Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-56

Now, let me start with the 2D geometric transformations. So, the way to kind of look at this is by which we mean that we want to have one operation right that we can apply right on let us say image i 1 in order to get i 2. What we are looking at is a single operation which we can apply on i 1 to get i 2. And that operation could involve something like you know that could be something like you know for all the guys here that are sitting here. So, each pixel moves by the same amount. For example, if I have I say that you know there is a translation of T x pixels and you know T y pixels between the 2.

It actually means that means that if I were to shift every one of these pixels in i 1 by T x and T y right it will you know if I shift this guy by T x and T y I will be able to reach the other reach the same pixel in the other image. Now, I mean I will not be going to say talking about when you can do this and so on okay. There is already a certain set of see what happens is right there is a constraint on when it will happen okay. It is not always true that such a thing will happen okay.

Again, we will not go into when it can happen. It depends a lot on the even if it is a planar scene right. There is something to do with the with the normal of the scene okay that actually has a role to play. Then the other thing that have a role to play is how did how did you know how did your camera move okay. Did it purely translate or did you do something else okay.

So, all of this is a function of so so right even though even though the way we will see it is you know well I can express one in terms of the other. But then when you can express it and all right it is not it is not something that I would answer okay in this in this course that I typically do when I when I teach you know this in image processing. So, there I there I sort of you know go into this go into this details of when you get that and so on okay. So, in this in this right if I do that then it will take a long time. So, we will skip that.

So, we will start with what is called what is called you know translation right which is the which is the simplest of things. So yeah because what can happen is and again it if you have if you have you know a rotation for example then what it means is you know if you are actually you know if you are doing a rotation at about some point then that point remains remains exactly at the same place in the you know second location it does not even move. Whereas there is something which is actually close to that will move less something that is farther off from the center will move more and so on right. So, this so this kind of a pixel motion okay that happens okay right between the images it need not be a constant right that is what that is what that is what I am saying. So, when you it is a translation it is a constant but then it does not always mean that right every time when you do this there is a there is going to be a constant motion or something.

You can have each pixel moving moving right according to a certain law right depending upon what you are applying. But then the final notion is that there is one law there is one uniform law that you can apply and that uniform law can cause a constant motion it can cause a motion that varies across image but then there is still one law okay. Whereas if you have a 3D scene there is no single law okay. If it is a planar scene okay yeah let me write that. So, we are looking at a planar scene okay.

It no so a planar scene it need not be fronto parallel or anything by a planar scene I mean it can be oriented at any angle. See for example I have a camera when you say fronto parallel right what it means is you need a plane exactly like that okay that is a fronto parallel plane. But then the scene itself it could be like this it could be like that it could be anything as long as it is planar okay. If I take if I if I could have taken image from here of the scene and then if I go there and take right one more scene wherever I am right. Typically if you have a complete 6D motion then it means that I can I can go from here to anywhere I want right I just have to do an R and D okay right in my in my 3D world.

So, if I go there and then and then I take a second image I can always actually relate them I can always relate them without even knowing right anything about the camera okay right that is that is interesting part of this 2D geometry transformation. If you go deeper right then you will start to start to ask right things like what was the camera camera matrix and then what was the what was the actual motion that the camera underwent what is the normal of the plane of the scene right that and all you can ask but not in this course right. But here we will just be interested in knowing can I actually relate the two and what does it take to actually relate the two. So that if I apply the transformation on the first image I get the second or if I do the inverse or whatever then if I apply it on the second then I can see the first right that is that is all that is all that is all we are interested in. So, translation right so we will say that that if I if I want actually a target and all of this right is that is that is that the as at the you know this one the spatial grid.

All these operations are are actually happening on the on the grid on your spatial 2D grid right so it is like it is like a purely you know geometric operation right there is no there is no photometric operation going on here we are not increasing intensities we are not we are not right decreasing intensity we are not doing anything right we are just having the same scene and we are trying to see how will this scene look like right if I were to look at it from here look at it from there strictly speaking the physics of the scene should all come into come into play but then all that we are ignoring right. Okay so this is the simplest form but the simplest form itself is enough for a for a lot of lot of situations. So what we see is if I want if I want a target coordinate right so for example this is the spatial coordinates in my in my say target image the spatial coordinates in in my in my say target image target if you are actually synthesizing or you can think about x t y t as your i2 and then write I mean I write x and y1 can be your i1 and therefore what you can think about is you can just write this as xs ys + translation tx ty okay and such a translation is called is called a global translation okay and again it is not true that it will happen every time under certain situations it will happen okay I am not here to talk about when it will happen and so on okay. So if such a thing is valid okay and if such a law holds then we can actually write or relate these two images by by simply knowing is here right you know tx and ty and of course on how do you know is it tx and ty because tx and ty will then have to be found out by your if you go back to your feature correspondences you can find out your tx and ty okay I mean most general situation I will talk about and these are all special cases of that okay. Now this let us see right let us see if you have a situation like this right then what does it so what things remain intact right when you do so for example it simply means that I have an image like this and then and then it if you do a translation right ideally that is the size and everything should remain the same so it moves there right that is what that is what we mean by a simple translation.

So what so in this case what happens is angles and lengths are actually right they actually remain intact are preserved okay or they remain intact okay I mean everything is on the is on this image plane okay everything whatever I say is on the you know image plane. Sir could you please explain this question like first I am using camera to take a particular image and then on the camera plane I mean the camera and take camera plane. Yes yeah so for example if you had your image plane like this right I mean and oh okay all right so so it is like this right so okay yeah one more thing that maybe I should tell this you know see there is something called you know a pinhole model for a camera right you know pinhole model which is what we are following here what is actually means is the following okay what does actually means is that means is that you know if you have a see if you have a you know 3D scene right and then if you have a if you have a you know a pinhole through which the light rays right from let us say each 3D point is entering then behind the behind this behind this is actually a was what is called an image plane okay and from here to here right you may have you may have a distance whatever f or something and then from here to here right this could be your see depth d of that point. So what does what this means is right means is right that this point comes and actually hits the image plane here and actually instead of seeing an inverted image right what we do is we actually think of the image plane as being in the front by a distance f okay and what we mean is if I can get an image like this okay let us say you know so it is like saying that if I keep my camera here and I actually see this image then if I see the scene sorry if you see this scene then I get an image and suppose I translate this point now right that is my camera center if I translate it on my so it should be like it should be it should be like it should be you know it should be a parallel to this image plane okay so my so my motion should not go forward or backward okay. So for example right if you have this as your image plane then I should then if I translate on the image plane and I take one more image and under some condition right these 2 can be related by a simple Tx, Ty okay not always like I said it is not always true that if you do this right you will be able to see a constant motion of Tx and Ty but there are situations when it can happen okay so the image form actually the same will also work if you had a lens under certain assumptions okay whatever we say is actually for this one a pinhole but if you had a focused image you see the moment you get a blurred image and all right then the math changes a little bit but then otherwise right if you think of a lens even if you had a lens right instead of this right if you had a lens here right okay if you had a lens there instead of you know a pinhole you see I am not kind of right going into why you need a lens and all that I think most of you know that a lens is to lens is to right I mean you know take more light see for example if you have a pinhole right then even though this point emits so many rays only one ray can actually enter through that right but this model is actually is a sort of a convenient model in order to see analyze what happens when the images are formed and this is an age old model and it works in more in almost every case but in a real camera you do not have actually a pinhole because of because the pinhole will mean that you will have to wait for a long time right because only one ray is coming right which means that even to gather light I have to wait that long that is the reason why you have lenses because you know a lens can take all of these rays into it and then it can focus it right back at the same point if you have a lens it is a focused image right so lens can gather all of the light quickly and then focus the image for you if you do not have a pinhole and again you need a lens because if you just expand the size of the hole what will happen we just make this hole bigger what will you see what kind of a distortion will happen blurred exactly right so it is not true that I can gather more light to gather more light I will just you know increase the size of the size of the aperture right that does not work then you will have a blurring artifact so what you need is actually a lens okay so lens will then kind of will can will then kind of bring in focus all the light and give it to you as a focused image I mean if the image is not focused which can also happen that then that then goes into a different relp we will not we will not we will not kind of enter into that okay that we do when in an image processing course you get to stay away from that that is a that is that is a lot more involved how depth is related and all that will come into play but here right since I said that right depth okay in this 2D geometry transformation we do not even want to say refer to depth simply say a planar scene okay we do not even want to you know talk about depth is that okay does that answer your question or is it still okay so yeah so always right think of and of course you know and of course you cannot have the image plane physically in the front right no I mean of course you cannot have it physically this is just a model actually it is always at the back right so you have a pinhole and then at the back should be or should be an image plane but then because you will see an inverted image just to make the math cleaner right we just assume that the image plane is in the front but but physically you cannot keep the image plane in the front because because the whole point of actually getting the ray through a pinhole is lost okay so there is only a model there is only a convenience thing right so you keep the think of the image as being in the front by f instead of being in the back end because then you will see an inverted image okay so just to just to handle the invertibility and all just say it is in the front okay now the point is this right now where was I huh so so then right in such a case what we mean is you know angles means all these is internal angles these internal angles and all do not change and similarly the I mean lengths and all right do not change whatever length you have here the same length you see after you have done the translation then you can have actually a rotation and this and whatever we are saying is all actually supposed to be in plane which means that if you're rotating then you're rotating like in plane again again it if you if you kind of think of the so the camera center again if you are rotating again it you're kind of staying within the image plane okay you're not you're not gonna sit rotating uh what do you say I mean so for example you're not you're not kind of doing like you're not doing like this right you are only doing like this okay you're staying in the image plane and actually rotating and normally that the way we kind of think of the X by Z axis is like this right the optical axis and the Z axis typically we align them.

So, for example, so if you are going to think of this image center right. So, this image center will be such that right I will have my x and y here and then my z will be along the optical axis. That is the easiest way to then you know. So, when you talk about depth right if you simply look at the z value that will kind of tell you the depth. You can have a different transformation it does not matter I mean then you just have to do the appropriate r and t to again align it back.

And to begin with it is always easy to think of the z axis being aligned with the, aligned with the optical axis. So, the optical axis is this guy. This is the optical axis by the way which is through the center of the through this one the pinhole and then through the center of the center of this image plane. That is the optical center that is the optical axis. Rotation right will simply again this is an in plane rotation right.

So, it is easy what you can do is you can have x t you can have y t and then you can say that this is like cos θ sin θ yeah I mean right depending upon clockwise or whatever anti clockwise minus sin θ cos θ then again right you can have source x s y s. This will be this will be this will be a rotation right where this that is θ right you can tell what this θ is. And then one more thing is a small aside is that right when you are when you are doing this you want an image finally right. This is only this is only you can see saying what will happen on the grid and it is only saying that if you want to go to if you right if you know x s and y s and where will x t be right if I tell you what my c θ is. But in reality right what we actually do is do is right we do not we do not get of we do not we do not actually we do not actually go go like this.

We do not take for example right there is something called called you know a source to target versus source to target versus a target to source mapping. So, this is the equation at which which which right I wrote target to source mapping. So, the equation that I wrote right I mean you know it looks like you pick up the source coordinate and then you see right where it goes. So, it is like saying that if I had an image which is the source image and then if I want actually a target image. Then according to that equation it looks like I pick an x s y s here which is on this grid find out find out right where it goes here that is like.

So, I pick an x s y s here and then and then I see right and then I see where it goes it goes to some x t y t and therefore, right you would say that once I reach here right I might be able to say assign whatever is this intensity I can assign to this point correct right which is what you would do because finally, you need to see an image right you have to assign intensities now. But if the problem is doing something like this creates what is called a situation of holes in the sense that in the sense that if you do this right there is no sort of a guarantee that that you will you will you will actually walk through or walk through all of them all these pixels you may actually end up skipping some of them. I mean I am not I am not going into this technical details of that, but if you want you can just you can just read it up source to target right there is no guarantee that see because of point is what every pixel here should be assigned an intensity right every pixel in the target image should get an intensity there is no there is no guarantee if you go from source to target that you will actually visit every one of them there is no guarantee what is called it is called a phenomenon of you know phenomenon of getting holes in your final target hole does not mean you have zero intensity hole simply means that you did not even you did not even visit that this is that pixel right it got skipped I would have talked about it, but then it is again digression and it takes a lot of just just read it up anywhere source to target why you do not do. So what we normally do is what is called actually a target to source what that means is you know because because I want to fill in I t. So, I can always pick I always go like on this integer grid and again right and again see remember that this t x t y everything can all be fractional it does not mean that you have to move by only one pixel or minus one pixel you can do like minus 2.

75 it can happen know when you move a camera there is no guarantee that you will get integer translations and all right. So, you need to be able to account for fractional shifts and all of them right. So, all that all that this target to source mapping right will actually take care of. So, all that you do is so for example, if you have an if you have an n cross n image. So, your first pixel is 0 comma 0 and then you are going all the way till the end which is like n minus 1 comma n minus 1 then you start with every pixel here you start with some pixel in x t y t.

So, for example, right. So, for example, if I had if I had you know x t y t if you if you if you take the case of a simple translation let us say I had x s y s what was it. So, and then I had + some t x t y right and suppose suppose suppose I knew by some means that t x was let us say right 1.25 and then t y was let us say whatever 1.75 something like that right if I had then what it will mean is I will actually I will actually subtract this from this from this from x t and y t right I will kind of take it to the other end we will see later right about how this entire option can be operation in terms of a matrix vector notation if you do it becomes a matrix inverse. You will see that later for the time being just assume that right this is simple as this is an algebraic thing right which you can do and then now now what you do is now you actually fix an x t y t right now you take let us say x t y t is 0 comma 0 then you see right what do what do you get I mean after you do 0 minus t x is 0 minus t y what do you get for x s and y s.

So, for example, right normally what will happen is if you kind of fix a coordinate here for which you want to assign an intensity correct every pixel in target image you want to assign an intensity. So, when you sit on a pixel which is now an integer value right there is no fractional here because in a target we know clearly that we are going from 0 comma 0 0 comma 1 right that is how we traverse. So, when I am sitting at an integer pixel, but then when I come back right using this kind of an inverse transformation I come and land here in my source, but then I could land anywhere there is no guarantee I land on an integer pixel. So, wherever I land, but then you know irrespective of where I land there will always be 4 4 see neighbors that I have right within within that within let us say 1 pixel 1 pixel square right I will land if at all if I am lucky I will land here or here or here or here if not if not I land somewhere inside inside a box right, but this is like 1 pixel width always right wherever I am right wherever you are you will always end up within a box of 1 pixel width. So, then now the point is okay I have landed here right somewhere here.

Now one thing is one one simplest thing which you can do is what is called you know nearest neighbor okay. So, for example, I am going to know here. So, for example, if you feel that you are kind of closest to this guy. So, whatever this is intensity which you know in the source right this intensity you know know. So, you can simply assign it assign it to this guy here or for example, if you happen to be closest to that that gets intensity, but normally right that is not what is done you do what is called a simple bilinear interpolation.

So, what that means is you know if you are kind of say ∂ 1 from this guy and see ∂ 2 from see this guy and let us say this is your Xs, Ys and let us say right this coordinate is X, Y what is X, Y in terms of Xs, Ys? Xs by ∂ 1, Ys. No, no, no let us say let us have it in a more simple form X is floor of Xs and Y is floor of Ys right just as you have seal round and floor right. So, floor will will so if it is 125.9 it will become 125, 125.1 will become 125 whatever it is it will become 125 right.

So, now so now this intensity right so now the intensity at Xs, Ys now the coordinate we know we want to say assign assign sorry not at Xs, Ys at Xt, Yt right that we want will then be so so so then we can write this as 1 minus ∂ 1 into 1 minus ∂ 2 I of I source right Is what did I write Is right Is at X, Y not not Xs, Ys right this one this coordinate is X, Y this is floor of Xs and Ys. I cannot write Xs, Ys right Xs, Ys is sitting here that I do not know what it is. Oh no, no sorry I wrote Is it is yeah yeah correct correct Is of X, Y right Is of X, Y + then I can write 1 minus ∂ 1 then ∂ 2 Is of X + 1, Y then I can write you know ∂ what is the ∂ 1 into 1 minus ∂ 2 Is of maybe X, Y + 1 and then + ∂ 1 ∂ 2 of Is of X + 1, Y + 1 okay this is this is a simple this is the simplest thing that you can do this is simply a bilinear interpolation this is called a bilinear interpolation right. So it means that so so if you see ∂ 1 and ∂ 2 is 0 then ideally you should you should be assigning Is of X, Y right this is what will happen right ∂ 1, ∂ 2 all the all the other terms will drop off right. I am taking some kind of linear combination.

Linear combination yeah that is all you just you are just doing a doing a you know weighted weighted linear combination instead of 1D grid if you would have 1D line you would have taken between 2 neighbors now it is on a 2D grid so it will take 4 neighbors you can you can increase the number of neighbors by the way if you go for you know a cubic interpolation or if you go for you know cubic spline and all right then then you can take more and more neighbors but then it is all about how much you gain by doing that versus versus rate doing something as simple as this. In fact most of the cameras and all that do right they do they do more sophisticated interpolation because they have to anyway give you a picture that looks better than and of course they have ways to implement it very fast and so on but yeah this is simply to tell you that even I can nearest neighbor also you can do but then then I then that is not the most ideal thing to do a simple thing that you can do is bilinear interpolation that is how you will get the image finally right you want to be able to get the other image right if you are synthesizing this is if you are doing some kind of you know you know a data augmentation or something right where you want to synthesize the other image given the first. Okay then scaling is another thing which you can which you can think of right which is a simple okay now when rotation it I still have not have not written the properties right I have to write in terms of so angles and lengths again again the same thing as this okay ditto what you said ditto so it is like this right you have an image like this then right it becomes like that all the internal angles stay intact if this is a right angle this remains a right angle lengths stay the same and you know and this is called actually actually actually no actually Euclidean Euclidean Euclidean motion okay. Then you go to the next level which involves scaling so when you get into scaling right then you have something like you know what is called uniform scaling okay you can also have non uniform scaling but we will just look at uniform scaling so it will be like S0 0 S and then XS YS all of that will still apply okay whatever I said this is a target to source and all of that will apply. So for example, suppose I said that right you wanted to do a soft zoom right in your lab assignment suppose you had to zoom in an image what will you choose as your S's should S be greater than 1 or should S be less than 1 if you wanted to zoom? Zoom yeah let us see right so what do you get so XT is S times XS S times XS and YT is S times YS right yeah so if you wanted to wanted to zoom in S yeah so then zoom in so what this means is this right so you have your IS already with you and then you have a target here right and you see what you are saying is if this portion right goes and sits as a smaller thing inside that will be what there will be zoom out correct because this so it is like it is like going back but then when you go back in a real camera you do not get zeros right you get something else right whatever else is in the scene that will enter but then in a simulation you do not I mean unless of course you know you have a boundary that you can extend and you know what lies outside then you can fill it up but normally right if you zoom out it will mean that you have zoomed out but so whatever is this portion that is appearing in a smaller area right in IT whereas if you zoom in right it will mean that it will mean that right a portion of this will occupy the occupy all of your target right that will be zoom in so in one case your S is going to be greater than 1 the other case S will be less than 1 so you can actually so you can actually create these effects right again all this will involve that the binary interpolation and then you can create a zoom in a zoom out whatever you want.

Then okay now with scaling right what will happen is so with this uniform scaling so you so when you scale it angles angles are actually preserved but not the lens right the ratios of lens are preserved okay ratios of lens are preserved and the combination of this is what is called a similarity in the sense that you can have X t Y t right you can have scaling you can have a rotation okay or else I will just simply instead of repeating all this and simply write this as S into R and then you have X s Y s + a t X t Y. Now the way I am still writing it as somewhat kind of say clumsy okay there is a much more neater way to do this but I didn't want to say bring that now at this stage I thought we will do it right in the next class so right now think about you know uniform scaling rotation and translation so how many unknowns are involved right how many unknowns are involved in this? So scaling there is one unknown for rotation there is one θ then translation t X and t Y four unknowns right so this is called a similarity transform or a similarity right that is what it is called and the similarity rate again so in so for this and all the all those are actually so the earlier ones are kind of a special case if S is identity and t is 0 then you get rotation and so on. So angles are actually preserved and so it is like saying that you have this right and then and then you can rotate and then it should be something like that right you scaled rotated translated angles are preserved and what is this so and ratios of not necessarily parallel okay links that also applies to this by the way links are preserved. No this is so the point is how you choose your S and R right so it is like it is like what is your eventual goal okay so how you choose your S R and t X and t Y right I mean it is like saying that in 1D right when you do X of a t + b okay that you can do either by say translating first and then you apply the scaling if you scale first right then it should actually translate you should scale the translation right so what a and b right there what you need right so what you need you have to fix right how you want it similarly here right depending upon what if for example for example I can also have a situation where

I want to translate first and then apply an R which will mean that my actual transformation will then become R into X minus t X Y minus t Y that is also possible okay but then if you want to write it in terms of a t X and t Y then the t X will become cos θ times this t X and so on right so it is just a matter of how you write it just an unknown so you have to fix it you have to know by what you want to do see normally what happens if you want to synthesize you should know what you want to do in which case you tell what should be the θ and the order you tell I am going to rotate first and then translate right R X Y + whatever + is t X t Y if you want to do it the other way then again you should choose your t X and t Y because they will then change because then you will be pushing t X t Y into XS and YS okay so we will stop here.