Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-57

Alright, so let us, so last class we did what is called similarity transform. And the next one, okay, which is a combination of uniform scaling rotation and, and translation. So another thing right which is of interest is what is called shear. Let us see that how this transformation looks like. So you have like Xd, Yd and then if it is an X shear right, you will get 1, k, 0, 1, Xs, Ys. So what this means is that your Xd right is going to look like Xs + k times Ys that is your Yd is equal to Ys.

Okay, this actually, this is something that you can see when you are traveling in a car and if you look at a zebra Xing and so the scene kind of sets itself up in a particular way in the sense that the normal sets itself up in a particular way, you are traveling in this kind of a direction, the normal of the scene is like this and your motion is along X. So you will get something like this. So what this means is that if you had a square right like this, then what this means is that you know you can get a shear like that. Or if it is the other kind of shear then it will look like this.

I mean depending upon whether the shear is along X or Y. This is called the shearing operations. The other kind of shear is that this becomes 1 0 and that becomes k 1. And this is like Xs Ys. So in this case you get like Xt Yt.

So Xt is equal to what is that Xs right and Yt becomes equal to k times Xs + Ys. Okay. So yeah, so which way the shear occurs right depends upon how we are moving, but typically right this is what will happen. Actually right this is all a special case of something which is more general what is called an affine, affine transformation. There is something more general than affine right which we will see shortly, but okay this center should be right there okay.

The center does not move. Just going to see tilts like that. So this is affine right this more general and it looks like Xt Yt. And by the end of this class right you should be able to stitch images. For example you should be able to take a camera, capture pictures and be able to stitch them.

So Xt Yt, so this looks like A, B, C, D, Xs, Ys + Dx, Dy. So as you can see that all the others that you saw previously were all kind of see special cases this is affine. So if you take A, B, C, D to be of a particular kind then you know right you can get the rest, but sort of a general transformation this is affine. And what it kind of see preserves is that parallel lines remain parallel, remain parallel. Angles are generally not necessarily preserved angles are generally not preserved.

And preserves ratios of lengths of parallel, preserves ratios of lengths of the parallel segments. I mean if you have a general affine, in some cases you may still end up keeping the angles and all intact. I mean that depends upon how the A, B, C, Ds are because the previous ones are all special cases of this affine. But in general if you take if you have like A, B, C, D in general if you have a general affine transformation then all that you can expect is parallel lines remain parallel, angles may not be preserved. See for example, in the earlier example when you had a shear right when initially the angles were all good as right angles, but the moment the shear happened right the angles changed.

And then the preserves ratios of lengths of the parallel segments, okay. There is something even higher than affine, okay. But then prior to going that right we will just look at what are called homogeneous coordinates. I mean there is a way to see if you see right I mean you know the way I am kind of say writing this is that in some cases I am able to write it as a transformation on this Xs, Ys. But in some cases right I am not.

For example, A, B, C, D into Xs, Ys + Tx, Ty right. It does not look like a direct transformation on let us say right Xs, Ys. It looks like I have to do something and then + add something right. So it does not look very say elegant, okay. There is something called I mean a projective space, okay.

I mean there is a whole area out there, okay which is called a projective geometry. Idea is not to go into the details of that, but just to take right elements from there which are actually relevant to us. And one of the things that is taken from you know a projective geometry is what is called homogeneous coordinates. So for example, the coordinates that we deal with are called actually heterogeneous coordinates, the ones that we normally use inside a Cartesian space, but then in a projective space these are called homogeneous coordinates. And one can kind of go from the homogeneous coordinates to your say heterogeneous coordinates.

And the way to interpret it is if you are say heterogeneous coordinate or what you normally use in a Cartesian space, heterogeneous, genius coordinate right. If it is let us say x, y okay, if it is like x, y then the homogeneous coordinate can is just you know you just have to add one more sort of a dimension to it, call it x, y, 1 or in general right it can be alpha x, alpha y, alpha, alpha not equal to 0. We will see later a little bit more about what these coordinates are and look at the last one is alpha. So, what this means is if you scale the first 2 coordinates by the last one right you should be able to go there, go to the heterogeneous coordinates. So, you can think of this as a sort of you know points on a line like this right and then a projection right of these points.

So for example, right every I mean these are called homogeneous because all of them say represent the same right I mean x, y, 1 I mean if you kind of think about it right. So, it is like x, y, 1 right. So, they all kind of I mean they all could have looked the same in the sense that except for a scale factor it is like saying that if you travel on this line and if you were to project them all here and if you had your x, y axis like this right where let us say this is your z and this point is like x, y, 1 okay this is z, this is x, this is y then all the points on this line okay if you kind of map them onto that plane right they would all map to the same x, y, 1. We will see this in more detail later but for the time being just get as you remember that if you are using the homogeneous coordinates and if you want to go to this heterogeneous coordinates because finally when you do image transformations right you have to come back to the to a Cartesian space. All that you need to do is the even third coordinate whatever it is that you have if you if you divide the first 2 by the third that you will come back to your original space where you are operating okay.

And for the time being okay this much is enough when we will kind of make use of this later I mean it also there are other things to this in the sense that right I mean you know points like points at infinity and all right which in a heterogeneous coordinate right I mean infinity is not really a number right in that sense whereas you know in a homogeneous coordinate you can write you know point at infinity as is x, y, 0 okay. So you can still use finite numbers to tell that you have a point at infinity and so also it helps you know so in terms of geometry right it helps explains points at infinity and all vanishing points as they are called but right now okay we will not get into that for our for our courtesy you know this one a temporary this one the purpose is what to know how to write this homogeneous coordinates because if you know this right then I can go back and rewrite many of the things that I wrote but now you know I kind of much more elegant way. For example if I go back to see translation okay because we need this prior to going into a projective transformation. So for example translation right if you remember we had simply xt, yt is equal to xx, ys + txt, y right. So instead of that what I am going to write is I am going to write this xt, yt extended extend this is a dimension by 1 and then I am going to write this as a kind of a 3 X 3 matrix and then I am going to write my source as xs, ys, 1 okay in this case I know that the last coordinate is 1 in general it need not be but in this case I know.

So now for example let so now if you have to write this how would you write it now how would you fill up this matrix if you have to represent a translation it will be 1 0 tx, 0 1 ty, 0 0 1. So if you see right this is like xt is equal to xs + tx, yt is equal to ys + ty and 1 equal to 1 right. So you see that exactly what we had earlier but now whenever it looks like a direct transformation and I mean otherwise right if I had asked you to write xt, yt and then I had what did you have xs, ys + tx, ty as a kind of a as a 2 X 2 operation can you write this as a 2 X 2 matrix multiplying xs, ys you cannot write it right. So this homogeneous coordinates allows you to allows that flexibility then this interpretation becomes much more easier it looks like you know you have the source coordinates you just want to apply on them and if it turns out that let us say right if this last row is not something like 0 0 1 in general it need not be then in that case whatever is this number that you get here okay if you use that number and scale these top 2 coordinates then you have back to this heterogeneous okay just have to remember that because if you want to go back to the original space that is like xy 1 and for that right you should just scale I mean if it so happens that the last number is not a 1 in this case it turns out to be 1 therefore you do not have to do anything you directly get your xt, yt if that is not happening then you should scale it. For example if you had to do a rotation right a 2D rotation how would you write it now you will again write xt, yt 1 then this is going to be xs, ys 1 then you will have like cos θ sin θ – sin θ cos θ know this one translation 0 0 1 right of course this even in a kind of a 2 X 2 you could do this actually even there you could write this is a 2 X 2 times xs, ys but just that right but translations you could not write at that time now you can write everything in this form rotation and then if you think about what was that scaling right so scaling would have been would be like xt, yt 1 then xs, ys 1 then you have like let us say your scale factor a then let us say 0 0 0 a 0 0 0 1 okay that would be a scaling and then you can also have a combination of these transformations right it depends upon the order in which you are going to apply them.

So for example you can have xt, yt let us say 1 and I can think of you know these 3 X 3 matrices so if you have a similarity transform I can have a scaling I can have a rotation I can have whatever a translation but then if you change the order then again the entire thing will change okay so this matrix operations do not commute okay so it depends upon what you are applying first right so for example if this is your xs, ys 1 and if you say that this translation suppose right this rotation and this scaling then it means that you are first applying a translation on xs, ys followed by a rotation followed by a scaling right and this last coordinate you have to watch out for okay it depends upon what is happening on this right hand side. So now if you know for example right affine, affine right if you had to write how would you write that now you have xt, yt 1 and then you will have xs, ys 1 and then you have something like a, b, c, d, e, f 001 right that would be an affine or you can write this c and fs t x, t y if you wish okay and so what this really means is right in an affine you have got like 6 unknowns. And all this right goes back to how we are able to solve for an image transformation finally why are we doing all this simply because if you had to do a data augmentation then we should know how to generate or synthesize an image if you are trying to align 2 images like in a kind of a panorama where you want to stitch then you need to know what is the transformation aX them which means that which means that right if you know a priori that it is an affine right then you need to only solve for 6 unknowns but in general if you do not know anything about it then it has to be higher than this right. We do not know what motion happened but if somebody gave you a priori information that it is only an affine then you can restrict yourself to 6 if they say that it is only a pure rotation then you can just restrict it to 1 if it is in plane rotation okay. And then this r the form of r will change okay depending upon whether it is okay yeah right which is what I am going to say going to come to next which is actually a projective a projective transformation a projective transformation a projective transformation or it is called a planar homography a planar homography this is also something called a rotational homography right which we will see at the end but then a projective transformation is the most sort of see general okay and this projective transformation is again a 3 X 3 matrix but which can be only estimated up to a scale because of the fact that these homogeneous coordinates right you know them only up to a scale right.

So if you think about it you will have like X t Y t 1 and you can write this as okay we will just use okay H 11 so this is a homography matrix okay so the standard notation is H 12 H 13 H 21 H 22 H 23 H 31 H 32 H 33 multiplying X s Y s 1 okay and now right this in general I just multiply this by λ because right it is there is no guarantee that this last number will be 1 now right because of the fact that this could come out to be anything right. So I will just say λ okay this is also the reason why this matrix itself has only 8 unknowns now because the matrix can only be found up to a up to a scale factor okay so you have actually 8 unknowns even though there are there are 9 entries in that matrix right you can only find it up to say 8 unknowns okay so this is the H matrix this is a planar homography and you know actually right this actually relates to see right now the way right we are actually thinking about it is if I had write 2 images and if I wanted to find out what was the relation between the 2 in general okay and this comes from a full 6D motion okay 6D motion what that means is I could set at a place like a so for example if I am at point C and I want to go to point C ‘ then I can translate along what is it so this is like X I can go like T X you can go like T Y you can go like say T Z okay I am allowed so this camera so this is in the this is the camera space okay all these coordinates that we are talking about are all on this here are all on the you know image space right when I write XT YT and all this is not in this 3D world right I am talking about what is happening on the image plane but all of this is happening because of because the camera is moving in a certain way right and the scene is going to say directed at the camera in a certain way again it is an interplay between the 2 and what this means is that if I am here and then and then if I want to go to C ‘ I can go through 3 translations and I can have 3 rotations. Rotations means I can have the whatever right I mean I could actually spin about the X axis I could rotate about the X axis or I could rotate about the Y axis or I could I could say rotate about the Z axis right. The Z axis is what is what aligns with the optical axis so your in plane rotation is typically this about the optical axis anything else that you do will cause will cause an is out of plane rotation I mean something like this that will cause an out of plane rotation or if you what is the other one this way right so. If you do out of plane rotation then the image plane is a free.

No no no it is like this right see your image plane is fixed right so what this means is that what kind of an image will get formed see for example, if your if your camera undergoes undergoes a rotation let us say right and then and then you go here then the idea is that how do you relate the image which you get in this view to this view. Projection. A projection. Projection of. Yeah, but we need to be able to show what is what is going on right when we know that we know that all of these have a have a role to play the how they how they actually come into the picture right you are doing a rotation 3D rotation you are doing 3D translations this homography helps you to relate the 2 images finally.

Because right I mean see eventually what do you want you want to be able to relate images right you see you may not even be interested in the camera motion actually most of the time you may not be interested in what is the camera motion right who bothers about it for example, when you stitch the images right do you think that you will be very interested in knowing how the camera what motion the camera went through you will be only interested in aligning the images see this homography it is was not the camera. Homography. Homography yeah. No, no what I am saying is the camera motion like this T x T y T z which is in the 3D world and the R x R y R z that may not be specifically of interest to you right.

Yes. That is what I mean that means those 6 unknowns that are actually sitting there you may not be interested in them or you may not be even interested in how the scene normal is see for example, if you take a if you take a unit normal of the scene that you got to see 2 unknowns there right and the third thing is how far away is the scene right none of this you may be interested in all that I know is anything like that can happen the optical axis of the scene could be completely front of parallel it could be inclined at some angle I could be I could be moving arbitrarily right you know in a kind of you know I could be having an arbitrary 6 D motion, but I may not be at the end of the day interested in knowing what how much I moved by my interest is only in being able to relate the relate the images because all that I have are feature points right. Finally, I have an image here I have an image here if I can get my feature correspondence is I should be able to align the images.. Due to motion all of this is because of motion by the way everything that we have said is because of camera motion none of this is anything other this is not a barrel distortion or.

A distortion. A distortion the lens or anything this is all pure camera motion. . You do not I will show you now know. So yeah so the idea is that in general see it is like this right

I think you know what you try to ask me is should I should I not know how the camera moved or something no the whole idea is you do not need to know. In fact, this homography matrix itself right this has a form this comes as some what is it r + 1 by d T n transpose k ‘ ok k inverse ok.

This is the I mean right I am not showing all this, but this is the homography actually if you have a planar situation right what this means is you have a normal for the plane which is this n you have a translation of the camera which is actually you know a 3 d translation. Then this vector if you take the outer product right this becomes a 3 X 3 matrix then there is a rotation which could be a combination of rx, ry and rz in some sequence and then that you have an you have what is called an intrinsic camera matrix which is k. So there are various factors here at play ok there is something called the intrinsic which is the camera matrix which is k I will talk about it right in the next class right how you arrive at the camera matrix and so on. But that is again there ok it does not mean that it is not that the camera intrinsic also matters then there is an r and t right which is a rotation and translate which is supposed to be extrinsics. So these are extrinsic parameters because they do not belong to the camera you just depends on how you move k is something intrinsic to the camera it is like saying I know where the center of the camera is it kind of deviated are the axis of the camera exactly orthogonal is there a skew all of that.

So that is all that is all related to the camera per say r and t are not related to the camera that is why you call them as extrinsics but then the idea is not to kind of go into this the idea is not at all to go into this we do not want to go into that ok that is what I am saying without going into that you want to be able to write relate images ok that is the idea. No, no, no actually I did not want to get into this but it is more or less like this right see for example I mean the 8 unknowns right if you want to get a say think about it will be like you know 3 for r 3 for t and then say 2 for n because it is supposed to be a supposed to be a you know unit normal. So you have 2 unknowns for the normal 3 for the translations 3 for r I mean if you want to think about it that way but the whole idea is not even to get into see it is not like we have got an edge and therefore we know we want to go back and work out r and t and we are not even interested in that the inverse problem there is something called you know finding out the camera motion given the homography matrix we do not want to get into that. It is like saying given this right can you can you arrive at this I mean that is a that is a harder problem we do not want to even get into that panorama stitching does not involve any of that ok. Now what this means is the following ok and I think I think we we spend too much time on that I was not planning to spend that much time.