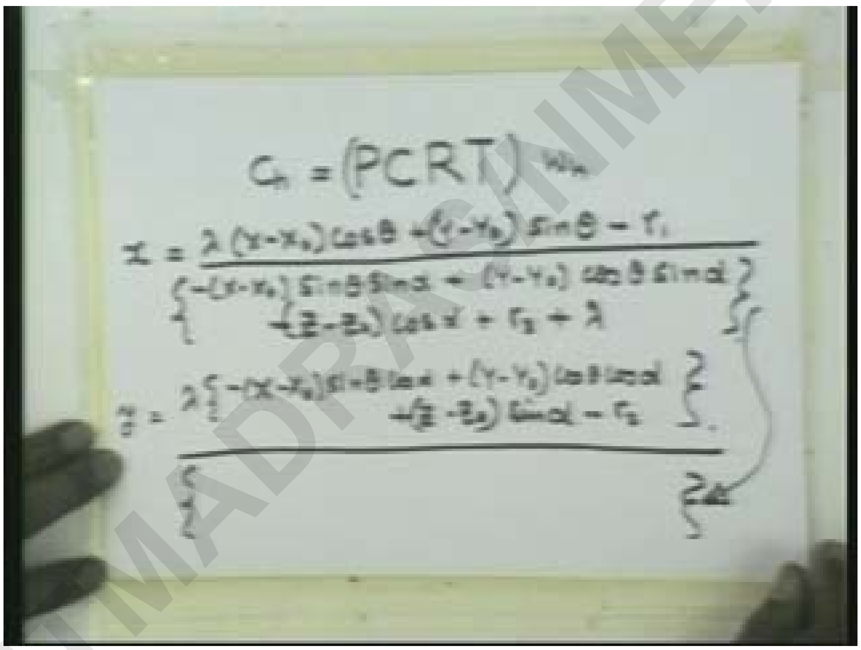
ROBOTICS

Prof. B.Seth
Department of Mechanical Engineering
IIT Bombay

Lecture No-27

Image processing

Apart from where we left off last time and we were looking at perspective transformation basically we were looking at the problem of finding the image point corresponding to a world point or a point in the real world when the camera axis and camera coordinates were not coincidental with the world coordinate axis and what we found was could write down [noise] our information to determine a point in the camera coordinate system c h(refer slide time05:09)



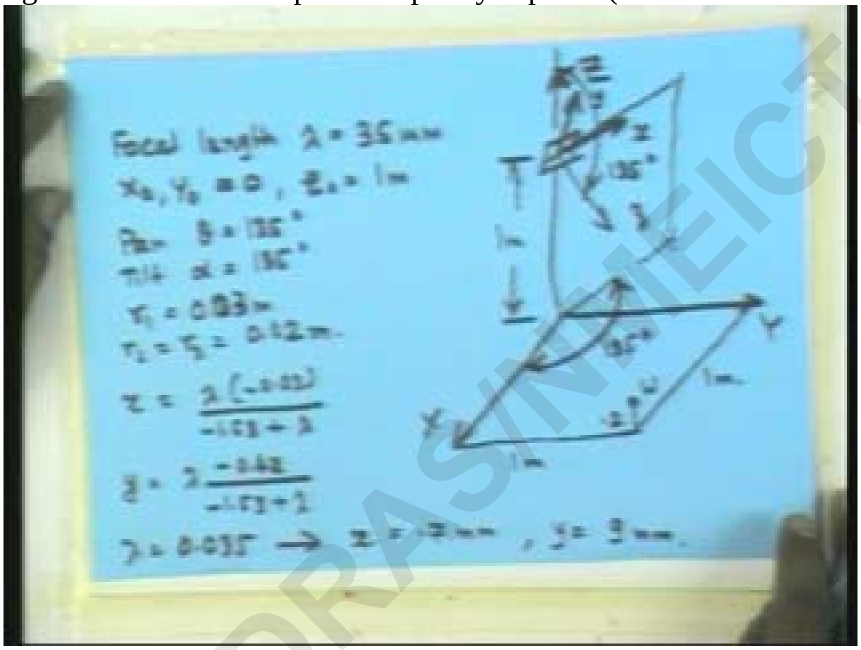
we had to multiply the number of transforms the perspective transformation the gimbal offset the rotation which can be clubbed together with pan or tilt and then the translation of the global axis to the gimbal center all this has to be multiplied to the world homogeneous coordinates right [noise]
one can write this all out each of the transformation is fairly simple and finally one can say that x that is the camera image plane x axis location of a point will be given by the focal length of the lens times x minus x not cosine of theta plus y minus y not sine of theta [noise] minus r one divided by minus of x minus x not sine theta [noise]sine alpha plus y minus y not cosine theta sine alpha minus z minus z not cosine of alpha
okay this determines so given any global coordinates x y and z right and the parameters which define the location of the coordinate frame of the camera with respect to the global coordinates which basically is determined by the offset of the gimbal center x not, y not,

z not the pan angle theta and the tilt angle alpha and the offset of the gimbal the image plane center with respect to the gimbal center which is r one r two r three

so here r one and r three are appearing lambda of course is the focal length of the lens

similarly you can write[noise] down for the y coordinate you will have lambda times minus capital x minus capital x zero sine of theta cosine of alpha plus y minus y not of cos theta cos alpha plus z minus z not of sine of alpha minus r two this in the numerator denominator remains the same

okay so this is the kind of relationship that you will get then we are looking at the camera which is not coincidental with the image plane axis okay now lets illustrate that so lets consider our global coordinates capital x capital y capital z(refer slide time09:41)



let me imagine that there is a point one meter in the x direction one meter in the y direction and say twenty millimeters above the ground plane so one meter one meter and "point two" meters this point i want to know where this will appear if i had a camera which is mounted with the image plane with axis so this is my x axis y axis and small z axis so my camera is put with the image plane center along the global z axis and i am looking at this point and i want to know where this will appear in the image plane

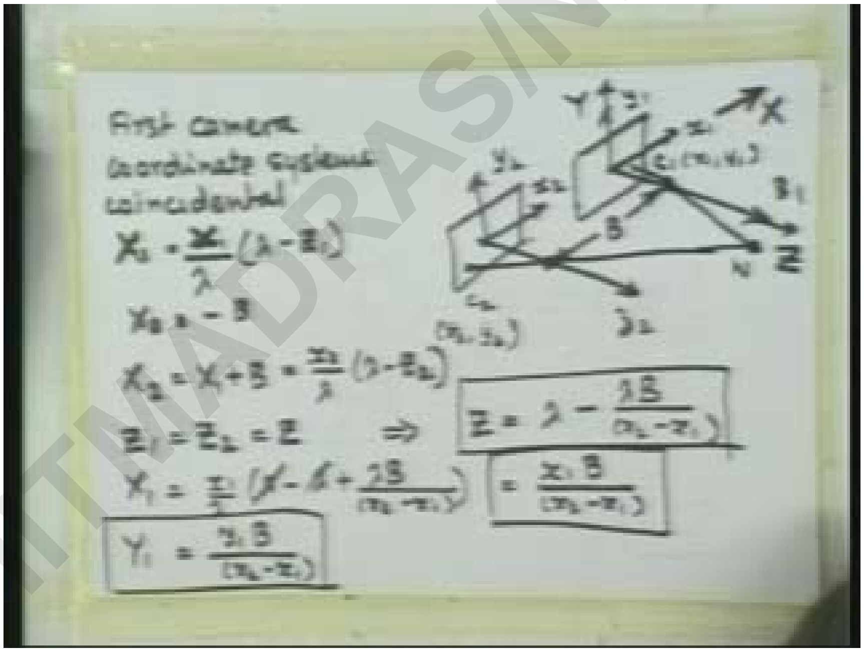
of course i will have to be given a few other things i will have to be given what is the focal length focal length lambda lets say thirty five millimeters

okay i need to know what are the other parameters x zero y zero both are zero z zero is say one meter so this is also one meter okay now what is the tilt and the pan angle here we will have to see how much it is x axis and small x axis is rotated

so this is the plane that defines the and so this is the angle that we are talking about this is hundred and thirty five degrees and the other angle we are talking about is the x z direction versus the small z direction so we are talking about angle here

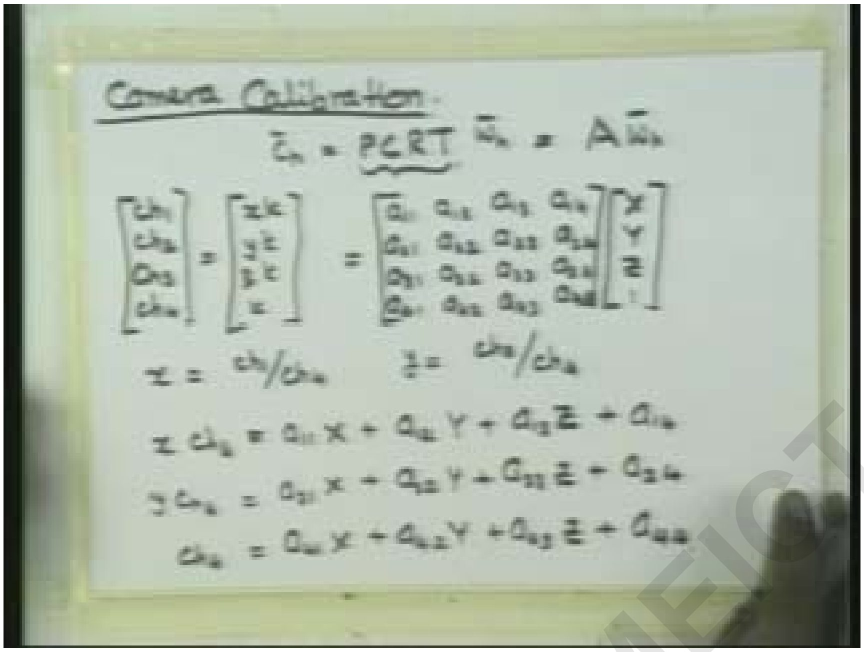
okay so that is also i have taken one thirty five degrees so we have pan angle theta is equal to hundred and thirty five degrees and tilt alpha is equal to one thirty five degrees alright now we also need to know r one r two and r three so lets say r one is equal to

“zero point three” “zero point” well “zero point zero three”meters and r two is equal to r three is equal to “zero point zero two”meters
so this will now i have all the information i will be able to compute what this point w h w looks like in the camera coordinates which will be somewhere on this plane right so one plugs in the thing with all the simplification that we have here so one can reduce this to lambda times minus zero point zero three divided by minus one point five three plus lambda and y is like wise given as minus zero point four two divided by minus one point five three plus lambda and if you substitute the value of lambda you get the [noise] lambda is equal to [noise] “point zero three five” it gives us x point seven millimeters and y equals to nine millimeters
so this is just to illustrate that once you have all the parameters you can determine what the coordinate image point would be of a given okay this is what we are interested okay now one other way we can go from image point to the real world we earlier when we look at the inverse perspective transformation or imaging transformation we found that there were some ambiguity if you have a image point you could not [noise] for sure what the coordinates of the external world point were because everything in a [noise] so one gets around this problem is by looking at what is called stereo imaging if you have instead of one camera if you have two cameras then we can have we can determine the z axis also and that is what we will look at now supposing i have two cameras(refer slide time17:10)



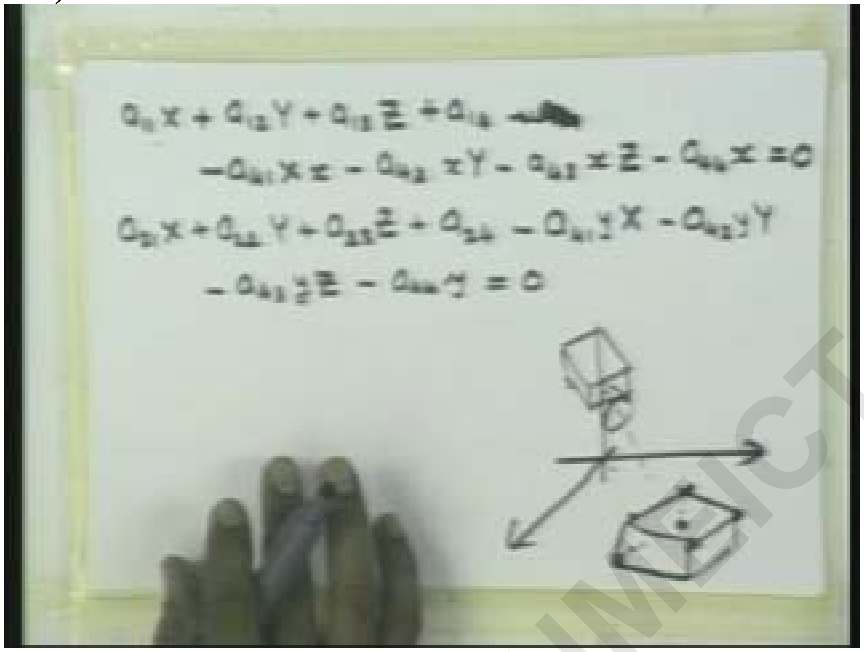
and i will only show the image planes of these two cameras one these are the two image planes with the axis these two axis are parallel so this is my say x one y one z one this is z two x two and y two
okay now the two cameras both are parallel they are displaced from each other by some amount lets say b along the x direction and lets define the axis global x global y and global z to be coincidental with the purse of the cameras right so now of course we have some focal point i am assuming both the cameras are identical so we have some focal

point lets say the focal point was here and here and lets take a let me take a image first and then correspondingly fine so i have got this and i have got some image here
so passing through here i get a i look at a world point here w small w and i have one image c two and i have one image c one here image point corresponding to this particular point right the idea is that if i know x one y one corresponding to this and x two y two corresponding to this then i want to know where this particular point is in space right so lets start out by looking at the first camera
okay for the first camera we have the two coordinates are coincidental okay coordinate systems coincidental therefore we can write down lets say for x one coordinate as capital x one divided i am sorry i am looking at the [noise] capital x determine small x one divided by lambda times lambda minus capital z one this is what we had determine earlier this is small x by the way okay
now how do we find out the coordinate we have some small x to here how do we to the coordinate of this particular point in the global coordinates well we have done[noise] much more complicated thing in the example and in the when we saw in general case so what we have is we have a displacement of this coordinate system with respect to the global coordinate system that it is displaced by a amount b so in terms of what we had seen earlier we have x not is equal to minus b if you refer to the what we have done so for y not [noise]zero r one r two
so we have a very simple case what we can determine is that x two x two which is the coordinate of this with respect to this coordinate frame now okay which is nothing but x one plus b right and this now simply can be written as x two divided by lambda lambda minus z two right both these image planes are aligned along the x axis so then z one is equal to z two right because [noise] the distance z and that lets say equal to z
so what i can do is now i can use this relationship i can subtract one from the other to get a relationship and I [noise]that two equals to z from all this i get that [noise] capital z equals to lambda minus lambda b divided by x two minus x one okay now i can use this z into this expression to find out that x one is infact equal to x one over lambda
okay times lambda minus z z is this which is lambda minus so it becomes plus lambda b divided by x two minus x one
okay so this cancels off and then you will see that this lambda will cancel with so what will be left with is x one times b divided by x two minus x one
right similarly i can find out the global y one coordinate as small y one times b divided by x two minus x one and of course i have already found z here so i have all the the complete global location of the points known if i had two cameras instead of one camera okay
okay so next thing that i want to look at is basically the issue of camera calibration what is camera calibration mean it basically means that i want to determine the transformation matrix that i have talked about which is the product of p times c times r times t now if i compute it from the pan angle the tilt angle every error um in that each of these quantities is going to lead to um a error in the transformation matrix right so from time to time i will need to calibrate it by actually knowing some points in the world frame which are known coordinates and looking at the image points and then trying to determine what this transformation matrix is
right so fact that we have camera coordinate vector is equal to(refer slide time22:16)



some p times c times r times t times the world homogeneous coordinate factor
so i can say that all this is collapsible into one single matrix because each of this is four by four matrix[noise] and let lets call that a that is a transformation matrix which i am interested now in general i can write so if i am given some world coordinate point
okay so that is x y z one and i have this matrix here which is a four by four matrix and it has some elements i dont know what the elements are
so let me write them as okay even though you cant read each of the subscripts very well you know what i am writing so a three one a three two a three three a three four and finally a four one a four two a four three a four four i have got this sixteen elements in this matrix this is multiplied and what i will get from this product is essentially some x times k some y times k some z times k and k which will correspond to the camera coordinates from this if i want to determine [noise] what the camera coordinate is i have to locate the four element in this vector and divide each of the other elements
so let let us say this what i will see in the camera is some c h one c h two c h three and c h four and i understand that if i want to get actually what the coordinate in the camera image plane is i will simply get it by c h one being divided by c h four and similarly y will be equal to ch two divided by ch four okay now let me multiply this rightern side out so we get c h one which is nothing but x times ch four
okay so let me write the first row as x times c h four that is equal to a one one times x plus a one two times y plus a one three times z plus a one four right
similarly i have y times c h four is equal to the second row which is a two one x plus a two two y plus a two three z plus a two four okay row i am going to ignore that is not significant i will look at the fourth row so ch four is given as a four one x plus a four two y plus a four three z plus a [noise] okay
so now what i can do is i can do some manipulation here i can use the fourth fourth equation which is the third equation that i have written here corresponding to the fourth row of the matrix equation and i can substitute that in the first two equations to eliminate ch four so i will have x and y relating to capital x y and z through the coefficients we have okay so [noise] now so what i will get is first equation that i will get is if i

substitute for ch four and take everything to the one side so i get a one one x which(refer slide time27:55)



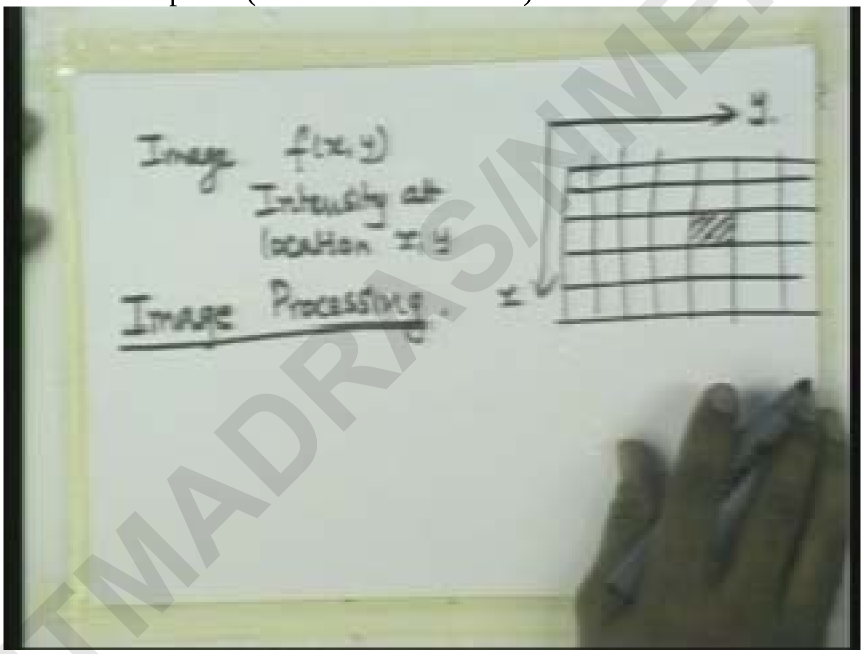
i had before plus a one two y plus a one three right plus a one four and now i am taking the what was in the left side which was x times c h four and ch four am [noise]
so therefore i will get minus a one four sorry yeah a four one [noise] okay let me write it here minus a four one times x times small x okay minus a four two times small x times capital y minus a four three times small x times capital z minus a four four times small x equals to zero
okay all i have done is i had this equation here so i am substituting for c h four from here right so c h four is dependent on a four one a four two a four three a four four and this gets multiplied by small x and i am taking it all to the rightern side
so we get this equation here so okay similarly [noise] i will get equation for the second row which is a two one times x plus a two two times y plus a two three times z plus a two four minus a times small y[noise] times x minus a four two times minus a four three times small y times capital z minus a four four [noise] small y equal to zero okay so much for you know algebraic manipulation that we have been doing what just look at these equations now i have two equations how many unknowns are there i don’t know the coefficients of the matrix that i am trying to determine
so i dont know all these a one one a one two a one three a one four a four one a four two a four three a four four right this eight plus four a two one a two two a two three a two four these are the unknown what i know is i will deter i will decide on a point in the global x y and z coordinates not only that i will decide a number of points say supposing i have some well known points so it could be a part of a box that i have kept here
now in the view of the camera right i am able to see this corner this corner this corner this corner this corner and this corner
so i can see six points in this size of this cube or rectangular parallel pipe it will be known to me so i know locations of each with respect to each other right i can also know the location of this whole box right to the global coordinate frame so knowing this i basically know x y z of six of these coordinates i have twelve unknowns i have two equations here

so if i write these equations corresponding to six points right then i will be able to solve for the unknown coefficients

so this is typically the way the calibration would be done you will take a some kind of a picture or which has known dimensions and the corner points can be identified in the image it could be done automatically using some kind of image processing that we will see for the in the course

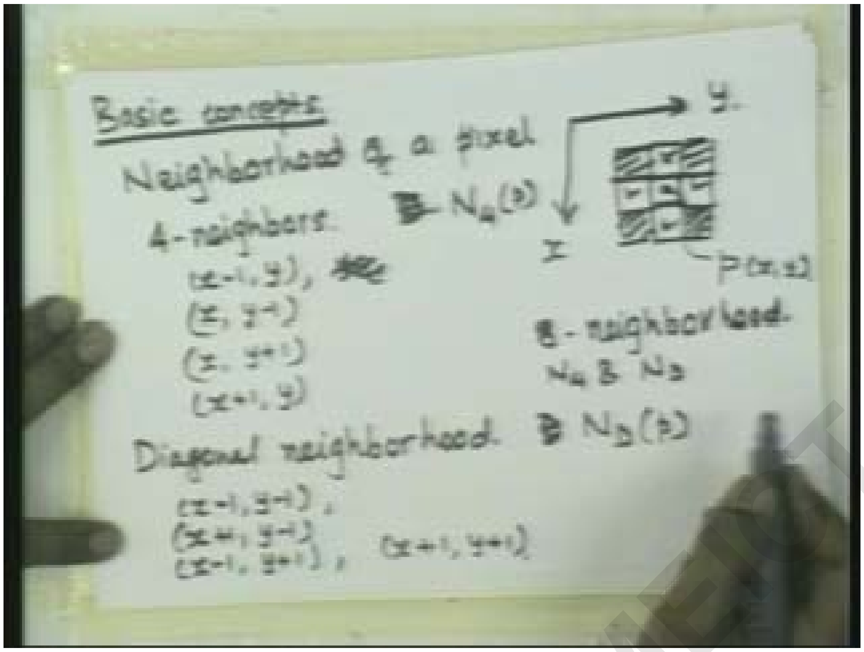
so we can basically find out the corner points knowing those x y z coordinates and knowing the corresponding image coordinates i can substitute them in these two equations so six times six point times equations twelve equations i will get and i have twelve unknowns its a linear solution problem of simultaneous equations alright so this is how cameras are generally calibrated it could be done periodically it could be done once in the beginning of the task and then next time when it is used okay

so now let us move on from what we have seen so for is we have seen that this is the image formation and transformation the imaging geometry is what we have seen so for now what we will concentrate is that we have a image which with okay we have a matrix with discrete number of pixels(refer slide time29:54)

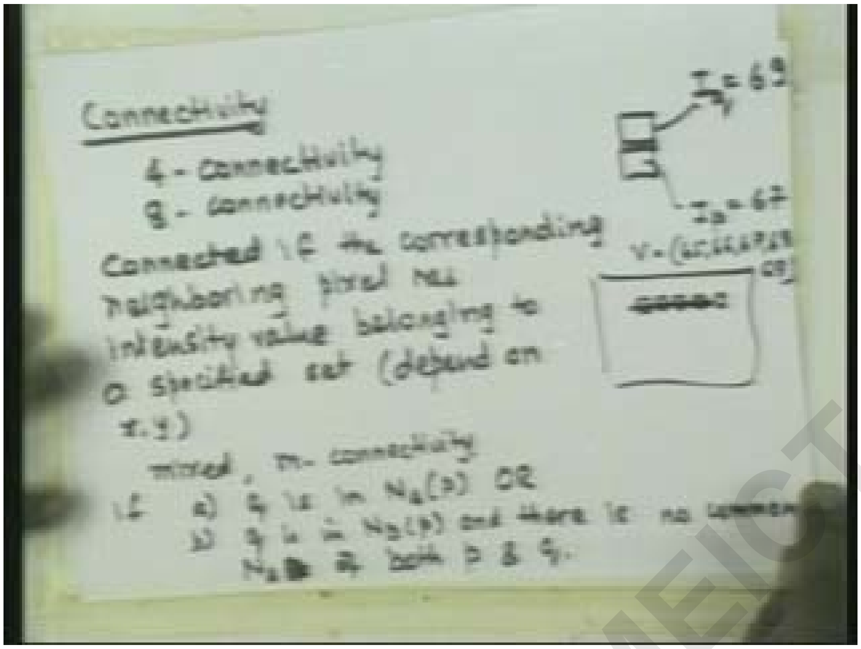


each of these pixels has some intensity value so my image is defined as is written as x y some function of that which is the intensity so this is the intensity at location x y x being this y being this so if i look at some particular pixel here then what is the intensity at this particular point [noise] specify all the image points then that is my image right now what we are going to do is we are going to look at how we process this information from the image okay

so we are going to look at now image processing but before we look at image processing we will look at some basic concepts some basic terminology and concepts(refer slide time33:59)



first thing is neighborhood of a point neighborhood of a pixel
so there are some neighboring pixels okay there are different types of neighbors i can this is my pixel say then i have a neighboring pixel here a neighboring pixel here neighboring pixel here neighboring pixel here
so i can define these as four neighboring pixels and these are called the four neighbors or four neighborhood so the point t the four neighborhood of point t is this one this one this one and this one right so obviously[noise] coordinates as x y the four neighbors are going to be x minus one y okay which will give me this pixel so just remember that i am keeping x axis here y axis here so x minus one this is x y x minus one is this one and y is this one
so this is the one that i am talking about then i have y minus one sorry x and minus one x and y minus one then i have y which is this so these are the four neighbors now i can also have other type of neighborhood which is the diagonal neighborhood
okay diagonal neighborhood and this is called this will be called d four or sorry n four neighborhood four of p this will be called neighborhood diagonal of p okay and
now diagonal neighborhood is basically these other four pixels which are not directly in the x direction or y direction but which are displaced by x and y or both right so the diagonal neighborhood is given by x minus one y minus one x plus one y minus one x minus one y plus one and x plus one
okay this is the four diagonal diagonally neighbor elements or pixels right together all these eight neighbors is going to be called eight neighbors just like we call four neighbors neighborhood is n four and n d of a given pixel this is just some definition this is what is commonly used okay then we have a set of connectivity (refer slide time39:24)



okay in the neighborhood we can define connectivity can be four connectivity or eight connectivity means is basically if

if i am given a pixel here okay and i have some neighborhood of this pixel so supposing i am looking at four neighborhood so i look at one of the four neighbors of this pixel and then i look at the intensity value if this has some intensity i p

okay and this is a pixel q and this is has some intensity i q if i p and i q are same or nearly the same then i may define this pixel to be connected to this pixel right just imagine what we have basically we are talking about some image right in this image say i have some some line existing here so let lets keep it very simple there is only one line here so there are various pixels here right i am looking at the four neighbor

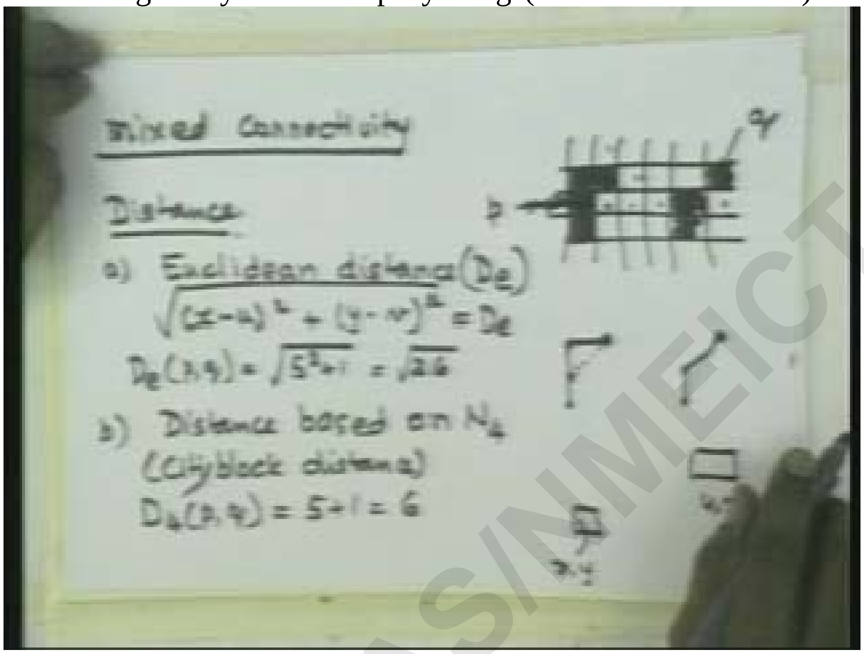
so basically this is the four neighbor of that and they have the same intensity value [noise] same intensity value so we will say is that these will be something is connected if the corresponding neighboring pixel

okay has intensity values belonging to a specified specified set okay this of course will specified set it may be depend on okay

it may depend on the pixel under consideration so it will depend on x and y location in general so what i am saying is supposing this is the pixel that i am considering right now it has a intensity value of sixty seven if it is a gray image from zero to two fifty five then it represents some amount of intensity or grayness here supposing this has sixty nine right its not identical but i may define that given this as the intensity here then i can define set v to include intensities sixty five sixty six sixty seven sixty eight sixty nine thats it then i am saying anything which is plus or minus two gray levels away from the pixel under consideration i will consider it to be the same body okay because there are going to be some variation from pixel to pixel because of noise because of variation of the source light etcetera

so there is going to be definitely some amount of variation and that i am saying okay i am going to define as set so if the pixel intensity belongs to that particular set and if it is in the neighborhood then i will say that these two points are connected right there is another concept of connectivity which is called mixed connectivity or m connectivity (refer slide above)

okay in this case you have m connected region if firstly q which is the neighboring pixel is in n four of p okay or q is in diagonal neighborhood of p and there is no common n four n four of both p and q i thing i need to illustrate this okay so what i am saying is lets consider part of a image okay let me simplify things(refer slide time46:05)



by saying that something is black or white rather than proceeding grays so it becomes simpler to see what has the same intensity value whether something is in the neighborhood or not so let me consider what one part of the image where you have this pixel is black this pixel is black and this pixel is black

okay then if i am talking about connectivity then if i consider this as my p pixel okay p as pixel then currently this is my q pixel i am looking at this this is in the four neighborhood so it satisfies the first condition it has the same intensity so they are connected now if i come to this one here it is in the four neighborhood but does not have the same intensity so it is not connected okay if i look at this as q pixel then this is in the diagonal neighborhood and it has the same intensity okay and further its four neighbors which are these four neighbors and its four neighbors which are these four neighbors so the common four neighbors are this and this and they do not have the same intensity they do not belong to that region right

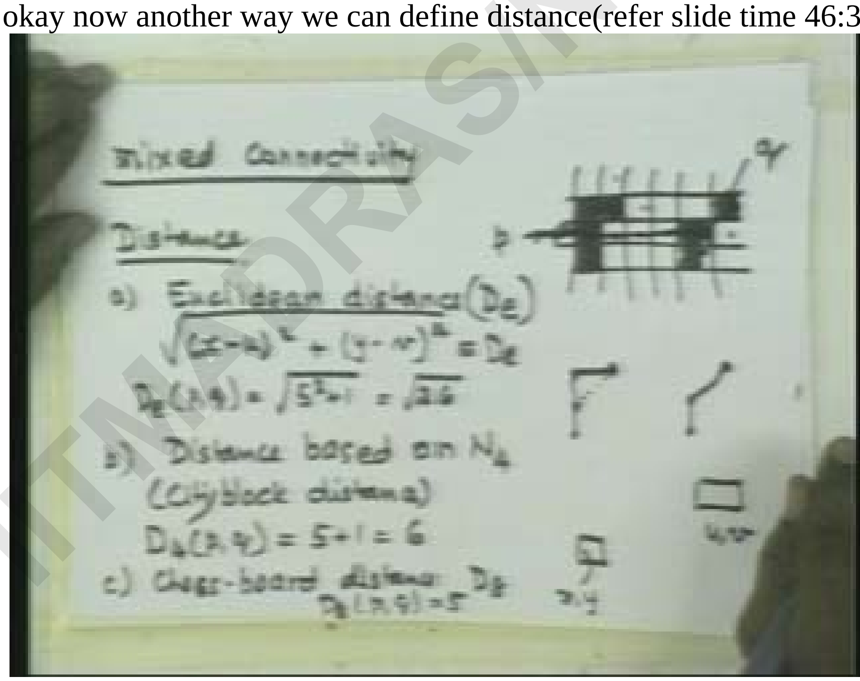
so then i will say that this is m connected this is also m connected okay well you could have simply said why not say eight connected so let us look at another situation where this gets this is also has the same intensity value right this pixel is m connected this pixel is m connected and this pixel is m connected to this so i have this is my [noise] but i will that this is i will not consider this connection because it will not satisfy my second condition that the common four neighbors of these two pixels that i am considering

okay they have the same intensity value so they they are connected therefore i will not connect this region here so this is just to make sure that you dont have duplicate connection because if you dont consider this type of neighboring or the connectivity then in general you will connect this you will connect this you will also connect this then there are duplicate connection and when you run when you run some of the algorithms you will run into problems because some duplicity has and the values will not come out correct so if in the first case when this was not black so which was the case if i show it again here you have this and this and this

then you will have the connectivity that this is connected to this and this is connected to this but when this is also present then it is taken as connectivity being this way okay so this is our mixed connectivity that i will illustrate it

okay then the next basic concept i want to deal with is concept of distance okay now distance is we are familiar with so what we are familiar with is the euclidean distance okay lets denote this by d e distance euclidean this of course we know is supposing i take two pixels of interest this is x y coordinates this is another pixel some where say this is u and v coordinates so the distance between these two pixels is nothing but is x minus u squared plus y minus v squared square root of that here this is how we define the euclidean distance we will take the positive square root of this so if i um consider distance say from this black and white is very hard to show okay supposing this my p pixel which i have already shown and lets say my q pixel is this one now

so the euclidean distance will be one unit two units three units four units five units along the y direction and one unit along the x direction so in this particular example okay d e of between p and q is nothing but five squared plus one square root that is square root of twenty six okay now another way we can define distance(refer slide time 46:31)

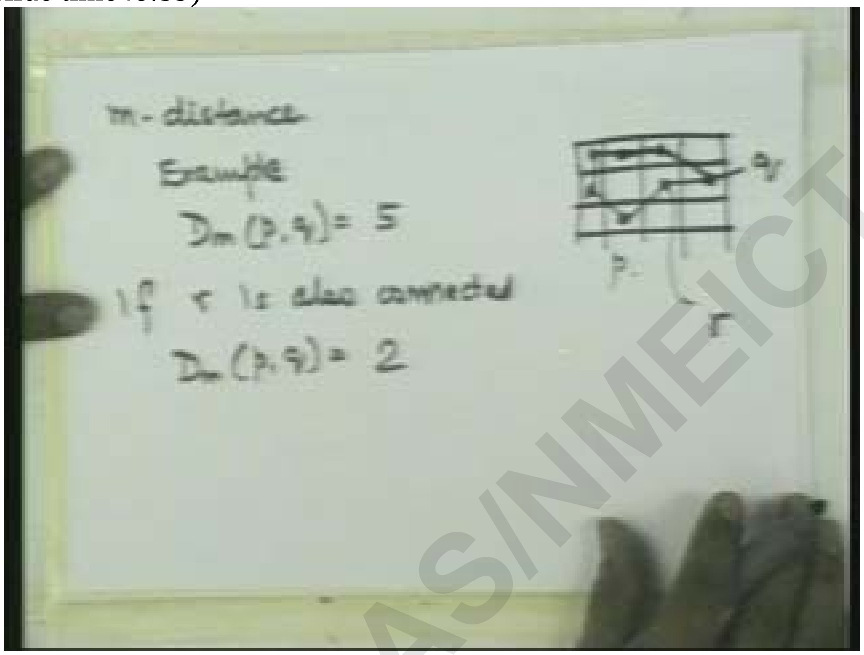


is if we look at the four neighborhood so distance based on n four right so if i look at that distance that will be what is also called city block distance

so imagine these two be streets and these pixels to be some buildings so obviously you cant go through the building so if you want to come from p to q then you have to go one block two blocks three blocks four blocks five blocks and six blocks

right you have to travel six blocks so this becomes if i d four between p and q will be equal to five plus one which will be the distance alright this way i can also define the

chess board distance oops chess board distance which can also be called as d a distance okay this will be how many moves i have to make either along the x direction or along y direction or along the diagonal direction
so here if i go one two three four and i can go from here to here so i have the d eight distance between p and q is only five units right
finally let us also look at the m distance which is the mixed mode distance criteria so this time(refer slide time48:59)



i will not paint them else then it will becomes very difficult to see so supposing i have this is my p pixel and say this is my q pixel but now you to find the distance it is not simply enough to look at this now i have to see whether connected or not only if the m connected i can find the m distance so it depends on what all supposing i put the dot instead of pulling it so these are both black then i have this one black this one black this one black this one black say i have and rest of them are not black which means connectivity wise if i look at my m connectivity this is connected to this this is connected to this this is connected to this this is not connected to this
right because it gets connected through the neighborhood four connectivity as a preference then i have this connected and this connected so therefore the distance here becomes one unit two units three units four units and five units in this particular example so what i have shown here as example here i have the m distance or d m between p and q is five units
just notice that if this pixel was also belonging to this set then the distance will reduce to only two units okay
so this when this is q this is a r if r is also connected then d m simply be only two units and not five units okay
so what we will do is now we will stop here next time we will start looking at some other steps that we require for processing the images starting from pre processing okay thank you