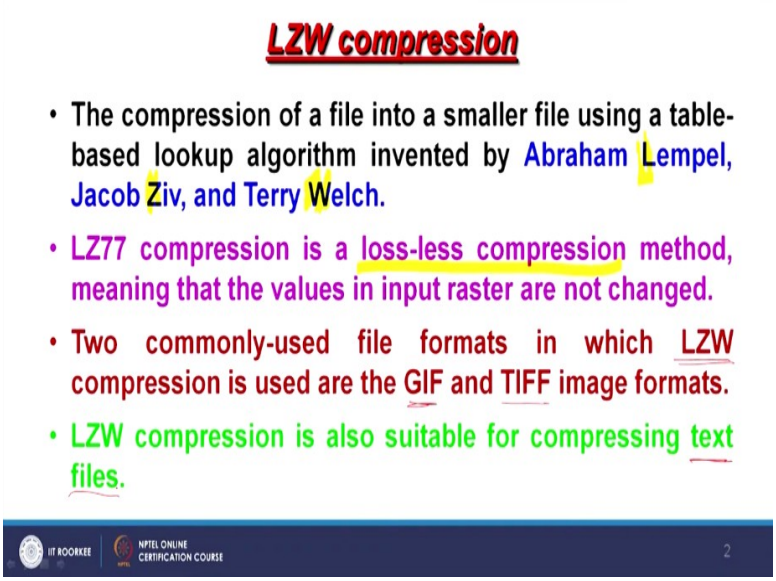
Geographic Information Systems
Prof. A. K. Saraf
Department of Earth Sciences
Indian Institute of Technology – Roorkee

Lecture - 12
Raster Data Compression Techniques - 02

Hello everyone! and welcome to the discussion on raster data compression. This is Part 2. So, in the previous discussion basically we have discussed the requirements of raster data compression. And also, we have discussed 4 very basic techniques with few examples. And we will continue this discussion with some other you know variants of compression techniques or maybe you know different techniques or some additional on the compression.

So, there are various ways of, not exactly the way we have discussed in the Part 1 but different ways of doing compression.

(Refer Slide Time: 01:10)



Like for example, this LZW is very famous compression technique which has been implemented by many-2 softwares of GIS and digital image processing or even graphic softwares as well. This algorithm of this compression technique which is a table-based or lookup table-based technique. Lookup means the mapping of pixel values.

And that is what table-based lookup algorithm which was developed by 3 mathematicians which were Abraham Lempel and Jacob Ziv and Terry Welch. And because they have used the initials of their surnames. So, for LZW, L comes from here and Ziv is going as Z and then

Welch, W is going there. So, this is called LZW compression technique. Now there are various versions are available and like very latest one is LZ77 compression.

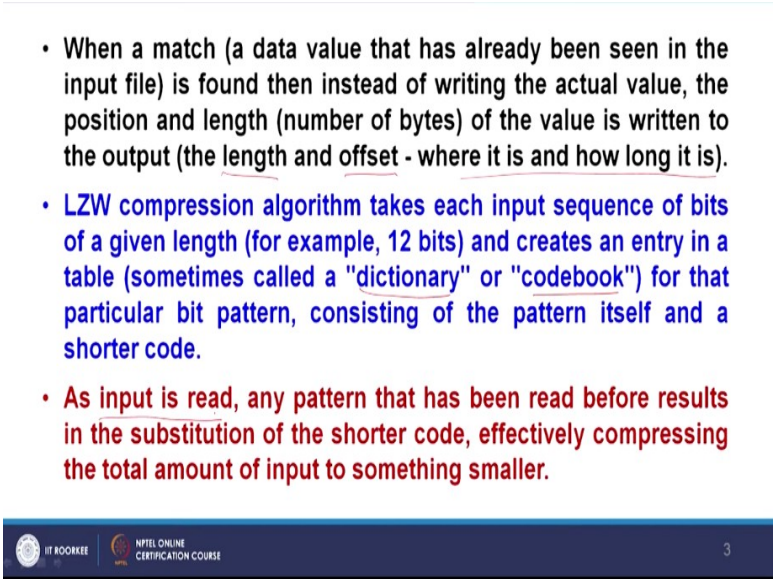
Again, it is a lossless compression technique. And that means whatever the pixel values which will be put to this compression, whenever you want to restore the original quality you should be able to do it. So, that is why it is very popular. Now, 2 commonly file formats which supports on which you can apply this compression technique.

One is the GIF one; that is Graphic Interchange File Format and then TIFF; Tagged Information File Format image. So, in these 2 image formats, you can still use LZW technique. But this is another important thing that on all image file formats may not support all kinds of compression techniques. So, you have to be little bit aware about that part also that which compression technique like. Img which is ERDAS imagine form file format for raster data.

Again, they may have their own compression technique but this particular technique cannot be applied there. And also, interestingly this is also suitable for compressing text files. Vector data compression we have learned because that is directly relevant with GIS and also raster data which is, we are going through. But just to add to our knowledge that LZW compression techniques can also be used to compress text files.

Generally, text files do not have much redundancy. But still there are redundancy about space, gaps, line, spacing and so on. And that can also be used to compress. Sometimes if a big file is there which we have to transmit through internet even if it is a text file, we look for some compression so that it can easily be transformed from one place to another. And when a match your data value that has already seen in input file is found.

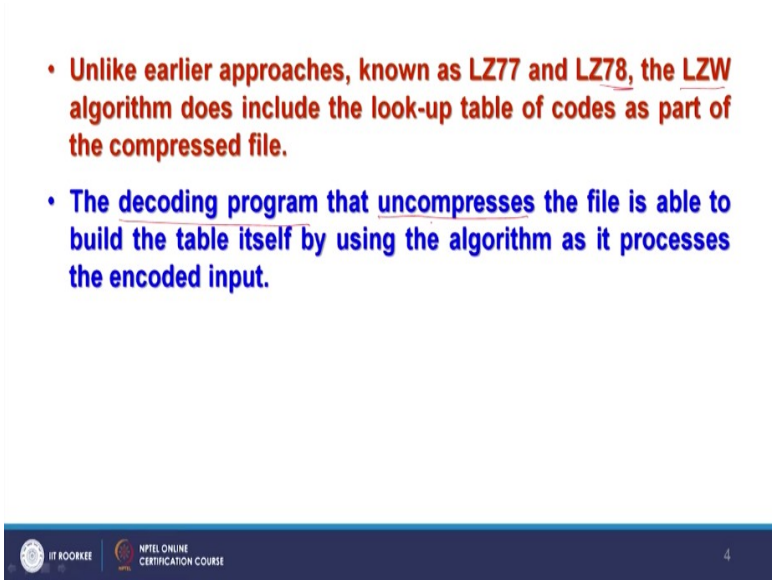
(Refer Slide Time: 04:32)



Then instead of writing the actual value, the position and length as different coding are there. Like in our you know run length code, the value is written to the output. And that is the length and offset where how it will long go. And this algorithm takes each input sequence of bits of a given length.

For example, maybe 12 bits or 8 bits and creates an entry in a table which is lookup table or dictionary also or codebook we can say, for that particular bit pattern consisting of pattern itself and a shorter code. So, by which we can achieve this LZW compression. And later on, this input table is read and any pattern that has been read before results the submission of shorter code and effectively basically compressing the total amount of input to some smaller size files.

(Refer Slide Time: 05:41)



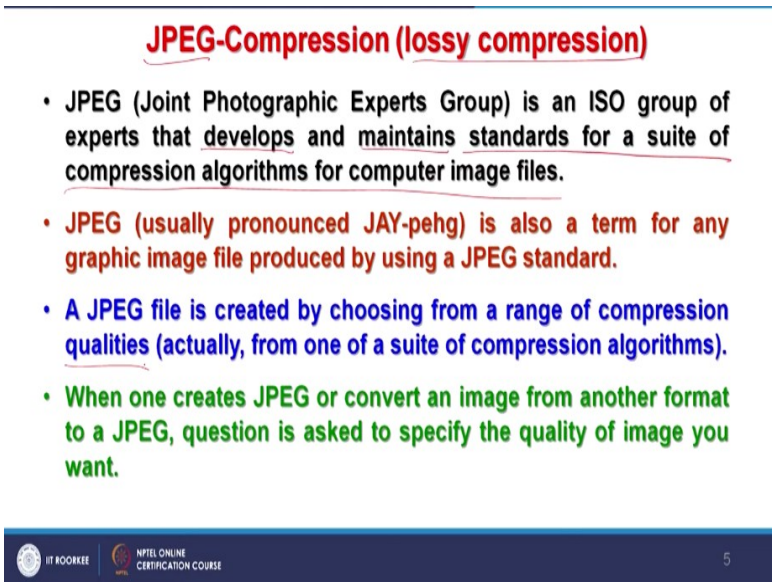
Unlike other approaches which known as LZ77, LZ78, the LZW algorithm does include the lookup table of codes as part of compressed file. And the advantage of including this is that when it is transported to some other place and someone would like to uncompressed it then there should not be any difficulty one and this will not be any loss to the quality of the data.

One important point which I want to mention here that there are many software's which allows us to use even compressed files which have been done using this LZW. So, you can do analysis part even on compressed file without uncompressing it. This facility is not available with all compressed files. This facility has been seen with LZW file. So, LZW in that way is very advantages.

That though your file on the hard disk may remain LZW compress but still, you can do the lot of processing on it. So, that is very good advantage with LZW. And whenever we go for uncompressing then of course, there has to be a program or tool. So, the decoding program that uncompress the file is able to build the table itself by using the algorithm and it process encoded input.

So it is a reverse process and that program should understand how the coding was done.

(Refer Slide Time: 07:26)



Now I will take another example which is a lossy compression or destructive compression, very famous file format is JPEG. In short, we say JPEG and it is stands for basically Joint Photographic Experts Group which is a ISO group of experts and they basically development

is going on continuously. So, that is why they develops and maintains standards for a suite of compression algorithms for computer image files.

So, you know like the latest version is JPEG 2000. So, this way, we get new development. The best part of JPEG is it provides a very good compression, very high compression compared to the other compression techniques which we have discussed. However, it is a lossy compression, destructive compression. It will really reduce the quality of your image.

So, if I do not have much you know, concern about the quality of image and still I want to transform or transform that image from one end to another through net then I may resort to the JPEG. But if quality is the prime for me, like in a professional photography or in a remote sensing or digital image processing or GIS operations where I do not want to compromise on the quality.

Quality means here, I do not want to change the pixel values or cell values. If it is a digital elevation model, no way I will accept to see the changed digital values. So therefore, for that purpose JPEG may not be there. But as I have mentioned, it is a very high compression technique but it is a lossy compression. So, there is a basically tradeoff between these 2.

Another point which I want to also mention before we go in little more discussion about the JPEG is that you know, there are people always going for higher and higher megapixel cameras. Whether it is a mobile or you know professional or digital SLR cameras. They go for higher and higher. But if you see that in what format they are storing their image which they are taking through using these cameras whether is a mobile camera or profession.

If they are storing image in JPEG format that means at a very high-resolution image was taken but at the same time when it was stored, the quality has been compromised and it has been stored in JPEG. So, in one way you are spending so much money for high spatial resolution cameras. But at the same time, in order to store more photographs or more images within a given memory, you are storing in default with JPEG in JPEG format.

And by which, you are definitely compromising on the quality of your image or photographs. So, if somebody is really serious about the quality of the image which they take from their camera then in their setup, they should not save the file in JPEG, instead they should save file

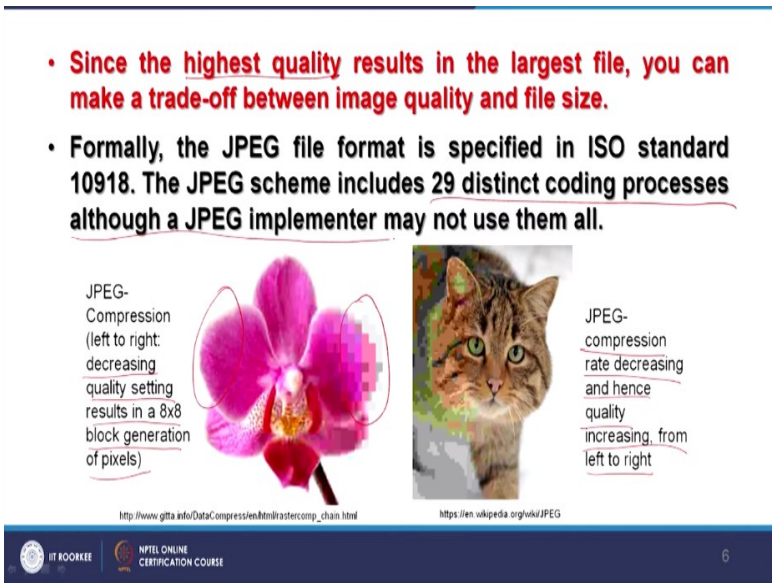
say for example in TIFF. Now the advantage would be the with the TIFF that is still you can imply some compression techniques but that is lossless compression.

And anytime you can restore to the original quality. But once you have saved an image in JPEG and same time you have deleted the original image or taken an image using high resolution camera and stored original images in JPEG then you have really compromise on the quality of image. Though you may achieve high compression that means you may be storing on SD card or in you know, your mobile memory lot of images.

But you are compromising on the quality. So, there is a definite tradeoff between choosing a file format and especially a file format itself is a compressed file format then one has to be very-2 careful about it because the quality matters in many-2 applications. So, one has to keep this thing in mind. Though software's will allow you to convert or save as suppose the input image is JPEG, you can save as a TIFF.

But it is not going to improve your quality. Because while saving, it has compressed your image and have introduced the loss in the quality of your image. So, one has to be very-2 careful on this aspect.

(Refer Slide Time: 12:48)



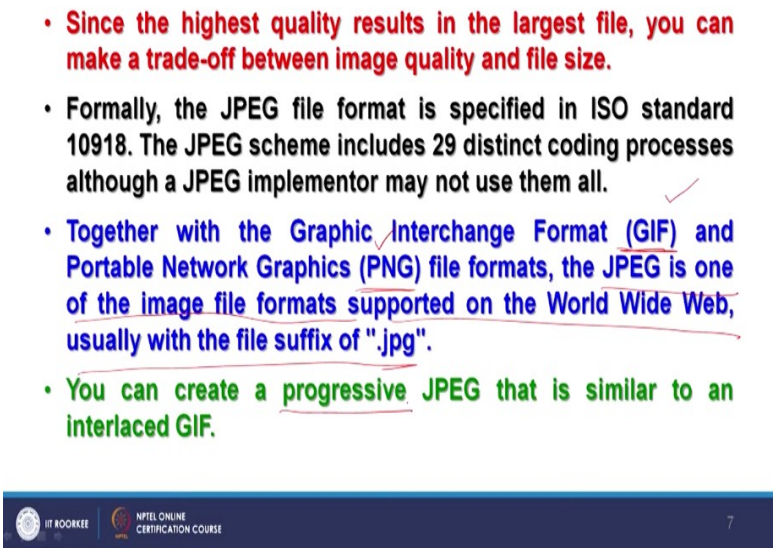
Here like JPEG compression decreasing the quality setting results in a 8*8 block generation of pixels. And here the compression rate decreasing, enhances the quality increasing from left to right. So, within one image like in this you know flower, here the quality is being

significantly compromised whereas here it is not. So, 8 pixels by 8 pixels cell or a block is given a single code and that is why you are seeing a blocky appearance.

However, at the same time you are achieving quite good compression. So, this is what is important that if quality matters then do not resort to the JPEG format. Since the highest quality results will create the very large file and we are discussing data compression. So, still you can keep highest quality but choose a different data compression.

And when you have to transport the file through net and you know, the size is important, quality is not then definitely one can choose JPEG also. JPEG is an ISO standard format and very 29 distinct coding processes exist with the JPEG. So, lot of development is taking place.

(Refer Slide Time: 14:25)



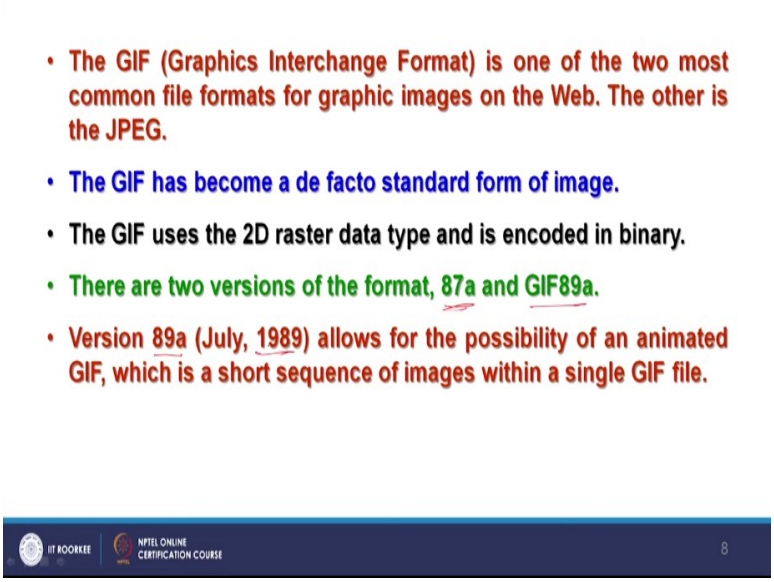
And people are trying to achieve best possible compression without losing the quality means lossless compression. So, highest quality results are in the largest and can make a tradeoff between image quality and file size. And then this the GIF file format which is basically used for animations. There is another file format which is also known is PNG which is Portable Network Graphics.

Specially these are designed for you know, transporting through network. And therefore, many times you would see on these portals or webpages for animations, they will use this file format GIF and for simple images, they will use PNG. Because for some sites just an image is required a very high-quality image may not be required. So, JPEG is one of the image file formats which is supported on the World Wide Web, usually with the file suffix jpg.

So, on the internet of course, all these formats are supported GIF, PNG, JPEG, etc. Although I gave the example of progressive JPEG that is similar to an interlaced GIF. Within a file, it can find that which area requires more compression and which area requires less compression. It is like the same concept you know, this like progressive lenses which we people use.

I am not using but some people use that progressive lenses who becomes bifocal. So that transition from near objects to the far objects is very smooth. Same way here that transition from one end of the image to another is very smooth but still you can achieve good compression.

(Refer Slide Time: 16:36)



Now this another GIF format is the very common format along with the JPEG format. And GIF has become a de facto standard for animated image. And GIF can also be used for 2D raster type and can be encoded in binary that means we can achieve quite good compression. And there are 2 known versions; one is the 87a and GIF89a. The maximum development currently is taking place with the JPEG not with any other file formats.

But anyway, this version 89a which was developed in July 89 allows the possibilities of animated GIF which is a short sequence of images within a single GIF file. I will show you an example also. Anyone can create. You need to have a sequence of you know, images of the same sizes having different poses of a person or a scene or whatever. And then you can use some tools which are available are called like GIF creators.

And you can create a small animated or a sequence or a movie kind of thing. Like for example, there are 8 images in this one. **(Video Starts: 17:59)** And here the illumination source which is a white circle is moving in all 8 directions that is north, south, east, west, northeast and southeast likewise.

And in the center, 8 simulated images are there which are showing a particular phenomenon which in remote sensing, it is called false topographic perception phenomena. So, in order to demonstrate that phenomena, some years back, I thought that I can create these 8 simulated images of the same size and then put in a GIF file format. Now here I do not have to use any window media player or any other thing.

GIF as soon as it is displayed, it is starts showing the animation. And so, most of these cartoons which we see or lot of messages **(Video Ends: 19:01)** which comes through WhatsApp or on internet, most of them are in the GIF file format. And remember that GIF file format can also be subjected to LZW like TIFF file format. So, even if you are having a big file of GIF, you can compress it in LZW.

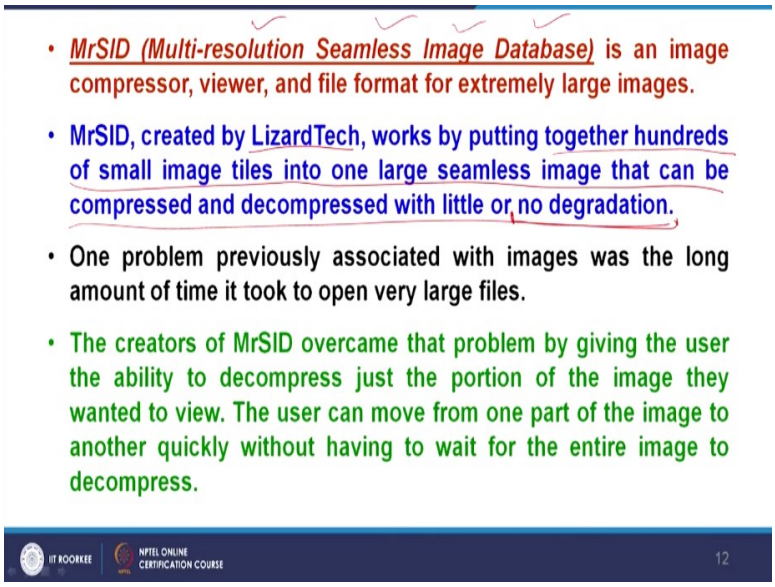
And still that file will work without doing decompression. So, that is very good advantage. Another one is you know illumination source is in one single direction but the height of the illumination source is changing. So, again these are sequencing every 5 degree and sequence has been put in a GIF file format. **(Video Ends: 19:47)**

(Refer Slide Time: 19:48)

- A GIF89a can also be specified for interlaced GIF presentation.
- A patent-free replacement for the GIF, the PNG format, has been developed by an Internet committee and major browsers support it or soon will.

So, this is a which we call as interlaced GIF presentation which you have just seen. And of course, this is patent free. So, copyright and other issues will not be there.

(Refer Slide Time: 20:03)



And the issues which are there with copyright and other issues is this file format which is called MrSID or in short, we say MrSID which stands for Multi resolution Seamless Image Database. And this is a very good image file format. However, it is copyright protected which has been developed by a company which is called LizardTech. So, lot of details about this are hidden, not known but what quality it provides that we will definitely will discuss.

So, this is basically MrSID is based on the wavelet concept or wavelet theory which has got wide applications in digital image processing and also in data compression techniques. And MrSID can put together hundreds of small image tiles into one large seamless image that can be compressed and decompressed with little or no degradation. Basically, my experience is no degradation whatsoever.

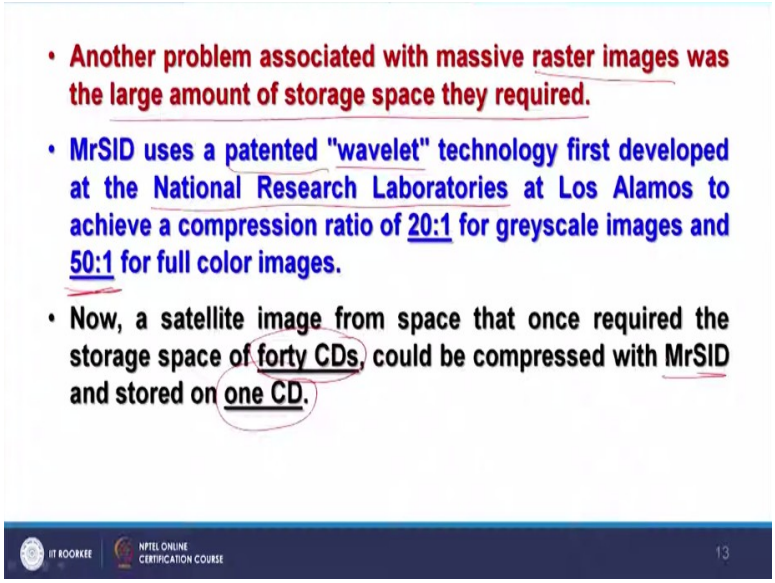
So, you see lossless data compression. And you know sometimes it can provide compression even up to 50 times. That means if I am having an image of 50 MB and I submit to this compression. I may compress that image to 1 MB depending on the redundancy present within the image. That means if more homogeneous area, more pixels are having the same value then I will achieve that 50 times.

But if heterogeneity is there, I may not achieve 50 times but maybe 40 times. But even 40 times is a great. There is a software which is digital image processing software ERDAS

imagine in which they have implemented this one; the free version of this. That means that up to 500 MB file, you can compress using MrSID without paying any money to this company.

But however, if you want to compress a file or files into MrSID which are larger than 500 MB size then definitely you have to pay the money. So, those who are familiar with ERDAS, they would know that in ERDAS, 500 or less image size can be compressed using MrSID. And that is free of cost.

(Refer Slide Time: 22:51)



Another problem we witness which is large image sizes because why large image sizes? Now we are going for higher and higher spatial resolution. And therefore, our image sizes are becoming larger and larger. This is one. Another thing is instead like when we started remote sensing, we had only 4 bands and say example Landsat MSS. Now people are talking in hyperspectral remote sensing. People are talking 256 bands, 512 bands.

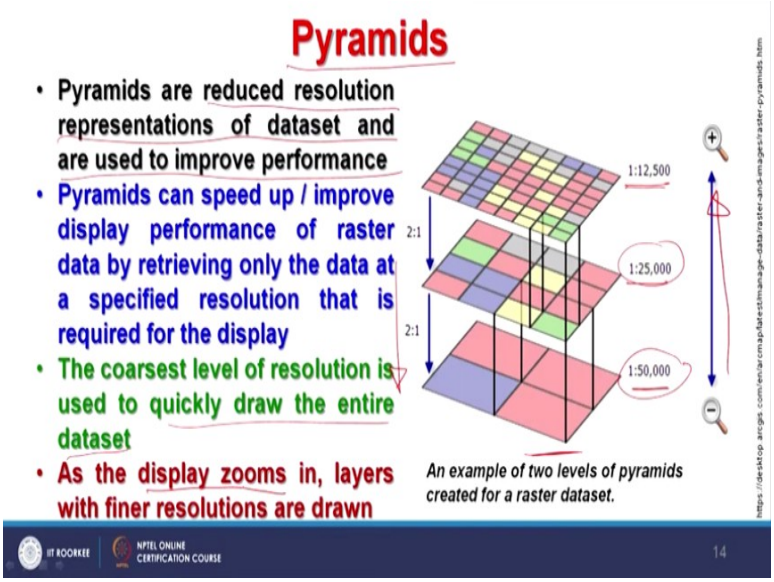
So, if you are having very high spatial resolution and very high spectral resolution images that means they are going to require large space to store. And therefore, requirements of data compression techniques become much more useful and important for us. So MrSID as I have already discussed is based on the patented wavelet technology and initially it was developed by NRL in USA.

And later on, this company got involved and they have achieved. So, you can achieve even 50 times compression to full image. That means that if I am having 40 CDs that I can

compress into 1 CD using MrSID. So definitely it provides a very high compression but how exactly does much details for that are not available.

Now another way of doing not really compression but a sort of compression, so indirectly we are achieving compression in a different manner and that instead of going for typical compression techniques which we have been discussing so far in Part 1 and Part 2.

(Refer Slide Time: 24:54)



This concept was introduced in ArcGIS software which is called pyramids. Pyramids in which what happens basically when you display a full-size image on a screen, at that time it is being displayed at low resolution. Like maybe something like this that only for this of course, schematic just to explain that at that time it is being displayed just 4 pixels. But if I zoom it then my scale is changing.

So, from small scale I am going towards the large scale. And then number of pixels will be more which will cover a less ground area. And if I zoom further then I am going for further high or large scale. More number of pixels will be displayed on my screen. So, this pyramid if as it has been constructed for any image then it will make my display very fast. And as I have said that ArcGIS have developed or have implemented this concept very successfully.

And first time when you know display any image or large data including your digital elevation model, it will ask would you like to construct pyramid. And if you say yes, it will create 2 more files which are not very big size. And next time of course, it will not ask if

these locations or files or names have not been changed or has not renamed then it will not ask.

Next time, the display after construction of pyramids becomes much faster; significantly faster. This I can say through my own experience. Now other big custom designed GIS software which we have discussed many times is Google Earth. They too have implemented so when you display say entire globe on your screen, at that time you maybe million or 10 million or you know 20 million or 50 million scale.

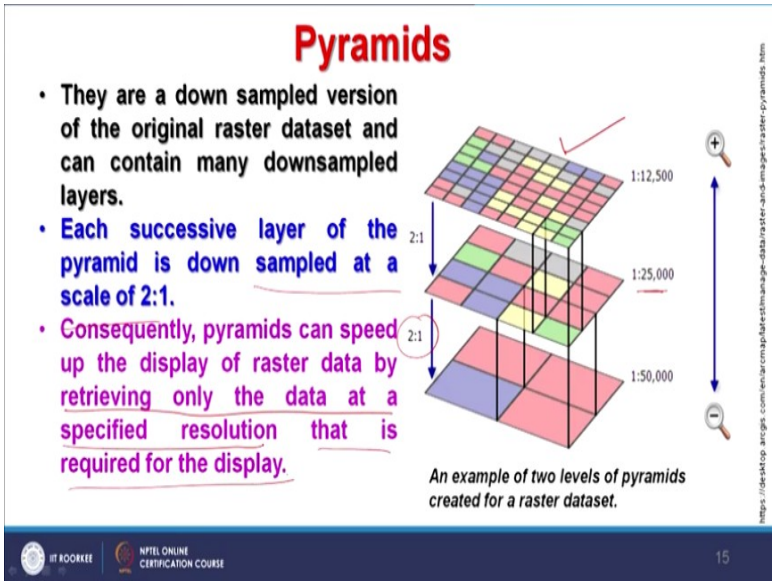
But as soon as you keep zooming in, zooming in, zooming in, you are going for towards the larger scale, less ground area coverage but at a high spatial resolution. So, this kind of concept is useful for displaying images on a screen because we are now handling large size images, large databases as earlier. So, then this concept was introduced. Since this a pyramid kind of concept that base is covering a large area of the ground whereas the top is covering a very small area but at highest spatial resolution.

So, pyramids are r educed resolution representation of dataset and are used to improve performance, reduced in downward side like this. And pyramids can speed up improve display performance of raster data by retrieving only the data at a specified resolution that is required for the display. And as you keep zooming in, it is reduced. It will display for a small ground area but a higher spatial resolution.

And the coarsest level of resolution at the bottom basically and quickly is draw the entire dataset. So that means if I am seeing the entire globe in Google Earth then very quickly it comes and then I keep zooming, I am reducing the ground area going for a higher end resolution and still it is fast. So, it is very-2 useful concept of this indirect way I can say is a data compression. That is why I kept in this discussion.

As you go for higher zoom in that is zooms in, layers with finer resolution are also drawn and then you get the advantage. And they are down sampled versions of original raster datasets are constructed when you go for construction of pyramid. Each successive layer of the pyramid is down sampled at a scale of 2 is to 1 as you can see.

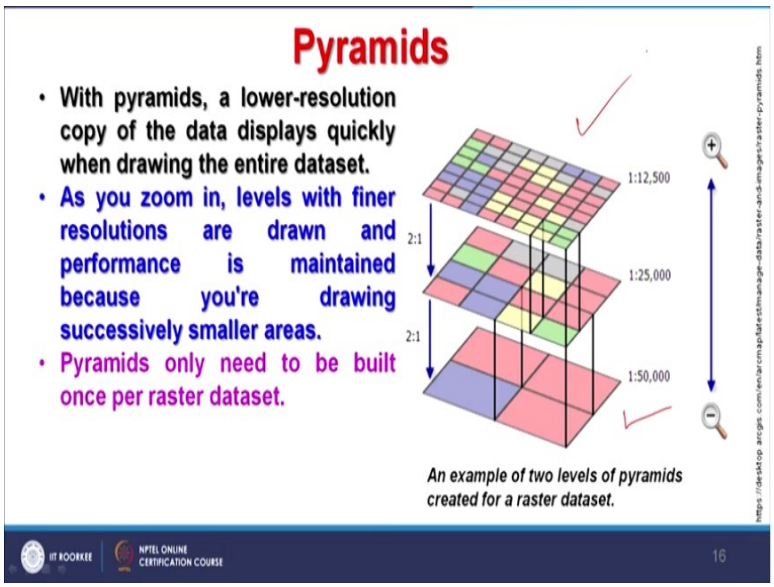
(Refer Slide Time: 29:27)



So here like in middle, you are having you know 4 by 4 pixels. Here you are having just 2 by 2 pixels. Here you are having you know in ratio 2:1. And this is how it is sampled down. And then consequently pyramids can speed up the display. The original image is this one. Remember this thing, original image at the top. Then rest these layers are constructed when you go for construction of pyramid.

And not really layers are stored in the system, just the codes are stored and that is why it becomes faster. So, consequently the pyramids can speed up the display of raster data by retrieving only that data at a specified resolution and that is required for the display. So, layer by layer, the data keep coming on your screen as per your requirements.

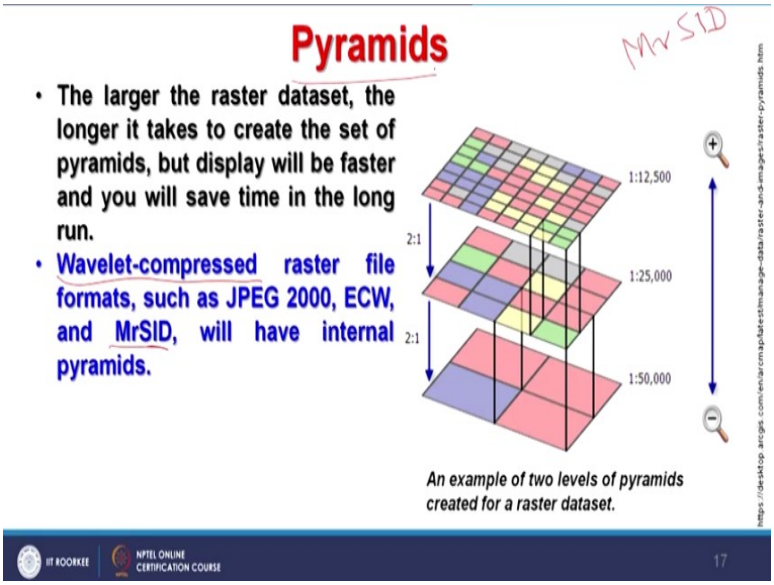
(Refer Slide Time: 30:27)



The lower resolution copy of the data displayed quickly. Of course, because at this level it covers large area; only 4 pixels say in this example have to be displayed it will come. As you keep zooming, it may take some more time to draw or to display your data because now it has to display say 8 by 8 instead of 4 by 4. And pyramids only need to build once per raster dataset.

As I have already mentioned like in ArcGIS whenever you use first time digital elevation model or grid or raster; any raster basically image or grid when you go for first time display, it will ask you would you like to construct a pyramid. If you say yes, it will take few seconds construct a pyramid. Next time it will not ask and your display becomes faster. And then you can access this thing as I have already mentioned.

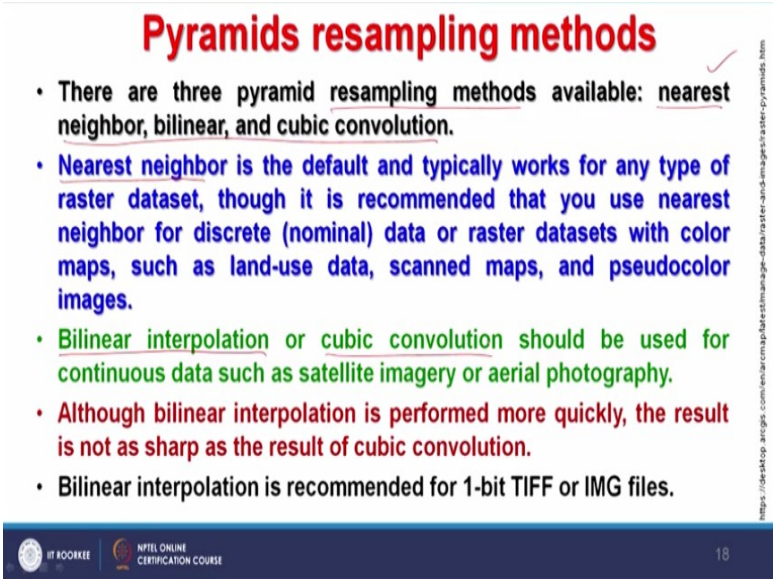
(Refer Slide Time: 31:28)



Now this also uses the wavelet compression technique. And this can be applied for various MrSID. As I have said also your Google Earth data is in MrSID format. And that is why you know, many high-resolution images many times on Google Earth are displayed at very fast. So, they are using; one is the MrSID data compression techniques and another one is the pyramid.

So, both have been implemented and that is why the display in the Google Earth is quite faster as compared to other systems. Larger your datasets of course, longer it will take to create the set of pyramids but displayed.

(Refer Slide Time: 32:15)



Now pyramids you know basically; you can say is a resampling method and 3 ways it can be resampled which we will be discussing also when we will discuss the georeferencing part in GIS in this course. And the simpler one is the nearest neighbor then bilinear and cubic convolution. And first in the nearest neighbor, it basically whichever the values which are nearest or whichever the pixels having near to that target one, their values are taken.

But in bilinear, you know all whoever are occupying or in cubic convolution, their values are taken. So, in that way you can construct the pyramid much easier way. So, this brings to the end of our discussion or completes the discussion on raster data compression techniques. Just to recap the whole thing what we have discussed. First 4 basic data compression techniques and many softwares have been implemented.

Their variants, I have also been implemented. More development has taken place. Sometimes we do not know much about because many-2 such compression techniques are copyright protected like your MrSID. We have also discussed you know on compressed files like LZW. When they are compressed using LZW, still you can do analysis and lot of operations without decompressing it.

And of course, indirect compression technique which is pyramid is a new concept which has been implemented. And I gave the example how it helps us to make a display faster. So, with this, I end this discussion. Thank you very much.