Data Model & Structure

Unit: II & III

Semester: II

Paper Code: GIS 07
Name of Paper: Geographic Information System

PG Diploma in RS & GIS

Dr. SHYAMA PRASAD MUKHERJEE UNIVERSITY, RANCHI

Data Model & Structure

2.1 Data Model

The data model represents a set of guidelines to convert the real world (called entity) to the digitally and logically represented spatial objects consisting of the attributes and geometry. The attributes are managed by thematic or semantic structure while the geometry is represented by geometric-topological structure.

There are two major types of geometric data model ; vector and raster model, as shown in Figure 2.1

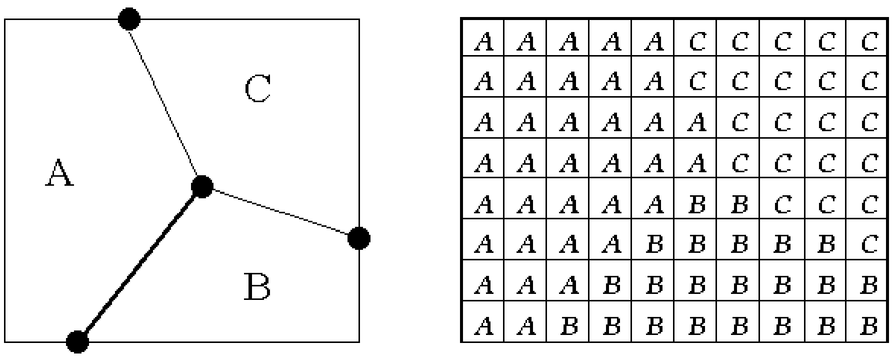


Figure 2.1 Vector and Raster Model

a. Vector Model

Vector model uses discrete points, lines and/or areas corresponding to discrete objects with name or code number of attributes.

b. Raster Model

Raster model uses regularly spaced grid cells in specific sequence. An element of the grid cell is called a pixel (picture cell). The conventional sequence is row by row from the left to the right and then line by line from the top to bottom. Every location is given in two-dimensional image coordinates; pixel number and line number, which contains a single value of attributes.

2.2 Geometry and Topology of Vector Data

Spatial objects are classified into point object such as meteorological station, line object such as highway and area object such as agricultural land, which are represented geometrically by point, line and area respectively. For spatial analysis in GIS, only the geometry with the position, shape and size in a coordinate system is not enough but the topology is also required.

Topology refers to the relationships or connectivity between spatial objects.

The geometry of a point is given by two dimensional coordinates (x, y), while line, string and area are given by a series of point coordinates, as shown in Figure 2.2 (a). The topology however defines additional structure as follows (see Figure 2.2 (b)).

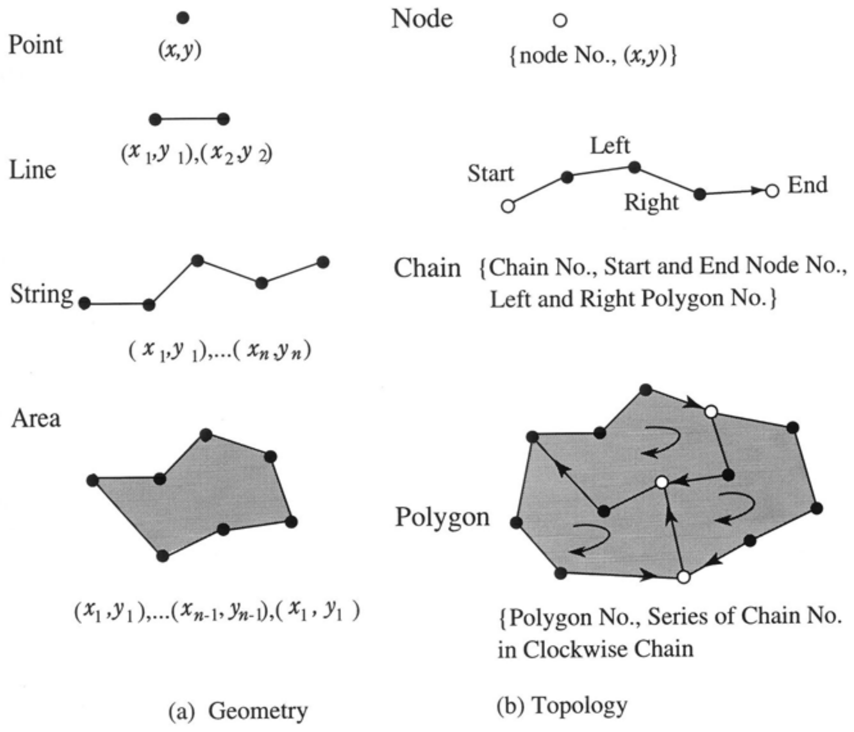


Figure 2.2 Geometry and Topology of Vector Data

Node: an intersect of more than two lines or strings, or start and end point of string with node number

Chain: a line or a string with chain number, start and end node number, left and right neighboured polygons

Polygon: an area with polygon number, series of chains that form the area in clockwise order (minus sign is assigned in case of anti-clockwise order)

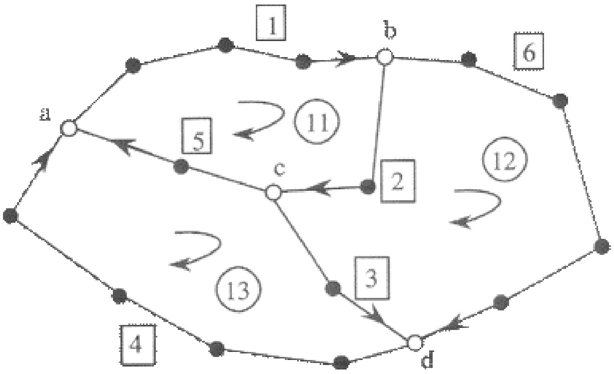
2.3 Topological Data Structure

In order to analyse a network consisting of nodes and chains, the following topology should be built.

Chain: Chain ID, Start Node ID, End Node ID, Attributes

Node: Node ID, (x, y), adjacent chain IDs (positive for to node, negative for from node)

In order to analyse not only a network but also relationships between polygons, the following additional geometry and topology are required as shown in an example of Figure 2.3.



Chain Geometry

Chain	Start	Coordinates		End
1	$(X_a, Y_a$	(X, Y) ……	(X, Y)	$(X_b\, Y_b)$
2	$(X_b, Y_b$	(X, Y) ……	(X, Y)	(X_c, Y_c)
⋮	⋮	⋮		
6	$(X_b, Y_b$	(X, Y) ……	(X, Y)	(X_d, Y_d)

Topology of Polygon

Polygon	Chain
11	1,2,5
12	-2,6,-3
13	4,-5,3

Topology of Node

Node	Chains
a	1,-5,-4
b	-1,2,6
c	-2,3,5
d	-3,4,-7

Topology of Chain

Chain	From	To	Left Polygon	Right Polygon
1	a	b	0	11
2	b	c	12	11
⋮	⋮	⋮	⋮	⋮
6	b	d	0	12

Figure 2.3 Building Topology

Chain geometry: Chain ID, Start Coordinates, Point Coordinates, End Coordinates

Polygon topology : Polygon ID, Series of Chain ID, in clockwise order (Attributes)

Chain topology : Chain ID, Start Node ID, End Node ID, Left Polygon ID, Right Polygon ID, (Attributes)

The advantages of the topological data model are to avoid duplication in digitizing common boundaries of two polygons and to solve problems when the two versions of the common boundary do not coincide.

The disadvantages are to have to build very correct topological data sets without any single error and to be unable to represent islands in a polygon.

2.4 Topological Relationships between Spatial Objects

In practical applications of GIS, all possible relationships in spatial data should be used logically with more complicated data structures.
The following topology relationships are commonly defined.

a. Point-Pont Relationship

"is within”: within a certain distance

"is nearest to”: nearest to a certain point

b. Point-Line Relationships

"on line”: a point on a line

"is nearest to”: a point nearest to a line

c. Point-area Relationships

"is contained in: a point in an area

"on border of area”: a point on border of an area

d. Line-Line Relationships

"intersects”: two lines intersect

"crosses" : two lines cross without an intersect

"flow into" : a stream flows into the river

e. Line-Area Relationship

"intersects" : a line intersects an area

"borders" : a line is a part of border of an area

f. Area-Area Relationships

"overlaps”: two areas overlap

"is within”: an island within an area

"is adjacent to”: two area share a common boundary

Figure 2.4 shows the several topological relationships between spatial objects.
Figure 2.5 shows geometric and topological modelling between point, line and area

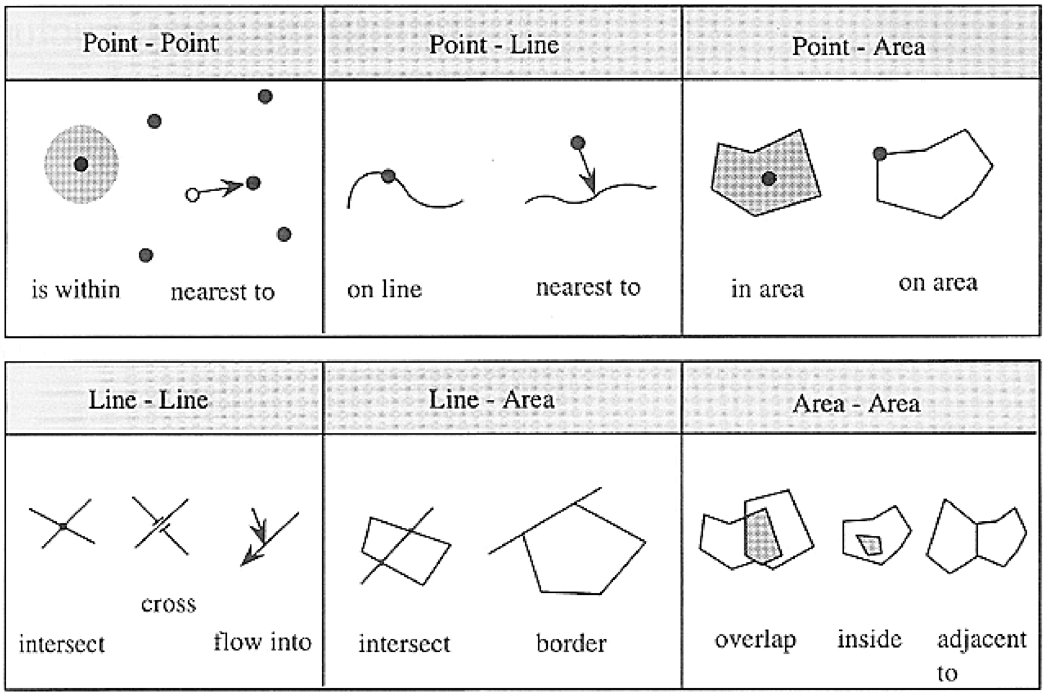


Figure 2.4 Topological Relationships Between Spatial Objects

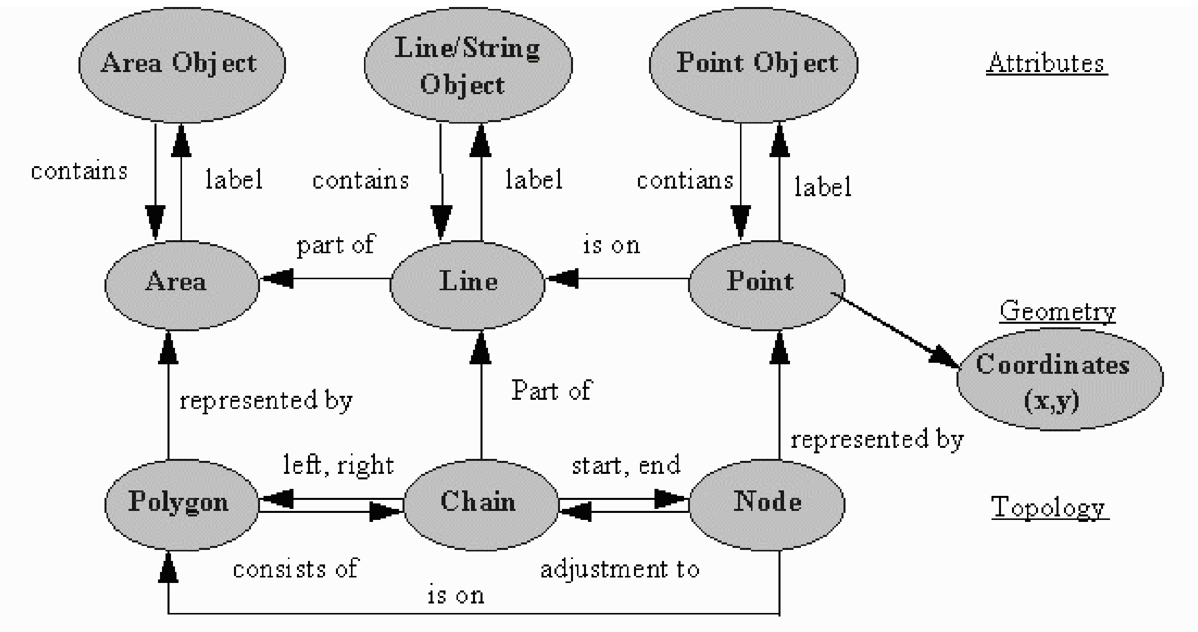


Figure 2.5 Geometric and Topological Modeling

2.5 Geometry and Topology of Raster Data

The geometry of raster data is given by point, line and area objects as follows (see Figure 2.6(a)).

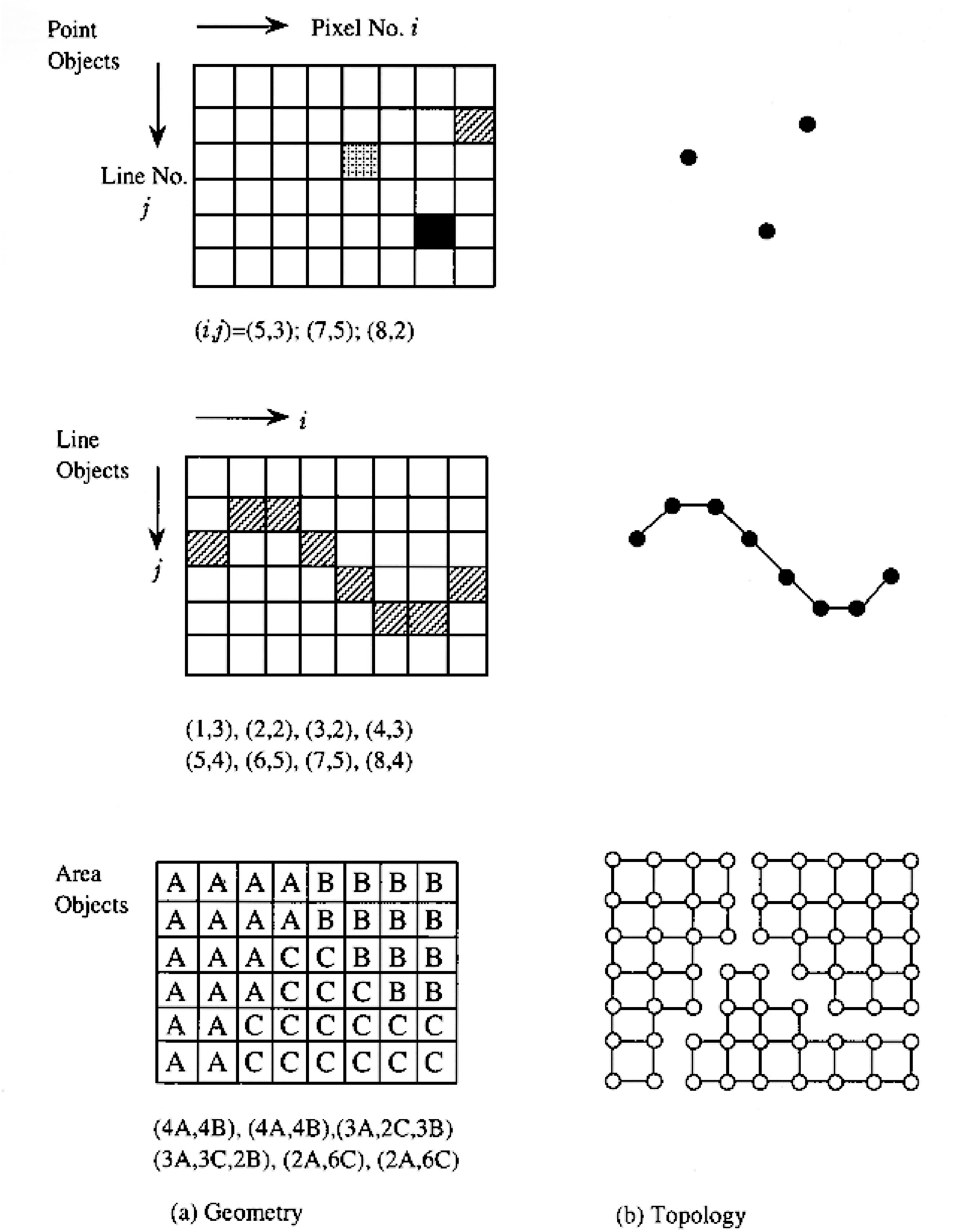


Figure 2.6 Geometry and Topology of Raster Data

a. Point objects

A point is given by point ID, coordinates (i, j) and the attributes

b. Line object

A line is given by line ID, series of coordinates forming the line, and the attributes

c. Area objects

An area segment is given by area ID, a group of coordinates forming the area and the attributes. Area objects in raster model are typically given by "Run Length" that rearranges the raster into the sequence of length (or number of pixels) of each class as shown in Figure 2.6 (a).

The topology of raster model is rather simple as compared with the vector model as shown in Figure 2.6 (b).

The topology of line objects is given by a sequence of pixels forming the line segments. The topology of an area object is usually given by "Run Length" structure as follows.

- start line no., (start pixel no., number of pixels)};
- second line no., (start pixel no., number of pixels)};

2.6 Topological Features of Raster Data

One of the weak points in raster model is the difficulty in network and spatial analysis as compared with vector model.

For example, though a line is easily identified as a group of pixels which form the line, the sequence of connecting pixels as a chain would be a little difficult in tracing. In case of polygons in raster model, each polygon is easily indentified but the boundary and the node (when at least more than three polygons intersect) should be traced or detected.

a. Flow directions

A line with directions can be represented by four directions called as the Rook's move in the chess game or eight directions called as the Queenís move, as shown in Figure 2.7 (a) and (b). Figure 2.7 (c) shows an example of flow directions in the Queen's move. Water flow, links of a network, roads etc. can be represented by the flow directions (or called Freeman chain code).

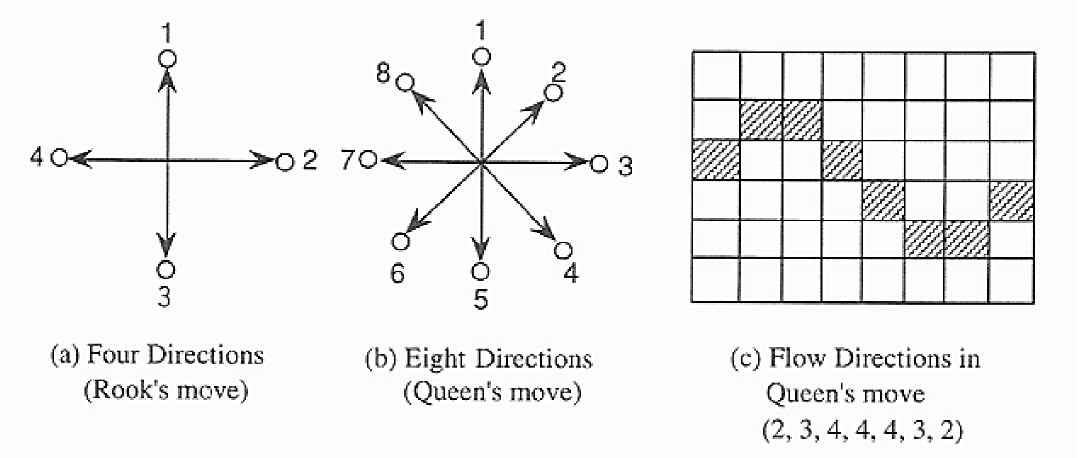


Figure 2.7 Flow Directions

b. Topological Features of Raster Data

Boundary is defined as 2 x 2 pixel window that has two different classes as shown in Figure 2.8 (a). If a window is traced in the direction shown in Figure 2.8 (a), the boundary can be identified.

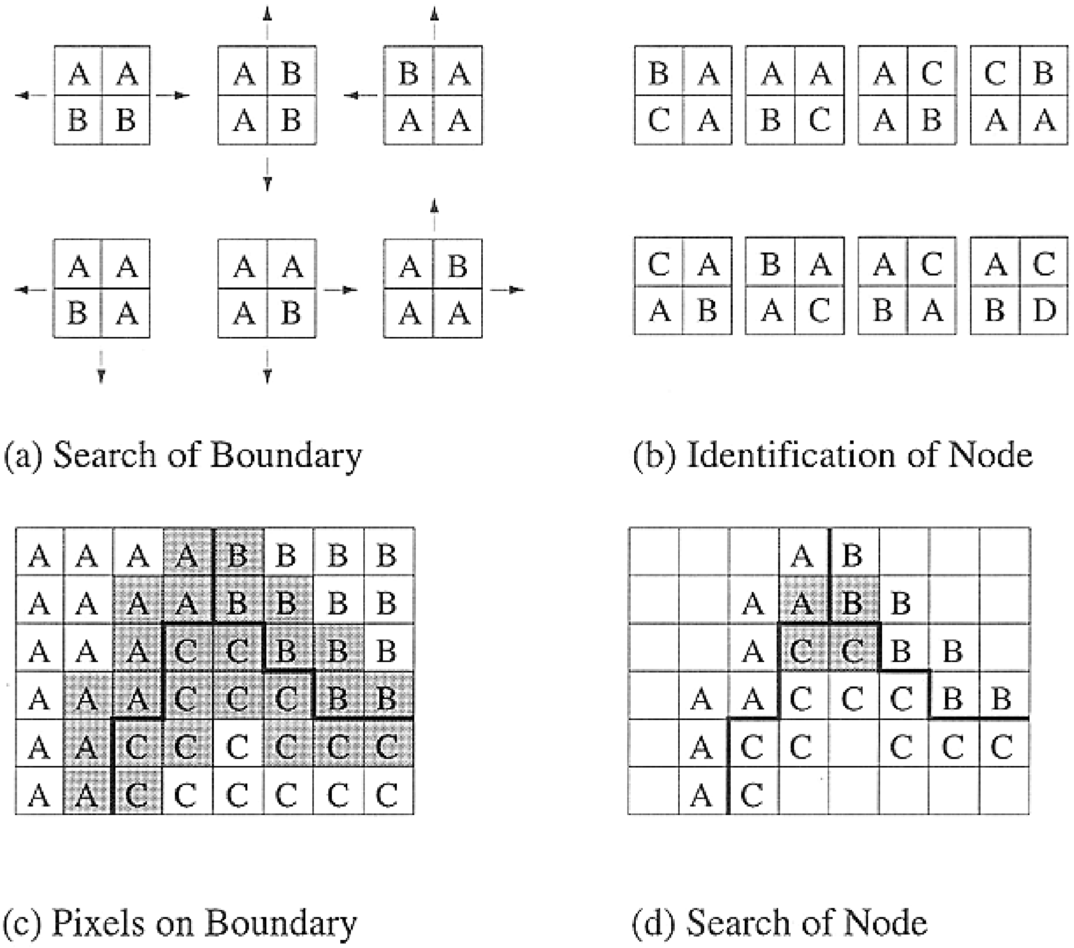


Figure 2.8 Identification of Boundary and Node

c. Node

A node in polygon model can be defined as a 2 x 2 window that has more than three different classes as shown in

Figure 2.8 (b).

Figure 2.8 (c) and (d) show an example of identification of pixels on boundary and node.

2.7 Thematic Data Modelling

The real-world entities are so complex that they should be classified into object classes with some similarity through thematic data modelling in a spatial database.

The objects in a spatial database are defined as representations of real-world entities with associated attributes.

Generally, geospatial data have three major components; position, attributes and time. Attributes are often termed "thematic data" or "non-spatial data", that are linked with spatial data or geometric data.

An attribute has a defined characteristic of entity in the real world.

Attribute can be categorized as normal, ordinal, numerical, conditional and other characteristics. Attribute values are often listed in attribute tables which will establish relationships between the attributes and spatial data such as point, line and area objects, and also among the attributes.

Figure 2.9 shows a schematic diagram of thematic data modelling.

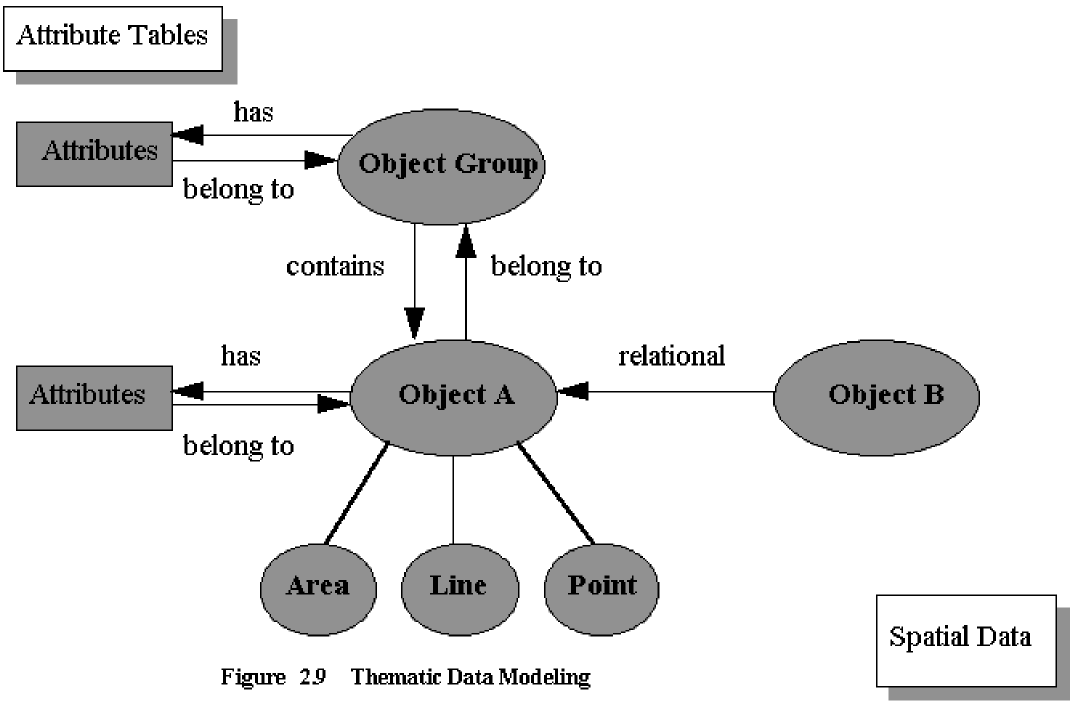


Figure 2.9 Thematic Data Modeling

Spatial objects in digital representation can be grouped into layers as shown in Figure 2.10.

For example, a map can be divided into a set of map layers consisting of contours, boundaries, roads, rivers, houses, forests etc.

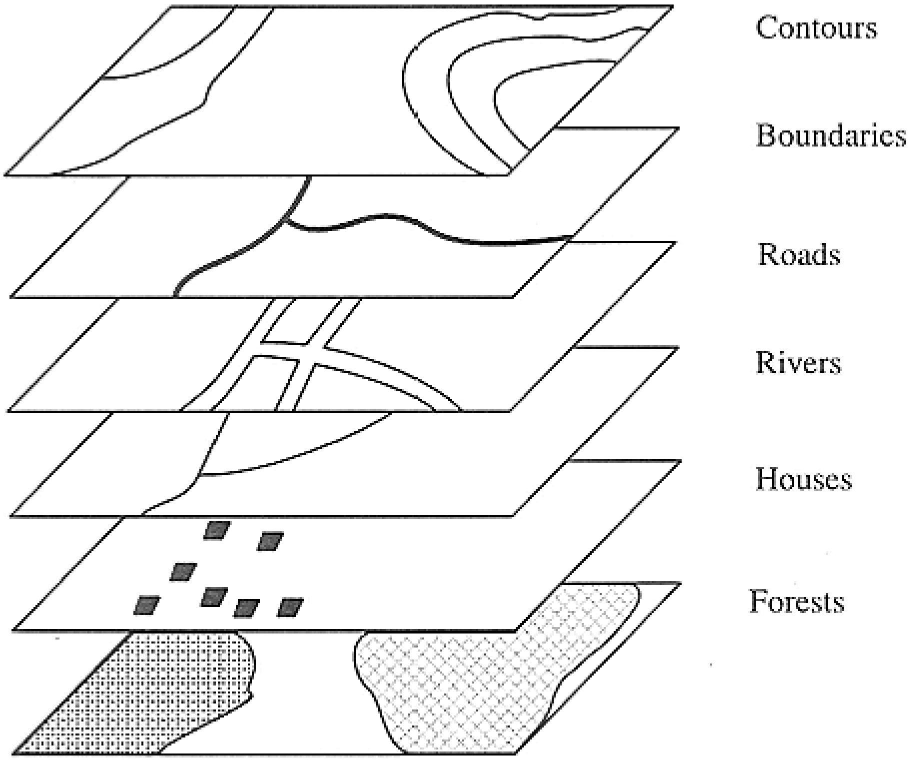


Figure 2.10 Map Layers

2.8 Data Structure for Continuous Surface Model

In GIS, continuous surface such as terrain surface, meteorological observation (rain fall, temperature, pressure etc.) population density and so on should be modelled. As sampling points are observed at discrete interval, a surface model to present the three-dimensional shape; z = f (x, y) should be built to allow the interpolation of value at arbitrary points of interest.

Usually the following four types of sampling point structure are modelled into DEM.

Grid at regular intervals:

Bi-linear surface with four points or bi-cubic surface with sixteen points is commonly used

Random points:

Triangulated irregular network (TIN) is commonly used. Interpolation by weighted polynomials is also used.

Contour lines:

Interpolation based on proportional distance between adjacent contours is used. TIN is also used.

Profile:

Profiles are observed perpendicular to an alignment or a curve such as high ways. In case the alignment is a straight line, grid points will be interpolated. In case the alignment is a curve, TIN will be generated.

Figure 2.11 shows different types of DEMs.

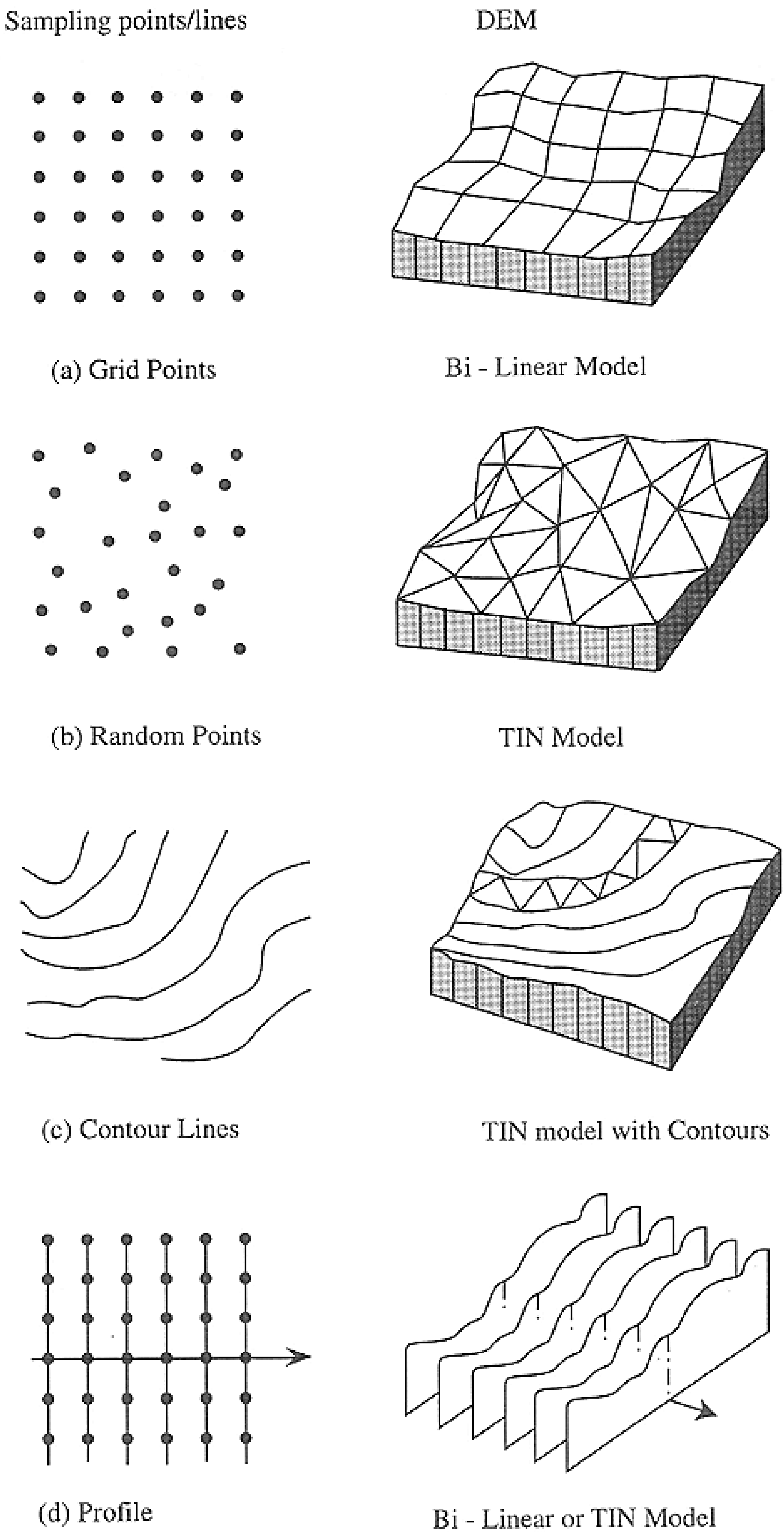


Figure 2.11 Different Types of DEMs