Image Signal Processing
Professor A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture - 12
Homography - Special Cases

And the idea is such that we would like to go back and know. So what we will do is we, next slide, what we will do is, we will actually look at how those things that we have already seen in-plane rotational, how do they turn out to be special cases of this. Because this homography is supposed to imbibe everything. You can ideally have whatever you want, you can have a 3D rotation that whichever, whatever normal you want to choose you can choose. So the plane can be inclined.

I never said the plane is fronto parallel. It is some n, it could be at whatever inclination it is. But then the fundamental fact is that it is still, have a still a planar world. There is still a planar world now what you are looking at. So, what this ideally means is that if you had an, if you had a scene that consisted of multiple planes, then there is no single homography that will actually relate one image to another.

Which is why a 3D scene, for example, in this room if you take, now there is some plane in the front and there is some plane in the back and so on. So you cannot, if I take the two views of the scene by translating a camera, there is no single homography that will actually relate the two. It has to be a planar world. So, but then you might ask why is it at all useful. Who wants to then use such a homography? The idea is that the farther you go, see many a time either you actually have a planar scene or imagine that you are sitting, you are flying from somewhere and then you are taking an image of let us say the land. Then normally what happens is we tend to ignore the changes in-depth given that we are sufficiently far away.

So even if there is a mound, let us say, there is a pit or there is a mound, then you will say relative to from where I am that difference should not matter so much, and then, in reality, it actually, it will actually work very well because those differences, which should ideally lead to some issues will become lesser and lesser as you go farther and farther.

So, either you already have a scene that is nicely planar, in which case you can, of course, straightaway go ahead and use it. In fact, there are also many situations when you have to see planar images, it is not like you do not have planar scenes. In fact, the whole world is supposed to be what is called a Manhattan, Manhattan World, that is what they say. What that means is the world is made up of mostly the buildings and made up of vertical and horizontal planes primarily, Manhattan World; that is what it is called.

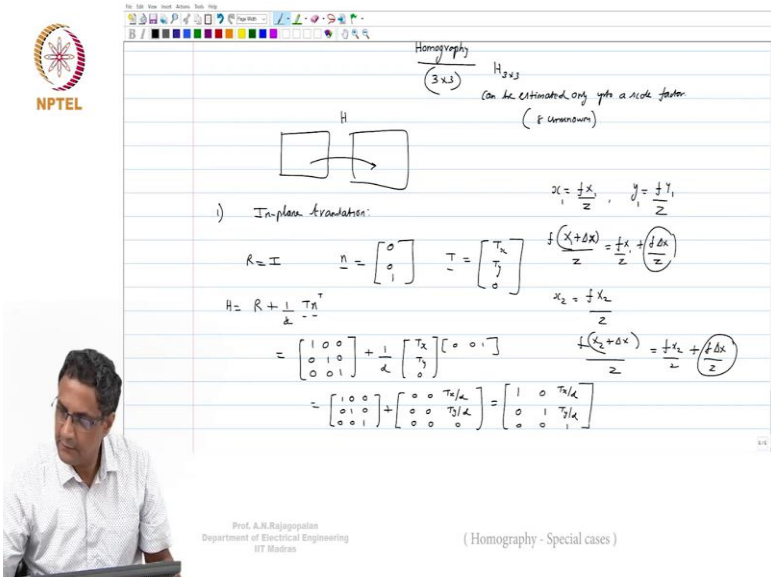
Which then means that, means that you could take let us say, one image of this plane, one image from that plane, and then you could actually relate the two. But if you take one common view, then you cannot, because no, there is not a single plane that you can use the model. So that way, even if you bifurcate a 3D world into different planes and on every plane, you can work with this kind of homography.

But ideally, when you say, you have a homography, you really means that you should be able to go from, go from one view to another with a single transformation. Now there is a special case where even if the world is 3D, you can still do this that is the only time you can do that is when you simply rotate a camera, you do not translate. Because the parallax, parallax comes only when you actually move, that is why, that is why our eyes are such that, I think I told you before, you have eyes that are separated. You do not have one eye over the other, wherein you could rotate the other or something and get another view. If you do that you will not get any information about depth. Because that parallax has to be there.

So, in fact, so we will see with some examples what this actually means, but at this point of time, so what you have to realize is that a planar world, you can very nicely map different views of a planar world through a single homography. So whenever any approximation of a surface can be, whenever you can approximate any surface as approximately planar, homography is the way to go.

But again, this will still, we still have not said how do we compute this homography. We have not talked about all that, just assuming that somehow we have the homography. In reality, we need to compute this, then we will have to robustly compute it and so on. So that we will do along the way, but for the time being just assume that, for the time being, because you just have to understand what this homography is going to do for us.

(Refer Slide Time: 04:14)



Now, let us come back to that, the special cases. What was the first one? Let us say we take in-plane translation. See each one of these will tell you something, something interesting. In-plane translation, so what does that mean? In-plane translation actually means that you have a, okay, no. See now, so at that time, we wrote down a very simple equation. We simply said xt yt is equal to xs ys plus n Tx Ty. We did not even bother of what, why are we writing the Tx Ty where is that coming from, what camera motion will possibly, we did not recall or bother to put any of that, great.

Now, now we need to worry about what will that be. So in this case, in this case, the way you imagine in-plane translation is that because you want all the image points to move by the same amount, Tx Ty is the same for all pixels, they do not change, so it has be a global shift. So when you want such a global shift, then it means that this camera that this plane has to be fronto parallel.

Because if you look at, how do you understand that? So if you take the perspective, sort of a projection model, what did you have? You had x equal fx by let us say, some z; and then y is equal to fy by z. So if you actually move the plane such that x becomes x plus delta x, then you delta, capital X. Then you will actually move to a new point which is fx by z plus f delta x by z; and similarly, we will get here which is x plus delta x by z.

But then this is not only true for this point, this is also true for every other point. This is x1 y1, let us say, which has x1, y1, z1 as its coordinates. So let me just put, but then delta x is the same because I am just moving the scene by the delta x. So delta x and delta y will be what they are. So we have x of fx1 and then delta x by z1 z1, here z1.

Now take another point x2, which is at some x2 y2, but z, because it is a fronto parallel plane. The z cannot change. So fx2 by z, so there is, so there is nothing like z1, so I will remove this z1, they are all at z. Because it is a fronto parallel plane, every guy is sitting at z. So now this is also, I will remove this. So it all becomes z. So now you can see that fx2 by z and then x2 moves by a delta x because the whole plane is moving by delta x. So if you see this will again turn out to be fx2 by z plus f delta x by z. So this motion, okay, there should be an f here, by the way. Yeah. There is this f, so f delta x by z.

So you see that they would all move by the same amount if it was a fronto parallel plane. Imagine if it was inclined, then this is no longer true because a point in the front which it will have a z that is smaller, a guy that is giving because his plane is inclined from some other point will have a different z, and therefore, you cannot have this equation that you say Tx and Ty is going to be a constant for all the, that is why I said, we have to go back and sort of imagine what situation will lead to, will even lead to a Tx Ty that is constant.

So in this case, what this means is that you have an n unit normal, because this is a unit normal of the plane, and because this plane is fronto parallel, that means it is aligned with respect to the optical axis with the camera. This is the optical axis of the camera and this plane sitting right orthogonal to it and so n will be 0 0 1.

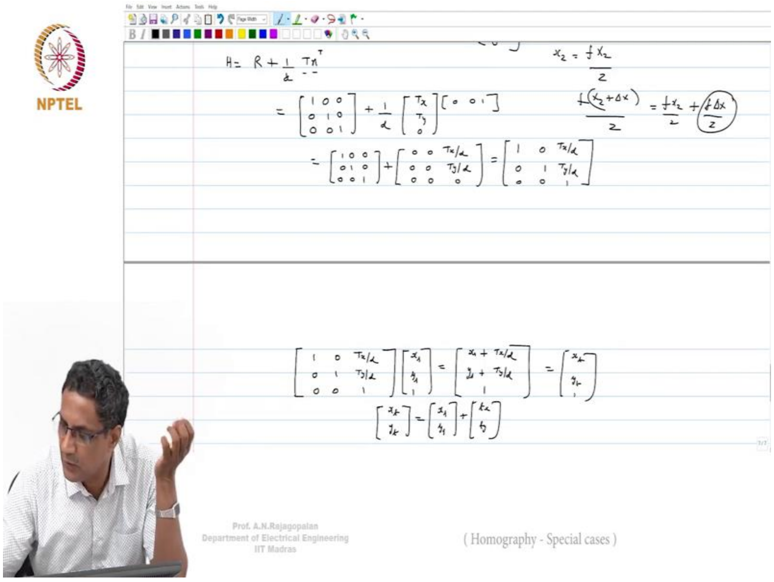
And since in this case, rotation, we are not doing any kind of rotation, so R is simply identity. It is simply an in-plane translation. And then, as far as your a goes, the a is that, well, anyway, a has no role because R is anyway, an identity. What about your T? See, because in order to complete the homography you need R, T, and transpose; three, and then d of course, d is something that we kind of do not know.

So, T is let us say that we are allowing in-plane translation that means, we can have Tx Ty that means this plane can go like that, it cannot come forward. Here is my camera, I can go like this, I can go like that, on the plane I can go but I cannot come forward and backward because that will be a Tz motion then. In-plane means in plane, you stay inside the plane. So Tx, Ty, and Tz should be 0 for the camera, this is for the camera.

Now, if you go back and do R plus 1 by d Tn transpose, here this is what we said is your homography, what will you get? R will be 1 0 0, 0 1 0, 0 0 1 plus 1 by d, where d is the amount by which this plane is away from the camera center. So d, which is something that we do not know, we do not have access to that information, because we can only see the image. So 1 by d and then T is what? Tx Ty 0 multiplying 0 0 1, which is the plane normal because it is fronto parallel.

So 1 0 0, 0 1 0, 0 0 1 plus what we will get here? Tx, sorry. The first guy is 0. So 0 0 Tx by d, 0 0 Ty by d, 0 0 0; add the two up, we will get 1 0 Tx by d, 0 1 Ty by d, 0 0 1. That is your homography matrix now.

(Refer Slide Time: 10:00)



And if you now multiply, what did you have? If you now multiply it 1 0 Tx by d, 0 1 Ty by d, 0 0 1 with a homogeneous source coordinate, let us say we multiply with some source image coordinate and we want to go a target, what will you get? What will this be? xs plus Tx by d, ys plus Ty by d, and 1. Now, this is xt, yt, 1; and this is what we had written as xt, yt is equal to xs ys plus Tx Ty. If you go back, this is how we started. Now, this Tx is Tx by d, Ty is Ty by d, and as we can see Tx, Ty is not the same as the camera motion.

So, what this actually means is that if you have, which is also the reason why just looking at the image coordinate you can never tell what was the camera motion, unless you know this depth unless you know what is the small d, you can never tell. For example, I can get the same Tx by having a small, if my d is small, then my Tx can be smaller. But then if I am very far away, then to get the same image motion I need to move a lot more, my Tx will not be higher.

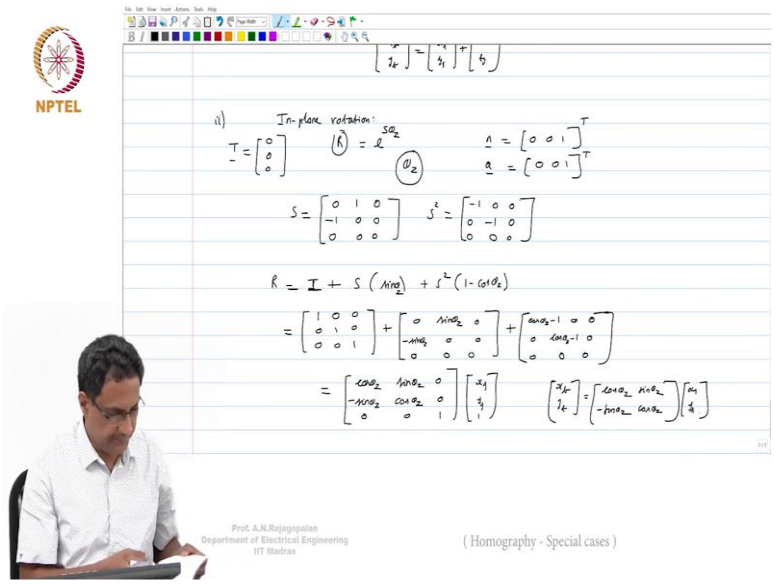
So, this sort of notion of depth is actually coming from there and you can see that. When you are closer, the image motion is less; when you are farther off, sorry, when you are closer, the image motion is higher because the d is small for the same camera motion. Imagine that I moved, let us say at 1 centimeter that is my camera motion, but if my d is small, on the image plane it moves like 1 by d, when d is small.

But if it is very far away and I move let us say 1 centimeter, on the image plane I will move much less because my d is now high, my motion, camera motion is still the same. It is only 1 centimeter. That is why unless you have a sensor on the camera that tells you the camera motion precisely, you cannot tell how much a camera moved because you do not know how far away is the camera when you image the scene.

That is why when we rotate, we simply said Tx, Ty. But now I hope you are able to appreciate the fact that you cannot simply write Tx and Ty without all that, behind that there is all this interpretation about what is the scene, which way is the camera moving, what would be the relative motion between the camera and the scene will lead to a global shift. And this is called actually a global shift. Because all pixels are moving by say, Tx and Ty.

So this is one simple case. Let us go on to the next one which will be a rotation, which is in-plane rotation.

(Refer Slide Time: 12:58)



Two; let us at least do in-plane rotation today. In-plane rotation, what is it like? So in-plane rotation means what? By the way, okay. Like we will also do that. So, in-plane rotation means that you have a scene and then in-plane rotation is about the optical axis you are simply rotating, that is z. About the z-axis, that is in-plane rotation. Out of plane would mean that you would actually rotate about x or you would rotate about y. That means you can do something like that or you can do something like this, but z is this. So you can only do this. That is in-plane, you are not going out of the plane again.

So when you have an in-plane sort of a rotation, then your R, you need to find out R which we said is e power s theta. And let us say that we are actually rotating by some theta. Let us call that theta z. This would indicate that this is a rotation about z-axis. Theta z, instead of theta I will write this theta z.

And again, now in order to get that kind of relationship which we had cos theta sin theta minus sin theta cos theta, you can actually verify that your n should still be a fronto parallel plane, that means it should be 0 0 1. And because you are rotating about z-axis; see, these two are completely independent, I can actually rotate about the z-axis, but then my plane normal could be whatever it is.

But in order to get that in-plane rotation, when you want to see what happens to that particular equation that we saw, which was a very nice form cos theta sin theta minus sin theta cos theta, if you want to imagine what the scene and what camera will take you there, then the one that will you take you there is a is equal to 0, 0 0 1. Because it is simply a rotation about the z-axis. I have it here, so I just wanted to verify. Now a is 0, of all these, are transpose. So these are all column vectors m is column, a is a column vector. So, please verify.

S is of course easy to see because you had 0 and then I know you had, what is it? az 1 0, minus 1 0 0, 0 0 0. This goes by that 0, az, ax, and so on, minus az, so this is a skew-symmetric matrix. Verify that a square, if you multiply this with itself, I am not going to multiply it all here. Let us just show that, I will leave it to you to show that this minus 1 0 0, 0 minus 1 0, and 0 0 0. I will just leave it to you to check this.

Now if you can verify that these are true, then we know that your R is, and this is all coming from, and anyway, by the way, because it is in-plane rotation, what about your T? Your T is simply 0 0 0 because this is just in-plane rotation, there is no translation. Therefore, irrespective of this n, you will simply have Tn transpose, so that whole term will drop out, all that all that remains is R. So R is I plus what was that S into sin theta plus s square into 1 minus cos theta.

So is this case, we will write this as theta z because we said this is actually a rotation about the z-axis. So what does this mean? So we will have I which is 1 0 0, 0 1 0, 0 0 1 plus sin theta z if you do, so 0 sin theta z, 0 minus sin theta z, 0 0 0 plus s square is this guy, 1 minus cos theta, so this would become cos theta z minus 1 0 0, 0 then again cos theta z minus 1, 0 0 0.

Now, add all the three up, then you will get, 1 minus 1 is 0, so you will get cos theta z, then sin theta z, 0; then minus sin theta z because of this addition then 1 plus 0 plus that is cos theta z, then 0; we will have 0 0 1. Multiply this with xs ys 1, and then you will be able to clearly see that this will come out to be of the form, I will not work it out here, you can verify that.

This will be xt yt is equal to, again the, and also we should also, I can show you in the next class that it does not always mean that the last entry after you multiply by it h will always turn out to be 1. These are special cases where it is turning out to be 1. There can be cases when after you multiply h with this xs ys 1; see as far as the source coordinate is concerned, you always take it as xs ys 1 because that is your actual coordinate.

But when you multiply that by h, it does not always mean that the last entry will turn out to be 1, you can have some number there. And then to get the actual target coordinate, target image coordinate, you should scale the first two guys by the third one. I will show you one example like that, it turns out that the first two cases are simple and then you are getting an entry 1 for xt yt. At the end, you are getting xt yt 1, but it may not always happen, you can get something else as a third entry.

So this you can easily show is the one that you saw earlier. Cos theta z sin theta z minus sin theta z cos theta z. So this would be a clockwise rotation about the z-axis, multiplies xs ys. So the following class, what we will do? We will also see how we can do shear; shear, probably I will just leave to you as an exercise maybe I should make you think. Scaling, for example, something recently I will show here so that we get a proper visualization of how these things happen.