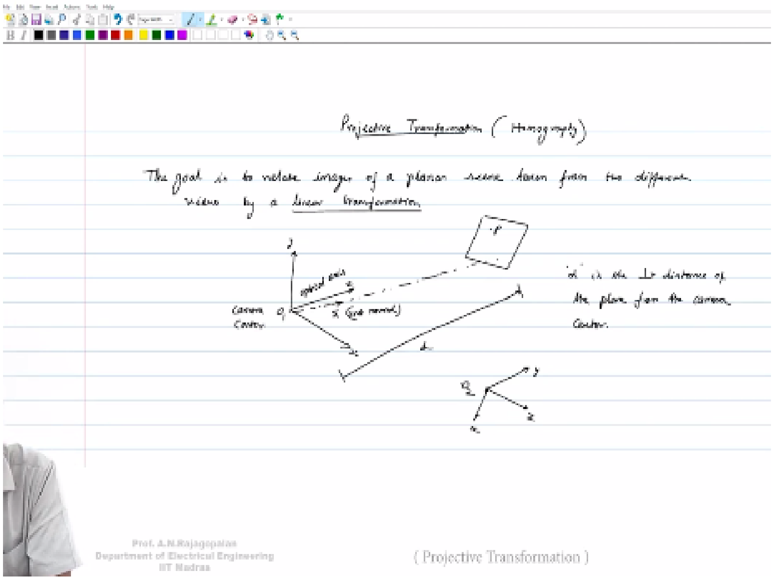
Image Signal Processing
Professor A N Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology Madras
Lecture 10
Projective Transformation

(Refer Slide Time: 00:16)



Okay, now let us come to a projective transformation, projective transformation, what is also referred to as a homography, called a homography. So it is a homography, homagraphy whatever, it does not matter as long as you know what we are saying. Now the goal is like this, I am going to draw now let me write our as a goal, the goal why we are doing this. Goal is to relate images of a planar scene now I will start with a planar world like I tell or like I told planar world is actually a nice world.

Planar does not mean it has to fronto parallel, by fronto parallel what I really mean is something like this is a very special case which is fronto parallel. So you have a camera whose sensor plane is like that, let me put it like this all of you can see then this is a fronto parallel scene and the way we normally call it, if this is your sensor plane think of this is your x coordinate of the sensor, this is your y coordinate.

And the one that come out is an optical axis of the camera that is the way we treat and the place where these axes are all originating that is supposed to be the camera center. So, all your light rays are going right through that center now so we can imagine a center through which is a pinhole through which your light rays are going and obviously that you can see that, if you even though from a point so many rays will kind of come out emanate, but only one ray goes in which is also the reason why you tend to use a lens.

Because for a pinhole if you want enough light you have to wait for the long time because you do not all the rays and if you have a lens then you can actually bring them all and then focus them, but right now with actually a pinhole model what it means is we are looking at a scenario like this where I have a camera like this and the scene could be something it is a planar world, but now this could be anything for that matter this could be something like that (()) (02:22).

Fronto parallel is actually a special situation which we will talk about this will be like this so this is a fronto parallel these are all other cases. Now we are not assuming fronto parallel here we are just saying that it is just a planar world. So let me write down where was I so the goal is to relate images of a planar scene taken from two different views by a linear transformation which means you want a matrix to be able to relate by a linear transformation that is the goal.

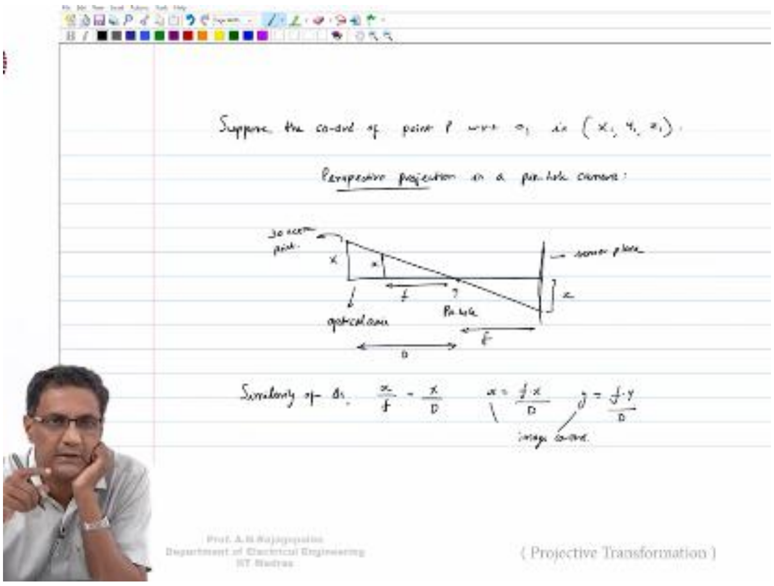
Now the way I am going to draw this diagram okay it is like this so I have let me say that my okay I am just drawing something let us say this is my optical axis, this is my camera plane so this is like xy and then that is my optical axis and normally the optical axis we tend to think of this as z axis we coincide the optical axis and we say z axis so this is your z and I will call this the optical axis all this just for simplicity.

We assume the optical axis to lie along the z or assume the z to be along the optical axis and imagine that here is a plane some plane is out here which we are imaging. We are trying to capture an image of this plane and let us say there is some point P sitting on that plane and let me draw a unit normal to this plane okay, so this is some n which is a unit normal and d is the perpendicular it is now d is from here to here.

So I mean how do I show that so from here to here that is my d and d really is the perpendicular distance, is the perpendicular distance of the plane from the camera center and then this let us call this as o1 here is where we are originally and this is the camera center. So all your light rays are going through, going to go through that and then assume that we actually move to some other location o2 there is some arbitrary y let us say some z and some x.

So I have gone, I have undergone some motion here of course it should involve translation also because I am not sitting right there if I was sitting right there and change the axis that would have been a pure rotation, but I am assuming that I have also rotated and I have also probably done some translation and arrived at o2. So, what we are asking is the image of this plane when seen from o1 when the camera was at o1., then the image of the same scene when the camera was at o2 assuming that the scene has not changed in between, how are these two images, images kind of related can they be related by a simple linear transformation that is what we are asking.

(Refer Slide Time: 06:29)



Now, let us say that suppose okay let us go on and suppose the coordinate of point P. Point P is some arbitrary point on the plane. Suppose the point P with respect to o1 is X1, what is that Y1, Z1 all in caps. So which means that this is your scene coordinate this is a 3D this is not image coordinate this is with respect to o1 that means in the world coordinate where is, so P is coordinate is like X1, Y1, Z1. Now at this point of time when we want to examine what will be the image of this 3D point on to the camera sensor when this ray some P hits the it goes through the camera center and then hits the sensor plane.

In order to understand that we should understand what is called a perspective projection model, a perspective projection in a pinhole camera. You must have heard of this, but then we will just see what is its relevance here and incidentally this particular model also holds even if we have a lens by the way which we will see later, so that way it is an important thing.

So what this means is that suppose we have let us say that this is our optical axis of the camera. Assume that our sensor plane is sitting somewhere here so this is the camera sensor plane here is where the image will get formed this is your optical axis of the camera and assume that the camera, assume that here is the pinhole through which the rays are going to come and assume that there is some point in 3D whose, suppose that this could be your point P for example, this guy let me draw it a little bigger. So this guy ray comes and hits the sensor plane here on the image.

Now in reality of course we should have like x, y, but I am just doing it on a page so I can only show one of the coordinate so imagine that it is like this you will also have a y correspondingly, I am just going to show x. So from here to here right this is your x coordinate for this point and of course treating let us say this as your origin and from here to here because you have focal length f, that is the focal length of the camera.

And let us say that with respect to this as a center this guy is let us say X a capital X and from here to here let us say it was at a distance D from the center of the camera which is your aperture or the pinhole. So here is your 3D point okay this is your 3D scene point. So a point will of course map as a point, but if it is a lens you cannot be sure it can actually become a circle.

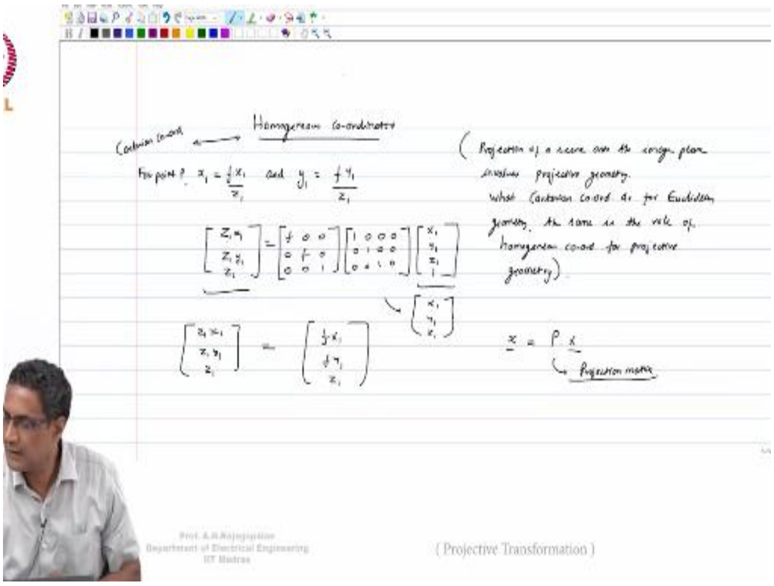
But because this is actually pinhole model so it is easy so a point will go as a point so this ray will travel and hit it here. Now, just to simplify things instead of looking at the inverted sort of image what we normally do is we just take the same F on the other side and then examine here, so here to here is still x whatever is here I have just taken it or this is other side just to ignore this, this is inversion which will, which you see here.

Now if you simply apply a similarity of triangles so what do you see, so we see that so from here to here is again F so we have kept it at, this was F this is also F this is x. So what we see is that x by f which is this is angle is the same as capital X by D or we see that x is f into fx by D and similarly you can show that y will then be fy by D and if this guy has a coordinate x, y, x, y, D actually this 3D world coordinate is like x, y, D and this will be some x, y which is your image coordinate.

So your small x and small y are the image coordinates, so where this guy is going to fall. x equal to this and then y is equal to this. So, what this also means is that the focal length also enters into the equation and so on. Now with this since we know that here is how this mapping will happen between the image and the scene coordinate now we can go and write down this kind of relation between the scene coordinate and the image coordinate.

But the unfortunately if you see here these two are kind of say entangled if you see this z or now in our case it is like x, y, z I call z as D here. In our case that point has a coordinate see this guy has a coordinate x, y, D for our P point in general it was some x1, y1, z1 so our equation would have been, the image coordinate would have been fx1 by z1 and y is equal to f y1 by z1 that is what we would have got for P image coordinate. Now we want to be able to write up an equation that will relate through a matrix the scene coordinate with the, with our image coordinate that is what we want to write.

(Refer Slide Time: 12:39)



So in order to write that what we, actually (()) (12:42) to what is called, what are called homogenous coordinate I will talk about what these things mean, we resort to what is called homogenous coordinates. See these homogenous coordinates just as, see this entire thing that we are doing which is really a projection of a scene what you are seeing the image is simply a projection of a scene on to the image plane involves, when you want to analyze this it involves what is called really a projective geometry.

And just as what Cartesian, let me write this, so what a Cartesian coordinate or what the Cartesian coordinates do for Euclidean geometry is the same is the role of homogenous coordinates for a projective geometry. Now our idea is not to go into the complete details because a projective geometry is a whole subject by itself.

But we would just like to kind of borrow important things from there. Now when you say homogenous coordinate what it actually means is, now the way we write it is I mean you can always go kind of back and forth between a homogenous coordinate and a Cartesian coordinate. If you have Cartesian coordinate in an n-dimensional space a homogenous coordinate simply an extension of the Cartesian coordinate to n plus 1 dimensional space, just adds on this extra sort of dimension

And why do you do that, (()) (14:53) and why this, I mean that is the reason why I wrote here that why you do this, why you require this homogenous coordinate, in fact I mean that is how it came actually when, let us say when these people wanted to explain a projective geometry and what you can see is that it allows you to go from a homogenous coordinate to a Cartesian coordinate.

And at the same time you can extend the Cartesian coordinate to a homogenous coordinate by simply adding one more dimension and it helps when you write the coordinate in the homogenous form then this matrix relationship that you are seeking will automatically happen how I will tell you. So, in this case we said that x1 which is the image coordinate x coordinate of that point is what f x1 by z1 for point P this I am writing for point P and then and y1 is f y1 by z1, this is what we have.

Now when you write in a homogenous form so what you will do is you will write this as z1 and you will write this as x1, y1, 1 or in other words you write this as z1 x1 z1 y1 okay now Z should be capital by the way okay this is not small it is looking small so Z1 x1 Z1 y1 Z1 is equal to now I will write this equation and then you will see that it makes sense f 0 0, 0 f 0, 0 0 1, it is a 3 cross 3 matrix and then I am going to add a 3 cross 4 matrix 0 1 0 0, 0 0 1 0.

And then I am going to write down my scene coordinate as x1 y1 z1 1. Just as my image coordinate now you see my image coordinate is what x1, y1 that is a Cartesian coordinate. Now from the homogenous I can go to the Cartesian by simply dividing the first two elements by the last. So z1 x1 by z1 is x1 z1 y1 by z1 is y1.

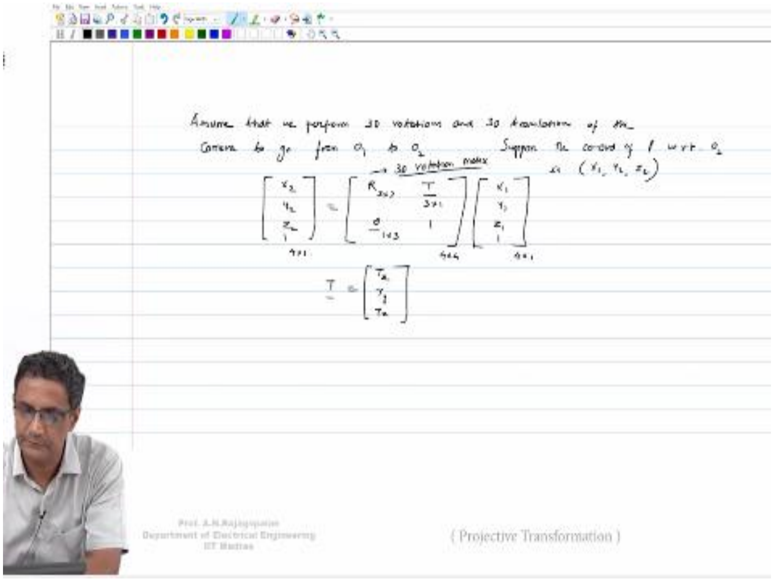
So I can go from here to the Cartesian by simply dividing the first two coordinates by the last entry and because I am in a homogenous world now my scene coordinate I am going to write as x1 y1 z1 1. I always append this additional 1 in order to be able to go from so which is what I mean by extending by let us say by one more dimension. So, in a simple form if you had let us say x, y and suppose you want to write a homogenous form it becomes x, y, 1.

You can also write it as zx, zy, z that is also fine because so the entire thing they are all like x, y, 1 because any alpha that you multiply will also help that is why this homogenous that word comes from that I mean x, y, 1 is the same as alpha x, alpha y, alpha, alpha-alpha should not be 0 of course so that is why this homogenous notion comes. And here, so this 3D scene is in a homogenous form.

The image pixels are in a homogenous form and you can clearly verify that if you multiply these so what you will get x1 what is that y1, z1 if you multiply these two guys and then if you multiply this with f what you will get you will get fx1 you get fy1 you will get z1 on the left you will have z1, x1 will get z1, y1 you will get z1 and you are (()) (18:21) saying this is equal to this and clearly x1 is fx1 by z1, y1 is fy1 by z1 and z1 is equal to z1.

So it is all fine, but then the advantage that you now have is you are able to write a scene coordinate. So in a way this is also represented as the image coordinate being represented as simply a P a projection matrix, P is not that point P, this is simply a project matrix multiplying the scene point. If you written in homogenous form you can write like that, you can relate a scene coordinate with respect to the image coordinate in a homogenous form where this is called really a projection matrix. This is exactly the reason why we resort to what are called homogenous coordinates.

(Refer Slide Time: 19:20)



Now assume that we actually assume that we perform 3D rotations and 3D translations and of course note that these rotations could be in any arbitrary sequence and the rotations do not compose like you know rx compose with ry compose with rz is not the same as rx compose with rz compose with ry and so on. Assume that we perform 3D rotations and 3D translations of the camera to go from o1 to o2.

So you have gone from o1 to o2 doing some kind of transformation and this again we can write, we will write in a homogenous form as x2, okay now and suppose let us say, suppose the coordinate of P with respect to o2 now because we have gone from you know we have moved the camera now with respect to o2 is x2, y2, z2 since we have moved now from o1 to 02. So now what we can do is right we can relate x2, y2, z2 1 which is actually writing the world, the scene coordinate in a homogenous form.

This again write as a 3 cross 3 matrix R and note that this 3 cross, see until now what you saw was simply R which is actually a 2 cross 2 that you saw only the cos theta, sin theta minus sin theta cos theta, but now when you actually talk about a complete camera motion the R in general we will write it as a kind of 3 cross 3 matrix and we will also see when will that, when will this R reduce to that and so on the one that you saw which is the in plane rotation and so on.

So this R will be a general 3D sort of a rotation matrix. So this is the 3D rotation matrix which means and it also involves the complete sort of a composition whatever order in which these rotations would have occurred, I could have occurred all that is encapsulated in this R now and then you have a translation vector T now I am going from the image coordinate to actually the way the camera moved now.

See there when I wrote small t xt and all I said that is on the image plane that is the image coordinate motion. Now this T is a camera motion now. This T is because you are coming from o1 you are bringing the camera to o2 by going through some translation in the world space and going through some rotation in the world space and reaching o2. So this is camera translation this 3D rotation matrix of the camera.

And then you will have a 0 vector which is like 1 cross 3 and you will have a scalar this T is of course 3 cross 1. So what this means is that your camera is actually undergone a translation Tx, Ty, Tz. I told you know in a general case camera can translate along x, y and z directions and this R will be a composition of these rotations about x axis, y axis, z axis in whatever order and there are ways to interpret this arc which we will see later.

There are, this is a very elegant ways to interpret arc and this is what we have and then this acting on x1 y1 z1 1. This is how we write if we were to write in a homogenous form. The motion in going from o2 to o1 I mean what is its effect on the point P how does it coordinate change. You can actually check that simply x2, y2, z2 is equal to Rx plus T that is what it will turn out to be if you simply solve this. So in this case this will be a 4 cross 4 this is 4 cross 1 and this is of course 4 cross 1 that is how it is.