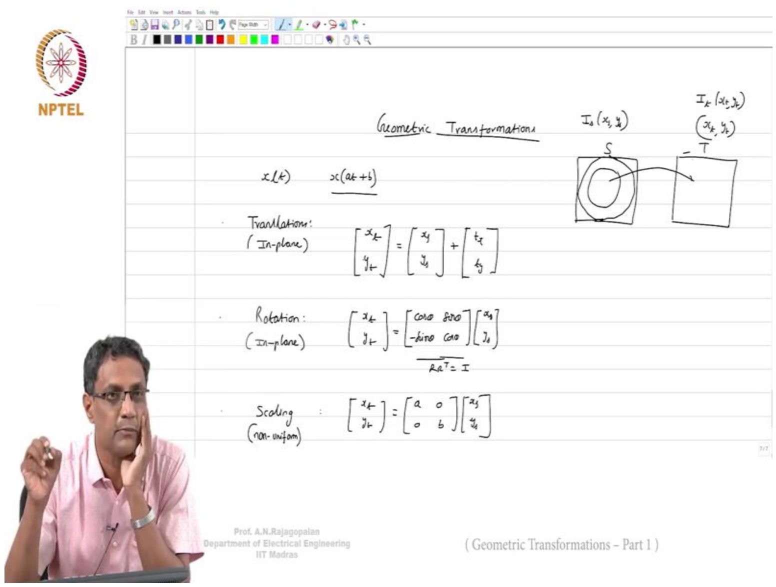
Image Signal Processing

Professor. A.N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras

Lecture No: 6

Geometric Transformations

[Refer Slide Time: 0:21]



Okay, now the next thing right that I want to what I want to talk about which will be a slight digression, but then a then a very, very important notion is one that is is one, which is of Geometric geometric transformations of of images. What this really means is that see, for example if you add a 1d signal right, you talked about suppose, suppose I gave you I gave you a 1d signal x of t, a then you know when you talk about the transformations, typically what kind of transformations we can do on the 1d signal?

You might do something like x of a t plus b, right? You could either scale, no you could squeeze or stretch the signal we can translate a signal you can do both. Then typically the most general cases x of a t plus b, this is what you normally do with 1d.

Now, if you had an image right and suppose you wanted to do something like a similar, now you might ask why is why is this why would somebody want to change an imagine to something else. For example, see as far as images are concerned, the kind the kind of transformations and this transformation is a very nice hierarchy okay. It is not, it is not, it is not as simple as we as we sometimes think, that is why that is why I would like to spend some time on this.

So, it is not, it is not, no it is not as straightforward as we as we assume it to be. So, one of the simplest things that you would have seen is what is called what is called translation right of an translation of an image. Translations what do we really mean by that? So, we mean, that if you are if you are source coordinates are x. So, it is like saying that I have an image call that call that a source image, I have another image which I want to which I want to which is which I would like to generate, which is really a target image, given the source image I want to be able to build a target image.

Now, under what circumstances do you want to do for example one simple example I can give you is this is a zoom effect right, all your cameras carry what is called there is something called a soft zoom and there is something called an optical zoom right, you would have seen in a camera those that carry an optical zoom are, they are more expensive.

Because an optical zoom comes with comes with a lens whose focal length can be changed, when you say it is a soft zoom and all that it is zoom it, that it is using an existing picture, that it has already taken and it is trying to trying to trying to let us say that is zoom in. What that means is let us say that, let us say that, if I had somebody say right occupying this entire source and suppose I said that you know what do you know, I would like to see the knows more and more detail or whatever.

Then it means that a part of this image right is going to is going to fill this entire thing. It is going to fill this target correct? That means your face will get cropped of course, you would not see the whole thing here. There is a whole idea of zooming in. So, a part of this will occupy those whole grid.

Now, as something as simple as this we all like to do, correct even in a camera you sometimes ask, where it is zoom in, a little more. But if you are asking for a zoom, that is that is soft zoom, then what it is doing is exactly one a what what we will be doing here that is the soft zoom. The other it is an optical zoom which I will actually talk about, zoom is just one such case you might want to do other things.

So, in a translation what do you do? If you have target coordinates, these are these are the special coordinates by the way, this is not the intensity, the intensity will be like I t, I t at x t comma y t, will tell you what is the intensity in the target image at the location x t comma y t, this is what you want to know given that you have a source image, where in all intensities are available at x s comma y s, that x s, y s, will go all over the grid.

These are your special coordinates x s and y s, x t and y t are your special coordinates, one corresponding to source, another corresponding to target. And I s is gives guys gives you the intensity in the source image at that location x s comma y s. Now, what you are asking is a transformation when you do a simple thing x t, y t, what you are asking is (())(3:53) x s, y s.

Suppose, suppose you wanted to translate an image you will say plus some tx ty. Now, there interesting thing when you when you write for 1d signals of, it really seems like a straightforward thing right it is right you know x of x of t plus b,(())(4:11) Now, in as far as images are concerned, even if you ask under what situations right, can I ask you if I take a camera and suppose I translate.

Under what situations do you think that you will get a shift that is actually a global shift. See I have write a one thing that you should notice is this tx and ty is independent of x s and y s. It is going to be applied on all x s and y s, right? Whether I take 0 comma 0, or whether I take somewhere in the middle of the image, everywhere I am going to, I am not employ the same tx ty, which is which is what is also true with respect to 1 d when I do x t plus (())(4:44) do I change b for every t, no right. I just translate the whole thing globally. That is what that is what we call as a global shift.

But as far as images are concerned there can be what are called local shifts and there can be what is called a global shift. And the whole idea you know when we when we when we try it and you try to talk about geometric transformations is that you would like to go from one image to another, even if there are local shifts you do not want to worry about them. But as long as there is one sort of a linear transformation that takes you from one to the other, it is like saying that you know.

So, let us just, just to set that tone, what it means is that you know if on a certain day if I if I want to take an image of the IIT campus, like they do, now even know somebody wanted to wanted to IIT Madras to prove, that that you know we have actually grown enough number of trees on campus. How do you going to prove that? Then you should have add some picture that was taken let us say 30, 40 years ago of the campus. And now you take one more image and now you want to compare the two.

Now, it is impossible that you know you go exactly to the same location where 40 years ago it was taken. You will take from somewhere nearby, you will know that okay right here is where the buildings are and roughly you will take from there. Now, you want to be able to match these two images right, you cannot directly subtract, because there are not even aligned. So, what you will have to do? So, we learn to do some kind of a rotation, do some something you have to do.

So, that so that there is one linear transformation, that brings you from one to the other. The whole idea in hinges on being able to write that one linear transformation. If there is more than one, we just need it, then it means that you know it is no longer the kind of thing that you would like.

So, so which is why which is why these translations rotations and scaling that we that you are familiar with is the most simplest of all like when there is a when there is a when there is a global translation all the pixels just the shift by a by a shift by ts and ty, then it means that I know that you know I knew that the camera was was right there, looking at the ground below and I simply move txt y.

But this tx ty which I am talking about is on the is on the image plane, the camera could have seemed moved by a larger amount. You go further for the same tx ty you will have to do more, you have to move more, as you go further and further this is like you sit inside a train. And then when you look outside as a train moves, what these things moves the fastest visually to your eye is the pole that is next to the train that it looks like it is fleeting, where is the hill right yeah no way but it looks like it is hardly moving right.

This is because this because which is closer will tend to move faster, the same way when the camera is let say close to the ground. And if you move on the image fully you will see a lot of movement, when you go higher if you want to say see the same movement on the image plane you will have to move much more. But the whole idea is you could move in a manner that tx and ty remains constant for all the for all the pixels, which is actually very, very nice and elegant thing. But then it may not happen always.

Now, the, the next thing is next thing is actually the rotation and all this, this in this is all actually called an in-plane translation in the following class I will tell you what I mean by that. Rotation I will specify that this is an in-plane rotation, please remember that I that a camera can actually undergo a 60 motion, no it does not have to just rotate like this, it does not have to translate like just like this.

It can go all the way like rx ry rz, that means about the 3 axis you can rotate and then about these 3 axis you can translate. You can go tx, you can gp ty, and you can also go tz, come back and forth. So, you can go like this, you can go like this and similarly rotation about each of the axis, addition be able to rotate an it may like this, which is in plane, out of plane is like this, when you do like this or for example when you do like this, these are all out of plane rotations.

So, but the ones that I am writing right now which you are most familiar with is the simplest one. Then I am going to start with, but then and then we will come up the and then we will go up the kind of hierarchy. So, the next case is like xt yt is equal to a rotation matrix, let us say it is simply cos theta, sin theta minus sin theta cos theta xs ys, this is this rotation you only all know that a rotation matrix is orthogonal.

So, if this is R then Rr transpose is simply identity, every, every, every rotation matrix is an orthogonal matrix. This is a very, very specific case by the way, this corresponds to in plane rotation. But even if you have a complete 3d rotation, where you have a sequence of rotations something going about X, something going along Y, something going along Z. If you kind of if you were to concatenate all of them together do a composition that rotation matrix will also be orthogonal any any rotation matric is orthogonal.

Then there is scaling in general you can have what is called a what is called a inhomogeneous scaling, which is like XT or a non-uniform scaling, which will look like xt yt is equal to let us say some a 0, 0 b. And then you have xs ys or other words x is going to ax s yt I s going to be b ys. This is this kind of non-uniform because a and b are not are not the same. But normally when you zoom and all these two are these two are identical.

Now, the whole point is ready now in the following class, what we would like to do is we would like to see how does one generate is, if you think that you can go from the source to target just like that you are wrong. You will end up with holes in the image which I will talk about.

So, what is normally done is you start from the target grid and come back, if you think that you know if you think that we will operate this equation right here put xs ys and try to get xt yt and simply assign intensities you will end up. There is no there is no right and you don't have a guarantee that every, every location here you will be able to you will able to hit that location, you might end up with holes in the target, these are things that we will talk about in the next class.