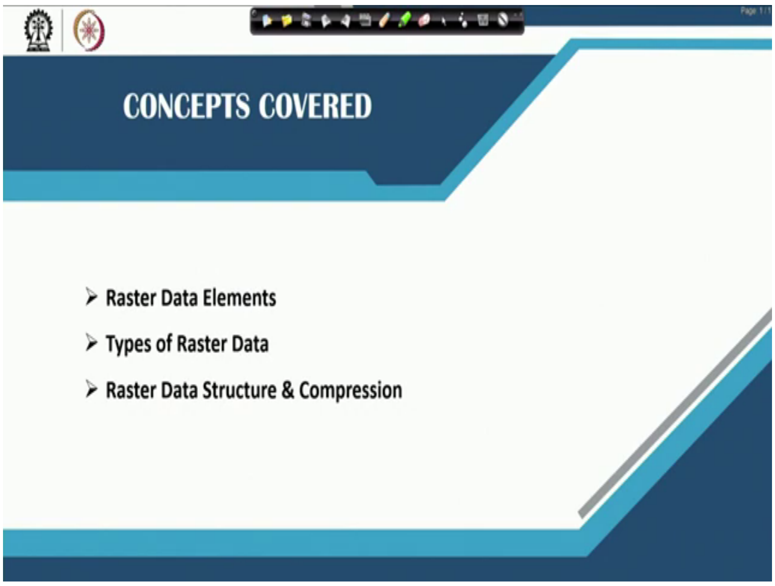
Geo Spatial Analysis in Urban Planning
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Module - 01
Introduction to Geographic Information System and Geographic Distribution
Lecture - 05
Raster Data Model

Welcome back. We are in the 5th lecture of this particular series and today we shall be talking about the Raster Data Model. We had earlier talked about the vector data model.

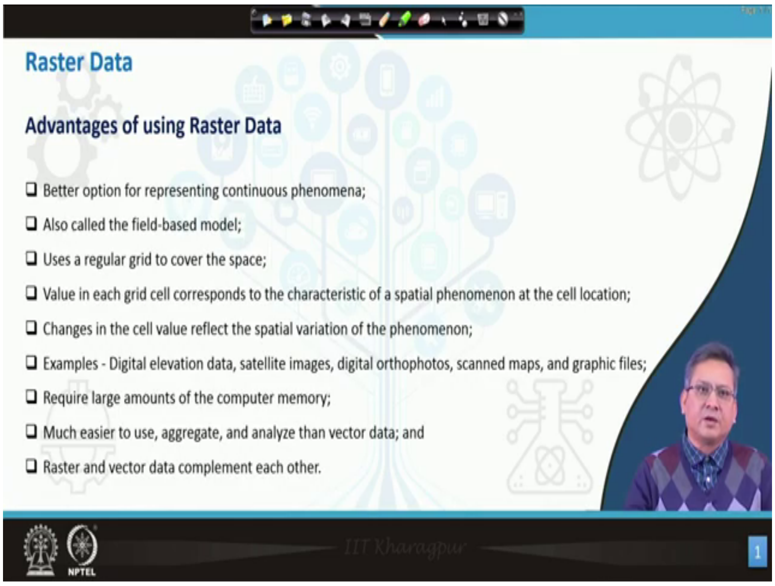
(Refer Slide Time: 00:30)



So, today we are going to talk about the raster data model. So, let us see the concepts that we are going to cover today. So, first we are going to talk about the raster data elements. What are the I mean different types of raster data and how a raster data is coded encoded as a file.

Next we are going to talk about the types of raster data. We are also going to talk about the raster data structure and raster a brief about or an introduction about the raster data compression.

(Refer Slide Time: 01:06)



So, in a GIS analysis as we had seen the I mean advantages of vector data of using vector data we have seen in our last lecture. Today we are going to see the advantages of using raster data. Now, both of these different types of data set has its own advantages and limitation. The most important advantage of the raster data set is its representation of the continuous phenomena.

What I mean to say is that whenever you have to record some physical phenomena or some kind of a observation which spans across a given spatial domain then and it is continuous in nature; that means, say suppose if we divide the entire region into grids say small squares contiguous squares then each of those I mean grids would have different values.

So, in such cases when we have continuous phenomena it is easier to represent those kind of data using a raster data approach. Now you can also do the output I mean when you are doing some kind of a processing in GIS, the resulting output if it is of a continuous nature can also be put in a raster framework. Apart from that it is also called a field based model since we are I mean modeling the continuous phenomena. It is also known as field based model.

So, it uses a regular grid to come cover the space which is contiguous I mean there is no gap between the grid elements. If we have a square grid you would see that all the squares are contiguous they are touching. The sides are touching it is neither overlapping or neither there is a space between those two squares adjoining squares. So, it is a regular grid which would cover the entire space and then we can code the number for each of the value.

So, the value in each grid or cell would be representing the spatial phenomena in that particular cell. So, say suppose we are measuring the temperature across India and how the temperature is changing in case you have a cold wave then how the cold wave is coming from the north of India you it can be represented as grid cell elements and the values would be contiguous those grids would be I mean a regular grid covering the entire Indian subcontinent and it would represent the continuous phenomena of temperature.

So, similarly we can talk about relative humidity or we can talk about say the rainfall pattern how much of precipitation is there, we can talk about urban I mean your some phenomena as well wherein. Say suppose you if you have a pollution due to a chimney stack in an industry I mean how what is the concentration of pollution across the city in each of the grid elements could be I mean coded as a raster data. So, I mean that is why we are also calling at a field based model.

Now, the changes in the value would be indicating if there is a spatial variation in the this physical phenomena; if you have a physical phenomena or a special phenomena if there is a variation it would be reflected across the cell values. So, definitely I mean if you have differences in the values you can then calculate the slope, you can calculate what is the rate of change you can calculate the aspect I mean in which direction the values are increasing or decreasing. So, we have some embedded advantages in the raster data model.

So, let us see an example in this case we are talking about a digital elevation data. Now whenever we want to I mean encode the height values we call a data type which is known as basically DEM digital elevation model and you might be familiar that you get in open source I mean if you can query in Google you can search for SRTM shuttle radar topography mission digital elevation model or you can look for aster digital elevation model or you also have the IRS based digital elevation model or carto digital elevation model.

It is known as carto DEM which is available from Bhuvan or you can I mean give an indent with the national data center which is the ISROs I mean agency which gives you the satellite images and other things and the processed products.

So, there you can requisition for the digital elevation data. So, this digital elevation data I mean if we are mapping a terrain say suppose it is a hill area no two adjacent grids would have the same value is not it. So, unless and until it is a plain area wherein there is not much of a change in terms of the elevation values in case of a hilly terrain the adjoining values would have different values.

So, we can I mean encode these heights as a raster data model and we call it as a digital elevation model. Now apart from this we can also have satellite images, we can have digital orthophotos, we can have also scanned maps like your topographic sheet or other maps your administrative area boundaries or if you have other maps can be scanned and those values can be encoded as your regular grid value and we call it as a raster data set or we can have graphic files also like you are seeing this particular image this is a video and we can have still images as well.

So, when you would have a still image you can zoom in and you would see that the pixels would get divided I mean it would become gridded and each would have a digital number numeric value attached to that I mean image and I mean or pixel or a grid of that particular image and I mean we can store it as a raster data file.

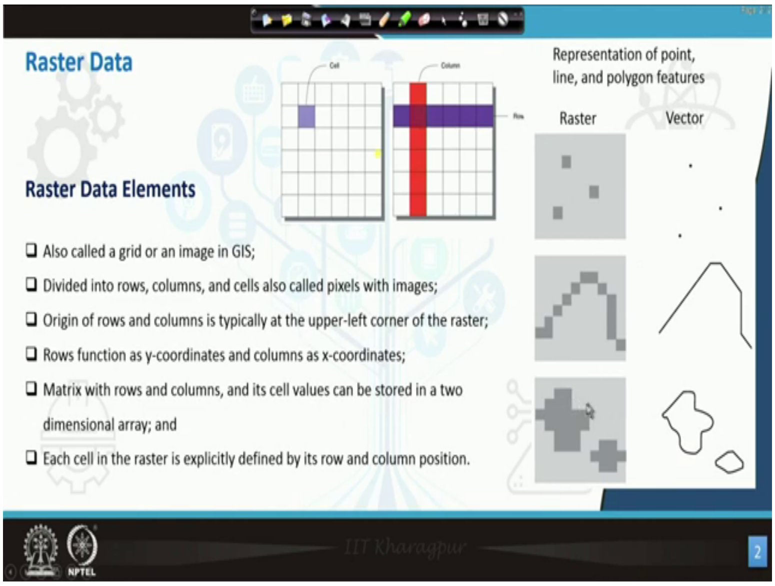
Now, the I mean the I mean the fact attached to this particular dataset is that since it would store each and every point data values, it would require some large amount of memory space in the computer. So, compared to a vector data model if you are storing a continuous phenomena it would entail a larger I mean image data size compared to a vector data, but same data if you are trying to encode it as a vector data you are creating a grid file of polygons and trying to encode it then it might take a still larger memory.

So, in such cases I mean raster data becomes more efficient when we want to process it. So, the biggest advantage with this particular dataset is that it is very easy to use I mean it is very easy to read this kind of data set using computers, you can aggregate this you can analyze in comparison to the vector data and you can visualize it.

So, but at the end of it whenever you are doing a GIS analysis you have to keep in mind that the raster and the vector data they complement each other they are complementary to each other and you have to code some of the linear feature or point feature as vector data and whenever you have continuous phenomena you should code it as raster feature.

So, this is the most important takeaway from these two lectures that we are going through the vector data model and the raster data model is you should know when to use what kind of data when you want to record or a given feature or an event or a phenomena you should know what kind of data structure to use whether you should use the vector data structure or you should use the raster data structure.

(Refer Slide Time: 10:13)



Now, talking about raster data elements. So, raster data as we had said we also call it as a grid or we call it as an image in the GIS parlance. So, this data is divided into rows and columns. So, I mean these individual elements they are also called as pixels I mean this term was abbreviated it originated from the word picture elements.

So, these two terms were fuse to create a new term called pixels. So, we have you are already aware of it. I mean when you buy a cell phone camera or a camera digital camera you want to know the resolution what is the I mean if it is a higher pixel resolution then I mean it would create better images.

So, you are already aware of these facts and you know what happens if you have more pixels in a given image if you have less pixels what happens. So, you can work it out with your friends camera if I mean you have images of I mean different pixels for your cameras in your mobile phones. You can try out taking pictures and zoom into those pictures and see what is the size of picture elements and thus it would clarify your idea regarding what is a pixel or I mean the element of your raster data what is its size.

So, we also talked about the origin because it is very important when you want to process this data. So, the rows and the columns we said which represent or I mean which would be used to code the raster dataset. So, rows would function as the y coordinate and the columns would function as the x coordinate. So, I mean if we see this particular I mean figure you can see that you have the columns you have column 1 column 2.

So, basically when you are talking about the column, you are talking about its distance from the origin I mean if I take the top corner as the origin. So, as I move this cursor your x coordinate will keep on increasing and as x coordinate increases the column number also increases. So, the biggest advantage with this type of a data model this type of a format the way we are storing this data is that it is you can easily store this data as a matrix as a numerical matrix as you do in your maths classes that you can you generate matrix.

So, you can we can have a matrix wherein we can have i rows and j columns. So, we can store the data as a matrix. Your I mean this matrix would have the cell values and it would be a two dimensional array and each array the position I mean say suppose we go to this position it would be defined by the number of the column and the number of the row.

So, ith column and the jth row. So, i and j I mean you can if you have a matrix you can process it as you can store the data as a matrix and you can call each and every element and do a numerical process on it. So, it is very useful in terms of our programming and encoding. So, you can see these are three examples of storing your vector data wherein we have the point data set, we have the line data set and we have the polygon data set.

So, this part I mean this image basically refers to the raster representation and vector representation of the points, lines and the polygons. So, you can see that whenever you have this points it is referred to in this grid as a blob I mean as a pixel. So, the difference between these two representation is that this point is more accurate in a way that the coordinate of this point it would be very accurate.

Now, this point could be located anywhere within this square within this square. So, it need not be necessarily at the centroid because when you are resampling it your this grid pixel size would be same for all. So, your point may not always sit in the centroid of this particular pixel and whenever we are trying to find out the coordinate of this particular pixel it may not give you the exact location of this point vector ok. So, this is the difference between locating I mean coding these two data types I mean this data type of your point line and polygon in your vector and the raster.

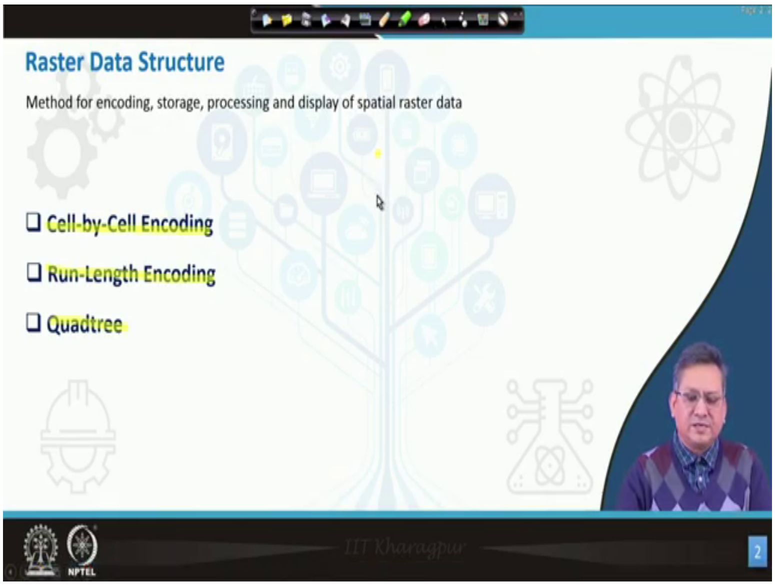
Next we go on to the line and you see that if we have this line encoded I mean it is a smooth line you can I mean do a lining of this particular line, but here the line gets pixelated and the contiguity there would be an issue in the in terms of contiguity. So, you see the gray values that is representing this line they have a very small joint over here. So, we have to have algorithms if we want to convert this pixel value into a line in this raster image to a line.

So, it is possible to in interconvert vector data into a raster data and a raster data into a vector data in the GIS mode. So, if you are coding some vector values vector entities you can always code it or convert it into raster layers those modules are always there in most of your GIS packages. Now going to the polygon areas you see suppose this is a land use parcel now say suppose this is a playground and this is a water body. Now this all the pixels within this particular playground with have the same values.

Now, when you see this, we will see that we have a very refined polygon out here which may be replicating a real world scenario that I mean your raster data this jagged pixels may not be able to I mean record because what happens is in some areas your the area would be extended beyond the area of this particular polygon in the vector and in some areas it could be less.

So, this is the limitation of representation in form of vector raster data set and the advantages of representing it in the form of vector data set, but as I told you that if you have continuous set of data where in your the data values are changing for every pixel contiguous pixel like you have height information or temperature information in that case it is always better to encode the data as a raster data set.

(Refer Slide Time: 17:37)



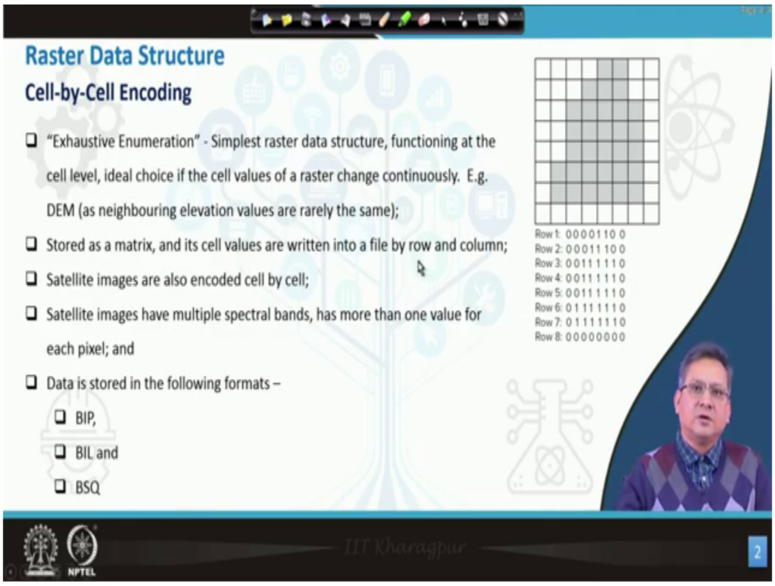
Now, let us see what are the methods for encoding or storage processing and display of the raster data. So, in this raster data structure I mean these are the aspects which I mean are I mean studied when we are talking about the raster data structure. So, how we encode how we code the data how we store the data I mean it has to be I mean done in a very efficient manner.

So, that in a minimal space you can store maximum amount of data how we process the data say suppose you run an algorithm on the data. So, how we can process the data and finally, the display of the data in the output I mean you may have a your monitor or you can take it as a map or a print. So, as a hard copy print. So, these are the important aspects when we are talking about raster data structure.

So, this raster data structure basically looks into encoding of the data. storing of the data processing of the data and display of the data. So, the different ways in which this can be done. The first one being the cell by cell encoding method. This method is the first method in which we do a cell by cell encoding and we shall see how we do this the next method is the run length encoding.

So, in the next method is the run length encoding and the last one that is being used here is the quadtree encoding. So, these are the three important methods by which I mean we encode, store, process and display raster data. So, let us see each of these processes one by one.

(Refer Slide Time: 19:41)



So, the cell by cell encoding. So, this method is the simplest method of storing data raster data. Now it is also known as exhaustive enumeration. It is quite evident from this word exhaustive and enumeration that each cell is encoded we are storing data for each and every cell at the cell level and I mean it is encoded as numbers.

So, the advantage being for this type of encoding is that you did not have any losses you do not lose the data. If you are taking an image you are storing each and every cell value, you would be able to reconstruct back the same image in its original form the way you have captured it.

So, when you are seeing an image through your camera it is an image analog image, but when you capture it became becomes nothing more than a matrix of numbers like in the earlier slide we have seen that you are recording images as an array of number as a matrix of number.

So, basically your image when you are storing it as a raster is nothing more than a matrix of number. So, you are recording a matrix of number. So, when you are doing a cell by cell encoding you do not lose on the level of details. So, it is the ideal choice when you want to reconstruct the original image. So, most of the image processing operations whenever you are doing planning say suppose we are doing a classification and you I mean order for a satellite image.

So, when you get that particular satellite image you should ensure that those data values are the original data values. The encoding that you do is you would be able to reconstitute or reconstruct the original image in its original form. There is no dilution in terms of the data value when you reconstruct the image. So, example I mean whenever we want to do some kind of an analysis on the elevation data we should have the original data and not a degraded data.

So, you can see that each of these rows have the data values I mean it has data values like your I mean 0 1 0 0. So, these are the representative data values for this particular row. So, you have four 0s out here four 0s which are these and then you have two numbers two 1s these two and then you have two 0s again.

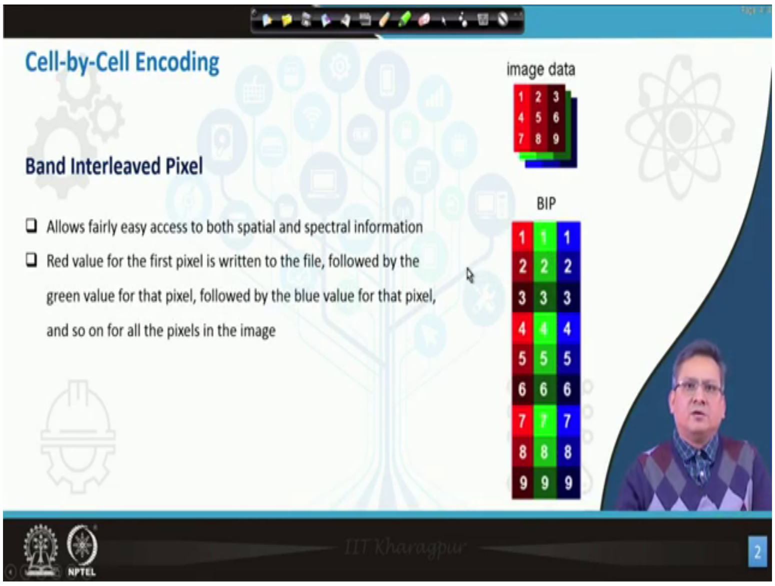
So, all these each of the cells are basically encoded I mean it has I mean coded as numbers and there is no loss in terms of data you would be able to reconstruct this data. Now satellite images as I was discussing I mean it is also encoded cell by cell and it would have a satellite images would have multiple spectral bands by spectral bands I mean that whenever you are taking an image say suppose you take an image from your mobile camera it would have different colors you would have a red color a green color and a blue color.

So, it would have three matrix it would have three matrix of numbers. So, whenever you are overlay overlapping this numbers it would result in other different colors. The secondary colors the tertiary colors would be I mean function of the data that is recorded in these three bands that is the red green and the blue band the r g and the b bands. So, similarly satellites record data in multiple band in multiple spectral band. We are talking about these satellite images and referring to this term called spectral band because of the spectral wavelength wherein the imaging has been done.

So, it could have more than one spectral band and it would have more than one value for each pixel because it has multiple bands. So, in different bands for each pixel you may have different reflectance for the red, green and the blue color. Now this data can be stored in any of these three formats. The first one is the BIP format which is the acronym for the term band interleaved pixel, the second one that is the BIL format is the term is the acronym for the term which is known as band interleaved line and the last one which is BSQ is the acronym for the term band sequential.

So, we shall see all three data formats how the data is stored in each of these formats and what are their advantages or disadvantages.

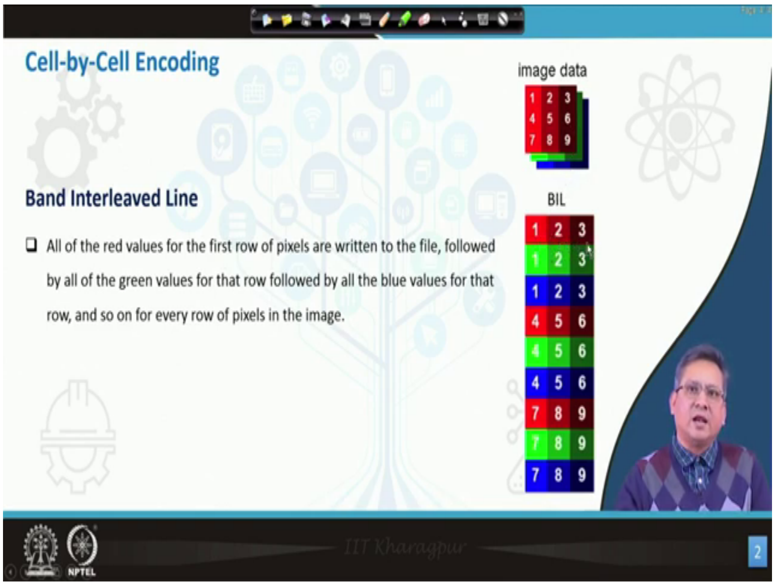
(Refer Slide Time: 24:52)



Now, this band interleaved pixel it allows easy access to both spatial as well as the spectral information like how this data is being stored. You may have three matrix three array of numbers the top one you can see is the red channel, the next one is the green channel and the third one is the blue channel or we can also call it as a blue band red band and the green band.

So, when the data is being stored you can see that we have first we have the first I mean your pixel value for the red channel then the green channel and the I mean the blue channel. So, then again we go back to the red channel, the green channel and the blue channel and you see the data that is recorded as two and similarly for the other data value. So, it is recorded in a sequential manner and it is interleaved that is each of the bands are interleaved with the pixel values.

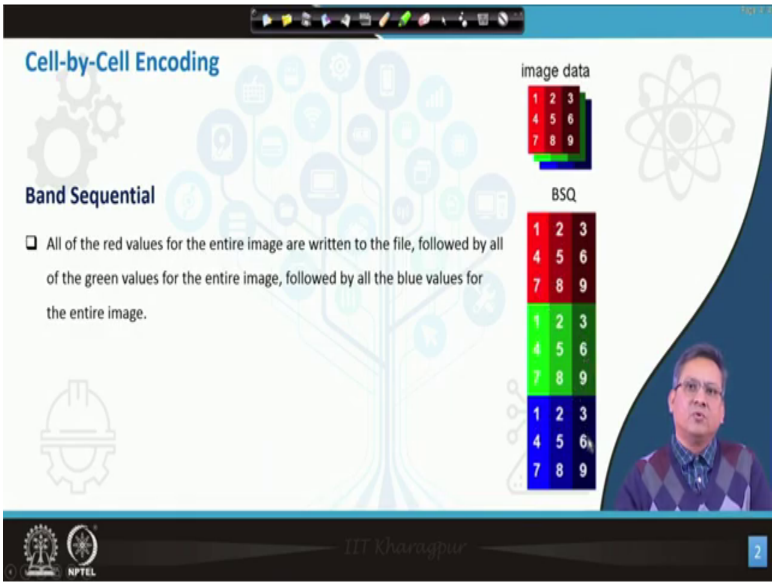
(Refer Slide Time: 26:01)



So, next one is your band interleaved line. So, you can go through the text meanwhile I will explain to you what this thing is. So, in this case what we take is we take the first line of the first layer that is the red layer and we write it down in the data in the sequence as we are storing the data. Next we store the; we store the second line that is in the green channel you will have this 1 2 3 values of the column 1 column 2 and column 3 values for the first row and the fourth I mean the next dataset would be the first row or the first three columns of the first row of the blue band.

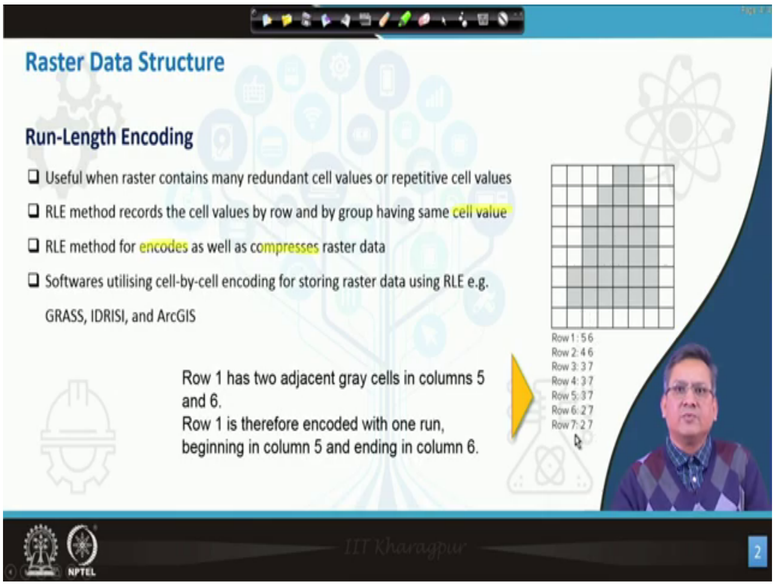
So, you can see that origin of the data set is this top left corner very important origin of these matrix or this data set is the top left corner and we are reading this data from the top. So, you can see the unlike the last time wherein we had we are picking up pixels individually sequentially for each of these bands together and writing it. In this case we are writing the entire line by bands and padding it with the next data sets.

(Refer Slide Time: 27:19)



So, this format is known as the band interleaved line then we have the band sequential wherein we have we write the data set pertaining to one band in the file and the next I mean the layer or the information in the first pixel of the next layer starts from after the end of the first layer first band or the first layer that is the red layer. So, we have the data for the I mean red band then we have the data for the green band and then we have the data for the blue band and so your retrieval method for this particular data has to be different.

(Refer Slide Time: 27:59)



Next we are talking about the run length encoding. So, it is useful when the cell has many redundant values. Let me explain this when we are talking about redundant value it means that there could be some background data values which are similar same data values. So, you may have few pixels having different data values, but background pixel having different data values. Say suppose I take an image of a your aerial image of a city in that city we may have a river.

So, you know that I mean those river would have the same pixel value because it is all water body. So, those contiguous values may have the same values. So, when you want to store that data since the pixel values are same we can devise a method wherein you can do some amount of data compression by I mean we will see an example of this.

So, what we do is it records the data values by row and by group having the same cell value. So, this is very important I mean when we are having the cell values we are trying to group it by rows and by group of pixels having the same cell value very important.

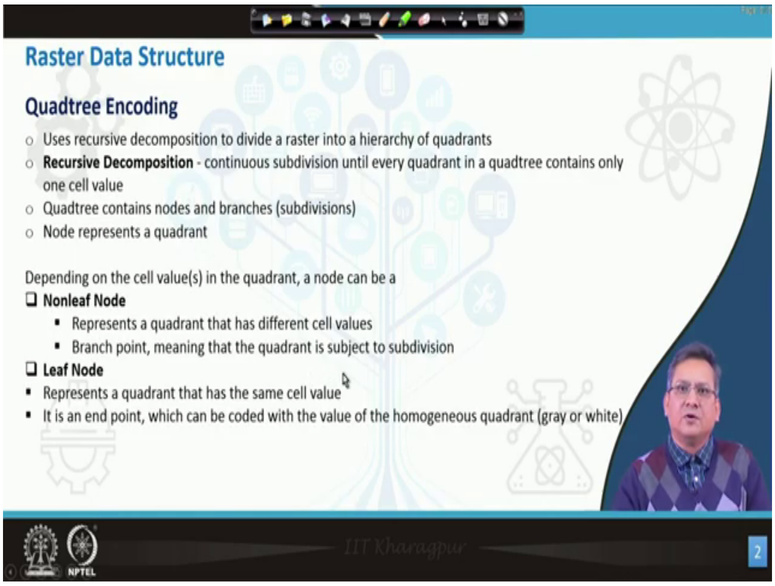
So, we also see that in this process what will happen is it will not only encode the data it will encode the data and as well as it will compress the data your data will get compressed in the given process as you are encoding the data it will also get compressed. There are some softwares which utilizes the cell by cell encoding such as GRASS, IDRISI or ArcGIS where in the store the raster data by cell by cell encoding process and uses the run length encoding.

So, in this first row you can see there are four cells which are white in color then we have two cells which are gray in color and two cells which are again white in color. So, what we do is in the row 1 we have data values only in the 5th and the 6th column this is one two three four fifth and the sixth column we see we have the data values rest these values the white data values are the background data values. So, we need not record those data values and we can only record the data values which are important for us.

So, we record only the data values which are there. In the line 3 row 3 or row 4 you can see there are four contiguous values which are of same numbers. So, we say we this for the 3rd row when we come to the 3rd row you can see that the for the 3rd row of data from 3 to 7 columns data from 3 to 7th column they are of the same value. So, we encode or record this data set.

So, likewise I mean you can see for other and you can see whether this has been written correctly or not. So, example say suppose row 7 when you go there you can see the data encoded for 2 to 7 has the same value.

(Refer Slide Time: 31:27)



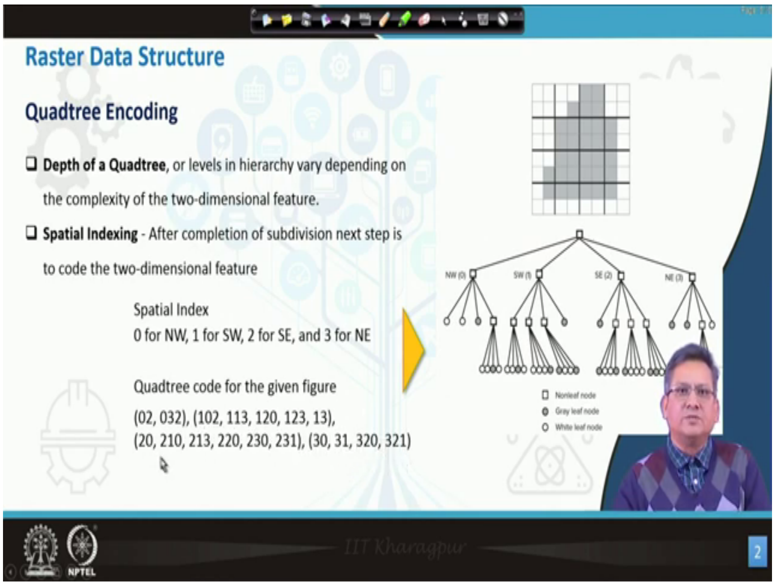
Now, the next one is the quadtree encoding which uses a recursive decomposition. What happens is it divides the raster into a hierarchy of quadrants of equal numbers. So, I mean this every it keeps on dividing or subdividing the quadrants until each quadrant contains unique cell value one cell value ok.

So, what happens is whenever I mean you have a quadrant with different cell value it will again subdivide. So, as a result your image may have different levels of quadrant I mean we will look at the picture and then it will be more clear to you. So, in this case quadrants quadtree would consist of node and branches that is the subdivisions and nodes would represent each of the quadrant and depending on the cell value in the quadrant a node can be a node leaf non leaf node or it could be a leaf node.

Now a non leaf node would represent a quadrant that has different cell values and there would be a branch point which basically would refer that the quadrant is subjected to subdivision I mean from that branch point whenever you have a branch point it means that quadrant is going to get subdivided.

Now, talking about the leaf node you it basically would represent a quadrant that has the same cell value. So, in the leaf node when we are talking about a leaf node each of the cells in that particular quadrant will have same set of value. So, it is the endpoint which can be coded with the value of homogeneous quadrant I mean gray or white.

(Refer Slide Time: 33:25)



So, if we see this particular image you can see this particular image has been divided into quadrants. So, first you divide this image into four quadrants then you can subdivide each of these quadrants again into four sub quadrants and then again each of these sub quadrants can be further divided into micro quadrants. So, you can see that these images are getting subdivided.

So, here we are talking about the concept of the depth of the quadtree. So, depending on at what level you are getting unique values. Say suppose in when you are subdividing this particular square or this particular image having this I mean data values you see after the first subdivision when we have this four quadrants it is further subdivided into four quadrants. So, this particular quadrant where my cursor is this quadrant this quadrant this quadrant and this quadrant work as leaf node.

So, you can see this quadrant this quadrant this quadrant these are the data quadrants I mean we have 1, 2, 3, 4 and 5 data quadrants at this particular level. So, we have 1, we have 2, we have 3, we have 4 and 5 data quadrants in the sub quadrant in the first level of hierarchy. So, these are the leaf nodes. In our last slide we were talking about non leaf nodes and the leaf nodes. So, you can see that leaf nodes are shown as circles and non leaf nodes are shown as squares or boxes.

So, you can see at the next level of hierarchy since you do not come across the all the pixels which are of uniform I mean values what you do is you do another subset of this particular I mean say this particular quadrant and divide it into again four pixels. So, you will have four quadrant within this particular pixel within this particular quadrant sub quadrant.

So, then again you can find out what are the leaf nodes and the value how it is to be encoded. So, for this encoding this we do a spatial indexing after the subdivision is completed and we arrive at all the leaf nodes we can do a spatial indexing and spatial indexing this rule is generally followed for a for Northwest quadrant we refer to it as 0, for the Southwest we refer to as 1 for Southeast we code a number 2 for Northeast it is encoded as 3.

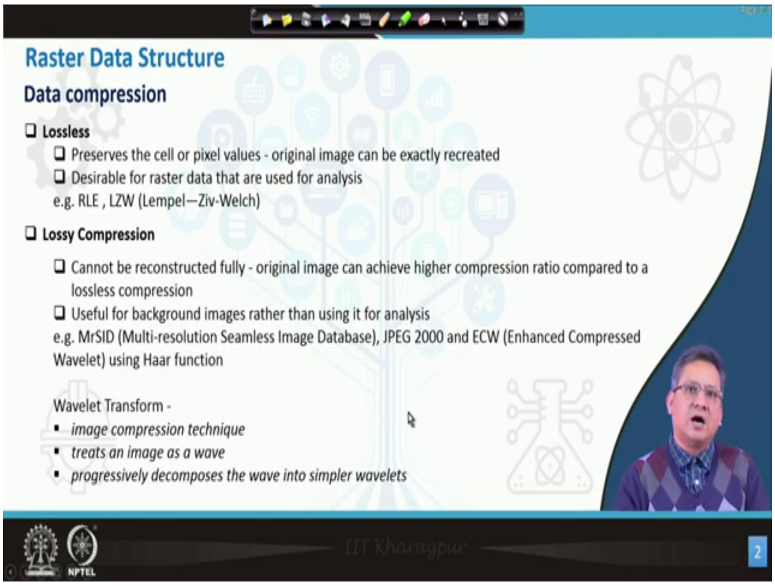
So, we start from this Northwest quadrant come down to the Southwest quadrant we come down to the Southeast quadrant and again go back to the Northeast quadrant. So, this is how it is encoded as 0, 1, 2 and 3 value. So, when we are subdividing this again you will see that it is having say suppose we are taking this quadrant we are when we are having this particular quadrant this would have a value of 0, 1, 2 and 3 say suppose we this we have all uniform values.

So, this is a leaf node. So, we do not need to subdivide it further, but when we are subdividing this particular semi I mean quadrant I mean sub quadrant you see there are again four I mean your I mean levels and in that each of these leaf would have values of 0 1, 2 and 3 depending on this spatial index.

So, we can code the quadtree and we can code the leaf. So, I mean you can read this particular thing I mean you have the first that is 0 and 2. So, that is the Northwest that is the 0 value and 2 is I mean the first subdivision that you do I mean in that first you are taking the Northwest quadrant that is 0 and in that 0 the 2 value that is 1 and 2 value has the data has the data I mean rest are all background values.

So, the gray one is the data value. So, you can see this entire quadrant is filled up with data. So, this is the leaf node. So, you can write it as value 2 because this quadrant is the I mean your Southeast quadrant and we said for the Southeast quadrant this value is 2. So, likewise you can do the coding.

(Refer Slide Time: 38:33)



So, if you have any problems you can get back to me, you can question me and we can do some exercises of how to do the spatial indexing when we are doing a quadtree encoding. Now last we come to the data compression. We have talked about data compression some concepts where we have seen that I mean you do a encoding I mean run length encoding quadtree encoding hm.

So, the those methods are lossless methods. The methods that we have talked about they are lossless methods and it preserves the cell values or the pixel values. So, you can again get back the original image you can recreate the original image. This is desirable for raster data wherein these data sets are going to be used for analytics if you are doing some kind of an algorithm applying some kind of an algorithm on this data set it is always preferable that you go for a lossless compression.

Now, examples of this compression is RLE LZW, it is also known as Lempel-Ziv Welch that is why this term comes from and RLE you are already familiar about it that is the run length encoding. Now next we have the lossy compression in which we lose the datasets I mean there is dilution in terms of the data value. So, you cannot I mean reconstruct the original image which you have used for compression.

So, I mean if you have a very high compression ratio you will lose on the data set. You can try doing a small exercise you can have a JPG image and you can compress that JPG image when you save as another filename it will ask you the I mean the level of compression.

So, what you can do is you can change the compression level to 80 percent, say 60 percent or 40 percent and see and open up the data in one window side by side to see what happens to the image after you decompress the data after you again recreate the data after compression I mean when you open that particular file what has happened to the image.

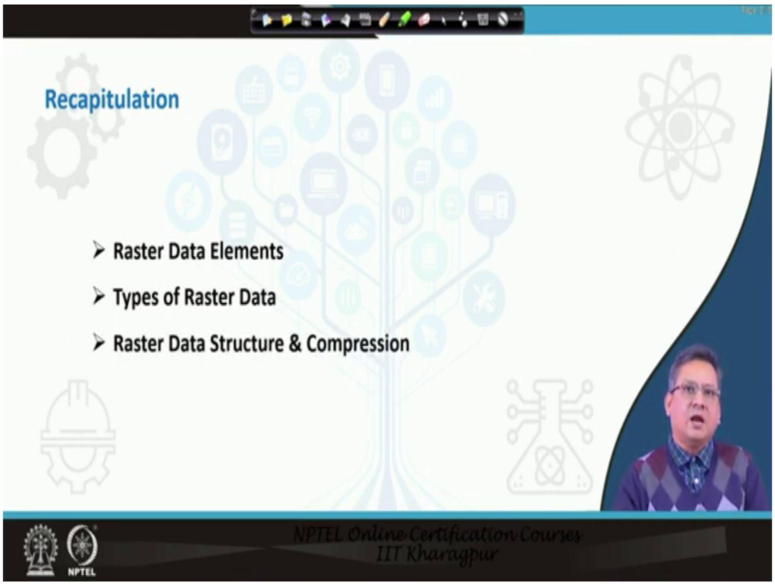
So, you would be able to see that the original image is lost if you do a lossy compression. Now this is useful I mean each of these methods have their own advantages lossless compression would the data size would be much bigger in comparison to the lossy compression. So, each of them have their advantages and lossy compression would be useful for us when we have background images. Say suppose you have to display some background image. So, they in that case we go for lossy compression and it is not being used for any sort of analysis.

So, examples of these type of data set is MrSID which is known as Multi-resolution Seamless Image Database. Its a more recent data format then we have the JPEG 2000 format and we have the ECW that is the enhanced compressed wavelet transform wavelet I mean image which uses the Haar function I mean it is the wavelet Haar function which is being used for data compression and if you want to learn about it is you can refer to it in books of image processing when they talk about image and image compression you will see the different types of algorithm that are being used for lossy compression.

So, we shall limit our discussion to only I mean getting a very brief knowledge about the different types of compression techniques. If you want to read about lossy compression you can read books on your I mean image processing. So, this is what we are going to cover till date and when we are talking about the this wavelet transform I mean wavelet compression, it converts the image into a wave it treats the image into a wave.

So, this is basically a compression technique wherein the image is treated as a wave and it progressively decomposes the waves into simpler wavelets into smaller or simpler wavelets. So, this is in just what your lossy compression is using the Haar or the wavelet transform.

(Refer Slide Time: 42:52)



So, a recapitulation of what we have done today we have talked about the raster data elements, we have seen the different types of raster data and we have talked about the raster data structure and how we can compress the raster data.

Thanks for your patient hearing.