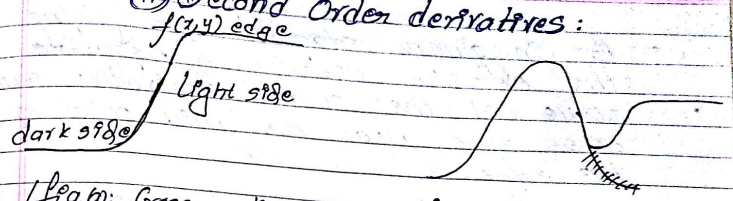
$$d_2 = G_y = [(Z_3 + 2Z_6 + Z_9) - (Z_1 + 2Z_4 + Z_7)]$$

The corresponding 3X3 mask for Slobel operator will be:

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

for d_1 for d_2

(ii) Second Order derivatives:



(fig (a): Cross section of a rampedge) (fig (b): second derivative of ramp edge)

The second derivative is positive at the transition into ie, associated with the dark side of the edge and negative at the transition associated with the light side of the edge ie out of the ramp, and zero along the ramp in area of constant gray level.

The sign of second derivatives can be used to determine whether an edge pixel lies on the dark or light side of the edge. ~~So,~~

So, there are additional properties of the second derivatives around the edge. for second order derivatives ~~four~~ following two points are important:

→ It produces two values of every edge in an image

Threshold can be viewed by operation that involves test against a function T of the ~~figure~~ form,

$$T = T[(x,y), P(x,y), f(x,y)]$$

where,

$P(x,y)$ = Some ~~local~~ local property of the point

$f(x,y)$ = gray level of point (x,y)

A threshold image can be defined as:

$$g(x,y) = \begin{cases} 1 & , \text{ if } f(x,y) > T \\ 0 & , \text{ if } f(x,y) \leq T \end{cases}$$

Thus, if T depends on both $f(x,y)$ and $P(x,y)$, the threshold is called as local and if T depends on spatial co-ordinates it is called dynamic or adaptive.

5) Basic Global Thresholding

In this approach, first partition the image histogram by using a single global threshold 'T'.

Segmentation is then accomplished by scanning the image pixel by pixel and labeling each pixel as object or background depending whether the gray level of the pixel is greater or less than the value of T.

The pixel are assigned and labeled are also assigned. This technique is especialy used in industrial areas where control of illumination is possible.

How to obtain 'T' using global thresholding

(i) Select an initial estimate for T

(ii) Segment the image using T. This will provide two group of pixel ie $G_1 \leq T$ & $G_2 > T$

(iii) Compute the average gray level values μ_1 & μ_2 for the pixel in regions G_1 & G_2.

The average gray level threshold value is given by:

$$T = \frac{1}{2}(\mu_1 + \mu_2)$$

(iv) Repeat the step (ii) through step (iii) until the difference in T in successive iteration is smaller than predefined parameter.

Scanned by CamScanner

Template Maching

It is a technique in DIP for finding small parts of an image that match a template image.

It can be used in manufacturing as a part of quality control as a way to detect edges in an image.

Here, we have a template $g(i,j)$ and we wish to detect its instances in an image.

The incoming image is compared directly to copies (templates) stored in the memory. So, it has various different applications such as: face recognition, fingerprint detection & medical image system.

In this technique, the behind logic is that comparing between pixels or pattern.

This method is also known as linear spatial filtering. This is done only through the use of correlation theorem.

Co-relation Theorem

It is the measure of degree to which two variable agree or not in a actual value.

Here, two variable are the corresponding pixed values in two image i.e template and source image. So, it is used to detect location of a certain object inside an image.

Therefore, correlation coefficient is defined as

$$\text{Co-relation} = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2 \sum_{i=0}^{N-1} (y_i - \bar{y})^2}}$$

where,

x_i = Template gray level image

y_i = Source gray level image

$\bar{x}$ = Average gray level template image

$\bar{y}$ = Average gray level source image

Edge or Boundary Linking & detection

(i) Using Template Matching

(ii) Using Gradient Model (Operator)

(iii) Compass Operator

Scanned by CamScanner

1) Using Template Matching:

Edge detection extracts edges of object from an image. Edge characterize object boundaries and therefore it is useful for segmentation, registration and identification of an object in the scene.

The variation of an image features usually brightness to edges. So, the edges are the representation of discontinuities of an image intensity function.

So, edge detection algorithm is essentially a process of detection of discontinuities in an image.

Hence, the change in brightness level indicates edges. Edge detection by using template matching can be derived as follows:

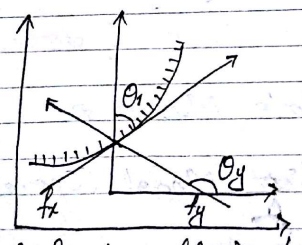


fig: Gradient of $f(x,y)$ along x-direction using template matching

For the continous image $f(x,y)$, its derivaties assumes a local maximum in the direction of edge. Therefore, measuring the gradient of f can be used for edge detection. It can be derived as:

Gradient of horizontal direction, $\Delta f = \left[\frac{df}{dx}, 0\right]$

Gradient of vertical direction, $\Delta f = \left[0, \frac{df}{dy}\right]$

Gradient in θ (any) direction, $\Delta f = \left[\frac{df}{dx}, \frac{df}{dy}\right]$

Now,

$$\frac{df}{dr} = \left[\frac{df}{dx} \cdot \frac{dx}{dr} + \frac{df}{dy} \cdot \frac{dy}{dr}\right]$$

$$\frac{df}{dr} = f_x \cos\theta + f_y \sin\theta$$

Then, the maximum value of $\frac{df}{dr}$ is obtained when,

$$\frac{d}{d\theta}\left(\frac{df}{dr}\right) = 0$$

~~Now,~~

Scanned by CamScanner

$$\frac{f_x \sin\theta_g}{\cos\theta_g} + \frac{f_y \cos\theta_g}{\cos\theta_g} = 0$$

$$f_x \tan\theta_g + f_y = 0$$

$$\tan\theta_g = -\frac{f_y}{f_x}$$

$$\theta_g = \tan^{-1}\left(\frac{f_y}{f_x}\right)$$

Hence,

$$f_x \sin\theta_g + f_y \cos\theta_g = 0$$

$$\boxed{\theta_g = \tan^{-1}\left(\frac{f_y}{f_x}\right)} \text{ (direction)}$$

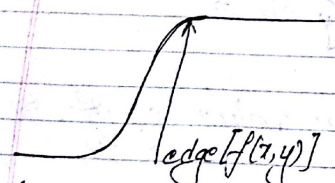
$$\boxed{\therefore \left(\frac{df}{dx}\right)_{max} = \sqrt{f_x^2 + f_y^2}} \text{ (gradient value)}$$

where,

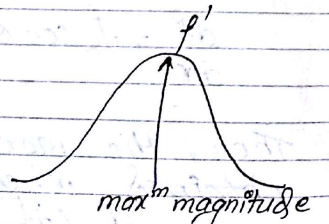
θ_g is the direction of edge & $\left(\frac{df}{dx}\right)_{max}$ is the gradient value that indicates the edges.

2) Using Gradient Model (Operator):

(B) First derivative (i.e difference Operator)



fig(a): Cross-section of an image edge



fig(b): Cross-section of an image edge using first derivative

The derivatives which yield high values at places where gray level changes rapidly are used to find gradient of an image as shown in figure above.

The figure above shows that first derivative produces thicker edge because the first derivative is positive at the point of transition into & out of the ramp as we move from left to right.

If (df/dx) & (df/dy) are the rates of changes of two dimensional function $f(x,y)$ along x & y-axis then the direction in which rate of change has the greatest magnitude is defined as:

$$\theta = \tan^{-1}\left(\frac{df/dy}{df/dx}\right)$$

and,

$$\text{magnitude} = \sqrt{\left(\frac{df}{dx}\right)^2 + \left(\frac{df}{dy}\right)^2}$$

The vector having this magnitude & direction is called the gradient of $f(x,y)$ & is denoted $f'(x,y)$ for digital image.

Scanned by CamScanner

First derivative is also called as difference operator which is used to find the differences like:

$f(r,c) - f(r-1,c) = d_1$

$f(r,c) - f(r,c-1) = d_2$

$\therefore$ The magnitude $f'(r,c) = \sqrt{d_1^2 + d_2^2}$

and, the direction of the greatest step is $\theta(r,c) = \tan^{-1}\left(\frac{d_2}{d_1}\right)$

There are various gradient operator only using first derivate. They are as follows:

(a) Roberts Operators
(b) Prewitt Operators
(c) Sobel Operators

Let the 3x3 area shown in figure below represents the gray level in a neighbourhood of an image

$$\begin{bmatrix} Z_1 & Z_2 & Z_3 \\ Z_4 & Z_5 & Z_6 \\ Z_7 & Z_8 & Z_9 \end{bmatrix}$$

(fig: 3x3 Image)

a) Roberts Operators

It is defined as follows:

$$d_1 = G_x = \frac{df}{dx} = Z_9 - Z_5$$

$$d_2 = G_y = \frac{df}{dy} = Z_8 - Z_6$$

The derivation can be implemented for an entire image by using masks, ~~mask~~ is given by:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} Z_1 & Z_2 & Z_3 \\ Z_4 & -1 & 0 \\ Z_7 & 0 & 0 \end{bmatrix}$$

Here, the masks of 2x2 are awkward because they do not have a clear center

b) Prewitt Operators

It is defined as follows:

$$d_1 = G_x = [(Z_7 + Z_8 + Z_9) - (Z_1 + Z_2 + Z_3)]$$

$$d_2 = G_y = [(Z_3 + Z_6 + Z_9) - (Z_1 + Z_4 + Z_7)]$$

The ~~8~~ prewitt operator mask will be,

for d_1, $\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ $\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$ for d_2

Scanned by CamScanner

→ An imaginary straight line joining the extreme positive & negative value of the second derivative would cross zero near the mid point of the edge.

Second order derivatives is derived by the Laplacian filter or operator.

The Laplacian of a 2D function (x,y) is a second order derivative defined as:

$$\Delta^2 f = \frac{d^2 f}{dx^2} + \frac{d^2 f}{dy^2}$$

i.e $$\Delta^2 f(x,y) = \frac{d^2 f(x,y)}{dx^2} + \frac{d^2 f(x,y)}{dy^2}$$

For a 3x3 region, one of the two features encountered most frequently as:

a) $\Delta^2 f = 4Z_5 - (Z_2 + Z_4 + Z_6 + Z_8)$

The Laplacian mask is:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

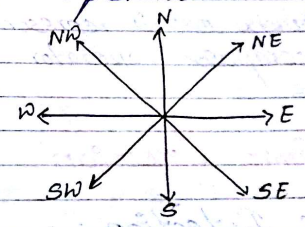
b) A digital approximation including the diagonal neighbours is given by:

$\Delta^2 f = 8Z_5 - (Z_1 + Z_2 + Z_3 + Z_4 + Z_6 + Z_7 + Z_8 + Z_9)$

The Laplacian mask will be:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

3) Compass Operator



(fig: compass operator)

Compass operators measure the gradients in a selected number of direction.

The matrix representation for different direction is:

(i) N ↑ $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$ ← W

Scanned by CamScanner

(iii) $\downarrow S$ $\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ (iv) $\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \rightarrow E$

(v) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \uparrow NW$ (vi) $\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} \rightarrow NE$

(vii) $\begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \leftarrow SW$ (viii) $\begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \rightarrow SE$

(fig: directionwise matrix representation of compass operator)

Line & Spot detection

Line/Curve Detection:

It is an important step in image processing and analysis. Lines are features in any scene taken by sensing devices.

We can detect the line by using following compass gradient i.e 4 line detection kernel as follows:

(i) $\begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$ $\rightarrow$ Horizontal

(ii) $\begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix}$ $\longrightarrow +45^\circ$

(iii) $\begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$ $\rightarrow$ Vertical

(iv) $\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$ $\longrightarrow -45^\circ$

Spot Detection:

These are most easily detected by comparing the value of a pixel with an average or median of the neighbourhood pixel.

The point detection mask or spot is given by:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

(fig: spot detection kernel)

The point has been detected at the location on which the mask is centered if $|R| \geq T$.

Scanned by CamScanner

where,
T = non-negative threshold
R = Response of mask at any point in image

The response R is given by:

$$R = \sum_{i=1}^{n} w_i z_i$$

where,
w_i = window or mask coefficient
z_i = gray level of ith pixel

The idea is that a spot will be detected easily from its surrounding by this method.

Chapter VIII

Chapter VIII Representation & Description

[~3 hours]

Introduction to Representation

The segmentation technique is a decomposition of an image generally referred to as raw data in the form of pixels along a boundary contained in a region.

Although, these data sometimes are used directly to obtain descriptors that compact the data into representation that are considerably more useful in the computation of descriptors.

So, representation refers to the dealing of naming for segmented image dedicated to the object point.

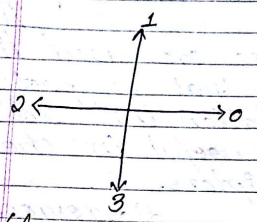
Representation can be classified into two types:
(i) Chain Codes
(ii) Signatures

1) Chain Codes:

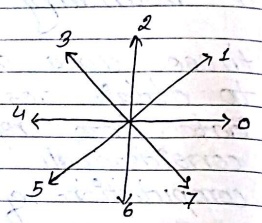
Chain codes are used to represent a boundary by a connected sequence of straight line segments of specified length and direction.

Typically, this representation

is based on four or eight (4 or 8) connectivity of the segments. The direction of each segment is coded by using a number. Chain code on this scheme are referred as:



(fig (a): Direction number for 4 directional chain code)

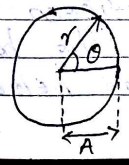


(fig (b): Direction number for 8 directional chain code)

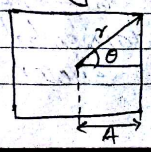
The chain code of a boundary depends on the starting point. However, the code can be normalized with respect to the starting point by treating it as a circular sequence of direction numbers & redefining the starting point so that the resulting sequence of numbers is a integer magnitude. We can normalize by using rotation i.e 90° or 45° for the codes using the difference of the chain code. Above figure shows the best example of 4 and 8 directional chain code.

2) Signatures:

Signature is a 1-D functional representation of a boundary and may be generated in various ways. One of the simplest method is plot the distance from an centroid point to the boundary as a function of angle. Signature generated by the approach just described are invariant to translation but they depend on rotation and scaling. Normalization with respect to the rotation can be achieved by finding a way to select the same starting point to generate the signature. The signature can be defined by using the following object:



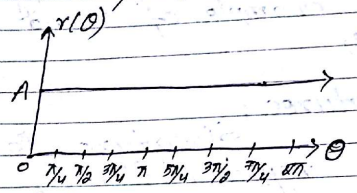
(fig (a): Circular object)



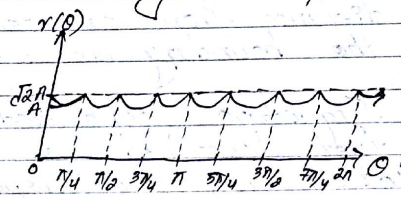
(fig (b): square objects)

Scanned by CamScanner

The corresponding angle signature can be defined as:



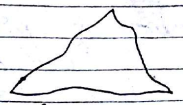
(fig (c): Corresponding distance Vs angle signatures of circular objects)



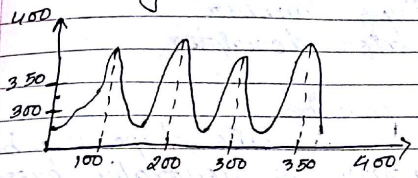
(fig (d): Corresponding distance Vs angle signature of square objects)

A ~~single~~ similar set of commands yielded the plot. Simply counting the number of prominent peaks in the signature is sufficient to differentiate between the fundamental set of the boundaries as shown in figure below:

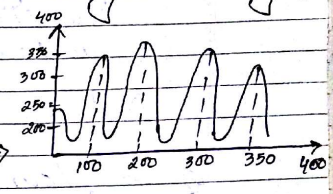
(fig(a): Boundaries of an irregular square)



(fig(b): Boundaries of an irregular triangle)



(fig(c): Corresponding signature of an irregular square)



(fig(d): Corresponding signature of an irregular triangle)

Introduction to descriptors

Descriptors are also called features of an image that are useful when working with region boundaries.

Many of the descriptors are applicable to region and grouping of descriptors in the toolbox. The length of a boundary is one of its simplest descriptors.

The length of a four connected boundary is defined as the

number of pixel - 1. Descriptor is usually categorized into two types based on the boundary. There are following types of descriptions:

(a) Shape Number
(b) Fourier descriptor

1) Shape number

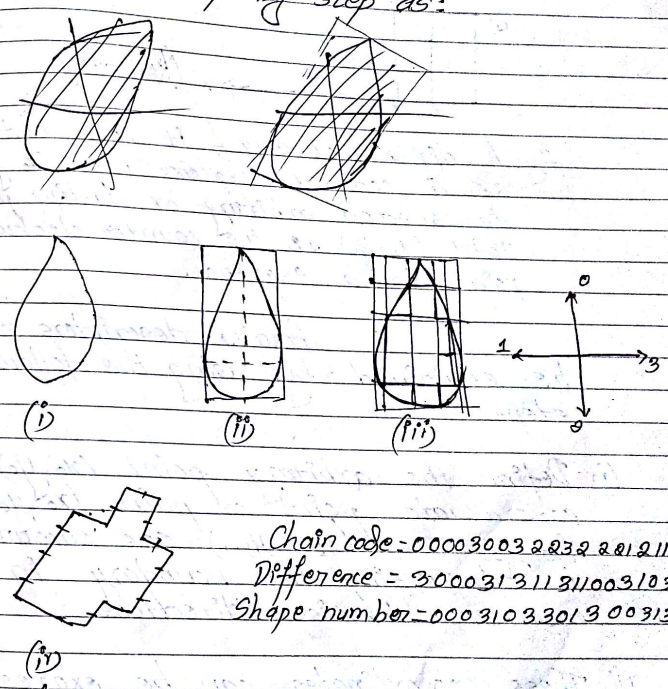
The shape number of a boundary is generally based on four directional chain codes. It is defined as the first difference of a chain code of the smallest magnitude. The order n of a shape number is defined as the number of digits in its representation.

Thus, the shape number of a boundary is given by the parameter and order of shape number (length).

The 4 directional chain codes can be made insensitive to the starting point by using the integer of minimum magnitude & made rotation that are multiples of 90° by using the first difference of the code.

Thus, shape numbers are insensitive to the starting point & rotation that are multiple of 90°.

Shape number is illustrated step by step as:



Chain code = 0000300322322201211
Difference = 3000313113110031031 0
Shape number = 0003103301300313 03

(fig: steps in the generation of shape number)

many dendrites which branch out tree like structures and receive signals from other neurons. The axon conducts electrical signals generated at the axon hillock.

Neural Network & Artifical Neural Network

Neural network is the ~~brach~~ branch of ~~the~~ AI that can be considered as blackbox ~~ie a~~ that is able to predict the output pattern when it recognizes the given input pattern.

Artifical Neural network (ANN) is the type of neural network which is inspired ~~as~~ artificially to show the human behavior.

ANN shows the brain process information system.

Q Why use neural network (NN)?

⇒ Neural network is only used for following:

(a) Self organization
(b) Real Time Operation
(c) Information is distributed over the entire network
(d) Adaptive Learning.

Applications of Neural Network

(i) Automotives, automobile and automatic guidance system

(ii) Aerospace and aircraft component simulation.

(iii) Banking cheque ~~return~~ reader

(iv) ~~Difference~~ Diff facial recognition

(v) Manufacturing product design (CAD)

(vi) Telecommunication Automated Information

(vii) Robotics Control System

Disadvantages of NN

(i) NN needs training to operate & lack of manpower.

(ii) The architecture of NN is different from the architecture of microprocessor.

(iii) Requires high processing time for large neural network

(iv) Hard to setup program.

Scanned by CamScanner

d) Fourier description

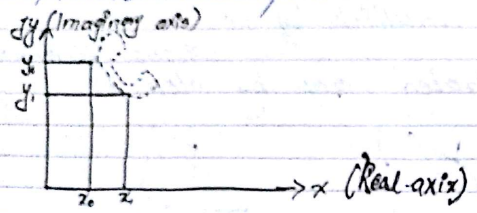


fig: A digital boundary & its representation as a complex sequence point (x_0, y_0) is selected arbitary or starting point. Point (x_1, y_1) is the counter clockwise point in a sequence.

Fourier descriptors can be explained by using the following steps:

(i) Define the arbitary point (x_0, y_0) & co-ordinate point or pairs $(x_0, y_0), (x_1, y_1), (x_2, y_2) \ldots (x_{k-1}, y_{k-1})$ are encountered in traversing the boundary in counter clockwise direction

(ii) These co-ordinates can be expressed in the form:

$$x(k) = x_k \quad \text{and} \quad y(k) = y_k$$

~~and~~

The boundary can be represented as the sequence of co-ordinates as:

$$s(k) = [x(k), y(k)], \text{ for } k = 0, 1, 2, \ldots, k-1$$

(iii) The each co-ordinate pair can be treated as a complex number so that

$$s(k) = x(k) + jy(k)$$

(iii) The discrete fourier transform of 1D sequence, $s(k)$ can be written as

$$a(u) = \sum_{k=0}^{k-1} s(k) \cdot e^{-\frac{j2\pi u k}{k}}$$

where,

$$u = 0, 1, 2 \ldots k-1$$

(iv) The complex coefficient $a(u)$ are called "the fourier descriptors" of the boundary. The inverse fourier transform of these coefficients as:

$$s(k) = \frac{1}{k} \sum_{u=0}^{k-1} a(u)\, e^{\frac{j2\pi u k}{k}}$$

(v) Therefore, for boundary descriptors, the fourier descriptors can be explained as fourier coefficients that are defined

Scanned by CamScanner

as:

$$\hat{s}(k) = \frac{1}{p} \sum_{u=0}^{p-1} a(u) e^{\frac{j2\pi u k}{K}}$$

where,

p = preceeding eq^n for $a(u)$ ~~where p=0~~

where $p = 0, 1, 2 \ldots K-1$

$\hat{s}(k)$ = Approximation to $s(k)$

Chapter IX

Object Recognition

(28)

[-3 hours]

Classification or learning of an Image or [Decision theoretic methods]

Object Recognition & Classification

1. Unsupervised classification (No target or teacher or sample signal)

2. Supervised Classification (presence of target or teacher or sample signal)

(i) Statistical supervised classification (Baysian Classification)

(ii) Distribution free (Neighbourhood classification)

fig: Decision Theoretic Methods

Unsupervised Classification

A very common task concerning in image processing is classification is done in order to use the image ~~fro~~ for mapping or futher analysis.

Scanned by CamScanner

In unsupervised classification learning, there is no teacher or target or sample signal. In the same way, it is also computerized method without direction of analysis. It means that the output are based on software without providing user signal or sample classes.

The computer uses the technique to determine which pixel is detected and what classes belongs together.

In this technique the user or human does not interface to the unsupervised classification.

Cluster is one of the unsupervised learning technique or classification.

for unsupervised classification the figure below shows the technique of object recognition:

101	103	105
101	101	105
101	101	101
103	101	101

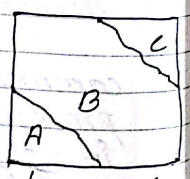
I/p Image

⟹

B	B	C	C
B	B	B	C
A	B	B	B
A	A	B	B

(A = water
B = land
C = rock)

⟹



class classification

Example of unsupervised learning using clustering

Supervised Classification

In supervised classification, it does not use the computer software to create the classes.

Here, the analysis identify several areas in an image which represents different features. These known areas are referred to as training sites.

The computer only do the assignment of pixel to the classes. It is based on the idea that a user can select sample pixel or target signal in an image i.e. representative of specific classification.

It means that the provided user signal or teacher signal can help to make the classification without using computer tool.

Supervised classification is categorized into two types:

(i) Statistical Supervised Classification

(ii) Distribution free

Scanned by CamScanner

In unsupervised classification learning, there is no teacher or target or sample signal. In the same way, it is also computerized method without direction of analysis. It means that the output are based on software without providing user signal or sample classes.

The computer uses the technique to determine which pixel is detected and what classes belongs together.

In this technique the user or human does not interface to the unsupervised classification.

Cluster is one of the unsupervised learning technique or classification.

for unsupervised classification the figure below shows the technique of object recognition:

101	101	105	105
101	101	101	105
103	101	101	101
103	103	101	101

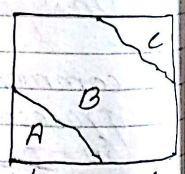
I/P Image

⟹

B	B	C	C
B	B	B	C
A	B	B	B
A	A	B	B

(A = water
B = land
C = rock)

⟹



class classification

fig: Example of unsupervised learning using clustering

Supervised Classification

In supervised classification, it does not use the computer software to create the classes.

Here, the analysis identify several areas in an image which represents different fitures. These known areas are refered to as training sites.

The computer only do the assignment of pixel to the classes. It is based on the idea that a user can select sample pixel or target signal in an image i.e. representative of specific classification.

It means that the provided user signal or teacher signal can help to make the classification without using computer tool.

Supervised classification is categorized into two types:

(i) Statistical Supervised Classification

(ii) Distribution free

(i) Statistical Supervised Classification

It is a type of supervised classification on which the hypothetical or probability of that classification is determined using presence of user signal (or sample signal or teacher signal or target signal).

This still can be extended to classification of features or pattern where similar type of pattern are grouped into one classes that same as the characteristic of user signal.

The term similar in the containing objects is closeness to the features between sample signal and segmented region.

To justify this classification we use Bayesian classifier.

* Bayesian Classifier

A Bayesian classifier is a simple and probabilistic classifier based on the Bayesian theorem.

The Bayesian classifier classifies features to contribute the probability of hypothetical evidence ie.

$$P(H_i/E) = \frac{P(E/H_i)\,P(H_i)}{\sum_{n=1}^{k} P(E/H_n)\,P(H_n)}$$

where,

$P(E/H_i)$ = The probility that we will observe evidence 'E' given that hypothesis 'H_i' is true.

$P(H_i)$ = A prior probability that the hypothesis 'H_i' is true.

K = The no. of possible hypothesis

$P(H_i/E)$ = The probability that we will observe hypothesis 'H_i' given that hypothesis 'E' is true.

(ii) Distribution free

It is a type of supervised classification & do not require knowledge of any density function on ~~pd~~ pdf based on the heuristic information or algorithm.

The example of distribution free is nearest neighbourhood

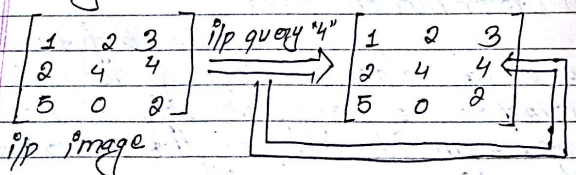
classification, so that it is also called as neighbourhood classification.

A very commonly used classification method is called nearest neighbourhood classification, therefore it works as follows:

→ We create a database of any object for which we already know that the correct classification should be known.

When the system is given a query ie new object classify the system. It finds the nearest neighbour of the query in the database.

→ Then, the system classify the query as belongs to the same classes as its nearest neighbours ie database images.



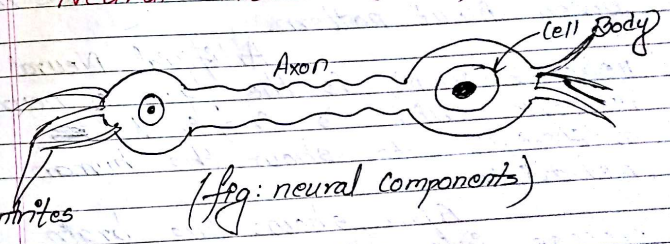
(fig: Recognizing digit using neighbourhood classifier)

→ Then, the system classify the query as as image of 4.

→ Now, the image is identified and classified by the distribution free supervised classification.

Here query 4 is the target signal by the user.

Neuron, Neural Network (NN) & Artificial Neural Network (ANN)



(fig: neural components)

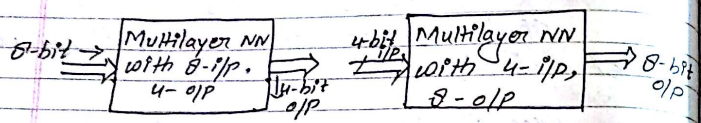
A neuron is a cell in the brain whose principal function is the collection, processing and dissemination of electrical signal.

The cell body is the part of the cell containing the nucleus and maintaining protein synthesis.

A neuron may have

Application of NN in Image Processing

1) NN for Image Compression



(fig: block diagram of Image Compression)

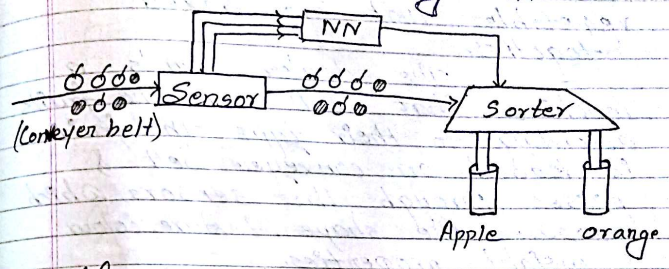
NN model have received much attention in many fields where high compression ratio is required.

Many NN approaches for image compression gives superior performance than discrete traditional approach.

Digital image generated with 8 bit may be reduced in size by ~~pred~~ feeding it to 8x4 multilayer neural network.

NN on the other side then ~~com~~ process the compressed image by 4x8 multilayer NN for reconstruction.

2) NN for Pattern Recognition



(fig: block diagram of pattern recognition (PR))

The NN is ideal tool for pattern recognition. Any recognition system needs to be trained to recognize different patterns.

NN is also a simplification of human neural network system. It is more likely to adapt the human way of solving the recognition problem than other techniques.

The design of NN system for pattern recognition starts from collecting data on each of objects that is to be recognized by the system.

A dealer has a warehouse

that stores a variety of fruits & vegetables that are mixed together.

The dealer wants a machine that will sort the fruits according to their type. The fruit is loaded on conveyer belt & passes through the sensors which measures the shape, texture, colour & weight properties.

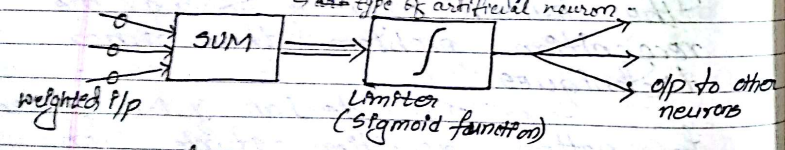
The o/p will be input to the purpose of NN ~~is~~ to decided what kinds of fruits are on conveyer belt.

So, fruit can be detected & directected to correct store.

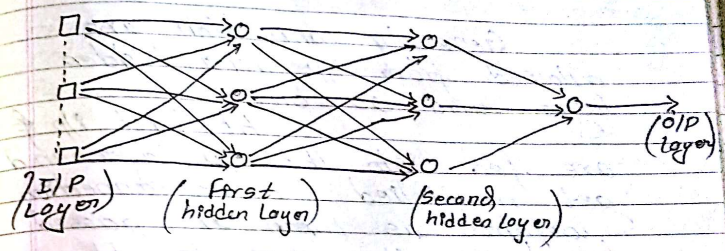
(computerized model to represent or simulate the ability of brain to recognize and discriminate)

3) NN for perceptron (perception)

↳ ~~a~~ type of artificial neuron.



(fig: Block diagram of perceptron)



(fig: Formation of o/p by perceptron)

It is one of the earliest neural network model. Here, inputs from one or more previous neurons are individually weighted and then summed.

The result is non linearly scaled between 0 & +1. Then, the result is passed to the neurons in the next layer for output result.

Several perceptron can be combined to form multilayer perception (MLP).

So, MLP is a development from the simple perception in which extra hidden layers are added.

Here, more than one layer can be used.

Scanned by CamScanner

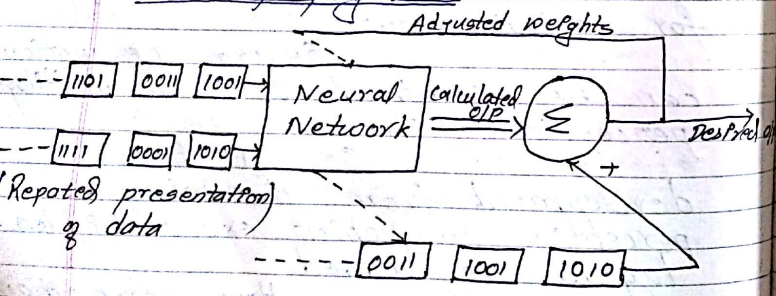
Generally, connection are allowed from i/p layer, hidden layer and so on.

At first i/p are fed into the i/p layer and get multiplied by interconnection weights as they are passed from the i/p layer to the first hidden layer.

After first layer they go the second layer & finally data is multiplied by interconnection weights & so on.

Back propagation never take the perception methods but at the time of sensing accuracy it is used.

<u>Back propagation</u>



(fig: Block diagram of Back Propagation)

The MLP (multilayer Perceptron) & many other neural network can be calculated using an algorithm, this is known as back propagation.

In back propagation the i/p data is repeatedly presented to the neural network. With each presentation of, the output of neural network is compared to the desired output & error is computed.

Then this error is passed as a feedback (back propagate) to the neural network.

It is used to adjust the weight such that the error decreases with each iteration. Now, the neural network calculates output and gets closer to produce the desired output.

This mechanism is known as back propagation.

Back propagation is mostly used to find the sensing accuracy which is shown in the above figure.

Scanned by CamScanner

Chapter X Pattern Recognition (PR)

[1 hour]

Introduction

A pattern is an arrangement of descriptors. Pattern is something that is set up an ideal to be followed.

<u>Example</u>: A ~~set~~ shape "A" is shown to a child & asked to copy that. Hence, that shape is ideal. The child draws the shape close to that shape. His teacher identifies as a 'A'. This step is known as recognition.

Pattern Recognition

It is defined as the classification of input data through the extraction of important features from a lots of ~~in~~ noisy data.

<u>Example</u>: PR involves the identification of faces, objects, words and melodies, etc.

Pattern vectors are represented by following representation method:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where,

n = total no. of descriptors

X = x-descriptors

$x_1, x_2 \ldots x_n$ = each component of descriptors

Applications of Pattern Recognition

PR has been applied in many fields such as Artificial Intelligence, Computer Engg., medicine, image analysis, etc. Some major applications of PR are as follows:-

(i) Computer Aided Diagnosis (CAD)

It helps in medical imaging such as ~~EG~~ ECG, EEG, X-Ray, etc

(ii) Handwriting Recognition

PR helps in handwriting recognition such as processing bank cheque and pen based input on PR.

Scanned by CamScanner

(iii) Speech Recognition

It is also used in speech recognition such as speech signal is classified into phonens and words.

(iv) Safety (Biometric Identification)

PR also used for biometric identification such as face, fingerprint & iris recognition.

(v) Astronomy

It is also used in astronomical fields such as classifying the galaxy by shape & telescope.

(vi) Agriculture

It is also used in agricultural fields such as soil evaluation and extraction of minerals.

(vii) Geography

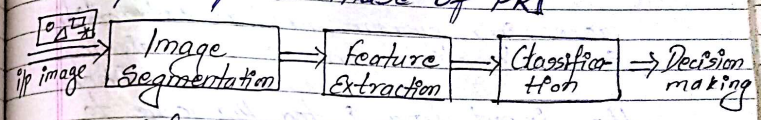
It is used for rock classification & identification in geography.

(viii) Engineering

It can be used for diagnosis of vehicle system and recognition of automobile.

Components of PR

[Imp. Steps or Phase of PR]



(fig: 3-phase of PR)

The phases for pattern recognition can be summarized as follows:

(i) Image Segmentation

The first phase of pattern recognition is segmentation in which the image is isolated from the rest of image.

Segmentation refers to the process of partitioning a digital image into multiple regions (set of pixels).

So, the goal of segmentation is to simplify and change the representation of an image into

meaningful information & easier to analyse.

It is used to locate object and boundaries.

(ii) Feature Extraction

In PR and Image processing system, feature extraction is a special form of dimensioning. When the input data is too large to be processed and is suspected to be redundant, then the input data will be transformed into a reduced & representable set of features. Thus, transforming the input data into the set of features is known as feature extraction.

Hence, objectives of feature extraction is data reduction by transformation method with a certain features.

Example: Spatial features, transform features, edge & boundaries features, shape features, moment features, texture features, color features, etc.

(iii) Classification

It is the mechanism to classify pattern according to the extracted features.

It is based on the availability of the set of pattern that have already been classified which is termed as training set.

(Design Cycle of PR)

Data Collection
⇓
feature selection
⇓
Classifier design
⇓
System evaluation
⇓
(output)

(fig: Classification Process)

S.N

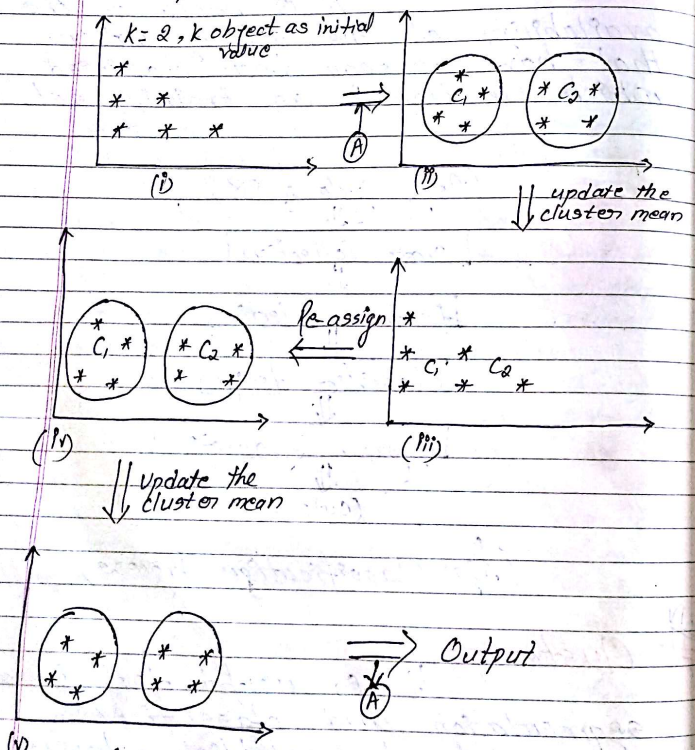
Clustering

It is useful for image segmentation and classification of raw data to establish classes and prototypes.

Scanned by CamScanner

→ Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

One of the most popular clustering technique is K-mean algorithm. Clustering is implemented as:



[fig: The process of clustering model (k-mean)]

Extra Material

Differences between DIP & CG (Computer Graphics)

DIP	CG (Computer Graphics)
(i) Emphasis is on analysis and enhancement of image.	(i) Emphasis is on synthesizing new image from geometry, lighting parameters, materials & texture.
(ii) It is a vector and raster technique.	(ii) It is a raster technique.
(iii) Example: zooming & histogram technique.	(iii) Example: 2D & 3D paging technique.

→ K-means clustering aims to partitions n observations into K-clusters ~~analysis in do~~ in which each observation belongs to cluster with the nearest mean, serving as a prototype of the cluster.

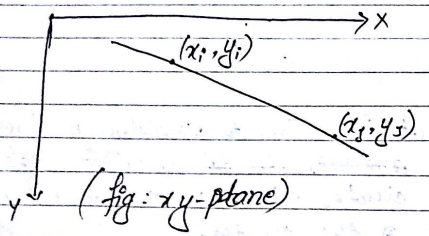
Scanned by CamScanner

Hough Transform

What is Hough transform & how it can detect the line edge?

⇒ It is a powerful global method for detecting edge. It transform between ~~partition~~ Cartesian space and a parameter space in which a straight line formulation can be defined.

Let us consider the case where we have straight line between two space on which number of pixel is placed.



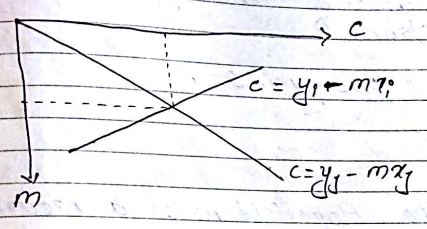
(fig: xy-plane)

We first note that for every point (x,y) in that image. All the straight line passing through the point satisfy equation (i) for varying values of lines slope & intercept (m, c)

i.e. $y_i = mx_i + c$ --- (i)

Then, inversely lines through the point in a Cartesian domain as:

$c = y_i - mx_i$ --- (ii)



(fig: parameter & Cartesian space)

Similarly, for j^{th} point,

$c = y_j - mx_j$ --- (iii)

Now, using $y = mx + c$ to represent line so that we use angular representation of line as:

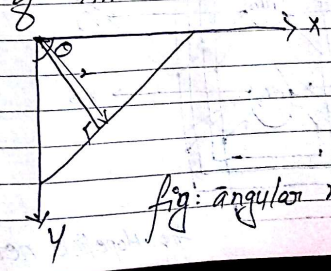


fig: angular representation of straight line

Feature extraction technique used in image analysis, computer vision & digital image processing done by global technique called Hough transform. It is a very short method to detect the edge of line by the following expression.

$$x\cos\theta + y\sin\theta = r$$

V.Imp

Hopefield Nets

Does the Hopefield nets a recurrent flow of ANN?

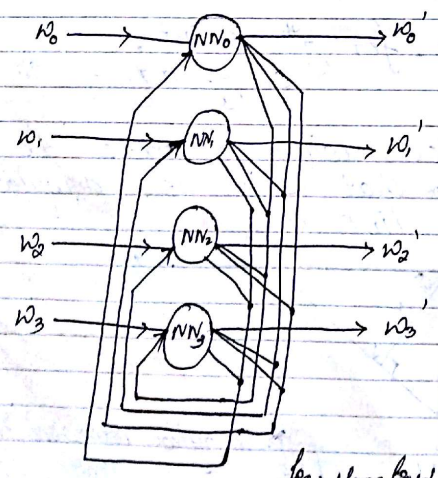


fig: Hopefield nets

It is the recurrent form of Artificial Neural Network designed by Hopefield on the basis of frequency domain.

The ~~limits~~ units in Hopefield nets are binary threshold units i.e the units only trained on two different values for their state and the value is determined by the threshold unit. Hopefield nets can have either values of 1 or -1, or values of 1 or 0. So, the two possible defination for unit is activation (a_i).

i.e,

$$a_i \leftarrow \begin{cases} 1 & , \text{if } \sum_j w_{ij} s_j > \theta_i \\ -1 & , \text{otherwise} \end{cases}$$

$$a_i \leftarrow \begin{cases} 1 & , \text{if } \sum_j w_{ij} s_j > \theta_i \\ 0 & , \text{otherwise} \end{cases}$$

where,

w_{ij} = strength of the connection weight from unit j to i (weight of the connection)

s_j = State of unit j

θ_i = Threshold of unit i

Scanned by CamScanner

The connection in Hopefield nets typically have following restrictions:

(i) ~~the~~ $W_{ii} = 0, \forall i$ (No unit has a connection with itself)

(ii) $W_{ij} = W_{ji}, \forall ij$ (Connections are symmetric)

Hopefield nets have a scalar value associated with each state of the network referred to as the energy E given by:

$$E = -\frac{1}{2} \sum_{ij} W_{ij} S_i S_j + \sum_i \theta_i S_i$$

Hamming nets

A form of recurrent artificial neural networks is invented by R.W. Hamming based on frequency domain.

It is the most useful channel encoding techniques through the neural networks. It is based on connecting number of bits to the data been encoded to ensure that some minimum number of bits must change between the valid code.

Hamming nets encoded the channel as:

$$h_1 = b_3 \oplus b_2 \oplus b_0$$
$$h_2 = b_3 \oplus b_1 \oplus b_0$$
$$h_4 = b_2 \oplus b_1 \oplus b_0$$
$$h_3 = b_3$$
$$h_5 = b_2$$
$$h_6 = b_1$$
$$h_7 = b_0$$

Hamming net shows that if 3 bits of redundancy are added to a 4 bit word, so that distance between any two valid code word is 3 and all single or multiple bit error can be detected and corrected.

~~Above~~ Above calculation shows that the 7 bit hamming code word $h_1, h_2, h_3 \ldots, h_7$ associated with 4 bit binary number b_3, b_2, b_1 & b_0. Hence, hamming nets encode the channel through ANN basis of frequency domain by checking parity of word. Therefore, hamming nets plays an important role as a recurrent formation for the ANN (Artificial Neural Network).

Scanned by CamScanner

2014 fall.
, 1.(a).

101 ⇒ 8 bit gray scale = 1 byte per pixel.
16 " " " = 2 byte per pixel.

1 bit line are = 1/8 byte per pixel.

∴ (800 X 600) is a pixel.

∴ storage for 800X600 ~~pixel~~ with 8 bit gray value is = 800 X 600 X 1
= 480000 bytes.
= $\frac{480000}{1024}$ KB
= $\frac{480000}{1024 \times 1024}$ mb
= 0.4577 MB.

∴ 0.4577MB = 1 image.
1MB = $\frac{1}{0.4577}$ image
= ~~2.02~~ = 2.1848 image.
∴ 512 mb = 2.1848 X 512
= 1118.63 image.

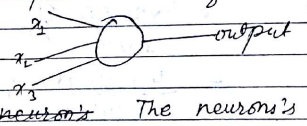
Note: 1 frame = 512X512 pixel.
⇒

Perceptrons:-

- Perceptrons is a ~~type~~ of artifical neural network neuron similar to biological neuron.
- Earliest neural network model developed in 1950 and 1960s. ~~by scientists~~.
- Today, it's more common to use other models of artifical neurons, eg. Sigmoid neuron.

Working.

- A perceptron takes several binary inputs $x_1, x_2, \ldots$, and produces a single binary output.



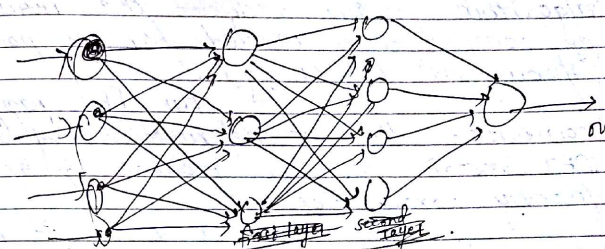
- ~~The neuron's~~ The neurons's output 0 or 1 is determined by whether the weighted sum $\sum_j w_j x_j$ is less than or greater than some threshold value.
- In order to compute the output, there are weights w_1, w_2, w_3, etc, real numbers expressing the importance of the respective inputs to the outputs.
- threshold is a parameter ~~to~~ real no. which is a of the neuron.

To put it in more precise algebraic terms:

$$\text{output} = \begin{cases} 0 & \text{if } \sum_j w_j x_j \leq \text{threshold} \\ 1 & \text{if } \sum_j w_j x_j > \text{threshold} \end{cases}$$

- ~~OR~~ We can think about the perceptron is that it's a device that makes decisions by weighing up evidence.

- The following figure illustrates how a perceptron can weight up different kinds of evidence in order to make decisions.



- In this network, the first column of perceptron - what we'll call the first layer of perceptrons - ~~what we'll call the~~ is making three very simple decisions, by weighing the input evidence.
- The second layer of perceptrons is making a decision by weighing up the results from the first layer of decision-making.
- In this way, a many-layer network of perceptrons can engage in sophisticated decision making.
- Perceptrons can be used is to compute the elementary logical functions we usually think of as underlying computation, function such as AND, OR and NAND.

Scanned by CamScanner

Chap-3 Continue...

- first rearrange the input element in bit-reversal order
- builds the output transform

Fast Fourier Transform (FFT)

The Fast Fourier Transform is a discrete fourier transform algorithm which reduces the number of computations needed for N points from $2N^2$ to $2N\log_2 N$.

An FFT rapidly computes the transformations by factorizing DFT matrix into a product of sparse (mostly zero) factors.

As a result, it manages to reduce the complexity of computing the DFT from $O(n^2)$ to $O(n\log_2 n)$, when n is the data size.

The FFT algorithm generally fall into two classes:
- decimation in time
- decimation in frequency.

Cooley-Tukey FFT algorithm

The basic idea for FFT (decimation in time) is to break up a transform of length N into two transforms of length N/2 using the identity as:

$$\sum_{n=0}^{N-1} a_n e^{-\frac{2\pi i n k}{N}} = \sum_{n=0}^{N/2-1} a_{2n} e^{-\frac{2\pi i (2n)k}{N}} + \sum_{n=0}^{N/2-1} a_{2n+1} e^{-\frac{2\pi i (2n+1)k}{N}}$$

$$= \sum_{n=0}^{N/2-1} a_n^{even} e^{-\frac{2\pi i k}{N/2}} + \left(\sum_{n=0}^{N/2-1} a_n^{odd} e^{-\frac{2\pi i k}{N/2}}\right) e^{-\frac{2\pi i k}{N}}$$

i.e., $X_k = E_k + e^{-\frac{2\pi i}{N}k} O_k$

It is sometime called Danielson-Lanczos Lemma.

The easiest way to visualize this procedure is perhaps via the Fourier matrix.

Applications

(i) Fast large integer & polynomial multiplication
(ii) Filtering algorithms
(iii) Solving difference equations
(iv) Efficient matrix-vector multiplication

Scanned by CamScanner

FFT (Fast Fourier Transform)
(On the perspective of Image Enhancement)

V.Imp # Redundancy (CH# 5)
- Coding Redundancy
- Interpixel Redundancy
- Psychovisual Redundancy

In digital image processing, three basic data redundancy can be identified and exploited as follows:
- Coding Redundancy
- Interpixel Redundancy
- Psychovisual Redundancy

1) Coding Redundancy
We know that average number of bits required to represent each pixel is given by:

$$L_{avg} = \sum_{k=0}^{L-1} L(k)\, p(k)$$

Hence, the number of bits required to represent the image is $n \cdot L_{avg}$ where, n is the total no. of pixel in given image.
Maximum compression ratio is achieved when L_{avg} is minimized. But, coding the gray level in such a way that L_{avg} is not minimized then resulting image is said containing coding redundancy

2) Interpixel Redundancy
It is related to interpixel correlation within an image. Usually the value of certain pixel in the image can be reasonably predicted from the values of its neighbours in the image.
Thus, the values of the individual pixel carries relatively small amount of information & much more information about pixel value that can be inferred on the basis of its neighbours value.
These dependencies between pixel value in the image is called interpixel redundancy.

3) Psychovisual Redundancy
The eye does not respond with equal sensitivity to all visual information.
Human perception

such as for important features (edges, textures) and does not perform quantitative analysis of every pixel in the image.

So, psychovisual redundancy takes into perception of the human visual system.

Scanned by CamScanner