REMOTE SENSING FOR NATURAL HAZARD STUDIES

Course Instructor: Dr. Rishikesh Bharti
Associate Professor
Department of Civil Engineering
Indian Institute of Technology Guwahati
North Guwahati, Guwahati, Assam 781 039, India
e-mail: rbharti@iitg.ac.in
Website: https://fac.iitg.ac.in/rbharti/

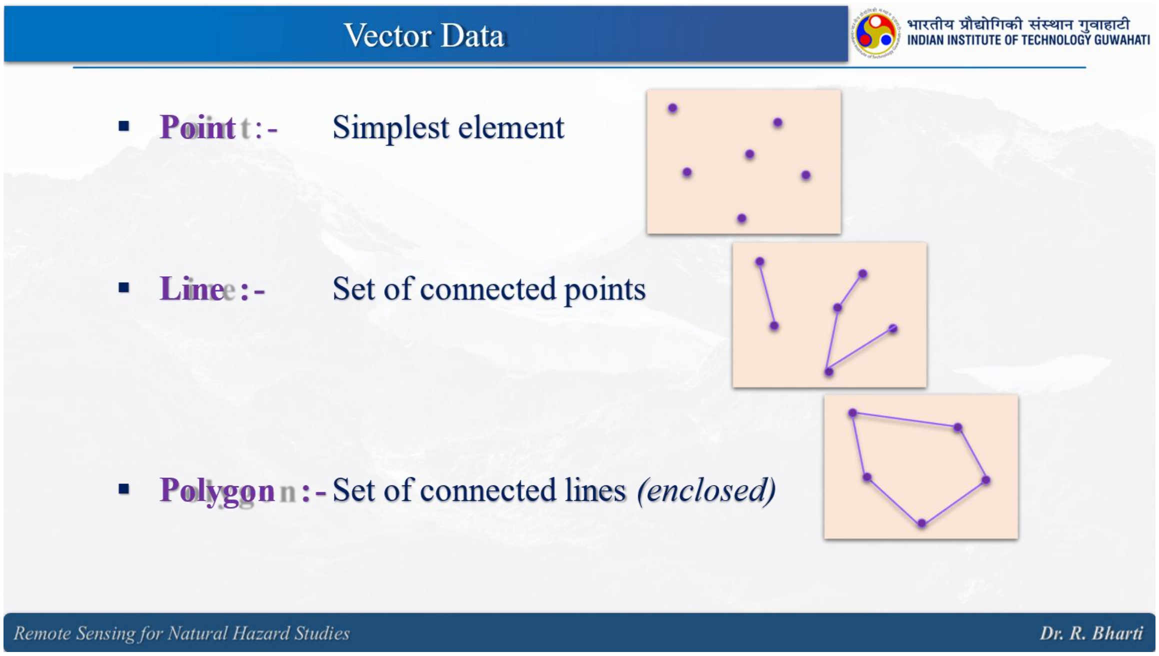
Lec 10a: Data Type and Remote Sensing Products-Part A

Hello everyone, welcome to Lecture 10 of the Remote Sensing for Natural Hazard Studies course. So, this lecture is on data types and remote sensing products. Like previous lectures, this lecture is also divided into two parts. So, this is the first part, and this is the first lecture of module 4. This is the potential of remote sensing in hazard studies. So, today we will learn about the data types and remote sensing products that are available in the public domain, and some of them are paid.

So, let us talk about the data types. So, before we go into the data types, let us talk about digital images. We all know that images are of two types: the first is an analog image, and the next is a digital image. So, what is the difference between the two? Analog images are produced using pencil and paper, whereas digital images are produced by electro-optical sensors.

So, here is one example, so you will easily understand this is an analog image and this is a digital image. Now, in remote sensing, we are talking about instrument-based measurement. So, here is one example of satellite data. So, here, when we capture this image using any electro-optical sensor, when you zoom in, you will be able to see the pixels, So, these are the pixels. These pixels are nothing but the digital numbers captured by your sensor, and these digital numbers are basically the reflected, emitted, or backscattered energy from the target, now, when we see the data, if you try to import it in any programming language, you will be able to easily see the associated digital numbers.

So, these elements of this matrix are basically the pixel values of your image. So, each pixel has an associated value known as the digital number, or DN value. And here you can see this is one example of how remote sensing sensors are capturing a particular area over time, and these are examples of remote sensing data and how we are utilizing it. So, we are able to monitor the glacial movement; we are able to produce the land use land cover. So, these are typical examples of remote sensing image applications. And here, this digital number, whatever we have captured, is basically helping you to extract the information, and remember, we have also discussed the classification, that is, how we are producing the land use land cover map. When we are going for continuous measurement, that means we have temporal images; we can have this kind of monitoring system, So, these digital numbers are very useful in extracting the information captured by our sensor, and they represent the target behavior. So, when we say data type, there are two major types: the first one is raster, and the next one is vector. So, here you can see that the raster basically has the pixels. So, that is why the grid cell, pixel, or elements are present in this particular image.



So, this is the raster data where the entire image or the data is divided into smaller portions or parts, which are called pixels, and these pixels are the reflected, emitted, or backscattered energies. Now, when we talk about vector data, you can see this is one line; it could be a road, or you can assume this is a river system, so these are represented with the lines, but some of the point measurements can be represented with this kind of point, so, points are here; you have points, a line, and the polygon. You can see here this is one polygon; this is another polygon. So, these 3 make the vector data point map, which is also a vector line map and a vector polygon map, and you can have a vector map of vector data that can include all these 3: point, line, and polygon together. When we talk about raster data, the area is divided into pixels, as we discussed earlier.

Stores images as rows and columns of digital numbers for each pixel. Remember, we talked about that when we export the image into any programming language, you will be able to see the background values associated with each pixel. So, these are basically your pixels, and these numbers are stored correctly. So this is very, very important when we talk about raster data; it is stored in rows and columns of digital numbers. Digital output from any camera or sensor is raster data.

So it does not matter whether you take the example of your handheld camera, your mobile phone images, or any instrument that is capturing the image that will have the raster data format. Based on the map scale or spatial resolution, each pixel represents the area on the ground. So, remember we talked about this satellite? Here you have a sensor; the sensor is basically an array of detectors, and each detector is looking at the ground. And then they are generating one pixel value, and these pixel values represent the area on the ground, whether it is 30 by 30 meters or 40 by 40 meters, or if it is map scale, maybe 1 centimeter on the image is equal to 1 kilometer on the ground. So, this is called scale, So, in this relationship, you have to remember the map scale or the spatial resolution that is represented by your pixel, which is very useful in representing the continuous variation of thematic or dense data.

Let us say I have a rainfall map, this is one state for which I have generated it. So, it is very ideal to use the raster data because it will show the slow variation in the values, so that you will understand better in some of the examples. Very simple data structure because here you have x, y, and z, so z is basically your dn value; in the case of a satellite image, z could be any measurement; it could be your rainfall, or it could be your temperature. It could be your sea surface temperature. So, anything is possible to have it in z value; , x and y will be fixed, that is your latitude and longitude.

But when you store these images and values, it requires a large storage space, because we are storing them in JPEG, TIFF, and GeoTIFF, which are commonly used data formats, but we have HDF, So, this is very popular for storing satellite images. Now, when we talk about vector data, we have points, lines, and polygons. Now when we talk about the point, this is the simplest element, here you have one value. So, this will have x and y location values, and the z can be any measurement that you have taken in the field; it can be rainfall or precipitation, temperature, sea surface temperature, wind speed, or anything you can store, or you can also have water table depth and air quality. all these are possible to store against this one point.

So, this is the simplest data structure. Now, when we talk about a line, it is basically two points that are connected by a line. So, essentially, it will have two points: one will be the starting point, and one will be the ending point. Here, this has 1, 2, 3, and 4 points, and these are connected with lines. So, a minimum of 2 points is required and a maximum can be anything, but it should not be enclosed.

When it is enclosed, we call it a polygon. So, here it is: the starting point and ending point are fixed. Now, here you can have an unlimited number of points in between. And here you can estimate the area. So, if you have a polygon, you can estimate the area; if you have a line, you can estimate the length of these two points or the distance between these two points. When you have point measurements, it is good for representing some of the values that you have captured in the field, so when we talk about this vector area separated by different polygons, the area here you can see is separated by different polygons. Statistics can be estimated for individual lines or polygons because you can estimate the perimeter, and you can estimate the area. So all of these are possible, and that is a very quick method. It is very useful in representing the discrete data. Discrete means having a very clear boundary.

So, let us assume this is a lake; this is a water body. So, the water body will end at this particular line. So, here if you have a raster, for example, then what will happen is that this raster may not match the boundary of this particular lake. So, in that case, this variation will not be shown properly. So, in such cases this vector is preferred. So, that is the meaning of discrete data. When you want to represent the discrete data, you should have the vector data, many attributes can be associated with each polygon. So, let us say for this particular polygon you have so many measurements; let us say you have measured the water quality. Or let us consider this particular lake; for this lake, you are going to have water depth, water quality, and a list of aquatic habitats. Then you can also find out the status of suspended sediment and what kind of flora and fauna are present here.

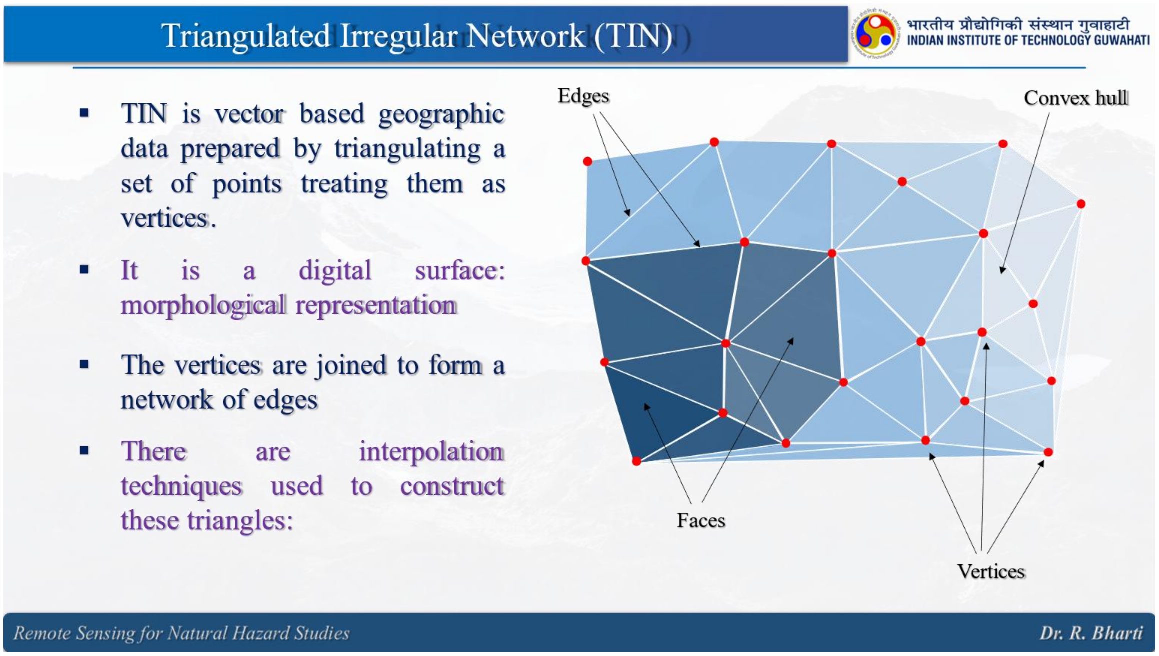
So, those things can be listed for this particular polygon. Whereas when we talk about raster data, it is divided into pixels, and here you can have only x, y, and z values. But when you are having this x, y, z, z can be reflected, emitted, or maybe some measured values that you have. But here, this list is not limited to this, so you can store maybe hundreds or thousands of parameters. So, that is why many attributes can be associated with each polygon; it is not only a polygon, it can also be a point, and it can also be a line.

It is a very complex data structure because you are going to store many parameters, and their relationships also need to be stored. Whether this particular lake is inside this or whether it is touching this particular boundary. So, how system will record this? So, that is why this structure is complicated. Good representation of data or maps requires small storage space because this whole thing will be stored in ASCII format. So, this requires much less space.

Now, when we talk about the attribute, let us say that this is one map, which is vector data, so here you have all these polygons. So now here, this is 1, this is 2, this is 3, 4, 5, 6, 7, 8, and we have this table. So, this is FID. So, this is 1 to 8; this is ID; this is your ID that you have given to this polygon. Now, we also have the information about the area. So, the area can be estimated. So, you can see that everything is listed here. Now, this is against 1. So, this is one. So, this is ID 2, the area is this much, and the value is this much. for 2, this is ID 3; this is the area, this is the value, and this is just for the example.

Now here you see all these are recorded, and now if in the future you want to know which areas are bigger in number. So, in the future, if you want to run a query that shows me the area which is greater than 4,345 square kilometer. So, except this all these 2, 3, 4, 5, 6, 7, 8, 9 that will be highlighted. So, here except this everything is greater than this area. So, similarly, if you want to run a query that shows me the land parcel that is the costliest in this particular area.

So, it will automatically highlight this 8, Id 8. So, this is 8. So, though this area is small, it is costly. So, such queries can easily be run on this particular attribute table. So, that is the advantage of having the vector data.

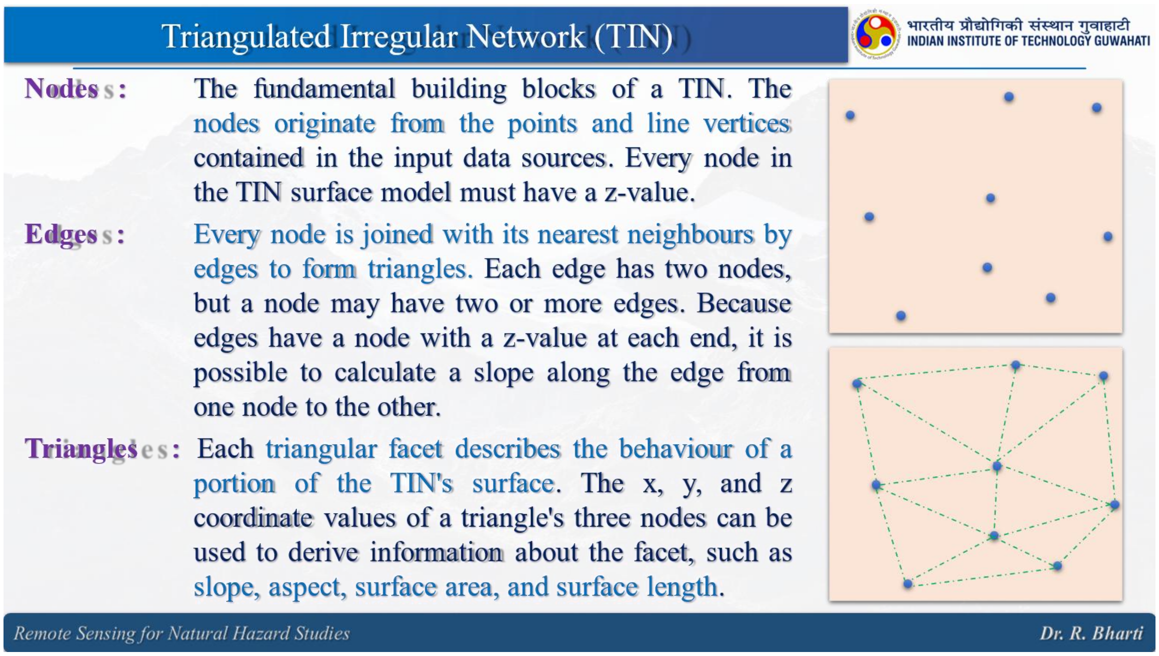


Now this is the third type of data, which is called a triangulated irregular network. So, here you can see that TIN is a vector-based geographic data set prepared by triangulating a set of points, treating them as vertices. So, we consider TIN, or triangulated irregular network, to be neither a vector nor a raster; it is the intermediate data type. It is a digital surface and basically the morphological representation. The vertices are joined to form a network of edges; you can see the vertices are joined to form a network. There is interpolation techniques used to construct these triangles because these triangles are the surface for which data is not available. We have the measurements for this, this, and this, but these are the places for which we are going to obtain the values using the interpolation technique.

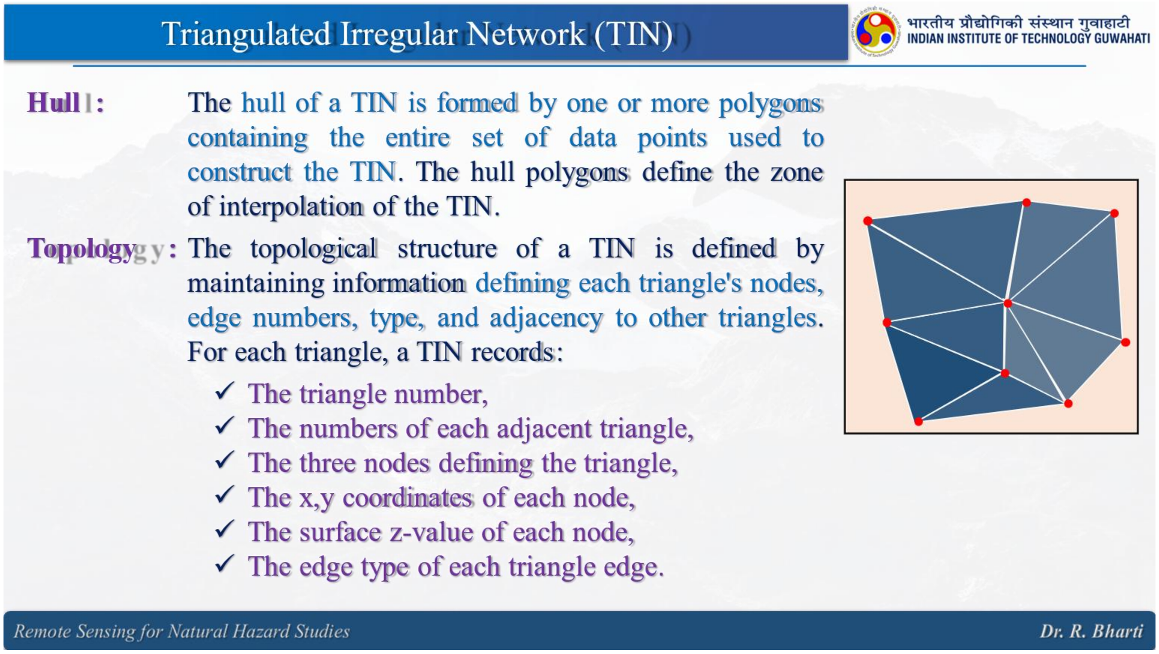
So, here you see this is the edge; these are the faces. And these are the vertices, and this is the convection, Now, when we talk about the tin, it is good for representing three-dimensional objects. So, here a three-dimensional parameter that is good, like one of the best examples, is elevation.

So, using methods of computational geometry, the points are connected into what is called triangulation, forming a network of triangles; the lines of the triangles are called edges, and the interior area is called the face or facet. So, here you can see this is the interior. While the TIN model is somewhat more complex than the simple point, line, and polygon vector model or the raster model, it is actually quite useful in representing the elevation. When we try to generate a digital elevation model, we are trying to represent the topographic variation of the elevation using the 2-dimensional images, So, here when we are converting the 3-dimensional information to 2-dimensional, we have to be very cautious or very careful, So, in such cases, we will have a greater number of points measured in the field. So, we can consider these the points that are measured.

And in between, we could not measure. So, we are going to have the interpolated values, and for that, this triangulated irregular network is very useful. So, there are different elements here. So, the nodes are the fundamental building blocks of a TIN. The nodes originate from the points and line vertices contained in the input data source. Every node in the TIN surface model must have a Z value.



So, here you should have the z-values; then only those values will be interpolated. Every node is joined with its nearest neighbor by edges to form triangles, and here it should not cross; we cannot have this kind of triangle. Here, the nearest point will be joined, and then that will make the triangles. So, this is the kind of triangle we will be making. Each edge has 2 nodes, but a node may have 2 or more edges because edges have a node with a z-value at each end.



It is possible to calculate a slope along the edge from one node to another because once you have the elevation, you can easily calculate the slope. So, this is the first derivative. And now here we have connected all the nodes with these lines, and we are making the triangles correctly. And remember, as I told you, none of them are crossing each other. So, all have separate triangles. So, now each triangular facet of the triangles describes the behavior of a portion of the tin surface. Now, when we talk about this particular area, this particular triangle should have a value closer to these 3 points. This is not very close to this point. So, if we consider this point to interpolate and derive the value of this triangle, that will not be correct. So, that is why the nearest points are connected and used.

The x, y, and z coordinate values of a triangle with 3 nodes can be used to derive information about the facet, such as slope, aspect, surface area, and surface length. So, all these values can easily be determined using these 3 values. Hull, the hull of a tin is formed by one or more polygons containing the entire set of data points used to construct the tin. The hull polygon defines the zones of interpolation for the tin. Topology is very, very important, So, topology is the structure of the TIN that defines each triangle's node, edge number, type, and adjacency to each other triangle.

For each triangle, a TIN records the following information. The triangle number, the number of each adjacent triangle, the three nodes defining the triangles, the x and y coordinates of each node, the surface z values of each node, and the edge type of each triangle edge are correct. So, to understand this topology, let us make this map. Here you have one rectangle, then you have some ellipses, here you have a triangle, then here you have a circle, and then you have a point. Now, how are you going to store how this will remember that I am inside this, this, and this, So, this is topology.

So, the adjacency is recorded here, which is very, very important, and suppose in another case you have this kind of map, or maybe you could consider the state boundary. So, this particular stretch of this polygon is common for this and this, but for this, this is the complete line. So, this kind of record will be stored in terms of adjacency. So, it will also remember whether this boundary is shared between polygon 1 and polygon 2; this is shared between polygon 1 and polygon 3, And this plus this makes the polygon 1 boundary.

So, all these will be stored. So, that is why when we say or compare this with vector data, the topology of TIN is complicated, So, when we use raster and vector data to represent our 3D surfaces. So, let us understand what advantages and disadvantages we will encounter. So when we talk about raster data, it is basically the pixel array of the pixels. And suppose that when we are trying to represent this water body, you can see the shape is getting distorted because we have the pixels here, and these pixels are not good for representing a discrete boundary. So, here if we have this, our lake will have this kind of shape, and this output channel will be like this.

Which is not correct; this is not possible on the surface, So, we can have this kind of flow of water; it can be irregular, but it will be smooth, So, this is not possible with the raster data. When we try to represent this water body with vector data, it is properly highlighted; you can see this is the output channel. But when we try to represent this forest, it will not have this kind of sharp boundary. So, here with vector data, that is the limitation because we cannot have gradual changes.

So, here what happens you will have this sharp boundary. Now, when we go back to the raster data, these pixels are good for representing this forest because they are slowly changing. So, if you compare these two, both have some advantages and disadvantages, but when we talk about the real world, neither of them is a true representation of it. So, what we do is use both of them in such a way that they represent the best of the real world. So, here you can see the forest is represented nicely; this is where we are using the vector data. So, this is properly represented. Represented, and here the other areas are also represented. So, when we compare raster and vector data, both have some advantages and disadvantages depending on the parameter or value you are going to represent or the information you are trying to represent; you have to choose raster or vector data. But if you are trying to put too much information in a single map. Maybe you can use both of them together, and then your area will have the best representation, Now, let us see some information about the digital elevation model, which is a very popular data product from remote sensing satellites. So, the digital elevation model represents height information in the z value, along with the latitude and longitude of the surface.

So, as I told you, the raster image has the x, y, and z values. Now here x, y is latitude and longitude; now z, instead of a digital number, represents the elevation. So, once we have the elevation, this kind of data is known as the digital elevation model. So, the digital elevation model represents height information in the z value along with the latitude and longitude of the surface. DEM products such as slope, aspect, basin information, etcetera, are widely used in many applications because once you have the elevation, you can easily calculate the slope, which is the first derivative of the altitude or the elevation.

Then, aspect what the direction of the slope can be; it can be between 0 and 360 degrees. Then, you can easily extract this kind of basin information. So, the most commonly used DEMs are Asatium, Aster, Ikonos, GTOPO30, and Cartosat. So, these are available in the public domain; you just have to create your login and then request the data; these products are freely available.

And you can have the elevation information about your surface correct. So, there are different techniques by which we generate the digital elevation model. The first one is radar interferometry, the second is satellite stereo pairs, then photogrammetry, laser altimetry, and the contour lines. So, the contour lines that are available in your toposheets can also be used to generate the elevation information, because these contours have the elevation information. So, along these contour lines, this height is fixed; maybe it is 100 meters from MSL, and this is maybe 80 meters. So, this difference is 20, and then you can use this to generate your digital elevation model, which can be used in your application.

So, one example of extracting information from the digital elevation model is given here; you can see this is the SRTM 90-meter DEM. I am going to extract the digital elevation values along two lines, So, the first profile is along this A, and the second profile is along this B, So, this is profile A; you can see this is 0. So, 0 is here. Now, along this line, what we have done is extract the values from this particular digital elevation model, those pixel values, and when you draw them, this shows this kind of profile, and you can calculate the slope, here you can see the height is in meters. So, this kind of information is very easily available when you have a digital elevation model of your area.

Now, this is for Profile B. Remember that when we see this digital elevation model, we generally use a blue to red color profile. So, here the red areas are basically the elevated areas. So, I suspect that there is some elevation here, So, these are some hills, these two, So, now I am going to extract the information along this B-line. So, here you can see these peaks are coming, which represent the hills of this particular portion, So, this is very useful in representing or identifying the elevation profile, calculating the slope, or determining the aspect, and it is very widely used in hydrological applications. So, when we talk about the digital elevation model, we have the satellite image, then we have the digital elevation model for the same area, and once we have this, we will have a better representation of the surface. So, here you can see it. That satellite image and DEM are both merged together, and this is the 3D representation of that area. So, here only the surface characteristics were available, here only the elevation profile was available, but when we merge them, we can see the 3D effects. So, with this, I will end part 1, and we will continue Lecture 10 with the next part.

Thank you very much.