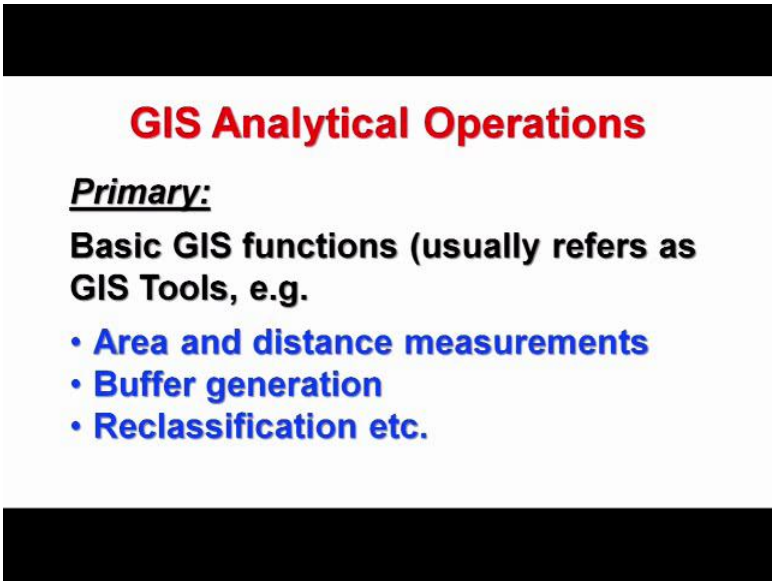
Introduction to Geographic Information Systems

Prof. Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture – 09
Pre-Processing of spatial datasets

Hello everyone, today we will be discussing pre processing of spatial datasets, as you know that. So, far we have discussed what is GIS different types data which are used GIS. How to handle them and how to even raster data compression techniques we have also discussed data base management specially, for the GIS. Now the in real sense the GIS is starts from here basically once everything is prepared in our system then we have to still go for analysis, but before that there are certain task which you we have to do it and therefore, we put them in the pre processing category, and processing or real analysis part will come little later.

(Refer Slide Time: 01:25)

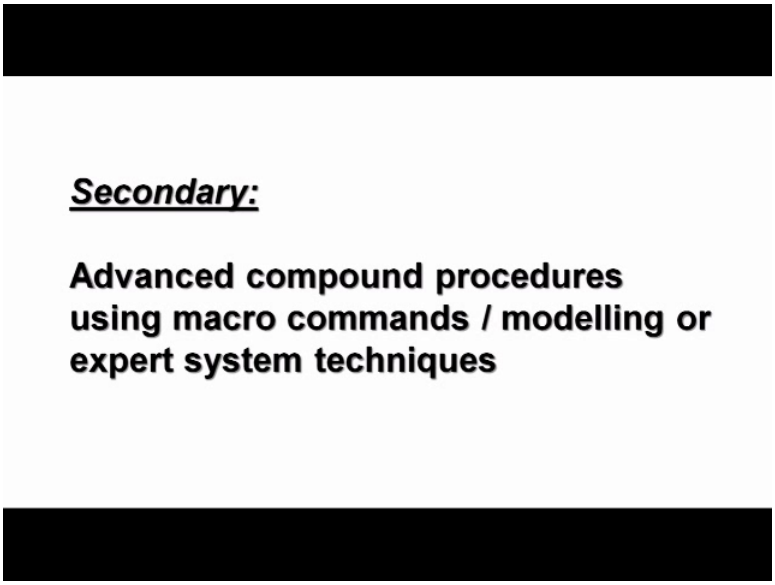


Now people have been dividing these operations and two major categories; one is primary pre processing steps and then there are second. So, we will see one by one all these steps both primary and secondary and see what are the process which are involve and most of these processes can be handled in any standard GIS software.

So, like primary functions which are basic GIS functions, for example, we refer as GIS tools as well depending on the software and terminology people use different. But the

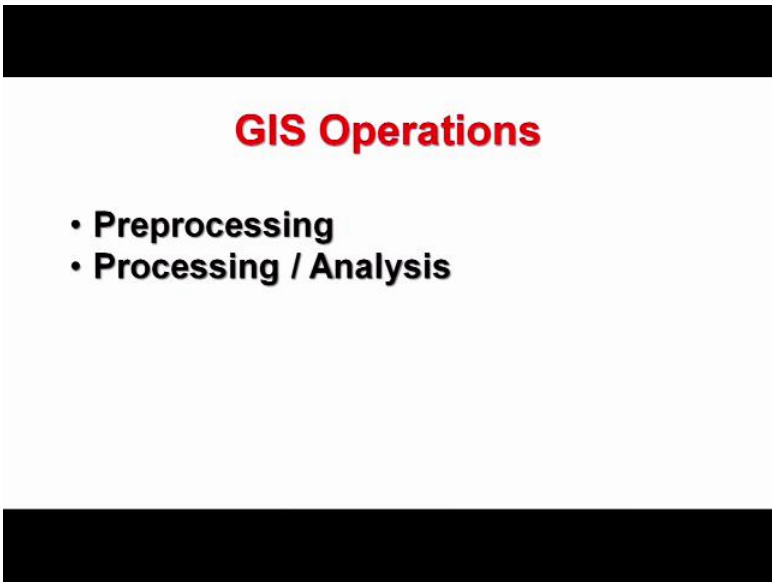
basic functions like area and distance measurements, if you are having polygon map you would like to have area distance perimeter and other things would like to store within the data base these operations can be done very easily in GIS, once your data base is ready. Buffer generation which is another important task is required for certain types of applications and modelling. So, we which will we also doing, but buffer we will have a separate discuss and all this, then reclassification that is also a part of pre processing that it requires a complete separate treatment.

(Refer Slide Time: 02:44)



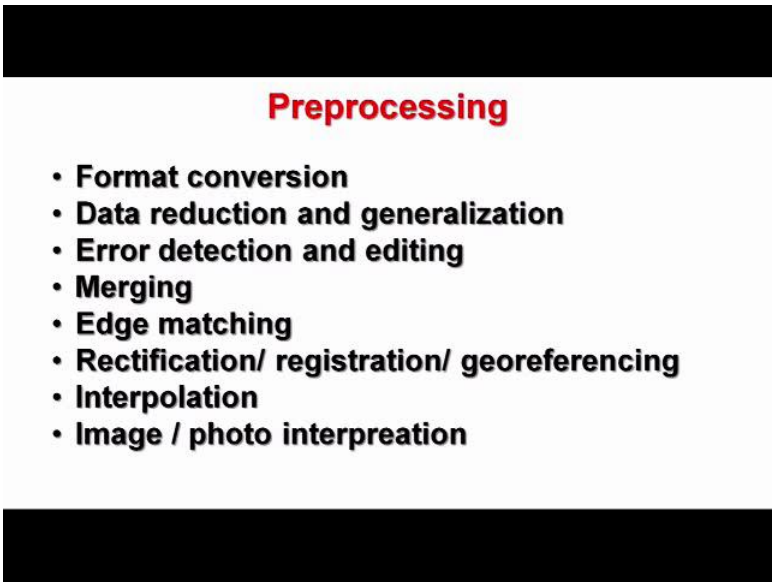
So, we will we discussing that one also now let us look the secondary one, secondary operations involves the advance compound procedures using macro commands or modelling or expert system technique. All this tools are also supported in GIS and that we can also handle very well here which is not very simple, but all this supports are levels so, this can be also done on your GIS systems.

(Refer Slide Time: 03:09)



Now, let us go in detail about the pre processing processes and processing or analysis 1.

(Refer Slide Time: 03:13)



So, pre processing involves the format conversion this is a very standard task which we have to do on regular basis. What here format conversion means that sometime we have to convert for our analogy to digital this is one kind of format conversion, sometime we have to convert one vector datasets into another vector dataset for example, the datasets might be coming from d x f format which is a auto cad format and then you would like to import into GIS may be say r GIS then you have to convert those tools sometime are also

available, modern days even the softwares which where you have developed the d x f file can also export to shape files which are suitable for easily products like r GIS or some other formats can also be done here, same with the image processing or image formats like you might be having data in tiff format, now you would like to handle in I m g format which is again comfortable with a most of the GIS software. So, you can convert.

So, the format convergence in (Refer Time: 04:23) so, many things same with the you textural files as well you may be having a simple table typed in a text mode. Now if there is a delimiter you can use it you can import in your data base and likewise you can handle. So, format convergence is very, much required in many, many steps it is required and another one data reduction in generalization this processing, pre processing step is a required as per depending on the target or the output is skilled of the map. So, we will discuss later in details little later about the data reduction in generalization, let us go first throughout this list of pre processing steps, error deduction and editing as I have already mentioned in previous lecture that errors I have to be handled very carefully, and after each and every step one has to check for errors. So, when once we find any error in over operations we must immediately correct it, edit it, and then go for the next operation.

A merging we will also see in much detail about merging. Merging sometimes you are having two adjacent maps which you want to merge or you merging in remote sensing domain or even in GIS merging may be you are merging two datasets raster datasets or two vector datasets. So, that that all merging we will discuss in little detail also later, and then edge matching edge matching is very much required because a map projection, a map might be in a different projection a another adjacent map might be in another map projection.

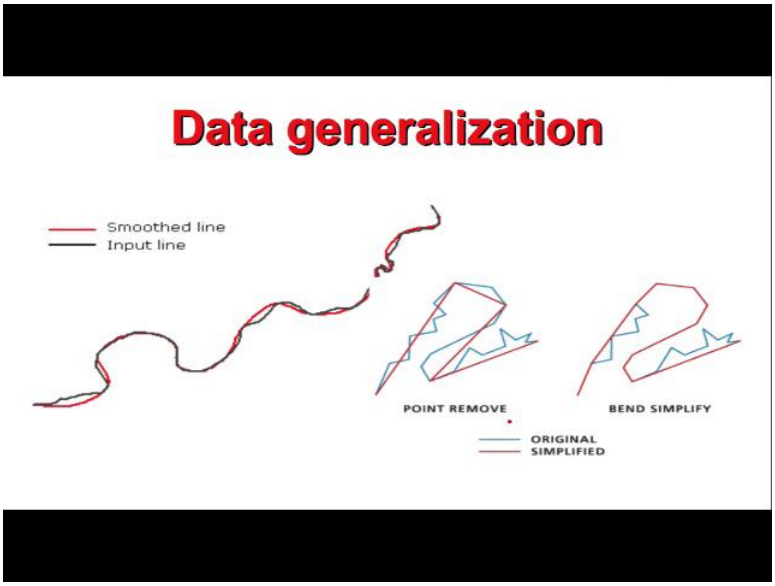
So, you have to correct that one and sometimes even if it cannot be done automatically then user interactions are also required for this, then reclassification registration sorry rectification registration and geo referencing that is also a common process between remote sensing, digital image processing and GIS very important one and we will be discussing in a separate lecture about geo referencing. Modern days we start we have a started calling as geo referencing; that means, your original image is or data sometimes when they are collected they are in geometric domain and which cannot be handle in GIS. So, we have to bring them into geographic domain and this process converting from geometric to geographic this is what the geo referencing mode processes.

So, we will deal this one in a separate lecture. Interpolation is another very important pre processing a step, because you might be having discrete data like point data or line data you would like to create a continuous data. So, further again interpolation is required. So, in future lectures we will be also discussing an interpolations different technique is interpolations their comparisons and how to handle them all those things, then image and photo interpretation because some lot of inputs are coming from satellite data or may be some other sources and therefore, we require image interpretation analysis of images, and may be classification and other things.

So, that also we consider under the pre processing category. So, let us discuss said in detail about data reduction in generalization. Data reduction that does not mean that you delete the data no data reduction means when suppose the output is required at a smaller a scale, where as you are having lot of detailed data for example, you say coastal line of India.

Now, you want to produce that map in a 5 million or 10 million scales just would like to show the boundary of India. You do not want to show all details of the coast line and therefore, you want to reduce the data and that can be done under this category data reduction and generalization.

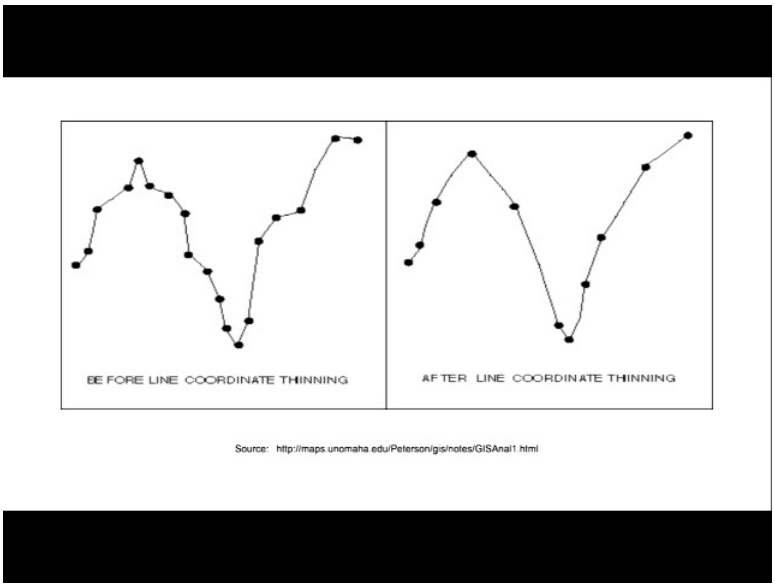
(Refer Slide Time: 08:37)



The examples are given here also, that day you know some points or some inter nodes are removed in from in between and then a new line has been created here. So, that that

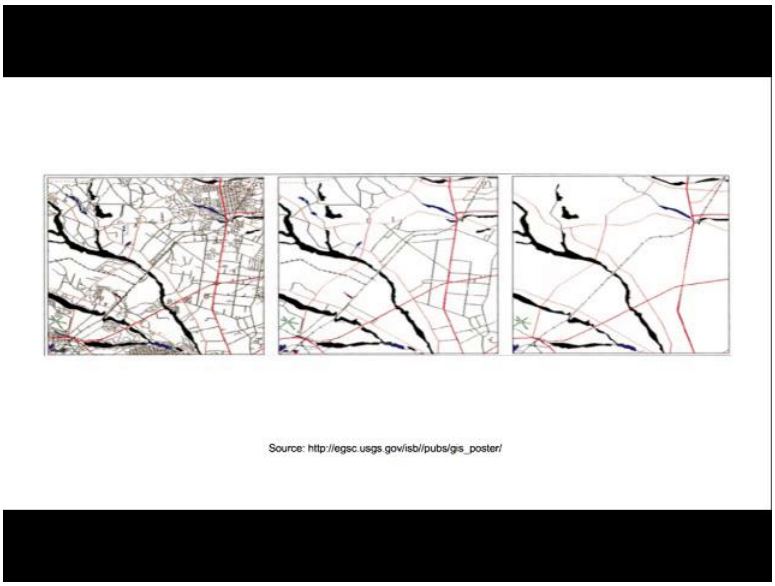
reduces the number of nodes and make your data more generalised form, 2, 3, 2 examples are given here one is this one. So, you can make a smooth line as well sort of best curve fit line instead of all a small small kings and bends which you can get rid of, and therefore, you can generalise the this is very much required if you are changing or a scale target a scale is much smaller than your available data and therefore, before that we go for data generalization. Always remember that original data should not be disturbed and whatever you do is you are going to create a new data set and therefore, they should not be any problem.

(Refer Slide Time: 09:45)



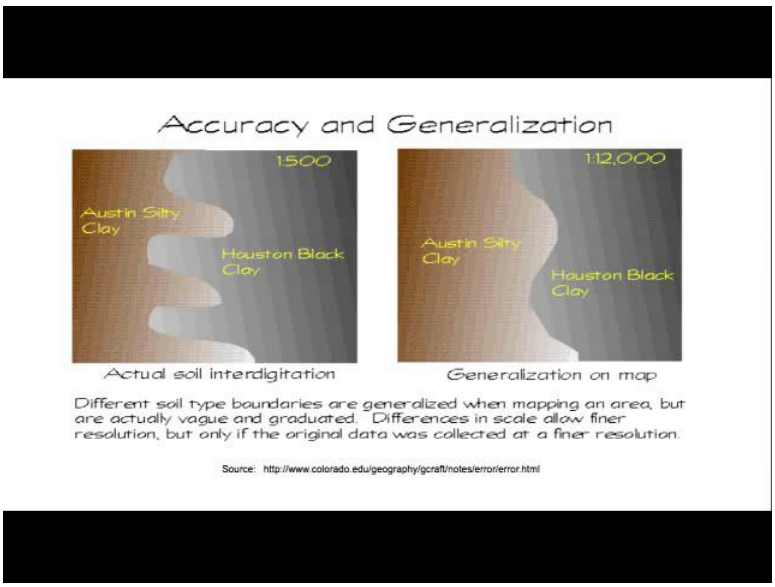
Another example here is like a thinning of a coordinate thinning is here. So, when small (Refer Time: 09:49) all these bends can be removed. So, one node as been removed here, and therefore, you got a almost very smooth line compared to what previous one.

(Refer Slide Time: 09:59)



Similarly here in real map example, here you can see that lot of details where there in one step the generalization was done and then lot of things have been removed and then further this smoothening as been done and very simple map as come. Because if this map I am going to print in a very small paper then I do need so many details, and they 5 I am having. So, many details the appearance of the map will also not be very nice.

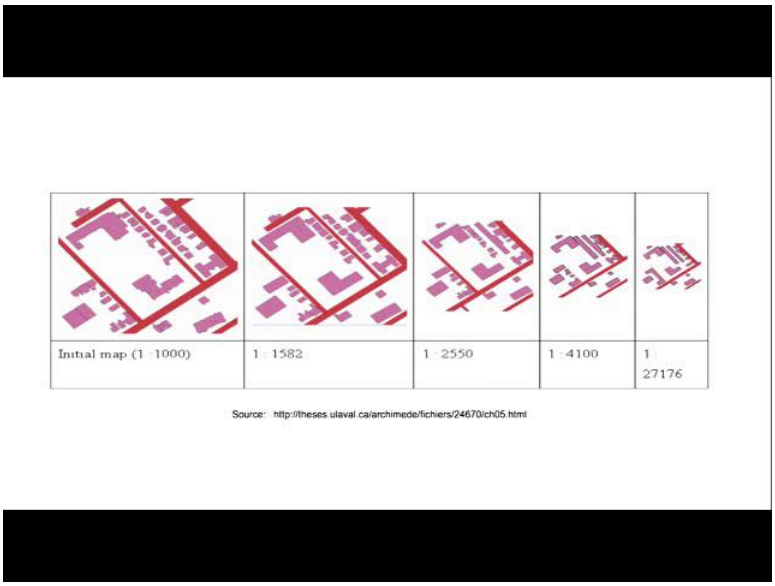
(Refer Slide Time: 10:30)



So, data generalization is very much required, one has to also remember because when you are going for a smaller in scale at that time probably you are not very much concern

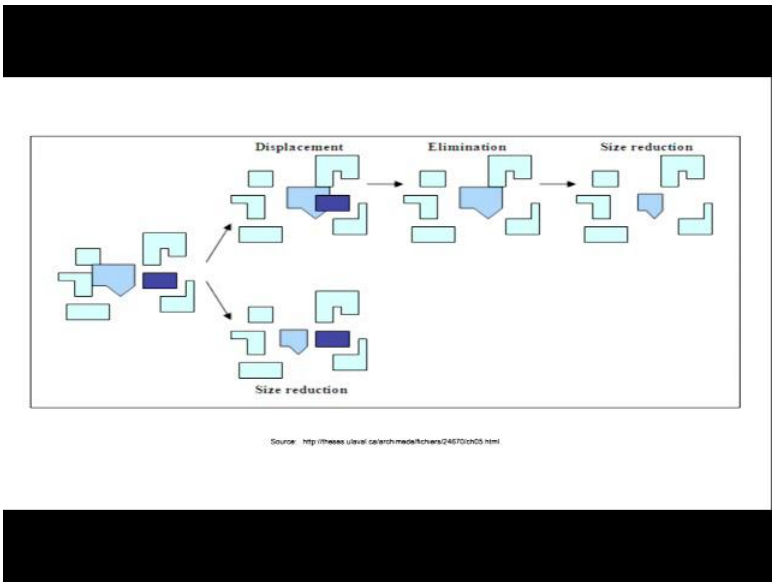
about the accuracy. So, generalization has in the coastal area this is how all details are not required. So, a very smooth line can be prepared here, as you can also see here that the scale was 501 raise to 500 here, 1 raise to 12000 and therefore, when you are going for a smaller in scale you need not to have lot of details there.

(Refer Slide Time: 11:05)



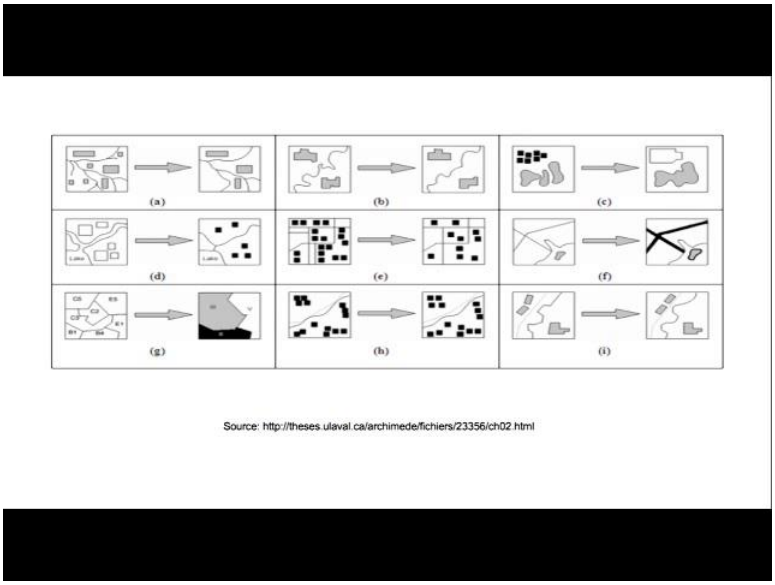
Similar examples are also here the scale is reducing here and therefore, you do not need all those details and all every things is getting generalised here.

(Refer Slide Time: 11:18)



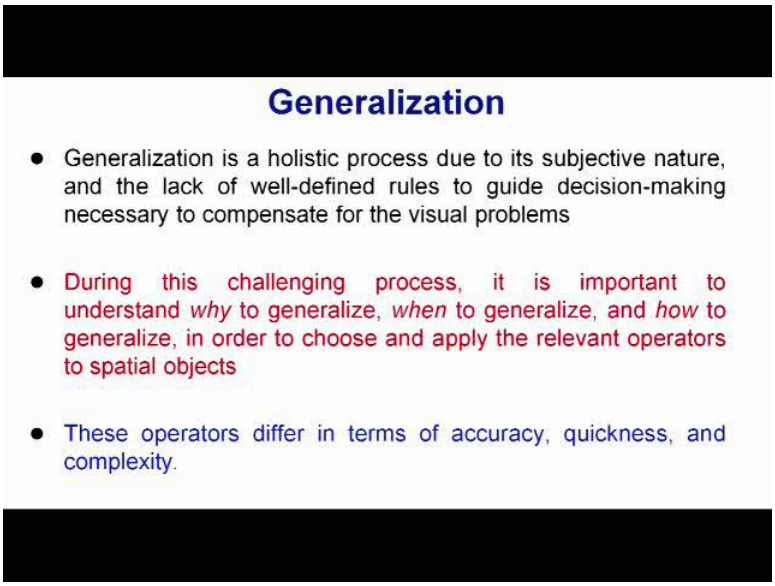
And other example size reduction sometimes you have to reduce the size. So, that can. So, be done under this category of processes.

(Refer Slide Time: 11:25)



They are few more example of similar even for the line data appalling on data point data these things like here point data too many dense points where there, if I reduce this map they will clutter. So, when as to get read out certain points as soon here, similarly line data if I go over this line data and if I reduce this then this line might disappear. So, we have to thicken the line and keep at that scale and generalization as now we can realise is a holistic process due to subjective nature and the lack of well defined rules to guide decision making which is necessary to compensate for the visual problems.

(Refer Slide Time: 12:11)



And during this challenging process it is important to understand why to generalize, when to generalize, and how to generalize in order to choose and apply the relevant operators to spatial objects?

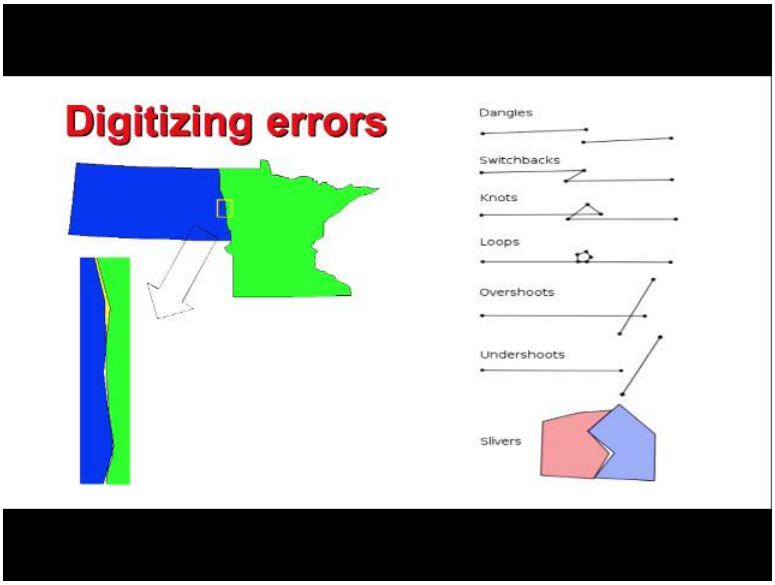
This is what the point which I have mentioned that, whenever you are going for large scale or detailed map or detailed boundary information, to a smaller one then you go for generalization. So, when to generalize you can you can understand how to generalize tools are available few things we have already discussed how to generalize, and it where you will apply this has to be decided by the user himself.

As for the requirements of your projects and these operators defer in terms of a accuracy, quickness and complexity. Because whenever you are going for generalization of course, you are compromising for the accuracy and that is why I said that the original data should not be disturbed. Whatever you will create a generalized map has to be a new data set, and it is not very difficult either to do in a any standard GIS setup.

A similarly error detections error lot of error comes generally during your digitization processes, earlier we used to have a digitizing tablet lot of errors used to come now we are having a this so, called the heads of digitization in which be mapped through a scanner and then as a we treat them that map as a raster and top of that whatever line or features are there we digitize using your mouse cursor and therefore, during the digitization lot of errors can come.

So, we will go one by one there are certain names have also be given to different data because suppose you we are suppose to digitize a steady line.

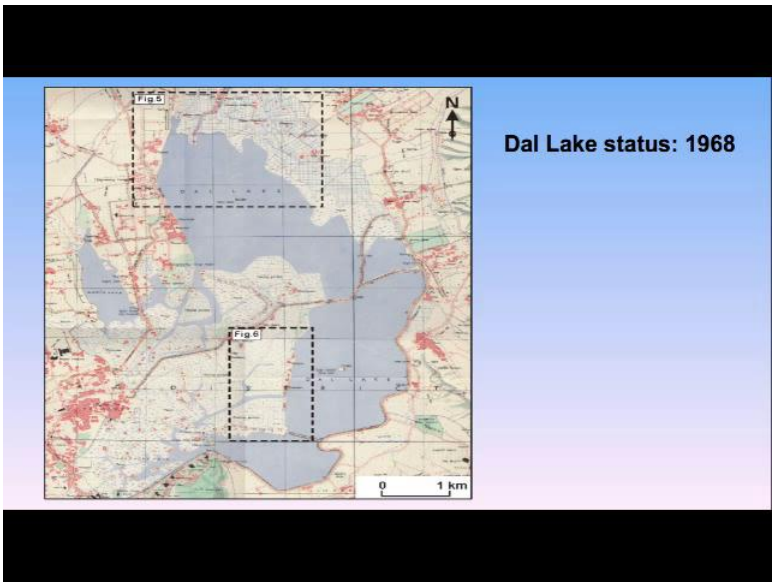
(Refer Slide Time: 14:03)



But you left here and then you restarted here. So, it gap was there it was not intended. So, this kind of error is called dangles, then there might be again you wanted to digitize steady line somehow you got another node and another node and you came back little bit and gone down back. So, there is a shift as well. So, you it is got switch backs and same there is a node which was developed there might be a loop which has been developed through intention was to have a straight line, over shoots are there and that day this would have started from here, but you over digitize it undershoots you left and no link where developed, though everything was intended here in this example to have a straight line.

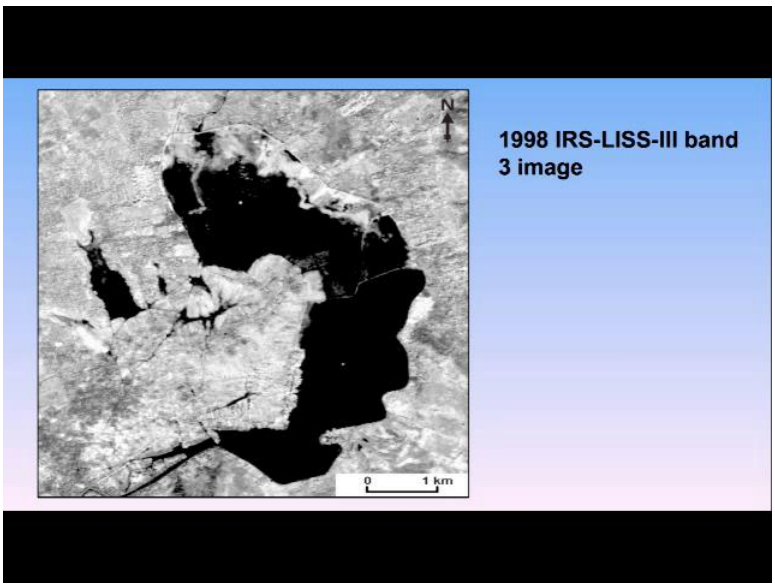
Similarly in case of a polygon data you might be having the problems of living a gap between two polymers, and here also a new set of polygons are their which are artifex which we are never intended. So, number of polygons in your data base will be more as not in as was where not intended and will give you problems in analysis. So, all this a problem should be removed before you go for any further operations in GIS.

(Refer Slide Time: 15:32)



So, if there is a overlap or gap like then you will tells to be removed another process which I have mentioned, I will go in detail is a merging. Merging has got different meaning in GIS, and merging may be merging 2 types of data sets. So, I am taking one example of merging map for a certain purpose which is the change detection. We wanted to detect the change is which have happened when 1968 and 1998.

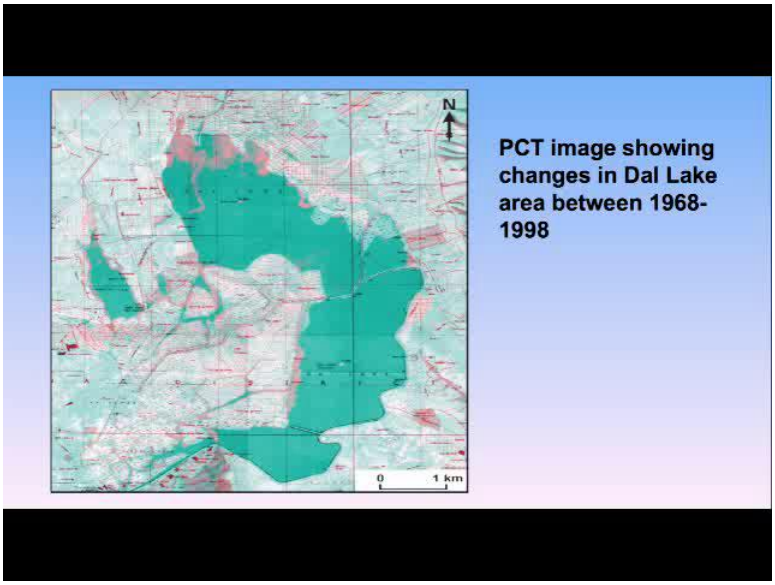
(Refer Slide Time: 16:07)



So, in this roughly 30 years we wanted to see what has what is the change which has occurred to the Dal Lake. So, we (Refer Time: 16:17) this toposheet and treated as an

image, as a raster and you know any coloured image is made from three colours. So, all presently all are here then we will also see their details of against this two boxes which are marked here, and this is the satellite image of 1998, which is infrared channel because we wanted to see the changes which as occurred in the vegetation, especially within the body of the lake or the water.

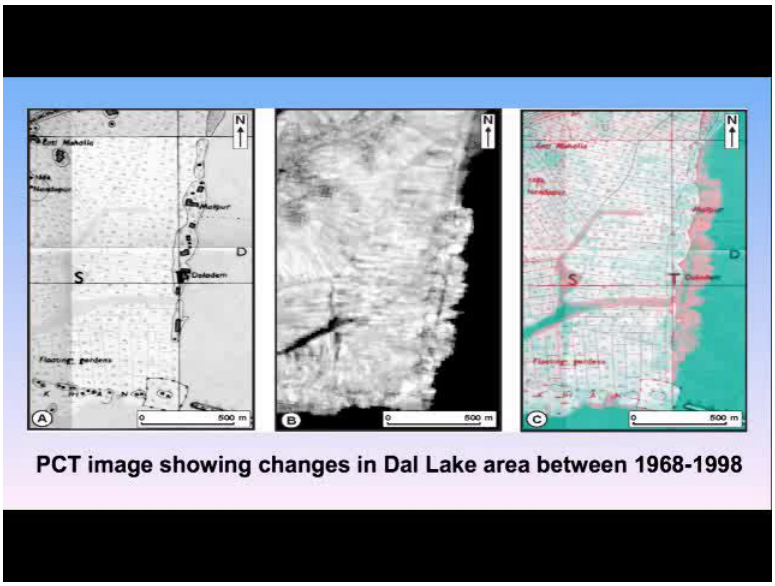
(Refer Slide Time: 16:52)



And when we merge these two then we get new product and which is also having in the background your map of 1968 and the satellite image of 1998, 30 years time difference in all this red a colours are appearing here as the vegetation which as incurred between 1968 to 1968.

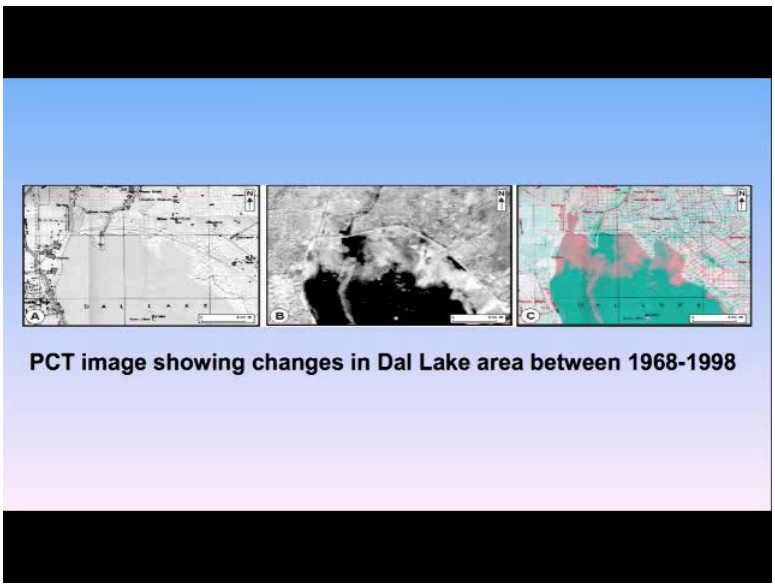
So, this is also one type of merging which is possibly GIS software, if it is not then we can also resort to some image processing softwares to achieve this. Because then we can use this a for a certain purpose here the in this examples the purpose was to first to identified what has happened in those last 30 years and where the encroachment has been done by the growth of vegetation or by human settlements which we will see here when we go over a larger one, and what remedial measures are required. So, exactly after such analysis exactly you know where things have changed drastically, where the corrections are required . So, let us see the detailed one.

(Refer Slide Time: 18:08)



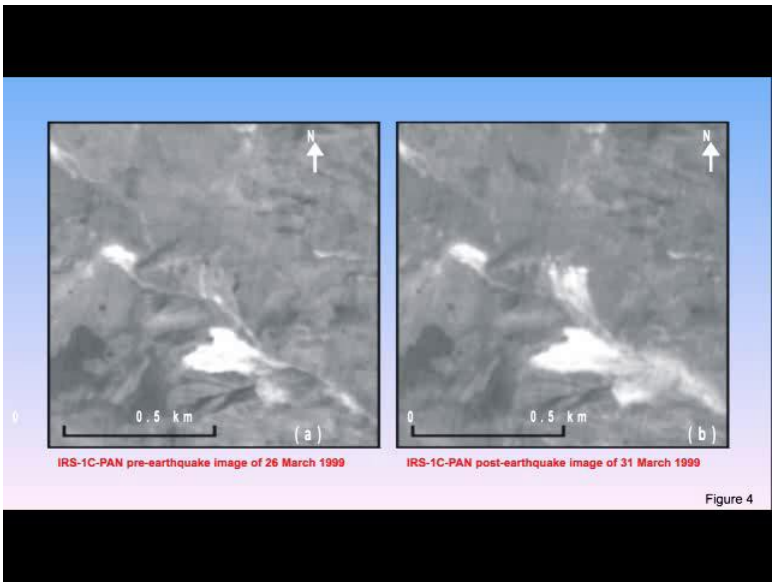
This is the input map this is the input satellite image have been 30 years age difference and this is the final product which is showing that on the this lake boundary these are the areas which has been encroached by humans, even the channels which had the connectivity with other lakes we are also block by the growth of extensive growth of vegetation all around here.

(Refer Slide Time: 18:33)



And another in northern part similar things again input map, input satellite image as you can realise that t his is the area which has been encroached by the vegetation.

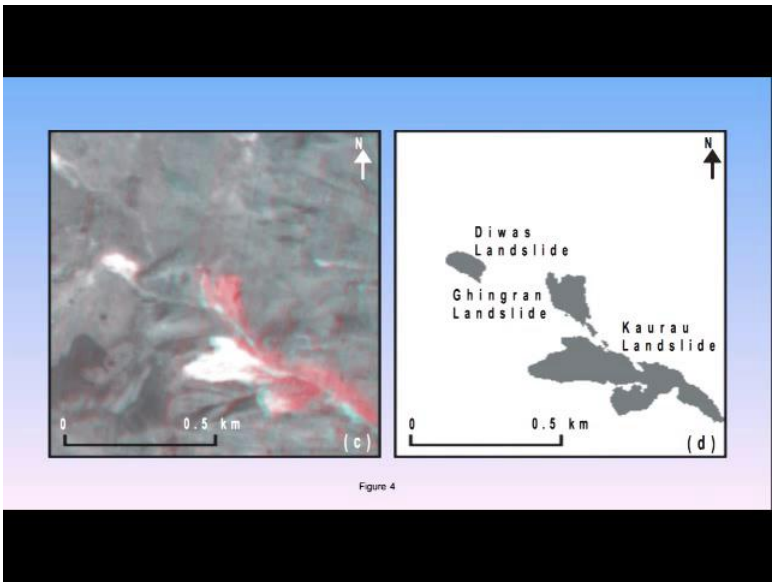
(Refer Slide Time: 18:43)



So, this is another merging technique merging techniques again for change detection can be, say in an example earthquake induced land slide. So, this is pre earthquake image this is post earthquake image. In previous example we have seen 30 years time difference, here what we are seeing just 5 days time difference. So, in between on 29 March 99, the Chamoli earthquake came.

We wanted to see that how which are the land slide which are induced by that earthquake event and it is very easy again almost the same way that you buy assign one color to this one, you assign red colour to this one, and again because if we go for pseudo color transformation then two channels can be assigned this pre earthquake image one channel, red channel can be assigned post earthquake image.

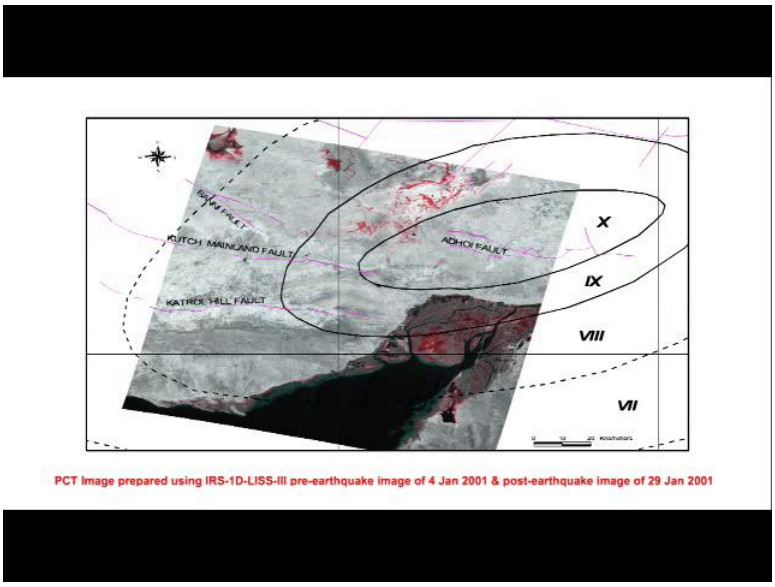
(Refer Slide Time: 19:35)



And you can create a product like this and when. So, here again all red areas are showing the changes between only 5 days, in the previous example changes where soon of 30 years, here only of 5 days and in between the earthquakes.

So, here we can attribute that all these landslides which have occurred are induced by that particular earthquake event. So, the confidence level becomes very high for in such analysis and a these tools can be exploited extensively in GIS the merging one.

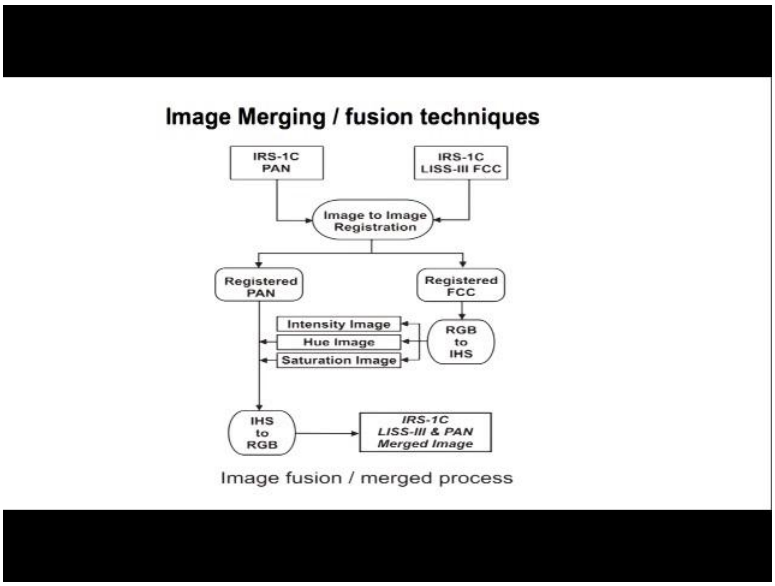
(Refer Slide Time: 20:13)



A similarly one more examples is also there about the liquefaction phenomenon which is again a cosize week phenomenon again, similar technique has been applied for pre earthquake image and post earthquake image that earthquake came on 26 January 2001 when we merge this image an pre and post earthquake and created a pseudo colour transformation all the red area are showing, where the water bodies where appeared and during and just after the earthquake and after few days these disappeared and a that is why you are seeing all this red and even you can see the clear channels which appeared having a fresh water bodies, and this phenomenon is called liquefaction since this is beyond of a scope of this lecture.

So, I am not going to detail about these phenomena, but what emphasis here is how to use your datasets. Whatever is available in your system to create new products which are very convincing, you will have high level of confidence and you can have very accurate results like examples I have shown for Dal Lake I have shown for Chamoli earthquake induced landslides I have shown for Burj earthquake induced liquefaction phenomenon.

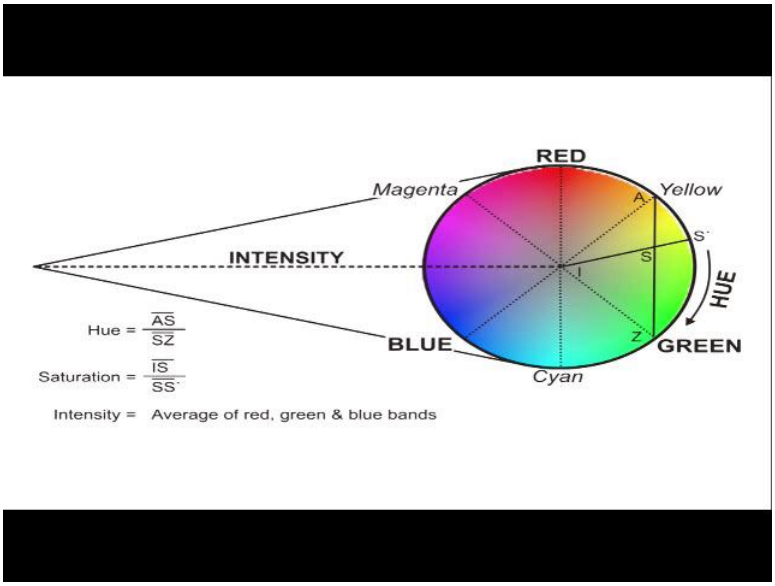
(Refer Slide Time: 21:33)



These are the way you can do it like in this examples, you know in merging techniques one of the standard image merging technique that you can take a high resolutions panchromatic data you can take relatively low resolution multispectral data, you do the image to image railway station the topic which we will discuss later and a then you register the you use the register pan data you register the FCC and then what do you need

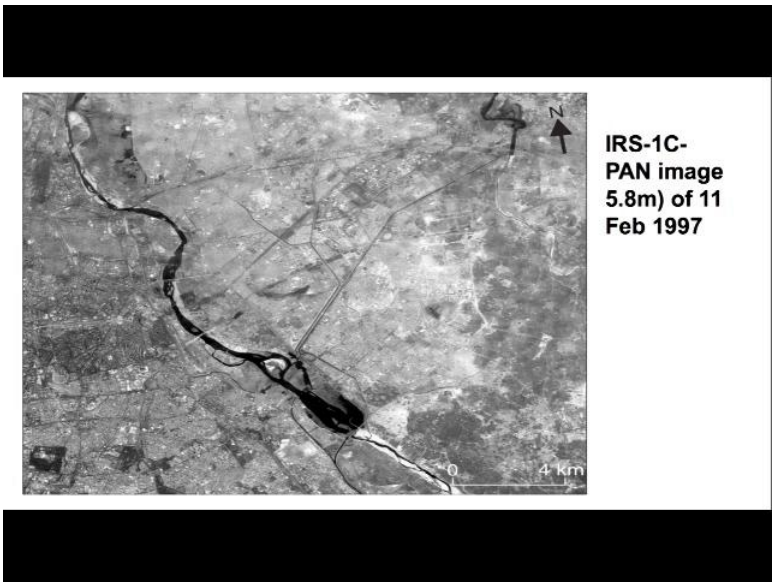
to do is splitting of the image from RGB to IHS, IHS stands here intensity hue and saturation intensity image is not carried further, only hue and saturation image is along with your registered pan image which is carrying high resolution high special resolution data and you go for back ward transformation from again IHS to RGB and you go for a high resolution multi colour or multi colour image, or a false colour composite high resolution. So, we started with relatively high resolution, but panchromatic single colour or single band image, here relatively low resolution, but colored image and we end up with the high resolution colored image.

(Refer Slide Time: 22:52)



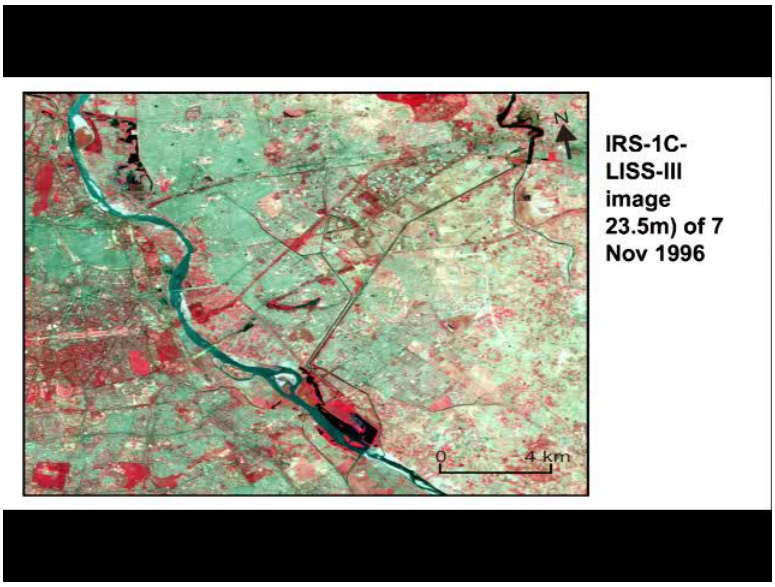
And examples is shown here that how this intensity, because you have to when you go inside this colour a space then you can play with the different planes instead of projecting your things on RGB plane one can project on IHS plane and so on.

(Refer Slide Time: 23:16)



So, this is the one cone inside of that one, and let me show you some examples of this merging of high resolution panchromatic image, that the relatively low resolution of multispectral image.

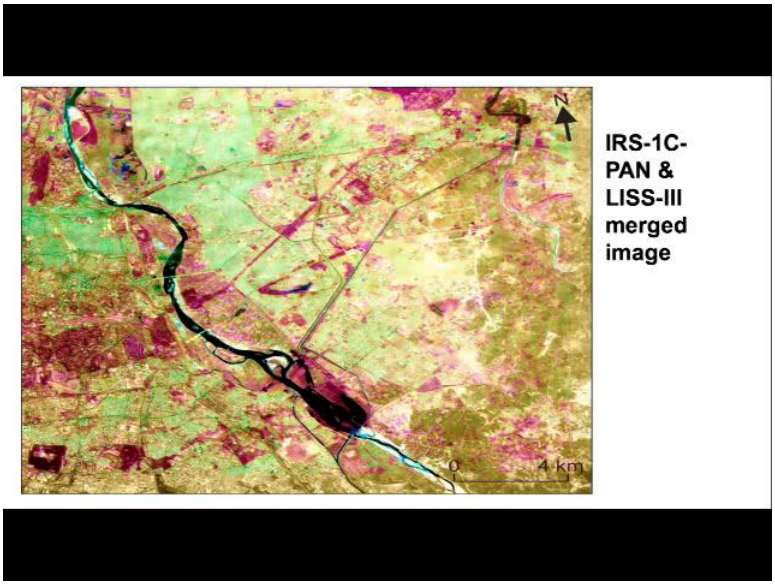
(Refer Slide Time: 23:31)



So, this is PAN image having resolution of 5.8 meter and this is the corresponding image list 3 having 23.5 meter, but this is colored, this is black and white, but it is having high resolution it is having colour, but relatively low resolution. So, we a through merging techniques you can explode the best out of this 2 inputs, images and then create a new

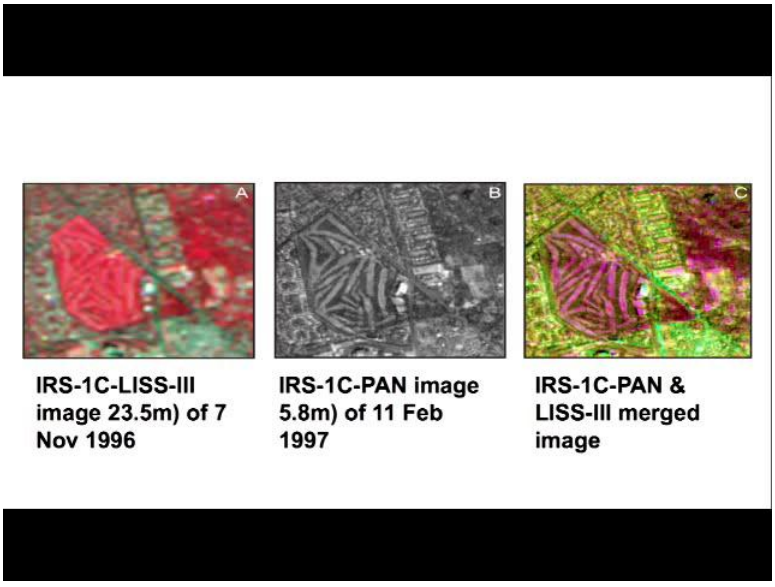
product and like this it has been created. So, now, you are having high resolution same time colored image.

(Refer Slide Time: 23:56)



So, likewise you can generate lot of merged data, which later on can be used for different kind of analysis and modelling purposes as well, this here the June parts of the that this is a colored image input image.

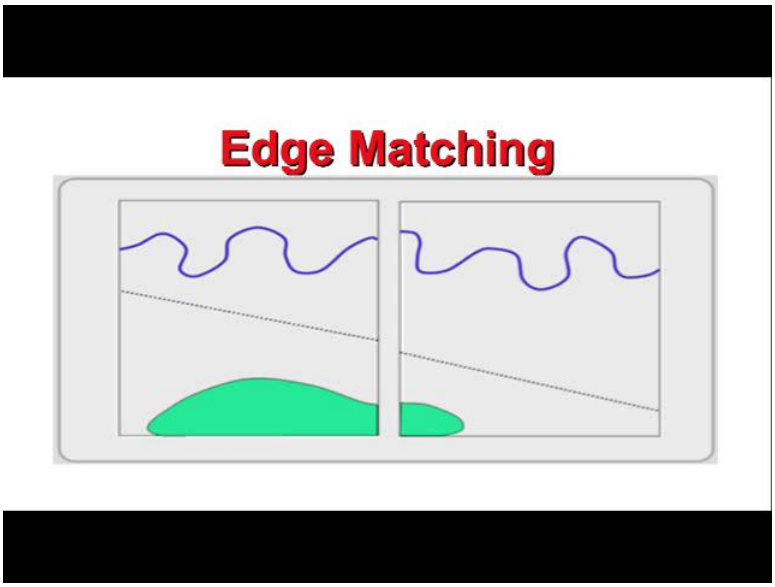
(Refer Slide Time: 24:18)



.

This is a black and white PAN image, but high special resolution this is the output as you can see this is the golf course. So, here the vegetation details are there, here are the spatial details are there both have been merged.

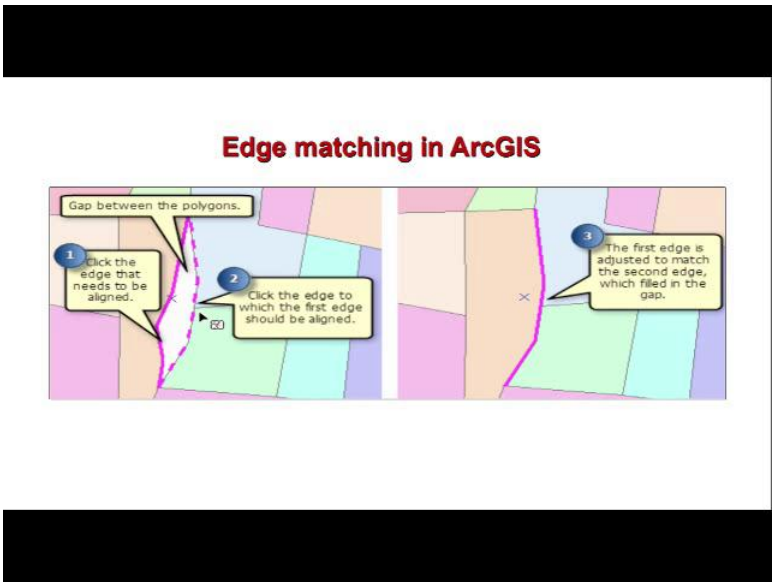
(Refer Slide Time: 24:45)



There are slide changes in the colours which can be adjusted as well, last topic in this one of pre processing there are other steps which we will discuss in separate lectures. So, edge matching is I have mentioned earlier, that when you are having two edges and maps and they are might be having even same projection or different projection, some time the boundary will not match. So, they are here the line may not match from one map to another.

So, some human interventions are required and then whoever is doing that part can bring this data to here and if there is some shifting is required that need to be also done. Similarly for a straight line, similarly for polygons edge matching is many times very much required. So, that your data becomes seem less otherwise it will create unnecessarily artiffects and problem during your analysis. Similarly here also that l matching can be done.

(Refer Slide Time: 25:41)



There are tools examples of ArcGIS is given that you can shift this one and bring here and then a perfect because, the white colored polygons was never intended, but because of solve over problems this problems came. So, it can be removed very easily from here. So, this brings the end of this presentation about the pre processing of our data into GIS.

Thank you very much.