Geo Spatial Analysis in Urban Planning

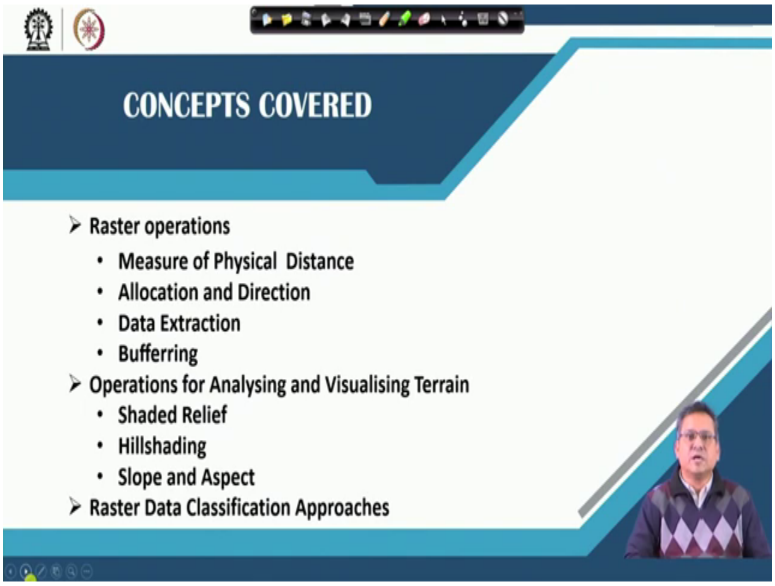
Prof. Saikat Kumar Paul
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture – 10
Raster Operations, Terrain Visualization & Classification

Welcome dear students, welcome back to this course on Geo Spatial Analysis In Urban Planning. So, today we shall look into Raster Operations, we are in the module 2 regarding I mean we are going to look into the GIS functionalities. And spatial analysis basics we have dealt with the vector analysis, today we have also dealt with 1 module related to raster analysis.

Today, we are in the 10th lecture where in we are going to discuss about the raster operations. How the terrain can be visualized if we have a digital elevation model which is also a raster data? How a terrain can be visualized? What are the different ways to visualize it? And how we can do a raster data classification? So, I mean a brief into all these would be discussed in today's lecture.

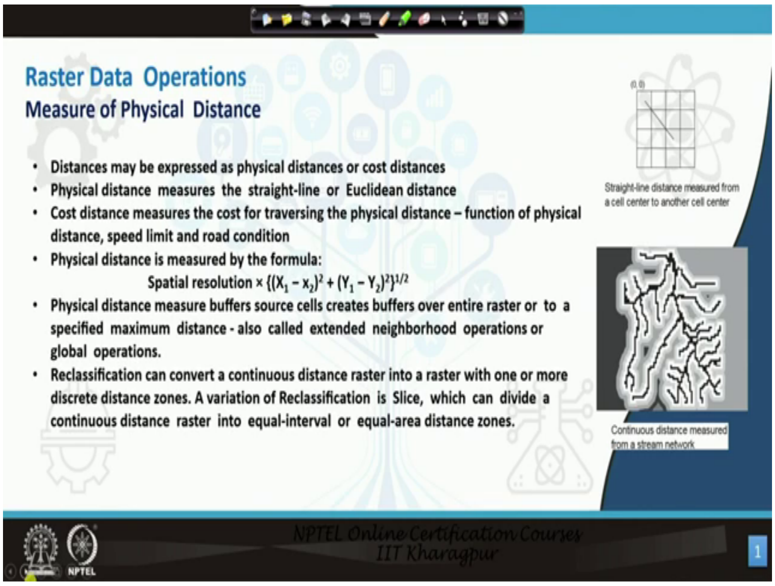
(Refer Slide Time: 01:15)



So, the concepts that we are going to cover today are the raster operations, we shall look into a measure of the physical distance. We shall look in to allocation and direction, we shall look in to data raster data operation regarding data extraction. We are also going to cover different aspects of buffering of raster data and see how it is different from your vector buffering which we had already I mean done earlier.

We shall also look into operations for analyzing and visualizing terrain using different approaches. Like, one of them is to calculate the shaded relief, I mean we can also do a hill shading or it is also known as hypsometric shading and then we shall calculate the slope and the aspect. And finally, we shall look into concepts of raster data classification the various approaches and how we can calculate error when we are doing a raster data calculation.

(Refer Slide Time: 02:26)



So, I mean regarding the this data operations we had talked about in our earlier I mean lecture regarding the data being analyzed in a local context in a neighborhood or context or in a regional or a global context. So today, first we are talking about the measure of physical distance which is I mean you already know how we can measure distances using vector, but we can calculate distances using raster’s as well.

So, these distances we can calculated as physical distance or as cost distance. Now a physical distance measure is a Euclidean measure it is a measure of the Euclidean distance that we had already seen and it is basically a measure of the straight line. And cost distance is a function of the physical distance it is also a function of the speed on a given segment of a road or also on the road condition.

So, these factors would be governing the cost of traversing a link say suppose, we have a road segment which is shown as a line either vector line or a raster line we can work out the physical distance as well as the cost distance. So, in today's I mean context we are going to talk about finding out the distance with respect to a raster data set.

So, I mean we generally measure the physical distance as I mean as a product of the resolution of each of the cells that is the spatial resolution what is the size of the raster cells in your particular raster data and we measure I mean multiply it with the distance I mean that is the square root of the centroid of the pixel origin pixel to the destination pixel.

So, I mean it is used of the distance is measured using your Euclidean formula to work out the physical distance in case of a raster what the difference between the vector data distance and the raster distance is that in vector you have the specific point coordinates which is more accurate in terms of measuring the distance wherein in case of a raster we generally measure the distance from the centriod of the pixel. So, there could be some inherent limitations in terms of the exact measure when we are measuring the distance using raster data.

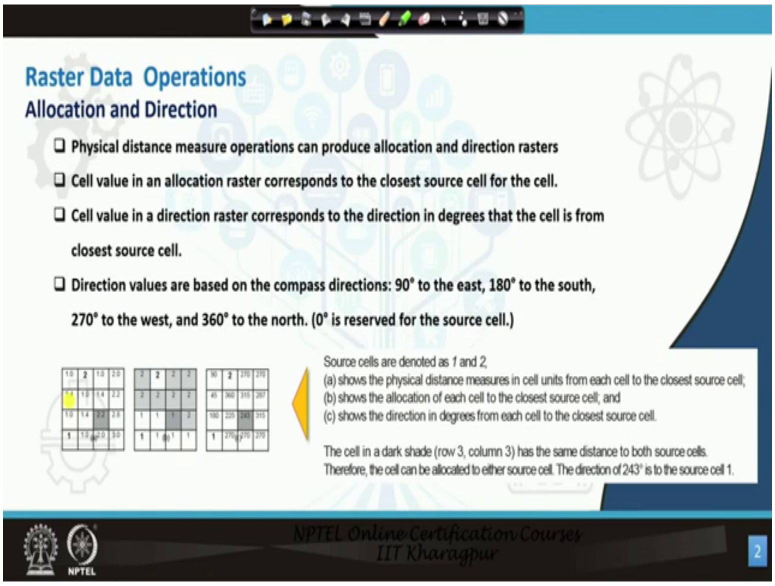
Now, the physical distance measure it buffers source cells and it would create buffers to the specified maximum distance and it is it could be a also neighborhood operation or it could be I mean a global operation. Now, in case of raster I mean data operation where we are measuring the physical distance we would have to reclassify the data or we would have to convert or regroup this physical measure because since this would be a continuous distance measure in case of a raster; so we would have to discretize the distance zone.

So, we can do this there is a operation it is known by different name's in different software's this operation is known as slicing which can divide a continuous distance function of raster into equal-intervals or equal-area distance zone.

Now, you can see this particular example wherein we have this particular raster where we have I mean these 2 cells which are connected by lines. So, we measure the distance centroid and then we I mean from the origin to the destination cell and then we calculate the distance using the equation that we had already talked about.

In the second image, you can see that there is a continuous distance measure from a stream network. So, in this case a slicing has been done to convert the continuous distance raster into different discrete distance zone. So, this is very important whenever we are measuring physical distance using raster data operation to slice the distance, once we have measured the distance from a linear feature.

(Refer Slide Time: 07:12)



Now, talking about allocation and direction we can measure the physical distance and produce allocation and direction rasters; for example, I mean we can have the cell value in the allocation raster which would correspond to the closest source cell for that cell we shall see in due course through an example. So, the cell value in the direction raster it would generally correspond to the direction in terms of degrees that the cell is closest from the source cell.

So, I mean we can use the compass direction suppose I mean; we take a clockwise measure. So, 90 degrees would be your east, 180 degrees would be your south, 270 degrees would be the west, and similarly 360 degrees or 0 degrees would be the north.

So, in this particular example you can see again that in the first image there you have the raster cells, which shows the physical distance measure in cell units from each cell to the closest cell. So, you have these particular cell values which are value 1 and value 2; the value 1 is the source cell and 2 is the destination cell. Now, the values in this particular cell; gives you a measure of the distance from the centroid of this particular cell to the neighboring cells.

So, from 2 to go to this particular cell you will have to traverse a distance of 1.4 because that is square root of I mean 1 squared plus 2 squared. So, that would give you root over 2 which is equivalent to 1.414 so, which is I mean rounded up to 1.4. So, you can likewise see this particular cell that this cell is displaced from 2. So, we find out the distance from each of these cells to the nearest pixel.

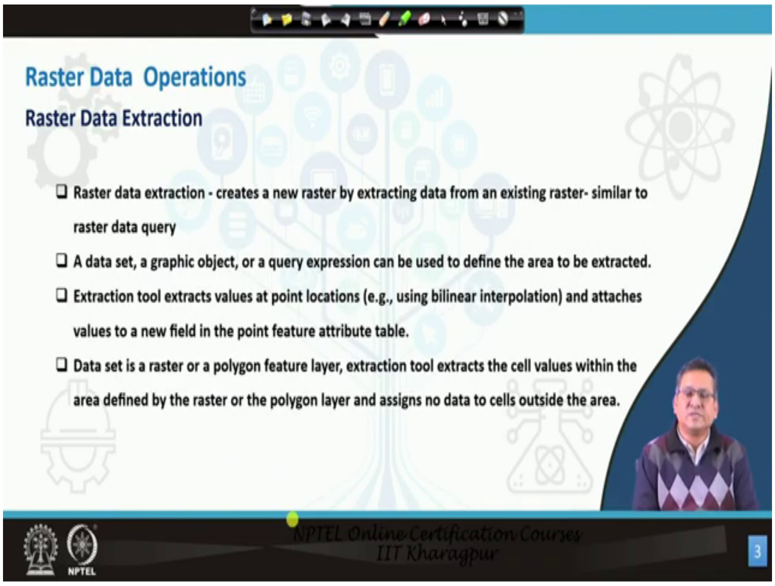
Now, in the second image you can see that it shows the allocation of each cell to the closest source cell. So, wherever I mean you have this source cell 1 and 2 this white cells it shows that they are allocated they are I mean in closest proximity to the source 1 and the cells mark 2 are in closest proximity to cell 2. Now here it is important to note that the distance from the centroid of this pixel 2 the destination cell; to this particular cell is 2.2.

And distance this distance is also same from the centriod of the source pixel, if this is the source and this is the destination pixel you see that this distance is same from both this source and the destination pixel. So, we have to allocate it to one of the either of the cells, so in this case there could be some kind of an anomaly.

Now, in the third image see you can see that it gives you the direction in terms of degrees from each cell to the closest source cell. So, I mean you would measure it from the I mean from 90 degrees and then we have said we can measure it clockwise. So, this would give you the directions in terms of degrees from each cell to the closest cell.

So, the cell in the dark that is shown in row number 3 and column number 3 it has we had said that the distance is 2.2 it is same from both these 2 cells I mean and the direction 2 to the 43 to the it is from with respect to the source cell 243 degrees is with respect to the source cell.

(Refer Slide Time: 11:22)



Now, talking about raster data extraction there could be cases where we want to extract the exact pixel value of a raster location; say suppose, you may have say temperature values or precipitation values of a continuous data set over India and you may want to extract the data value for a particular city or a given location. So, we might need to do a data extraction of a specific points or regions; so in that case a raster data extraction operation would create a new raster by only extracting data from the existing raster I mean and it is this operation is similar to a raster data query.

Now a data set or a graphic object I mean we can use a query or a expression to define the areas to be extracted you can have a very complex query when you are running it or you can also put it in a equation or you can see if the data so you can have various ways of giving a query expression when you are doing a raster data expression extraction.

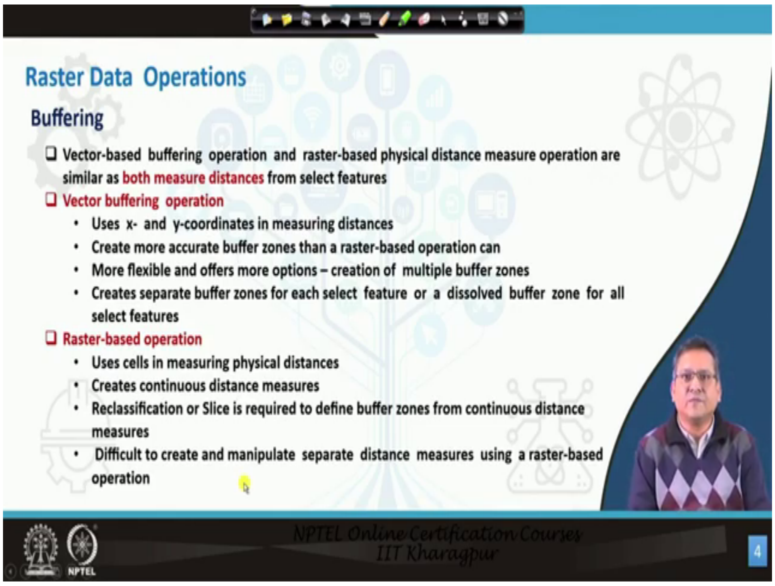
Now, the I mean it would extract the point locations I mean for example, I mean you can have a bilinear interpolation we had already talked about bilinear interpolation when we were studying the sampling approaches resampling approaches so the nearest neighborhood I mean operator the bilinear interpolation, the cubic interpolation, so we had seen what is the what are these types of interpolation.

So, we can extract the value using to I mean interpolation techniques such as bilinear interpolation we can use other interpolation technique as well and it would attach this particular value to the new I mean new field in the new layer. So, it could be a point feature I mean we can extract create a point vector data set and we can extract the values from the raster data set and tag it as an attribute to the point attribute feature table.

Now, the data set could be a raster layer that is the input layer or it could be a polygon feature layer and extraction tools would basically extract the cell values which is defined by the raster or polygon layer like; in our earlier point we were talking about extraction using point feature we can also do the same operation using a raster feature or a polygon feature using a vector polygon feature or raster area feature.

So, what we can do is we can extract that cell values within a defined cells of raster which is like a mask or a polygon air polygon layer and assigns no data to cells which are outside these mask layers of the raster or the polygon feature layer.

(Refer Slide Time: 14:38)



We can also do buffering like we had done the buffering in case of vector layer where in the buffering of a operations would give you the physical distance measure the physical distance and it is also similar to I mean both the vector and the raster operations where we can measure the distances from selected feature. Now, when we were talking about vector buffering operations we had used the x- and y-coordinates while measuring the distances.

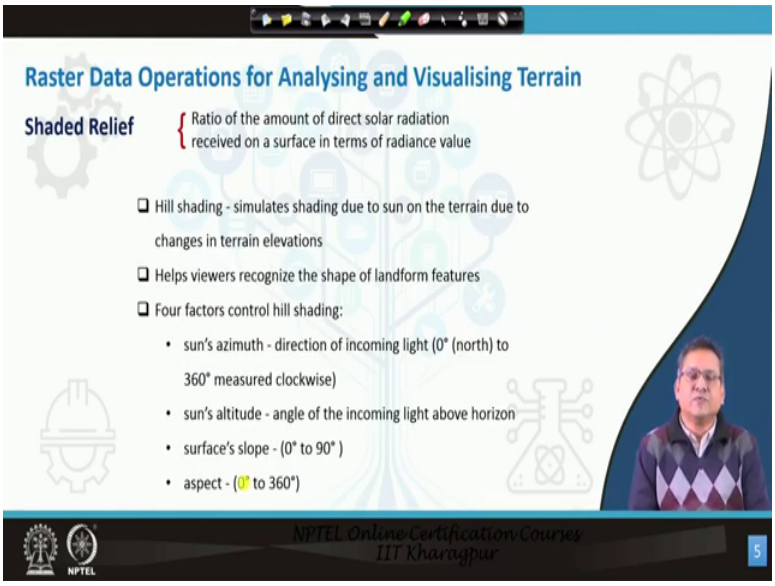
Now, we can also I mean in the case of your vector buffering we had seen that we can create very accurate buffer zones in comparison to a raster buffer I mean when we are doing a raster buffer we cannot have such accurate buffers because your line coordinates would be very specific or point or polygon coordinates would be very specific, so you can create very accurate buffers, but with raster data sets you may not be able to create such exact buffers.

So, the vector buffering operation was more flexible and it gives more options in terms of creation of even multiple buffers zone not just one single buffer, but multiple buffer zones. So, I mean we had seen that we could be we could create separate buffer zones for different selected features or we can dissolve the buffer zones in the boundary for the all the selected features.

So, I mean these options are available in the vector buffering operations; but when we are talking about raster buffering it uses cells in measuring the physical distances and generally it creates continuous distance measures. And in the measure tool we had seen that we generally, splice those distance measures I mean when we have these continuous data values we would be splicing those and grouping it into different distance regions.

So, we said that slicing is required to define the buffer zones because this operation would give you a continuous distance measure. So, it is difficult to create or modify separate distance measures like in vector I mean buffering mode we had said that we can create separate buffer zones for each select feature. But in case of raster data it would be difficult for us to create separate distance measure using raster based operation.

(Refer Slide Time: 17:22)



Now, let us talk about I mean representing data or analyzing and visualizing data in case of a terrain where we are talking about height information. So, we had said one such example of data could be a digital elevation model or it could be a digital terrain model so generally, I mean we can visualize the terrain using a shaded relief model.

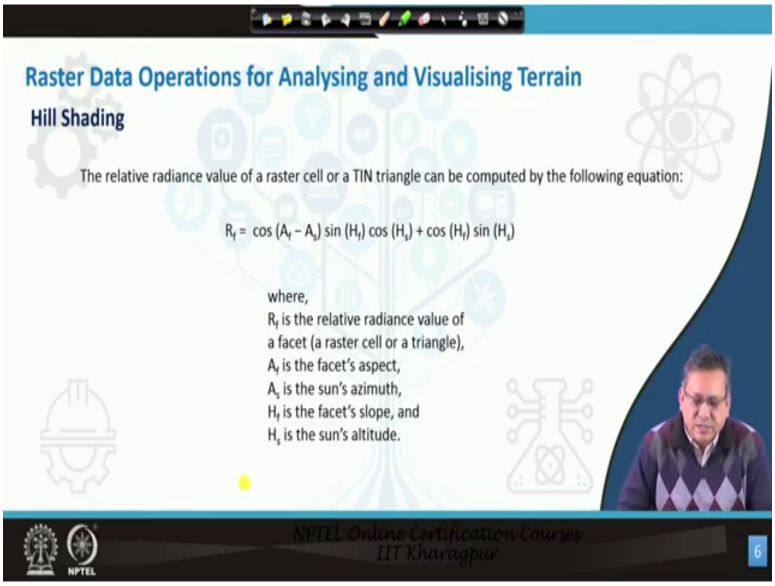
So, in this we basically work out the ratio of the amount of direct solar insulation or direct solar radiation received on the given surface and it is generally work it can be worked out in terms of your radians values. So, you can see that this method could be very well used for working out the physical I mean quantities of radians values.

So, this is the very interesting tool wherein we create the hill shading which would simulate shading due to the effect of sun on terrain because there could be changes in the terrain elevations, so you may have hills or mountains and you can see those the effect of the hills in the image and you can perceive the third dimension because of the this aspect of shaded relief. It would help viewers to recognize the shape of the landform features.

Now, there are four factors which would generally control the hill shading the first one is the sun's azimuth or the direction of the light I mean from which the sun's light is coming to the given sight. So, as we had said earlier the convention is to measure I mean we assigned 0 degrees to the north and we measure the direction in a clockwise fashion. So, again I mean your east would be 90 degrees, south would be 180 degrees, west would be 270 degrees and again north would be 0 or 360 degrees.

The next factor which I mean controls your hill shading or shaded relief would be the sun's altitude that is the angle of the incoming light with respect to the horizon there is also the effect of slope the topography may have a kind of a slope. So, depending on the nature of the slope the surface would be affected and also the aspect the direction in which the slope extends. So, we shall see in our later slides the different concepts of the slope and aspect.

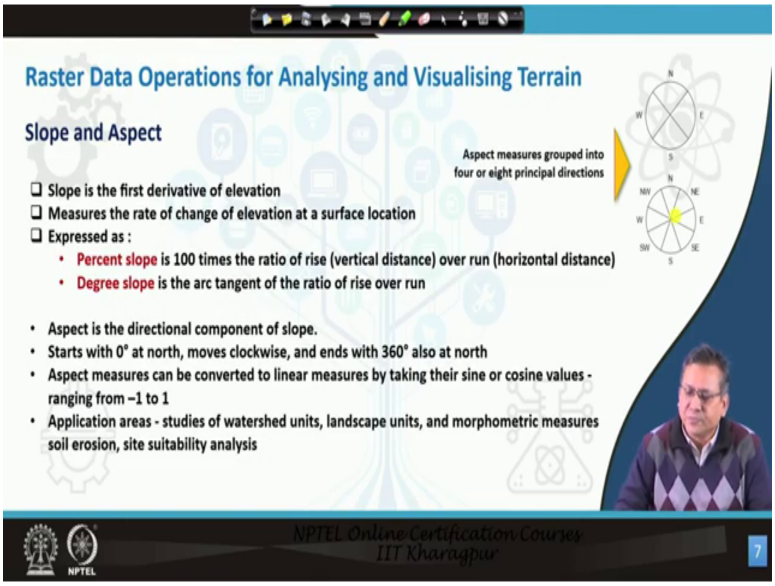
(Refer Slide Time: 20:17)



Now, we can also do a hill shading so this is done using this particular equation I mean wherein you have this factors which are R f is the relative radiance values, a is the facet or a raster cell or a triangle. Now we have A f which is the facets aspect I mean we can have a facet when we are having TIN a triangulated irregular network, so we can have a facet or it could be a raster cell also so we can; we can; we call it as a facet your A s is the sun's azimuth, H f is the facet's slope and H s is the sun's altitude.

So, these values are put in this particular equation to work out the value of R f which is the relative radiance values of the raster cell. So, this I mean equation can be run to generate the radians values of the raster cells or it could also be run on a triangulated irregular network or TIN data set which is a vector data set to calculate the effects of hill shading.

(Refer Slide Time: 21:36)



Now, talking about aspect and the slope we were talking about slope and aspect been one of the factors which basically has an impact on your hill shading. So, talking about slope first it is the first derivative of the elevation that is it is the rate of change of elevation with respect to distance. So, I mean it is expressed as either percent slope which is 100 times the ratio of the rise in terms of your vertical distance to the horizontal distance that is also known as run.

Now, the second way which in which your slope is expressed is degree slope in which we calculate the arc tangent as the ratio of rise over run. Now, we can also work out the aspect which is the directional component of the slope say suppose; if you have a pixel or rectangular cell and it has a particular inclination if you put a drop of water at the centroid of that particular pixel the aspect would be the direction in which the water would flow, the drop of water would dyne or flow.

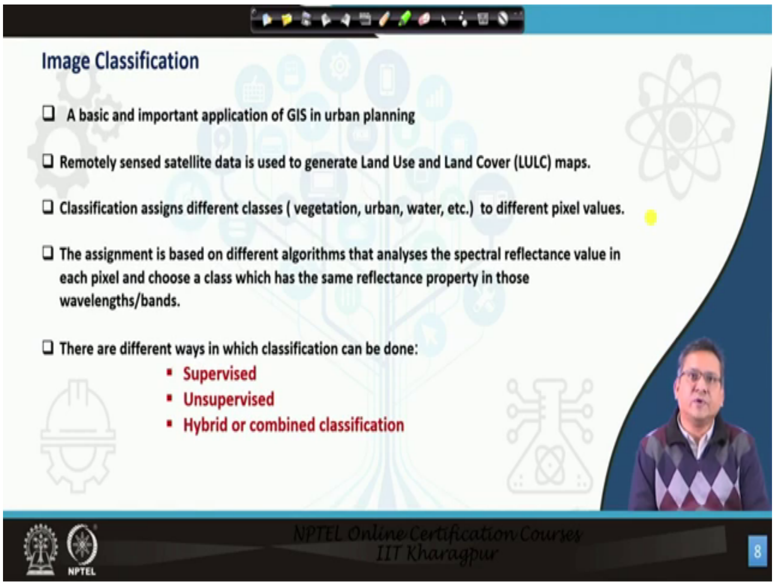
So, it is basically the it gives you the directional component of the slope because from slope you can only measure the first derivative of elevation that is the rate of change of elevation, but it is only the aspect which will give you the directional nature of the slope. So, again we use the same convention that 0 is the north and we measure it clockwise.

Now, yours aspect measures can also be converted I mean you can easily convert it into linear measure since these are these terms are in degrees. So, you can either take a sine or a cosine of these particular degree values and you would get values ranging from minus 1 to plus 1. So, we can also convert the this aspect values into linear measures. So, I mean we generally use slope and aspect extensively whenever we have DM data and we can run it on different types of analysis.

So, it is used extensively when we are studying watersheds, when we want to generate watersheds from a digital elevation model we can use the slope and the aspect. So, I hope all of you are aware of watershed, if you are unaware of watershed please look into the definition of what a watershed is. So, we can also work out the landscape units, we can also do morphometric measures, we can find out I mean this morphometric measures can be used for studying soil erosion.

This slope and the aspect can also be used for site suitability analysis in case of a urban planning job. So, I mean we have the this visuals in which you can see that the aspect measures are grouped either into four or eight principle cardinal directions. So, in the first top image you can see it is grouped as north east west and south and in the next one you can see there are eight principle directions in which you have sub element of north east, south east, south west and north west.

(Refer Slide Time: 25:21)



Now, let us go into image classification that is the last part of this lecture. So, when we are doing the raster analysis this is a very important tool which would give us some kind of slicing of the data based on the input values. So, these can be done on a statistical data sets, it could be done on multiple thematic layers and you can classify it all the different data sets and the attribute values based on the inputs.

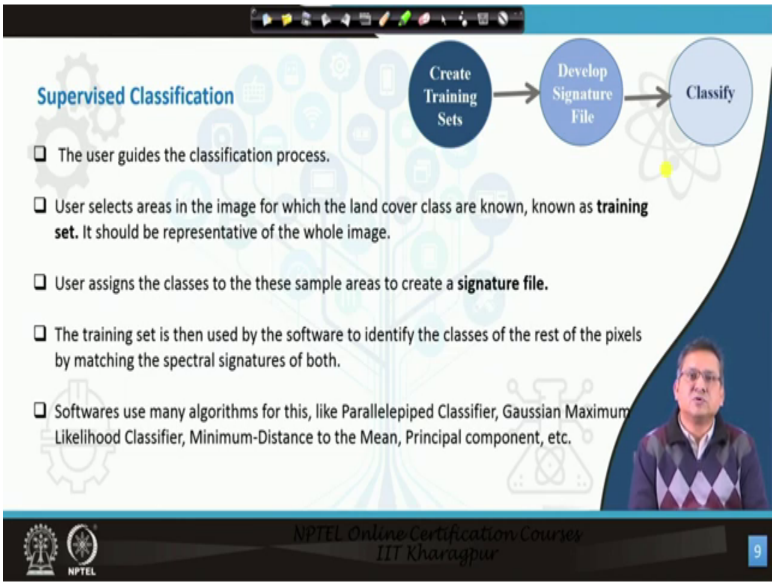
So, you can run this I mean the same algorithm the same concepts on vector data as well as raster data. So, in case of raster data we generally do image classification of remote sensed images, so it generates the land use and the land cover maps. So, I mean whenever we are running the classification it generally would assign different classes to the input pixel values. The assignment the way this image classification is done is based on different algorithms.

So, some of these algorithms are either supervised or some of the algorithms are unsupervised. So, we can find out the spectral reflectance values and distance from the class means which are used as a guiding tool while we are running this classification algorithms. So, I mean we would choose a class which has the same reflectance properties of the class means in the wavelength bands.

I mean when we are talking about remote sense image we have different wavelengths or bands in which the data is captured or stored. So, we can have your the visible bands, we can have the near infrared bands, or we can have your thermal infrared bands, or the microwave bands. So, we can have different bands in which the image can be acquired and then we can choose an algorithm to classify the data.

Now, there are different ways in which the classification can be done that we had talked about. It could be either done in a supervised mode, or it could be done in a unsupervised mode, or it could be done in a way in which we can combine both these two methods which is also known as a hybrid classification approach.

(Refer Slide Time: 28:01)



So, generally in supervised classification what we do first is to we create a training set in which we have a raster data. And we select some regions or areas where the ground conditions are known to us or the ground cover is known to us, so we use that as training sets and then we create a signature file. So, the from this signature files we try to calculate the statistics of your mean, variance, standard deviation.

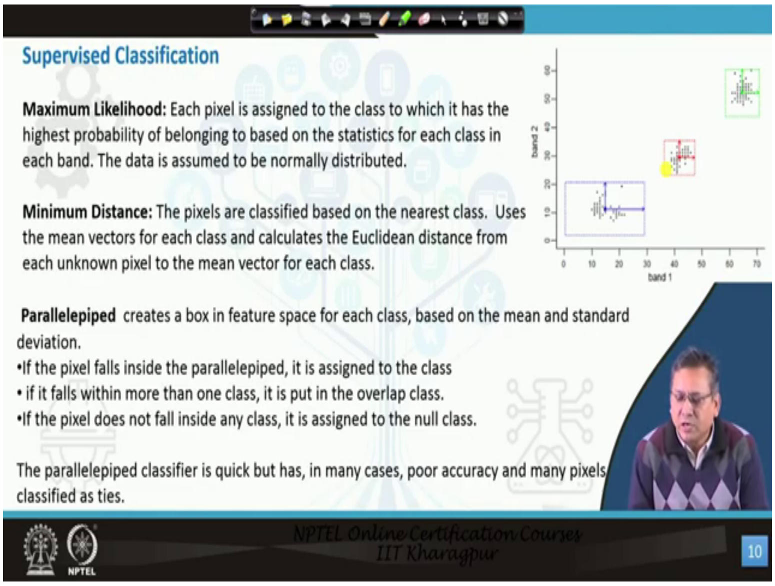
And we create the signature I mean values for these I mean statistical parameters. Then what we do is, we try to take each and every pixel and try to measure the distance to the means of these training samples or classes. So, wherever the distance is minimum it is used I mean the pixel is basically assign to that particular class, so that this is the last operation that we do on the data set.

So, I mean in this case u as a user would guide the classification process. So, it depends on your ingenuity of how good you are at interpreting the data or about your knowledge of the ground conditions. Now, I mean we had already talked about acquiring the training sets I mean it has to be representative of the entire image. So, that the pixels that you are using as training sets, I mean would have representation on the entire image.

So, I mean we can assign classes and create a signature file. So we can create a name or we can assign a color to the signature samples that we had taken. So, those output pixels would also have similar color; they would have similar same pixel values as your class number in which you have acquired the signature training sets. And it will also have the same name; either name that you are specify it that same nomenclature would be retained in the output image.

Now, the training set is used by the software to identify the classes and I mean we have different types of algorithms for classification which is used in the software's. So, you can read about parallelepiped classifier, you can I mean read about Gaussian maximum likelihood classifier, you can we also have the minimum distance to mean classification scheme, or we can also have a principle component.

(Refer Slide Time: 31:07)



So, there are different types of algorithms which are used or approaches which are used when we are doing a supervised classification. So, in the maximum likelihood classifier we try to identify pixels and find out the minimum distance to the mean of the different classes to that pixel value. And wherever the highest probability is there based on the statistics for each class in each band the I mean it is assign to that particular class.

And in this case the basic assumption taken is that the data is normally distributed. So, if you see the I mean histogram of a satellite image generally you would see an kind of a inverted bulk of. So, which is representative of a normal distribution and that is the basic premise of this supervised maximum likelihood classification algorithm that the data is normally distributed.

Now, the second one second approach is the minimum distance. So, I mean we create the classes based on the nearest class and we I mean calculate the mean vectors work out the Euclidean distance. So, there are different distance approaches we have different names for different distance and your equations. One such way of finding out distance is named on our great statistician that India has produced Professor Mahalanobis.

So, it is known as the Mahalanobis distance which is also used to do the supervised classification. Otherwise you have different types of distance such as Euclidean distance. We have the next one which is known as parallelepiped in which you specify a box which is basically the bounds of a particular class the upper and the lower bounds of a particular class. And it is based on the mean and the standard deviation of this particular parallelepiped trainers.

So, I mean if the pixel has values in different bands and it falls within the parallelepiped we call it has a parallelepiped, because say suppose you have two bands in your input raster layer you may have two bands. So, if you plot that pixel value in x and y you will get a feature space you will get the pixels will lie in that particular feature space.

Now, you can create parallelepiped parallelograms, so and in case you have more than two bands, so your data would extent to the third dimension. So, in that case you can extrude those parallelograms into parallelepiped, so that is how the name is derived. So, I mean we would work with multidimensional data and this is a approach wherein we give the bounds the upper and the lower bounds of the training data values.

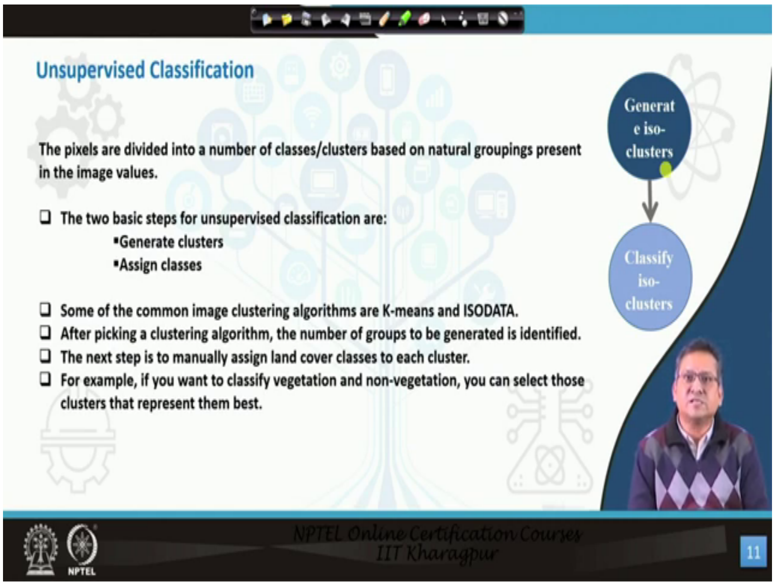
And wherever the pixel if it falls within those parallelepipeds, so it is assigned to the particular class and it may so happen that the training parallelepiped classes may not be adequate to cover the entire range of pixel values. So, there would be some pixels which would remain unclassified in such a case. So, I mean it may also happen that the bounds of the parallelepiped could be overlapping one another.

So, in case a pixel falls within such an overlap it would put it in one of the two classes depending on the statistical mean and the standard deviation values or it could also be put into a overlap class. So, we said that if the pixel does not fall in any of these parallelepiped classes, then it is classified I mean it is given or assigned as unclassified or it is assigned as a null class.

So, among these the parallelepiped classifier is the least computationally intensive and the quickest. So, I mean if you want to do run I mean quick classification then in that case the parallelepiped classification would be good. But, it has its inherent limitation in terms of poor accuracy and that your many pixels might remain might be assigned to the null class or it may remain unclassified.

So, I mean we see these this particular image which is the feature class which is the feature space of band 1 and band 2 data values. So, we have these class boundaries in which you see a blue class, red class, and a green class. And this arrows give you the distance vector mean vectors which gives you the distance from the mean of this particular data cloud.

(Refer Slide Time: 36:35)



Next we can also use a unsupervised classification method. So, in which the data would be divided into number of clusters depending on how many number of clusters you want and later on you can edit those clusters. So, I mean there are two basic steps wherein it is used for unsupervised classification.

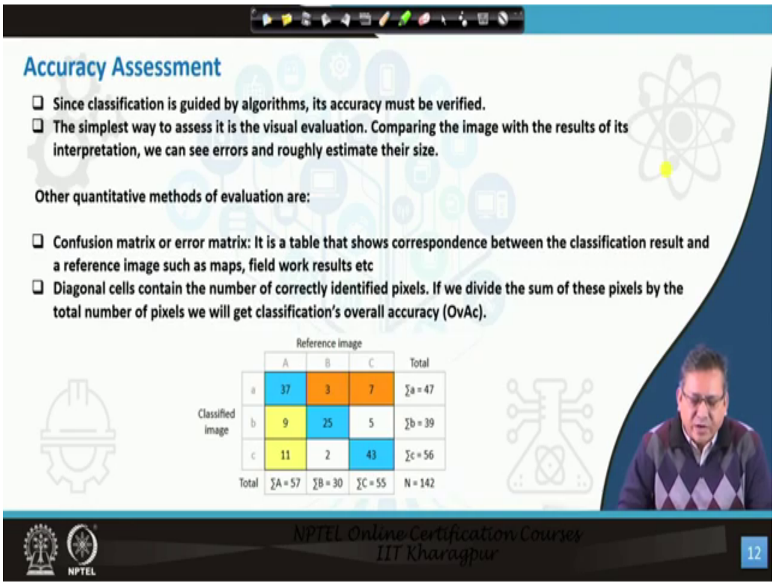
So, first is we generate the clusters and the second is we assign the classes of the pixels to this particular clusters. So, some of the I mean algorithms which come under this unsupervised classification are the K means or the ISODATA classification which are basically iterative algorithms. That basically your the entire data of a particular band would be partitioned into the number of classes that you desire.

So, then what we do is? We try to work out the statistical parameters of the means, standard deviation, and variance. And each of the pixels are then I mean assigned to this particular classes depending on the distance vector the mean distance vector to these particular classes. And again a second iteration is worked out taking into consideration the mean of this particular pixel.

So, in the second iteration your mean values of the classes will shift. So, we keep on doing this particular iteration till there is a time wherein the results of two successive iterations are I mean converging I mean the difference in the outcome of two subsequent iterations are within your acceptable tolerance limit.

So, I mean as we said after picking the clustering algorithm I mean we identify how many numbers of groups are to be done. So, I mean we assign the land cover classes to each of the clusters and I mean we can I mean generate your statistics of the this cluster values. And as I had told you we work it out in a iterative fashion.

(Refer Slide Time: 39:02)



So, the next part or the last part of this particular lecture is the accuracy assessment of the particular lecture of the particular classification scheme. So, I mean the most of the classification schemes are done through say a signature assessing signature which is not exhaustive for a given area which are basically representative signatures. So, it is important that we should test the accuracy of the classification.

So, we can do a visual evaluation, but the we it is it only gives rough estimation of the kind of errors it does not give us a statistical estimate it does not give us a quantitative estimate. So, to do a quantitative estimation we create a confusion matrix or it is also known as an error matrix. So, it is nothing but a table which shows the correspondence between the classification result.

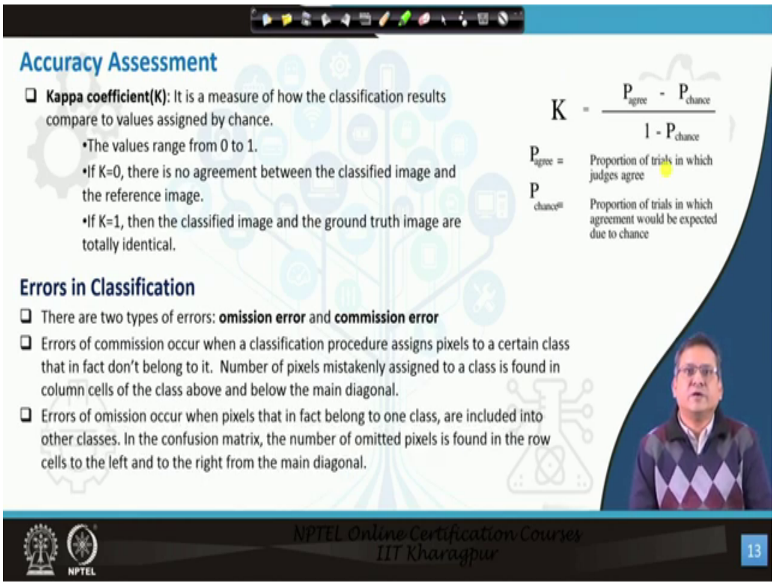
So, we can have the field data, we can create a stratified random samples depending on the area of the classified I mean data classes. So, we can create a stratified random sampling points and we can do a field work to find out what are the actual classifications or the classes of those points data points. And then, correspond it to the classes that have been derived through your classification algorithms.

So, these algorithms we had talked about this could be through your unsupervised classification algorithms or using say supervised classification methods. So, in this confusion matrix or error matrix the diagonal cell would contain the number of correctly identified pixels the correctly classified pixels would be located in the diagonal element of this error matrix.

Now, if we divide the some of these pixels by the total number of pixels we would get the classifications overall accuracy; overall accuracy. So, this is an example where in you can see image we have a classified image and we have points of the reference image. And we have the two classes that is the class three classes A, B and C.

So, in the reference image you would have gone to the sight to identify this all these points these 142 points to see what is the actual pixel value what is the class existing on ground. And the small a, b and c are the classes which are created through the process of classification. So, the diagonal pixel elements are the correct values correctly classified values wherein the half diagonal elements are the incorrectly classified pixels. So, we had talked about the how we can work out the overall accuracy.

(Refer Slide Time: 42:13)

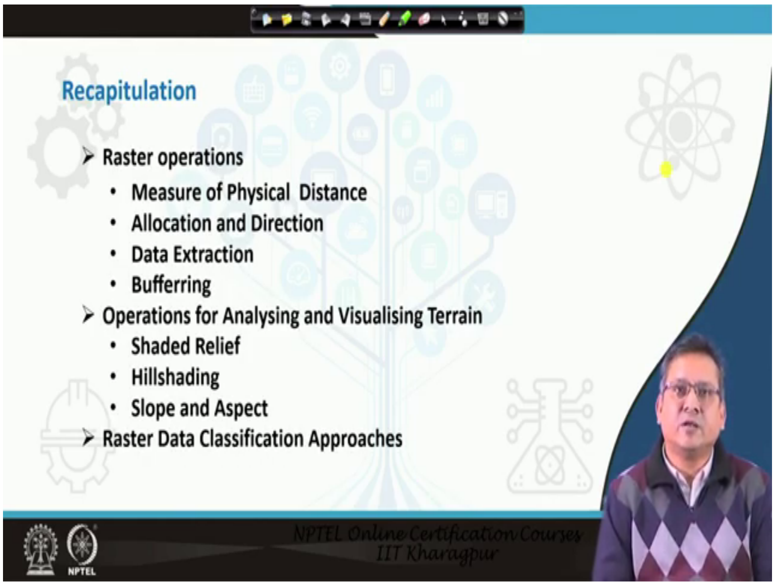


So, we have a measure which is known as Kappa coefficient which is used to compare the values I mean assigned and the values assigned by chance. So, this value would range from the Kappa value would range from 0 to 1 in case the Kappa value is 0 there is no agreement between the classified image and the reference image.

Now, if the Kappa value is 1 then the classified image and the ground truth image are exactly identical which is not possible. You know, I mean whenever somebody says that the K value is 1 for his image classification I mean you have to take it with the pinch of salt. So, you can also calculate the K value using this particular equation where wherein it is the ratio of the proportion of trials it is the difference of proportion of trials in which there is an agreement in terms of the data and the I mean existing values.

And the proportion of trial in which the agreement is there would be expected due to chance to the difference of chance I mean it is subtracted from a unit value. So, when we do this ratio we would generate values ranging between 0s and 1. So, there are two types of errors, it could be errors of omission or it could be errors of commission. So, I mean when the classification pixel assigns a pixel to certain class which do not belong to it is known as the error of commission and otherwise.

(Refer Slide Time: 44:08)



So, in this particular lecture if we do a recapitulation, first we had talked about raster operations in which we talked about the measure of physical distance. We had talked about allocation and direction, we had talked about data extraction buffering. And then we had I mean seen how we can analyze or visualize the terrain and what operation can be used.

So, the first one that we had seen was the shaded relief, second we had explode the hill shading concept of hill shading. And then we calculated the slope and the aspect and the last I mean topic that we had touched upon today is the raster data classification approaches.

Thank you.