IAT3 Question Paper

1. Explain Erosion, Dilation, opening and Closing.
2. Briefly explain Hit or Miss Transform
3. Write short notes on Convex Hull and Pruning.
4. What are Image Pyramids? Explain with system Block diagram for creating it.
5. What is Subband coding? Briefly explain a two band Subband coding and decoding system.
6. What are the derivative operators useful in image segmentation? Explain their role in segmentation.
7. Illustrate and explain how chain code is used for compression of monochrome images.
8. Explain how polygon approximation approach can be used for morphological shape approximation.

IAT 3 Solution

1. Dilation

Dilation of A by B and is defined by the following equation:

$$A \oplus B = \{z \mid (\hat{B})z \cap A \neq \emptyset\}$$

This equation is based on obtaining the reflection of B about its origin and shifting this reflection by z.

The dilation of A by B is the set of all displacements z, such that $\hat{B}$ and A overlap by at least one element. Based On this interpretation the equation can be rewritten as:

$$A \oplus B = \{z \mid [(\hat{B})z \cap A] \subset A\}$$

Dilation is the set of all points in the image, where the structuring element “touches” the foreground.

It adds pixels to the boundaries of objects in an image. It increases the size of object & fills gap.

Dilation “grows” or “thickens” objects in an binary image.

The extent of thickening is controlled by the shape of the structuring element used.

Dilation is used for expanding an element A by using structuring element B.

Erosion:

With A and B as sets in Z2 , the erosion of A by B, denoted A, is defined as

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

This equation indicates that the erosion of A by B is the set of all points z such that B, translated by z, is contained in A where B is the structuring element.

The statement that B has to be contained in A equivalent to B not sharing any common elements with background. The equivalent expression is

$$A \quad B=\left\{z \mid (B)_Z \cap A^c = \varnothing\right\}$$

- Erosion is used for shrinking of element A by using element B.
- Erosion is the set of all points in the image, where the structuring element “fits into”.
- Good for, e.g.,
    - Noise removal in background
    - Removal of holes in foreground / background
- One of the simplest uses of erosion is for eliminating irrelevant details (in terms of size) from a binary image.
- These are the other two important morphological operations.
- Opening – smoothes contours , eliminates protrusions
- Closing – smoothes sections of contours, fuses narrow breaks and long thin gulfs, eliminates small holes and fills gaps in contours as opposite to opening.

- These operations can be applied few times, but has effect only once
- The Opening of set A by structuring element B, denoted $A \circ B = (A \ominus B) \oplus B$
- First – erode A by B, and then dilate the result by B
- In other words, opening is the unification of all B objects Entirely Contained in A
- The Closing of set A by structuring element B, denoted $A \cdot B = (A \oplus B) \ominus B$
- First – dilate A by B, and then erode the result by B
- In other words, closing is the group of points, which the intersection of object B around them with object A – is not empty

2. Hit or Miss Transform

The Hit or Miss transform is used for detecting shapes. It uses two structuring elements.

The 1^{st} SE is the foreground shape of the object to be detected.

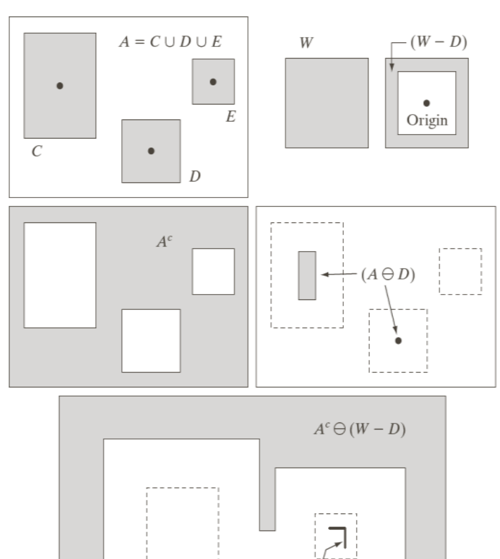
The 2^{nd} SE contains the background shape around the object which is to be detected.

At any point on the given image,

- If the foreground matches with the 1^{st} structuring element

  And

- If the complement(i.e. background)matches with the second structuring element , then we say that the object shape exists at that point.



a b
c d
e
f

FIGURE 9.12 (a) Set A. (b) A window, W, and the local background of D with respect to W, $(W - D)$. (c) Complement of A. (d) Erosion of A by D. (e) Erosion of A^c by $(W - D)$. (f) Intersection of (d) and (e), showing the location of the origin of D, as desired. The dots indicate the

- Let the origin of each shape be located at its center of gravity.
- If we want to find the location of a shape , say – D , at (larger) image, say – A :
- Let D be enclosed by a small window, say – W.
- The local background of D with respect to W is defined as the set difference (W - D).
- Apply erosion operator of A by D, will get us the set of locations of the origin of D, such that D is completely contained in A.
- It may be also view geometrically as the set of all locations of the origin of D at which D found a match (Hit) in A.
- Apply erosion operator on the complement of A by the local background set (W – D).
- Notice, that the set of locations for which D exactly fits inside A is the intersection of these two last operators above.
- This intersection is precisely the location sought.
- If B denotes the set composed of D and it's background, the match of B in A, denoted $A \circledast B = (A \ominus D) \cap [A^C \ominus (W - D)]$
- Generalizing the equation B = (B1,B2) ; B1 = D , B2 = (W-D).
- The match (or set of matches) of B in A, denoted is:
- $A \circledast B = (A \ominus B_1) \cap (A^c \ominus B_2)$
- The reason for using these kind of structuring element – B = (B1,B2) is based on an assumed definition that, two or more objects are distinct only if they are disjoint (disconnected) sets.
- In some applications , we may interested in detecting certain patterns (combinations) of 1's and 0's and not for detecting individual objects in this case <u>a background is not required</u> and the *hit-or-miss transform* reduces to <u>simple *erosion*</u>.

- This simplified pattern detection scheme is used in some of the algorithms for – identifying characters within a text.

3. Convex Hull

A set *A* is convex if the straight line segment joining any two points in *A* lies entirely within *A*.

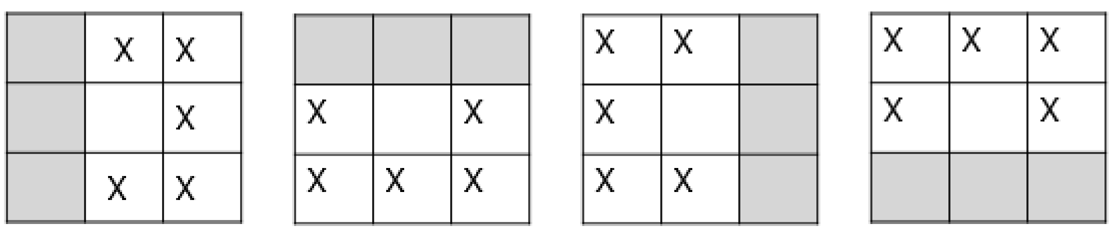
The convex hull *H* of an arbitrary set *S is* the smallest convex set containing *S*.

The difference *H-S* is called convex deficiency.

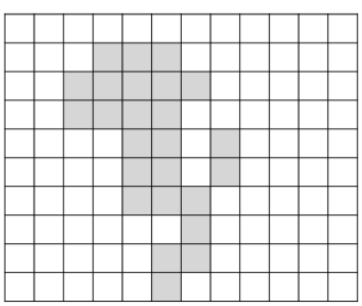
The convex hull and the convex deficiency are useful quantities to characterize shapes.

We present here a morphological algorithm to obtain the convex hull *C*(*A*)of a shape *A*.

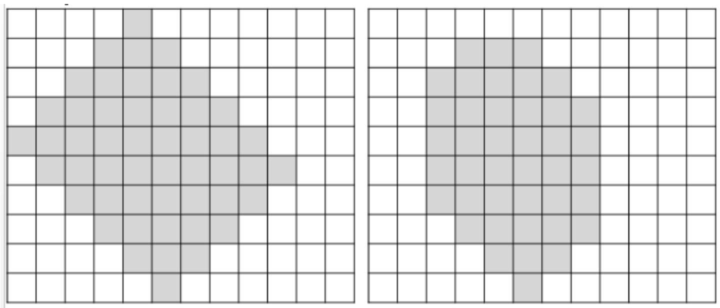
- The Convex Hull of a set A is denoted as C(A).
- The method consists of iteratively applying the hit-or miss transform to *A* with B^i(set of structuring elements).
- $B^i,\ i = 1,2,3,4$ →4 structuring elements.
- $X_k^i = \left(X_{k-1}^i \circledast B^i\right) \cup A\ , i = 1,2,3,4\ and\ k = 1,2,3, \ldots\ldots\ with\ X_0^i = A$
- The iteration stops when $X_k^i = X_{k-1}^i$(In two subsequent iterations the output does not change).
- $D^i = X_k^i$
- As i=1,2,3,4 , we get 4 different point sets. The Union of all point sets is the convex hull.
- $C(A) = \bigcup_{i=1}^{4} D_i$
- The structuring elements are



B^1 B^2 B^3 B^4



Set A



C(A)

C(A): Result of limiting growth of the convex hull algorithm to the maximum dimensions of the original set of points along the vertical and horizontal directions.

Pruning:

a. Thinning and skeletonizing tend to leave parasitic components

b. Pruning methods are essential complement to thinning and skeletonizing procedures

It involves 4 steps

a. $X_1 = A \otimes \{B\}$

b. $X_2 = \bigcup_{k=1}^{8} \left(X_1 \circledast B^k \right)$

c. $X_3 = \left(X_2 \oplus H \right) \cap A$

$H: 3 \times 3$ structuring element

d. $X_4 = X_1 \cup X_3$

4. Image pyramid is the simple structure for representing images at more than one resolution.

An image pyramid is a collection of decreasing resolution images arranged in the shape of a pyramid.

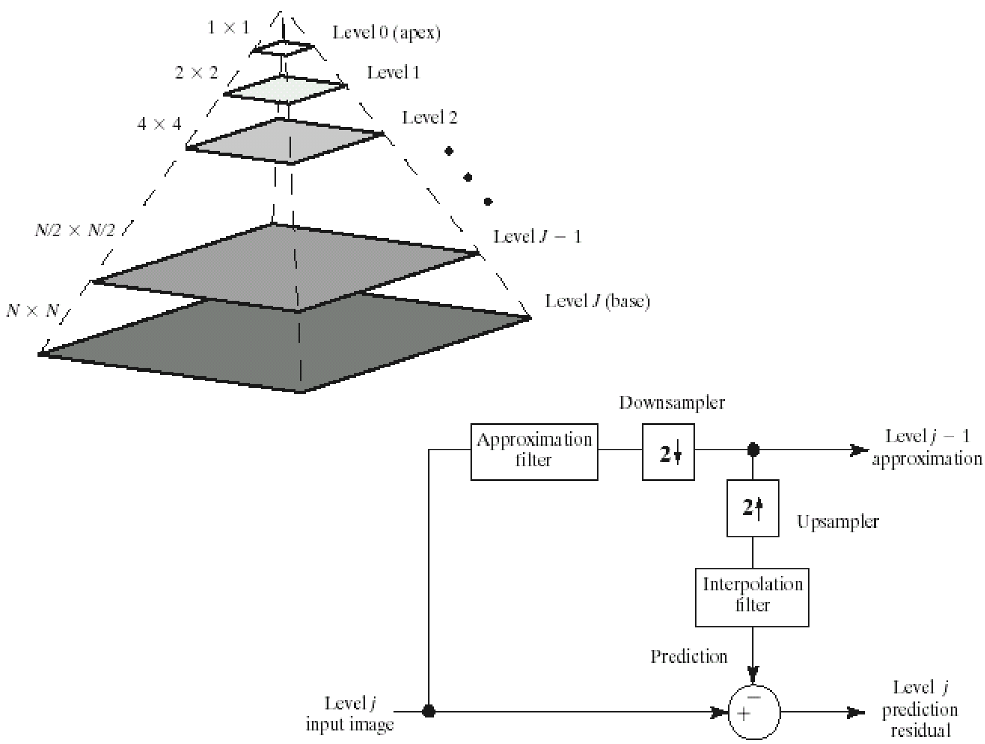
It is devised for machine vision and image compression algorithms.

The base of pyramid contains a high resolution representation of the image being processed.

The apex contains low resolution approximation.

As we move up the pyramid, both size and resolution decreases.

- Base level J is of size $2^J \times 2^J$ $or\ N \times N,\ where\ J = \log_2 N,$ apex level 0 is of size $1 \times 1,$ and general level j is of size $2^j \times 2^j$, $where\ 0 \leq j \leq J.$
- The image pyramid shown is composed of J+1 resolution levels from $2^J \times 2^J$ to$2^0 \times 2^0$.
- Most image pyramids are truncated to P+1 levels, where $1 \leq P \leq J$ and $j = J - P, \ldots, J - 2,\ J - 1, J.$
- The total number of pixels in a P+1 level pyramid for $P > 0$ is
- $N^2 \left(1 + \frac{1}{(4)^1} + \frac{1}{(4)^2} + \cdots . + \frac{1}{(4)^P}\right) \leq \frac{4}{3} N^2$



a b

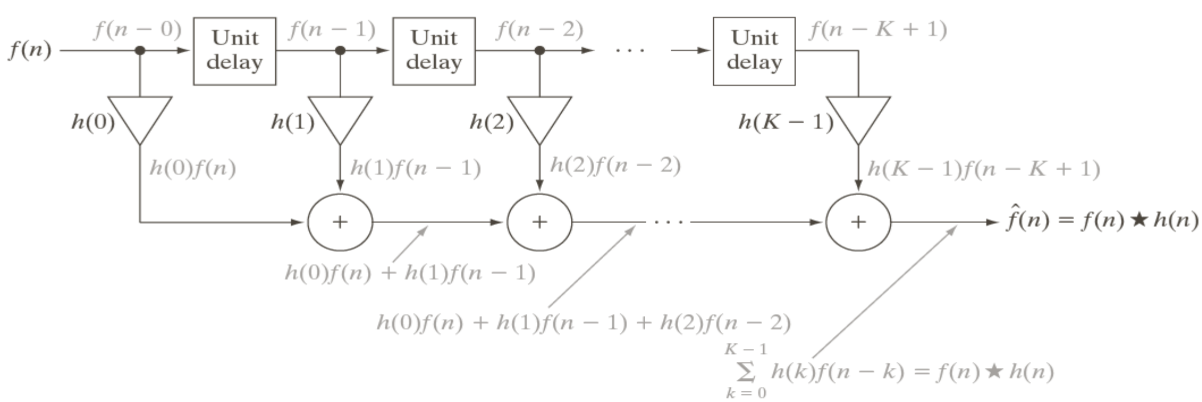
FIGURE 7.2 (a) A pyramidal image structure and (b) system block diagram for creating it.

- At each level we have an approximation image and a residual image.
- The original image (which is at the base of pyramid) and its P approximation form the approximation pyramid.
- The residual outputs form the residual pyramid.
- Approximation and residual pyramids are computed in an iterative fashion.
- A P+1 level pyramid is build by executing the operations in the block diagram P times.
- During the first iteration, the original $2^J x 2^J$ image is applied as the input image.
- This produces the level J-1 approximate and level J prediction residual results
- For iterations j=J-1, J-2, …, J-p+1, the previous iteration’s level j-1 approximation output is used as the input.
- Each iteration is composed of three sequential steps:

1. Compute a reduced resolution approximation of the input image. This is done by filtering the input and downsampling (subsampling) the filtered result by a factor of 2.
    - Filter: neighborhood averaging, Gaussian filtering
    - The quality of the generated approximation is a function of the filter selected.
2. Upsample output of the previous step by a factor of 2 and filter the result. This creates a prediction image with the same resolution as the input.
    - By interpolating intensities between the pixels of step 1, the interpolation filter determines how accurately the prediction approximates the input to step 1.

3. Compute the difference between the prediction of step 2 and the input to step 1. This difference can be later used to reconstruct progressively the original image

- The upsampling and downsampling blocks are used to double and halve the spatial dimensions of the approximation and prediction images that are computed.
- Given an integer variable n and 1-D sequence of samples f(n), upsampled sequence $f_{2\uparrow}(n)$ is defined as
- $f_{2\uparrow}(n) = \begin{cases} f\left(\frac{n}{2}\right) \; if\ n\ is\ even \\ 0 \; otherwise \end{cases}$
- The downsampling by 2 is defined as
- $f_{2\downarrow}(n) = f(2n)$
- Upsampling is inserting 0 after every sample in a sequence and downsampling is discarding every other sample.
- Subband Coding
- An image is decomposed into a set of bandlimited components called subbands.
- The decomposition is performed so that the subbands can be reassembled to reconstruct the original image without error.
- The decomposition and reconstruction are performed by means of digital filters.

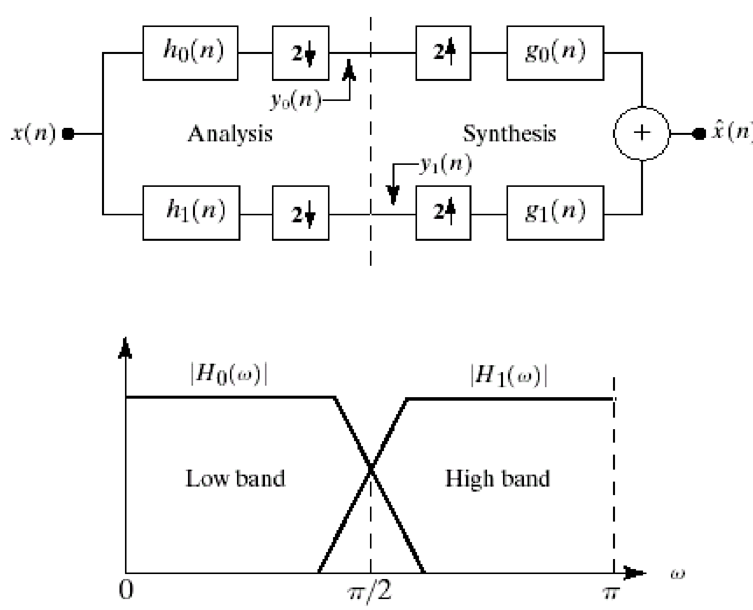


a
b c

FIGURE 7.4 (a) A digital filter; (b) a unit discrete impulse sequence; and (c) the impulse response of the filter.

- Consider a simple digital filter which is constructed from three basic components – unit delays, multipliers and adders.
- The unit delays are connected in series to create (K-1) delayed(right shifted) versions of the input sequence f(n).
- The delayed sequences $f(n-0),\ f(n-1), \ldots,\ f(n-K+1)$ are multiplied by constants $h(0), h(1), \ldots, h(k-1)$ respectively and summed to produce filtered output sequence
- $\hat{f}(n) = \sum_{k=-\infty}^{\infty} h(k)f(n-k)$
- $\hat{f}(n) = f(n) * h(n)$

- denotes convolution.
- The K multiplication constants are called filter coefficients.
- Each coefficient denote a filter tap.
- The order of the filter is K-1.
- If the input to the filter is the unit discrete impulse then the output of the filter is
- $\hat{f}(n) = \sum_{k=-\infty}^{\infty} h(k)\delta(n-k)$
- $\hat{f}(n) = h(n)$
- That is, by substituting $\delta(n)$ for input f(n)and making use of sifting property of the unit discrete impulse, we can find the impulse response of the filter.
- The unit impulse is shifted from left to right across the filter, producing an output that assumes the value of the coefficient at the location of the delayed impulse.
- The filter is called the Finite Impulse Response(FIR) filter.
- The following figure shows the six functionally related filters.
- Filter $h_2(n)$ is a sign reversed version of $h_1(n)$.
- Filter $h_3(n)$ is a reflection of $h_1(n)$ about the vertical axis.
- Filter $h_4(n)$ is a reflected and translated version of $h_1(n)$. Neglecting Translation,the responses of the two filters are identical.
- Filter $h_5(n)$ is called a modulated version of $h_1(n)$. Because modulation changes the signs of all odd indexed coefficients[i.e., the coefficients for which n is odd in $h_5(1) = -h_1(n)$ *and* $h_5(3) = -h_1(3)$, *while* $h_5(0) = h_1(0)$ *and* $h_5(2) = h_1(2)$].
- Filter $h_6(n)$ is an order reversed version of $h_1(n)$ that is also modulated:
- $h_6(n) = (-1)^n h_1(K-1-n)$
- This eqn is to illustrate the fact that sign reversal, order reversal and modulation are sometimes combined in the relationship b/w two filters.
    - A two band Subband coding and decoding system
- The system is composed of two filter banks → Analysis filter Bank and Synthesis Filter Bank.
- Each Filter bank consists of two FIR Filters.



a
b

FIGURE 7.4 (a) A two-band filter bank for one-dimensional subband coding and decoding, and (b) its spectrum splitting properties.

- The analysis filter bank, which includes filters $h_0(n)$ and $h_1(n)$ is used to break input sequence f(n) into two half length sequences $f_{lp}(n)$ and $f_{hp}(n)$, the subbands which represent the input.

- The idealized characteristics of $h_0(n)$ and $h_1(n)$ are H_0 and H_1 respectively.
- Filter $h_0(n)$ is a lowpass filter whose output, subband $f_{lp}(n)$ is called an approximation of f(n).
- Filter $h_1(n)$ is highpass filter whose output, subband $f_{hp}(n)$ is called the high frequency or detail part of f(n).
- Synthesis bank filters $g_0(n)$ and $g_1(n)$ combine $f_{lp}(n)$ and $f_{hp}(n)$ to produce $\hat{f}(n)$.
- The goal in subband coding is to select $h_0(n), h_1(n), g_0(n)\ and\ g_1(n)$ so that $\hat{f}(n) = f(n)$. That is, if the input and output of the subband coding and decoding system are identical then the resulting system is called perfect reconstruction filter.
- The resulting system is perfect reconstruction filter if the synthesis filters are modulated versions of the analysis filters- with one synthesis filter being sign reversed as well. i.e,. The impulse responses of the synthesis and analysis filters must be related in one of the following two ways:

$$g_0(n) = (-1)^n \quad h_1(n)$$

$$g_1(n) = (-1)^{n+1} h_0(n)$$

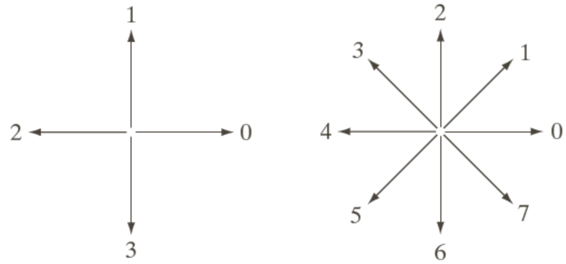
or

$$g_0(n) = (-1)^{n+1} h_1(n)$$

$$g_1(n) = (-1)^n \quad h_0(n)$$

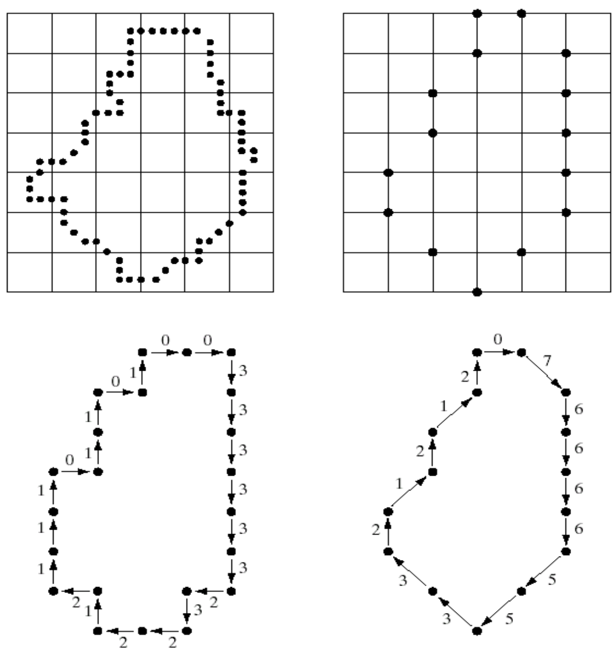
- Filters $h_0(n), h_1(n),\ g_0(n)\ and\ g_1(n)$ are said to be cross modulated because diagonally opposed filters in the block diagram.

7. Chain Codes:

- A chain code is a lossless compression algorithm for monochrome images. The basic principle of chain codes is to separately encode each connected component in the image.
- It is a Boundary based representation and from this we find boundary descriptors possible from this.
- Chain code – describe the sequence of steps generated when walking around the boundary
- Chain code 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 4- or 8-connectivity of the segments. The direction of each segment is coded by using a numbering scheme, as in Fig.11.3
- 

a b

FIGURE 11.3 Direction numbers for (a) 4-directional chain code, and (b) 8-directional chain code.



a b
c d

FIGURE 11.2
(a) Digital boundary with resampling grid superimposed.
(b) Result of resampling.
(c) 4-directional chain code.
(d) 8-directional chain code.

- The descriptor obtained from chain code must be scale invariant, translation invariant and rotation invariant.
- The Chain code is translation invariant, and it can be made scale invariant by choosing the sampling grid properly. But the chain code is not rotation invariant.
- Hence first difference is found from the chain code.
-

8.Polygonal approximation: approximate a boundary using a polygon(a plane figure with at least three straight sides and angles, and typically five or more).

- For closed boundaries:
- Approximation becomes exact when no. of segments of the polygons is equal to the no. of points in the boundary
- Goal of poly. Approx is to capture the essence of the shape in a given boundary using fewest no. of segments.
- One method of polygonal approximation is using Minimum Perimeter Polygon(MPP) algorithm.
- Firstly enclose the boundary of the given object set by a set of concatenated cells.
- Think boundary as a rubber band. As allowed to shrink, it will be constrained by the inner & outer walls of the bounding regions.
- This shrinking produces the shape of a polygon of min. perimeter. Size of cells determine the accuracy of the polygonal approximation.
- In the limit if size of each cell corresponds to a pixel in the boundary , the error in each cell between the boundary & the MPP approx. at most would be $\sqrt{2}d$, where d-min possible pixel distance.
- This error can be reduced to half by forcing each cell in polygon approximation to be centered on its corresponding pixel in the original boundary.
    - Data for the MPP algorithm:
- Form a list whose rows are the coordinates of each vertex and an additional element denoting whether the vertex is W or B.
- The concave vertices must be mirrored.
- Arrange the vertices in sequential order(counterclockwise direction).

- The 1st vertex should be an uppermost, leftmost and W vertex. Let V_0 denotes this vertex.
    - Algorithm uses two crawler points.
- A White crawler: W_C → crawls along convex(W) vertices.
- A Black crawler: B_C → crawls along mirrored concave(B) vertices.
    - The algorithm starts by setting $W_c = B_c = V_0$.
    - Let V_L denotes the last MPP vertex found.
    - Let V_k denote the current vertex being examined.
    - One of the three conditions can exist between V_L, V_k and the two crawler points.
- If V_k lies to the positive side of the line through pair(V_L, W_C): that is sgn$(V_L, W_C, V_k) > 0$ *then* next MPP vertex is W_C, and let $V_L = W_C$, reinitialize the algorithm by setting $W_C = B_c = V_L$ and continue with the next vertex after V_L.
- If V_k lies to the negative side of the line through pair (V_L, W_C) or is collinear with it; that is sgn$(V_L, W_C, V_k) \leq 0$. At the same time, V_k lies to the positive side of the line through (V_L, B_C) or is collinear with it; that is, $(V_L, B_C, V_k) \geq 0$ *then* V_K becomes a candidate MPP vertex. Set $W_c = V_k$ if V_k is convex(i.e., it is a W vertex) otherwise set $B_C = V_K$. *Continue with the next vertex in list*
- If V_k lies on the negative side of the line through pair $(V_L,\ B_C)$; that is sgn$(V_L, B_C, V_k) < 0$ *then* next MPP vertex is B_C, and let $V_L = B_C$, reinitialize the algorithm by setting $W_C = B_c = V_L$ and continue with the next vertex after V_L.
    - The algorithm terminates when it reaches the first vertex again.
    - The V_L vertices found by the algorithm are the vertices of the MPP.
    - This algorithm finds all the MPP vertices of a polygon enclosed by a simply connected cellular complex.