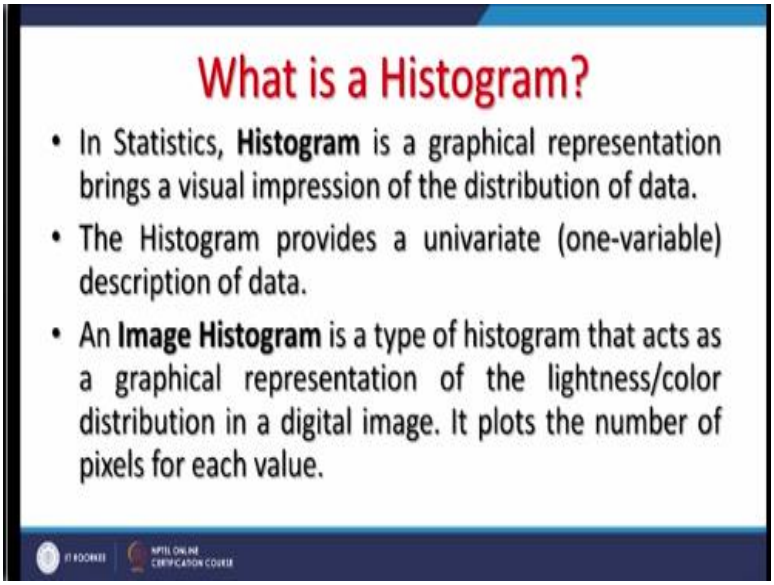
Remote Sensing essentials
Dr. Arun K. Saraf
Department of Earth Sciences
Indian Institute of Technology-Roorkee

Lecture-20
Image Histograms and Statistics

Hello everyone, and as in the previous lecture we have been discussing about histogram. But in this one we will be also going in detail about the different properties, characteristics of a histogram along with a 2 dimensional diagram that is bivariate plot. And also we would be discussing some statistics related with a satellite images, so how we can use that information.

(Refer Slide Time: 00:56)



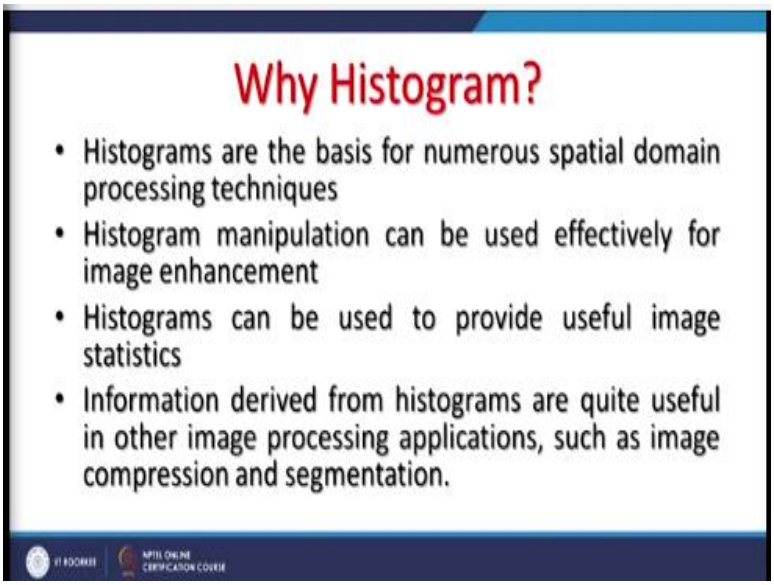
So, I will be going quickly through this histogram, because some part we have already discussed. Basically histogram is a graphical representation and that brings a visual impression of the distribution of data. Rather than looking just a digital numbers, which becomes difficult to understand how the pixel values are distributed against a image. But if we plot a histogram, which is easy to do it, on y axis we get the frequency on x axis we get the pixel value.

And once it is done, then lot of information as also discussed in previous lectures can be gathered about a particular image. So histograms provide a basically univariate because histograms of individual bands, individual images there, that is why it is a univariate. And if we want to see the

histogram of 2, then it would be bivariate or multivariate or also possible. So one variable description of data, how data is distributed and that tells also how the quality of an image taken by a sensor, so that can be seen

An image histogram is basically which is a not only graphical representation, but also it tells you that whether there are lighter objects or darker objects present in the image and basically it plots the each pixel have that image

(Refer Slide Time: 02:26)



Histogram, why histogram need to be done or studied, assessed, that is the basic information it provides. And of which is a start of a image processing or image enhancement techniques. And because there are various spatial domain processing steps we might be taking. And to understand or choosing a best one, first we have to understand the histogram.

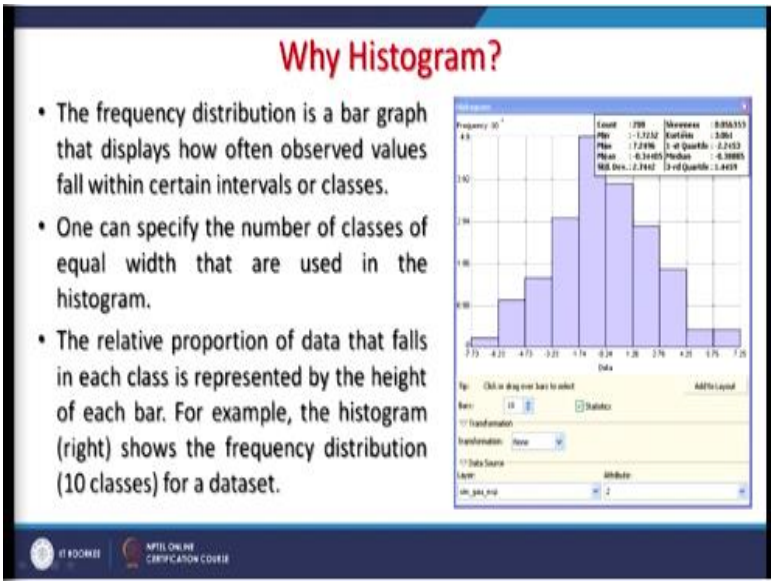
And therefore as also mention in previous lecture, sometimes should always be spent on image histogram and statistics, before we go for any kind of image enhancement techniques. When we enhance images basically what we are doing we are manipulating with the histogram and we will be seeing some examples also. And histogram provides the information about the image statistics as well mean, median and mode.

Minimum value, maximum value what is the maximum frequency that kind of information can be retrieved with the histogram. If there are 2 peaks, it tells something else, if there is single peak that also, if distribution is Gaussian distribution, then it tells different thing, so that is why it is important. And the information derived from histogram are quite useful in many other applications of digital image processing or in remote sensing maybe in a image compression and segmentation.

So, image compression, why because this image compression is required to reduce the size of image without compromising the quality of the image, that is the purpose here for us. And if there is a homogeneity present in the image, lot of compression image compression can be achieved. And if there is a lot of heterogeneity is present among the pixel values then high image competition cannot be achieved.

So, if I study histogram, I can assess whether the distribution is heterogeneous or homogeneous. If it is homogeneous, I know now, that image will provide a particular compression technique will provide me better results. So, that is why it can be used also for image segmentation in the same way it can be used. Though understanding histogram has got various applications.

(Refer Slide Time: 05:10)



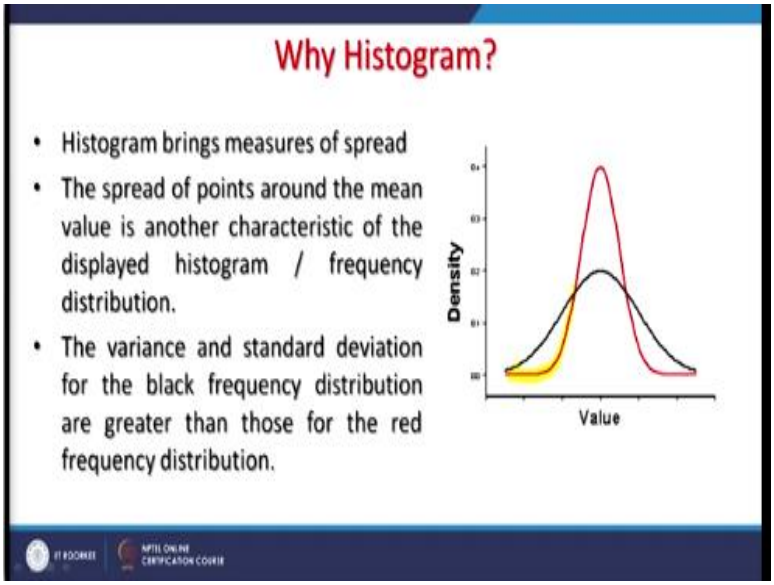
Now, basically it is a frequency distribution in a bar graph and that displays how often observed values fall within certain intervals or classes. So, these towers which you see are basically

intervals or different classes are there. We can also control number of classes, how these thing should be displayed. And relative portions of data that falls in each class is represented by the height of each building or bar.

So, for example histogram is shown here, and that is showing basically frequency distribution in 10 classes. So, it depends on my requirements I want more number of classes or even if it is 8 bit scenario. Then I can have even 256 classes, variations between 0 to 255. And I can see the things accordingly, same time also you get the minimum value, maximum value, mean, mode, median, standard deviation.

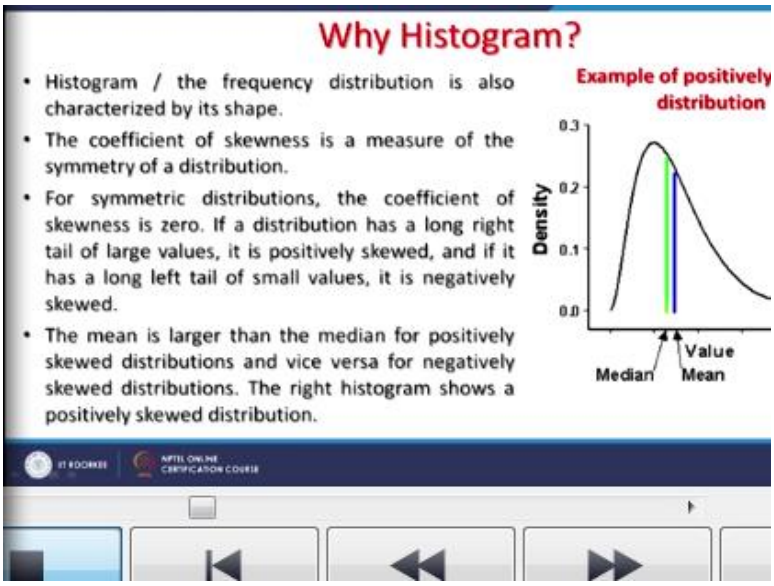
Other statistical parameter like the skewness, kurtosis and all those things can also be identified from there, so that is why histogram is important.

(Refer Slide Time: 06:26)



Histogram brings measures of a step basically how pixel values are distributed in a image, spread of points around the mean is another characteristics of display histogram or frequency distribution. As you can see that they these red curve is showing less distribution around the mean whereas the white curve is showing more distribution around the mean. And the variance and standard deviation for the black frequency distribution as you can see is the greater than those of the red frequency distribution.

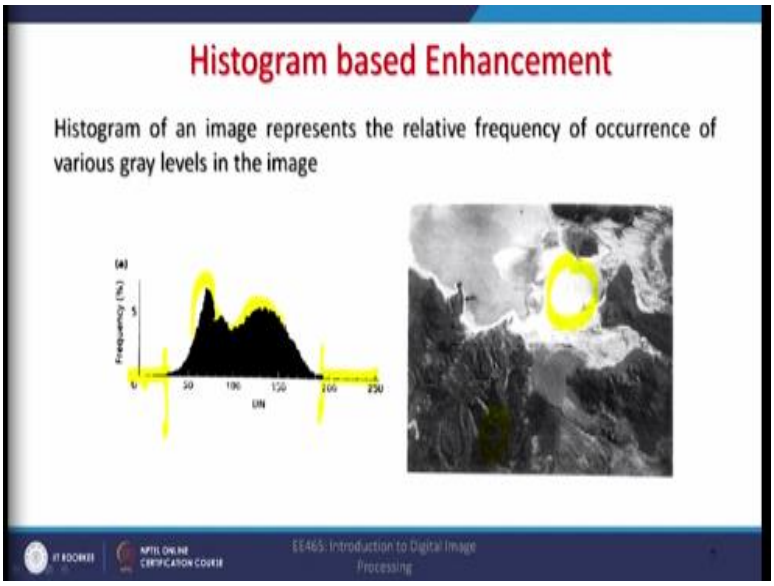
(Refer Slide Time: 06:59)



If you are having skewness present in your image, some a bright part or more or dark part are more. Then in histogram you would see a skewed representation skew distribution in the image, so that can be also the shape of histogram tells about the skewness present. And this coefficient of a skewness is basically measure of symmetry of distribution. So, generally, you do not get symmetric histogram generally you get systematic histogram of distribution.

And therefore, there will be skewness either towards the you know smaller values or large values. In this example skewness towards the lower value and therefore, mean, median and mode and will be located at different places. So, that also tells that there are a like a I can while going through this histogram, I can tell that. There are large number of pixels which are having lower pixel values compared to the values which are having high pixel values. And if it would have been reversed, then a different interpretations will come out from this one.

(Refer Slide Time: 08:24)

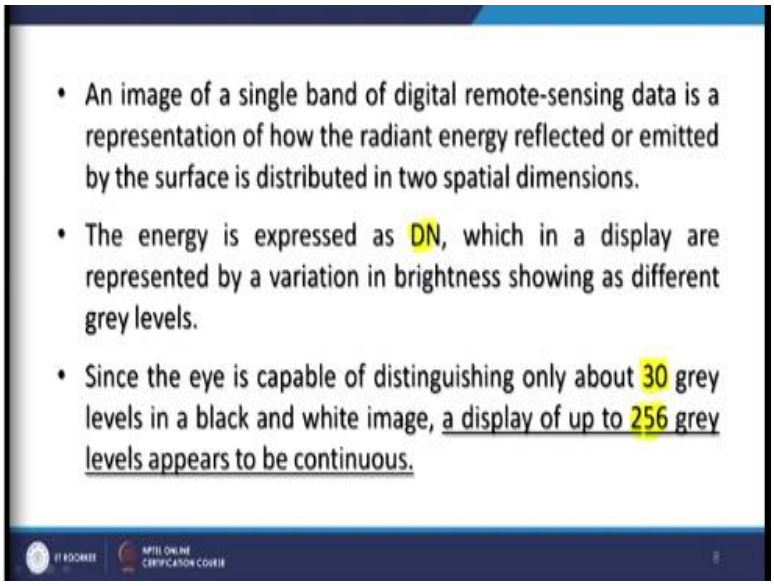


Now, we see in a real example, that histogram and it represent a relative frequency of occurrence of various gray levels in that image. On the right side you are seeing the image, on the left side you are seeing the histogram distribution. In this case, the raw image histogram is occupying the full range available to us. That is between 0 to 225 right. Again, it is a 8 bit scenario and that means if I go for simple linear stretch, I do not have much (()) (08:54) available to me.

Because I am having brighter parts, I am having darker parts as well, brighter parts as well. And that is why there are 2 areas or 2 peaks in my histogram there and some values are also here. So, if I want to improve the contrast in the image then I have to ignore certain the tail values on both end to improve the contrast. So, that way histogram also becomes important to assess that situations.

Image, generally for individual bands we have to see because individual band will have different histograms, this one has to remember. So, instead of even if you are working on a colour composites, but while is studying histogram individual histogram must be seen. Because different when will have a different distribution of pixel values. So, in a image which has of a single band image and remote sensing data is a representation of how the radiant energy reflected or emitted by the surface is distributed in 2 spatial dimensions.

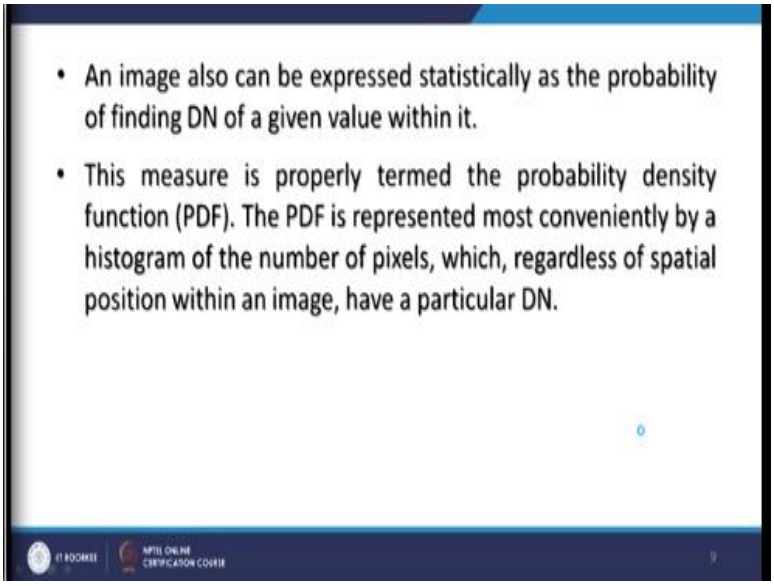
(Refer Slide Time: 09:58)



And the energy as you know is express as DN or digital number or pixel value which is displays at the representation by a variation in the brightness showing. And since our eye is capable of distinguish only about 30 grey levels in a black and white image and about 256 whereas we are having display of 256. So, we might be displaying but our eyes are sensitive only to 30 grey levels does not matter.

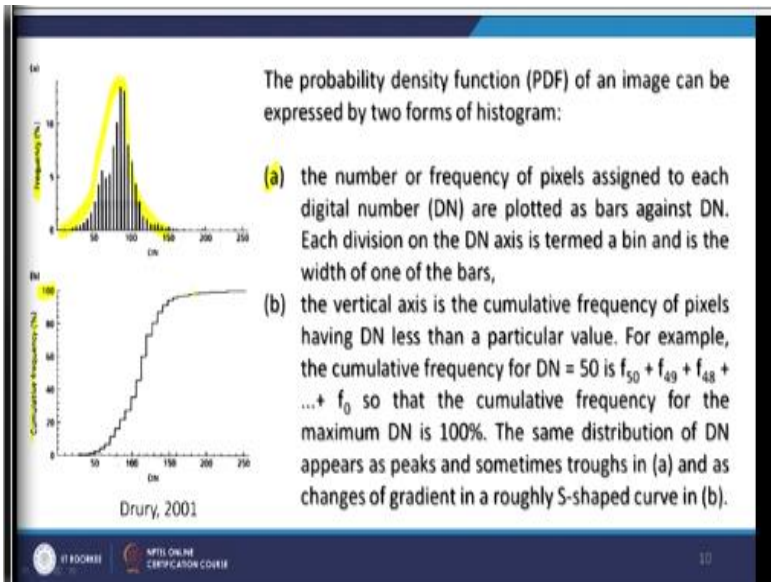
Because when we go for image classification or other processing of images, then the machines will take care about this distinguishing 256 grey levels rather than 30. So an image can be also expressed statistically as probabilities of finding digital number of a given value within it.

(Refer Slide Time: 11:08)



And this measure is properly termed as probability density function PDF, and this PDF is represented most conveniently by histogram. Of number of pixels, regardless of spatial position within an image have a particular brightness value or pixel value or digital number.

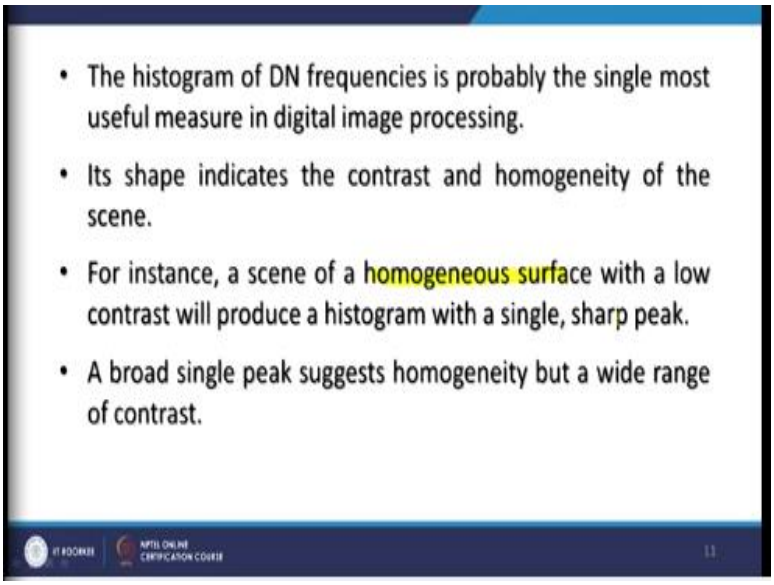
(Refer Slide Time: 11:30)



Now, here this is PDF the probability density function of an image can be express in 2 forms either a through this histogram or maybe cumulative way like on the bottom side. So, one has to be careful this is frequency and this is cumulative frequency. So, the first one this a the top one is the number of frequency of pixels normal histogram. And whereas the bottom one the b part is the vertical axis is the cumulative frequency of pixels having DN number.

And this a cumulative frequency, so the maximum DN is 100 here, and here you will have the real frequency distribution. So, the same distribution of DN appears as peaks and sometimes troughs as in case of a and as changes the gradient roughly as saved curve in b because this is how the distribution in original histogram is. Now, this curve in cumulative frequency versus pixel values will vary depending on again image to image.

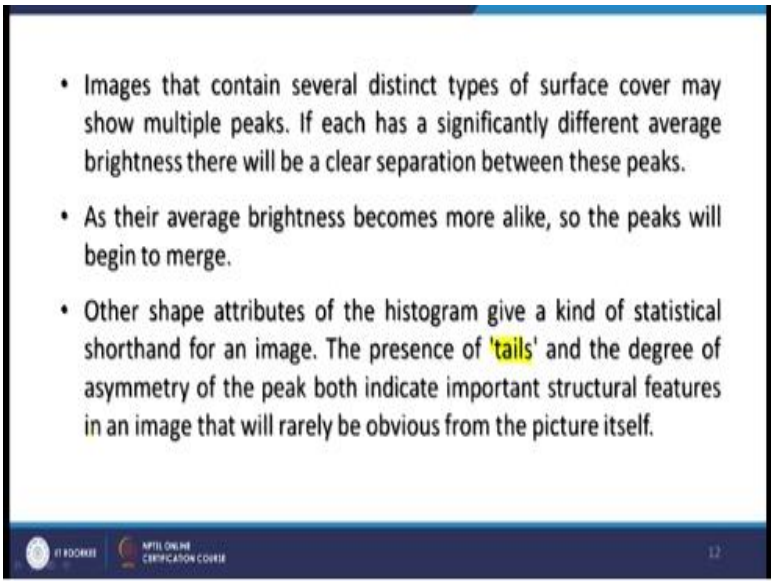
(Refer Slide Time: 12:45)



So, one can also plot like this and try to understand the distribution of pixel values through this PDF also. So, this is indicates basically the shape indicates the contrast and homogeneity of the scene, that is important. Because homogeneity is useful, not only in image enhancement, but also in image compression, for image compression purposes.

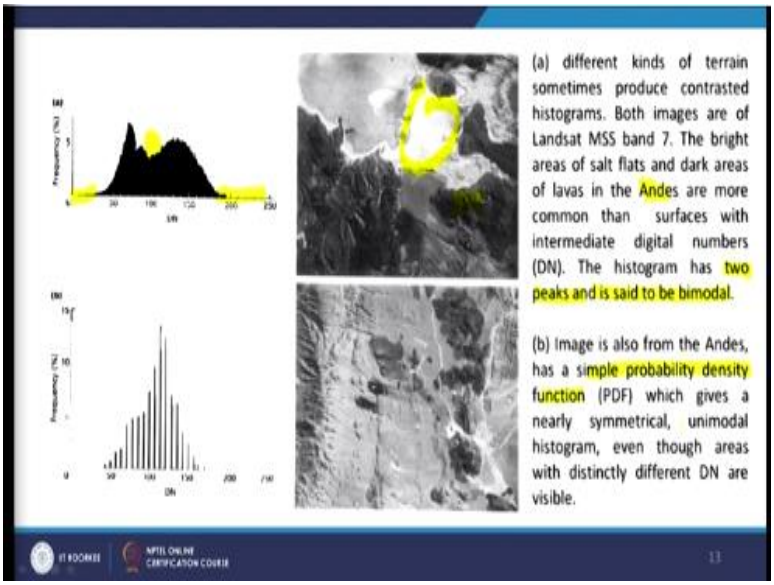
So if a scene is a homogeneous surface like of a desert or a snow covered area, or ice shield areas. A low contrast with a low contrast will produce a histogram with a single or sharp peak. And if you are having a broad peak, single broad peak, which suggest that homogeneity but a wide range of contrast. So, a histogram tells lot many things that is why I have been saying that histogram is important.

(Refer Slide Time: 13:46)



There are some other values or other advantages are there with 1 peak, 2 peak and 3 peaks and tails also. Like the presence of tails and degree of a symmetry of the peak both indicate important structural features an image. That will rarely be obvious from a picture itself, so this is structural features might be there because of some features or a symmetry or tails.

(Refer Slide Time: 14:22)

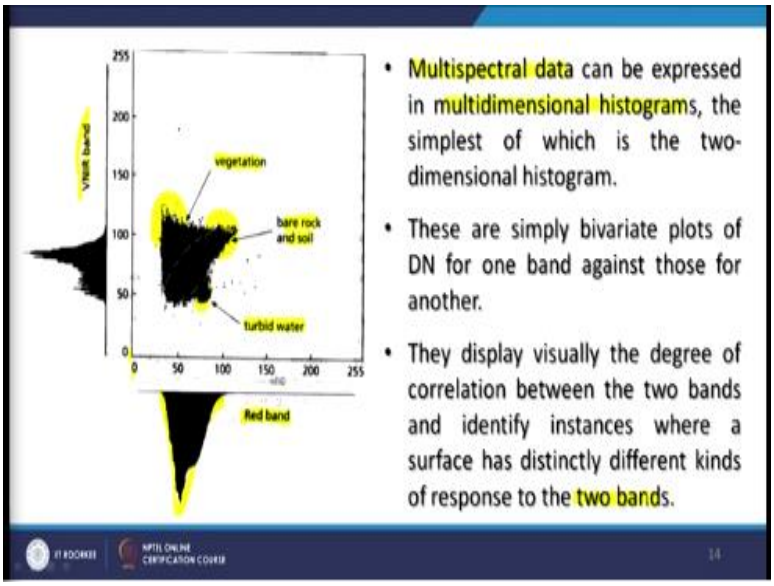


Like in one example, you have seen the tails also at like this one, you are having tales also having pixel values. So, different kinds of terrain, natural conditions or such scenarios sometimes produce contrasted histograms. And both these images are of Landsat MSS band 7. The bright areas here, the salt areas salt flats and dark areas are the lavas, which you are seeing here, 2 different ground features.

In the Andes are more common surface than intermediate digital values. So, in intermediate digital values which are falling here are less compared to low values or high values, so the histogram, as you know has 2 peaks and we say it is a bimodal. In case of the bottom example, here you do not have that kind of contrast. So see the distribution single p, so the image is also from the same area but as a simple probability distribution function.

Here which gives a nearly symmetrical, unimodal histogram which you can see on the bottom left, even though the areas with distinctly different digital values are there. So not only it is telling the distribution, but it is also kept telling inbuilt contrast within an image.

(Refer Slide Time: 15:54)



Now when we go for multispectral then we go for a scatter plot or 2 dimensional histogram if you wish to call. So, this is multispectral data example, which can be expressed or this histogram through multi dimensional histogram 2 axis are having 2 histograms and each pixel value is distributed in this space. So, both axis are having value dynamic range between 0, 255 by axis is also having 255.

So, pixels values of 2 bands, one is the red band in this example, and VNIR are very near infrared band, these have been used and displayed. So, far about we have been seen of histograms of single band. Now, when we go for 2 bands in this example, then now we can even

we can identify how 2 winds are correlated. That is the first purpose of doing this exercise **is** do plotting a scatter plot or seeing you know, multi dimensional histogram.

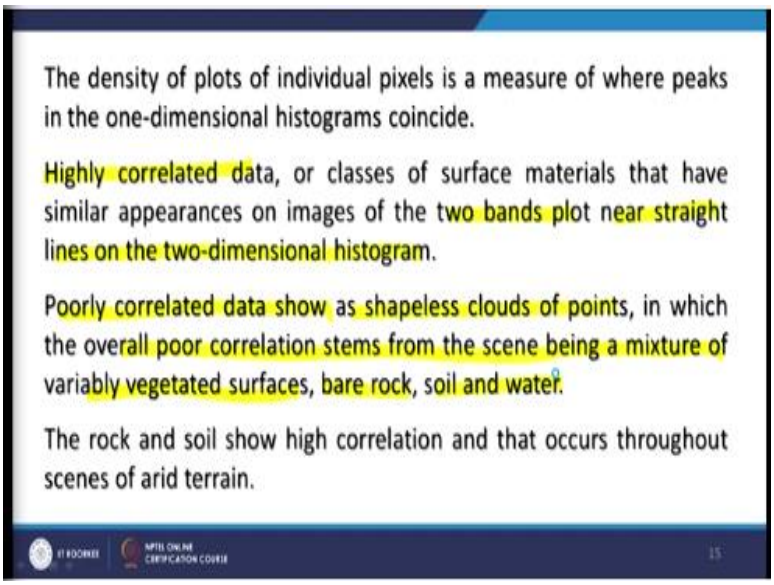
So, in this one by identifying the areas because represented by the low pixel values in one band that is in red band. And relatively higher pixel values in a infrared band, we can identify that these pixels will might be belonging to the vegetation. Similarly, this area because in both bands they are having higher values relatively higher values, so might be bare rocks and soils.

Because if would have been vegetation cover, then those signatures will come here, those pixels will be plotting their position here. And if I am having very low values in infrared channel, and relatively higher values in red band. Then I am probably looking a turbid water, water with a lot of suspended particles or turbidity. So, moving from one dimensional histogram when 2 dimensional histogram many more things can be seen in these 2 dimensional histogram.

So, these are bivariate plot, scatter plot, 2 dimensional histogram again very useful. So, whenever you are going for multispectral analysis, instead of single band analysis, histogram is sufficient. But when you go for multispectral analysis, it is better first to plot a bivariate plot and try to understand not only the distribution of individual bands which can be seen here through these histogram.

But how they are related, how pixel values are related in different bands, here, 1 visible band, 1 infrared band of course things are completely different in these 2. So, this will allow you to identify different surfaces as I have just discussed vegetation bare rock, soil and turbid water, by looking these 2 bands histograms.

(Refer Slide Time: 19:12)



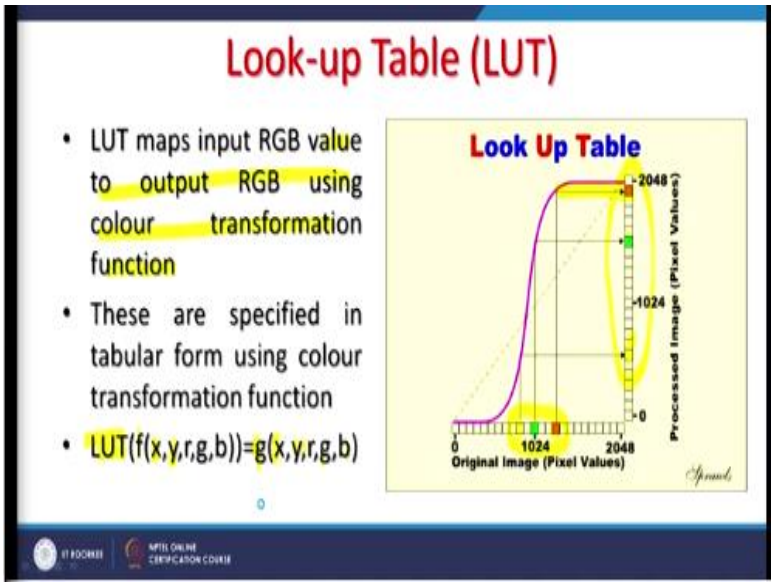
Now, density of plots of individual pixels is a measure of where peaks is one dimensional histogram coincides. So, if there is a very good relation between 2 bands, then there will be high density of pixels plotted in a small area. That means, they are correlation will also come when they are in a linear fashion. So, highly correlated data or classes of surface materials that have similar appearance on different bands or images or 2 bands plot nearly a straight lines on 2 dimensional histograms.

So, if there is a choice to drop one band for my colour composite, I can do it very easily, if I find the 2 bands, say Landsat MSS 4 and 5 are highly correlated. And in colour composite I need only 3 band, so one band I can drop it because they are highly correlated in there is no need of putting them in my colour composite. So, this correlation among bands is also important in multispectral analysis.

So, whereas poorly correlated data, so as a shapeless clouds of points, no linear distribution of pixel values in 2 dimensional histogram. And which poor overall poor correlation stems from the scene being a mixture of variably vegetated surface, bare rock, soil and water and that example you have seen earlier like this one. So, there is no linearity in the distribution of pixel values is there.

Now the rocks, different rocks and soils, so high correlation that occurred throughout the scene of arid terrain. If you are having dry terrain where you do not have the vegetation cover, then they may show almost similar characteristics.

(Refer Slide Time: 21:09)



Another thing which is used in digital image processing, a concept which is basically lookup table or LUT. The purpose here is not to modify the original image, but when image is displayed in between a lookup table is created. So when you want to save a stretched image, you generally do not save the image, you just save the lookup table. And therefore you do not occupy a lot of space on your hard disk, so that is the advantage of lookup table.

So if you image is occupying 100MB whereas lookup table may require to maybe 100 kilobyte, so by which you can save lot of space and it becomes faster for display as well. So, lookup table concept have been implemented in almost all softwares and it becomes very advantageous not to flood your hard disk with a lot of data. Or duplicate data or original image then enhance image, one more enhancement, there might be 10 versions and you are have of the same image.

So rather than saving like this, you save the original image and look up tables for all other process to part, so that makes things a bit easier. So basically it is a 2 dimensional table and look up maps which is input RGB values to output RGB using colour transformation function as shown here. For example, if I take this red values here, then it is projected here.

So in original distribution, I am having values distributed, say between only this dynamic range. But when I go through the lookup table, now I have distributed these values between this much of dynamic range. If I see the values in that one, then they are around 1024 values are around 1024 this yellow, green and red. Whereas, now these values are distributed, here is the yellow, here is the green and here is the red.

So, they are occupying now, large dynamic range compared to my input and only through this table lookup table, I will save in my machine, my result only the lookup table not the original image. Original image will be always there but process image will not be saved only the processed lookup table will be saved. So, these are specified in tabular form using a colour transformation function which is given like this LUT function x, y and RGB, x, y the location of individual pixels.

And when output you are having again x, y the location has to be intact and then RGB how changes have occurred and likewise you get this lookup table. So, this brings to the end of this discussion about histograms. We have completed and discussion on histograms simple statistics, this is also using image analysis, image enhancement technique. Then bivariate plot, a scatter plot or 2 dimensional histogram and their applicability, how it is useful.

And finally, the concept of lookup table which have been implemented by various standard softwares, very useful not to flood your hard disk with a lot of image, original image, enhance image and then filtered image and other thing just save the lookup table, so this brings to the end of this discussion, thank you very much.