

## Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-60

So, what we can say is, I mean that I am just writing a little bit loosely here in the sense that okay this is small  $z$ , okay. So, if I have homogeneous coordinates and I want to go back to my heterogeneous, I just have to scale by this last guy that is  $z$ , right. If I want my original  $xy$ , I am still calling this as  $xy$ , if you want you can call this  $x' y'$ , but that is okay. But you know that if I wanted to get back to my image coordinate, I just have to scale by this  $z$ , right. And this is equal to, see for example, I am going to write this in terms of a matrix, a  $3 \times 3$  first to start with, okay. This is the simplest camera, camera intrinsic that you can have  $\begin{bmatrix} 0 & f_0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$  followed by, so see this is  $3 \times 3$ .

I am going to follow it up with, with a  $3 \times 4$  matrix which is  $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$  and this matrix, its importance will, I mean right now, okay, well its role is this  $x$ , what is that, what did I say,  $x_c, y_c, z_c$ , right. So the same thing, when you want to express the 3D coordinate in terms of its homogeneous form  $x_c, y_c, z_c, 1$ , right. So you got this as  $4 \times 1$ , this is  $3 \times 4$ , this is  $3 \times 3$ , right. So effectively you will get  $3 \times 1$ , right on the left and if you multiply these two, right, you get  $x_c, y_c, z_c$  and then here, right, you have  $\begin{bmatrix} f_0 & 0 & 0 \\ f_0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ .

This in turn will give you  $f x_c, f y_c$  and  $z_c$ , right, which then means that small  $x$  equal to  $f$  times  $x_c$ ,  $x_c$  small  $y = f$  times  $y_c$  and small  $z = z_c$ , right. And I know that if I want to go back to the original image coordinate, I just had to simply scale this by  $z_c$ , if I scale this by  $z_c$ , I kind of get back my, that is a small  $z$ , right, I just had to scale it and I get back my, let us say, original, original image point, right. So if I scale  $x$  by  $z$ ,  $y$  by  $z$ , small  $z$ , okay, it is my, it is my  $c$  image coordinate, right. No, small  $z$  is not  $f$ , small  $z$  is simply, is simply this representation of the, say, 2D coordinate in terms of its homogeneous form and that will be equal to capital  $Z_c$ , because you have to scale by that, right, I mean, you have  $f x_c$  and you have to scale that by  $z_c$  to get your small  $x$ . See, the, the, the, no, you need to get your small  $x$  and small  $y$ , okay, here I mean, okay, that is why I said it, if you want to, if you want to think about it, think of this as  $x', y'$ , okay and you want to go, so, so it is like saying that I mean, I have, I have my, I have my, see, original coordinates as  $x, y$ , this is my image coordinate, right.

I want to transform, transform into this homogeneous form. The general homogeneous

form is  $x'$ ,  $y'$ , let us say, let us say  $z'$ , where I will express my  $x$  as  $x'$  by  $z'$ , I will express my  $y$  as  $y'$  by  $z'$ . Now, the question is what should be the small  $z'$  and what should be the, what should be my  $x'$  and  $y'$ , right and those have to be related to the camera parameter because everything finally boils down to what the camera is doing, right. So and we know that eventually we want this equation, no, which is actually, I mean, this is the one that is happening on the image plane, right,  $f x c$  by  $z c$  is what is happening, now  $y$  is  $f y c$  by  $z c$ , right, that is what you want and therefore the only way you can write it is this way, I mean, where, but there is a reason to write it because, right, this power, this portion that we will see later how to extend this, okay and this portion is the camera. Right now it is looking simple, right, it just has a focal length but we will actually make it a little more interesting, okay.

So then  $f x c$  by  $z c$ , so therefore, right, yeah, I mean, if you want it you can write it as  $x' y' z'$  and say that, right,  $x y$  is  $x'$ , so  $x$  is  $x'$  by  $z'$   $z'$  and if you want you can even think about it, it does not matter,  $y'$  by  $z'$  but in general you just write it as  $x y z$ , right, with the understanding that you had a scale by the last coordinate, right. Okay now the, okay now the point is, see, right here one of the assumptions, okay, that we made is this, see for example, right, we assumed that, I mean, I have not drawn it exactly, but then the assumption is that the center, let me draw it by different color, so this center which is the optical center, right, it, you know, it exactly coincides with the, with the center of the image plane, I mean, here I have not drawn it like that, but it is like saying that I have my image plane and this optical axis, right, that is actually passing through it, this center and, right, and this center are exactly aligned, that means the center of the image plane, but it could happen that they are actually not aligned, that means this guy could be off a little bit, I mean, this is all a manufacturing thing, we are assuming that it is all perfect, right, but then, okay, DOB, so there could be a small shift here, but that is a kind of a, you know, a permanent shift, it is a manufacturing thing, right, and to accommodate that what we do is we actually use that offset, right, that might be there, if it is there you have to account for it, if it is not there, not a problem, then it will be, so that offset, right, we actually call that, let us say,  $P_x$  and  $P_y$ , okay, so let us say that the camera offset, the camera offset, these are things that can actually happen, okay, and they do happen, that is why you need camera calibration, because, right, that is also the reason why you do a camera calibration, because if you do not calibrate a camera, that means if you do not worry about what the intrinsics are and simply go ahead and do some triangulation and all, you will go out wrong, because the actual place where it is imaging, right, you have to express it correctly, right, if this matrix is not correct, then all your calculations are going wrong, okay, so you have to account for them, so that is camera calibration. So the camera offset, let us say, right, is given by, suppose we say that the camera offset is given by, you know, what is this, you know,  $P$ , what did I say,  $P_x$ ,  $P_y$ , that means the  $x$  offset is  $P_x$ ,  $y$  offset is  $P_y$ , okay, then the way to accommodate that is, is to actually you know, throw that into this  $K$ , okay,

so let us put that into this, so, okay, this is an intrinsic parameter, right, so  $F_0$ , so it comes here, it does not come anywhere else, it is a part of the camera, just a focal length is a part of the camera, then you have  $O_F$ , then you have  $P_y$ , then you have  $O_0$ . Now, if you notice that if I again multiply the right hand side that  $3 \times 4$ , that  $4 \times 1$ , you still have  $x_c, y_c, z_c$ , right and therefore, if you multiply the 2, you get like  $F_x c + P_x, z_c, F_y c + P_y, z_c$  and then you have  $z_c$ , right, so again if I want my image coordinate, I should do  $F_x c + P_x, z_c$  by  $z_c$ , but  $F_x c$  by  $z_c$  is my, you know, original  $x$  and  $P_x, z_c$  by  $z_c, z_c$  goes away, so I get  $x + P_x$  and similarly,  $F_y c + P_y, z_c$  by  $z_c$ , it will give me  $y + P_y$ , this is what you want, right, so you want to account for that  $P_x$  and  $P_y$ , okay, if there is a, okay, if there is a manufacturing thing, right, that is a manufacturing stuff, okay, but that has to be accounted for and during camera calibration, these are the things that you actually find out and these are frozen, right, for a camera this is fixed, okay, you do not have to keep changing it, so the intrinsic once and for all if you calibrate, it is kind of fixed. This  $F$  itself, right, looks like, looks like, right, that is, okay, now you can just keep it like that but actually even there, right, there is a little bit of adjustment needed, okay, there is something called the field of view, right, I mean you all heard about field of view of a camera, field of view, right, we keep saying field of view, right, that field of view depends on this  $F$  and the sensor dimensions, okay.

So for example, okay, if you think about the, so if you think about the sensor, okay, now let us say, right, we have something like this, okay, let us say I have a width  $W$  and then I have a height  $H$ , okay, this is the sensor, okay, this could be in millimetres or something, right. Now what we are saying is that, that, right, we have the optical centre, right, going through this, let us assume that there is no drift arranging, okay, just going through the centre and now if you actually look at this ray, right, that is, I mean if you look at this outermost ray that kind of hits the top of the sensor and goes and if you kind of look at the bottommost ray that also hits, so for example if you look at this line and then look at that cone that is going, right, that gives you the field of view, right, so this angle, right, if you think about this as  $\theta$  then the field of view is actually 2 times this angle  $\theta$ , okay, that is your field of view. And right simple thing will kind of show that, you know, so if I do  $\tan$  of  $\theta$ , okay, so this  $\theta$  is of course half the field of view, okay, the full field of view is  $2\theta$ , so  $\tan \theta$  is  $w$  by  $2$ , okay, because this height by  $f$ , right, because from here to here is  $f$ , from the camera centre or the optical centre to the image plane centre, that is  $f$ , right,  $h$  by  $2$ , yeah,  $h$  by  $2$ ,  $h$  by  $2$ , so it is  $h$  by  $2f$ . Therefore ref looks like, what is this,  $h$  by  $2 \tan \theta$ , right,  $h$  by  $2 \tan \theta$  and similarly for the other one, right, if that field of view, okay, so for that, right, what will you have, if it so happens that  $h$  is not equal to  $w$ , it can happen, right, you can have a sensor which is not square, right, I mean here we have taken a general sensor, so for that you may have  $w$  by  $2 \tan \phi$ , which means that your field of view along the height is not the same as the field of view along the width, it can happen, which generally does not happen, but most general case you can take  $h_0$  to be

equal to  $w$ , okay. But then what happens is even if  $h = w$ , right, there is still a small sort of, you know, a problem that can arise because, you know, if you look at this, right,  $\theta$ s are after all an angle, so  $\tan \theta$  is no units, but look at  $h$  and  $w$ , right, so they are in actually millimetres or something, right, you have a sensor of some 2 millimetres by 2 millimetres, something you will have or 1 inch by 1 inch, right.

Now, so this  $f$  turns out that  $f$  will also be in terms of inches or millimetres or something, but if you actually look at the  $xy$  coordinates, right, that you have, there is a small  $x$  and a small  $y$ , we actually typically represent them in terms of a pixel location, right, we do not say  $x$  is 0.01 millimetre from the centre, right, we want it in terms of the coordinate, right. So, and therefore, and also read  $P_x$  and  $P_y$  either you can express it in terms of, you can do, there is nothing wrong, for example, the intrinsic camera matrix can be expressed entirely in terms of just, what do you call, you know, in terms of the actual length, okay, but then that is not normally done, right, pixels are more useful because otherwise every time you will have to do this conversion. If I say 1 millimetre, then you have to worry about 1 millimetre how many you see pixels right, does that contain then and so on. So, so normally we would like the camera, that intrinsic matrix to be such that if I just use it, then it will tell me where the point is and not like I have to do a conversion after that, okay, that is, that is, that is the, you know, normal thing.

So now this  $f$  if you look at it, right, this  $h$  even if, even if let us say, right, even if, okay, right, it does not really matter whether  $h = w$  or if not, but then the point is what could matter is if I have, what do you say, I mean, so for example, right, I mean, if I, if I kind of, if I kind of look at, look at the, see density, the density of these, see pixels, right, see, I mean, I can have, I can have a sensor which is perfectly square, okay, this is square let us say, right, I have  $hh$ , assume that as soon as I have a square sensor, but what can happen is, right, I may have a density like this, I mean, you know, again, right, this is all, it may not happen, but then, right, I could have something like this. That means for the same length, right, along one direction I have let us say, right, 3, 3 pixels, but along the other direction I have got like whatever, right, 9 pixels or something, it can happen, I mean, you know, you may have a camera that has like, you know, 496 by something 1024, even though the sensor size is probably a square, okay, that means that your density is actually varying  $\alpha X$ , see,  $\alpha X$   $X$  and  $Y$ , it is not the same. So when we write that  $F$  and  $F$ , right, and you know that the second  $F$  is actually multiplying for the second, the  $Y$  coordinate, the first  $F$  has an influence in the  $X$  coordinate, right, in that matrix, that  $K$ , see this guy that had, no, that  $F_{00}$ ,  $0F_0$ , right,  $001$ , so this  $F$  is for the, is responsible for the  $Y$ , this  $F$  is actually responsible for the  $X$  and if the pixel density is not the same, then that has to be accounted for, right, and therefore, what is normally done is we write this as  $\alpha F$ , if  $\alpha$  is 1, then it means that the pixel density is the same, but if  $\alpha$  is,  $\alpha$  could be less than 1, could be more than 1, if  $\alpha$  is different, right, then it means that your pixel

density is actually not the same and therefore should be accounted for when you do this conversion, is it okay? So more general form, right, then is like this, right, so the intrinsic now changes to  $F$ , some people write this as  $F_x$  and  $F_y$ , but actually it simply means that, you know, and then they will say there is an aspect ratio between the two, it is all one and the same, but the simplest way to understand is in terms of the density. So  $F$  then  $0 \leq \alpha < 1$ . Sir, so density is higher  $\alpha$  is greater than 1.

Alpha, if it is higher along that axis, along  $Y$ . Along  $Y$  axis  $\alpha$  is greater. Alpha, if  $\alpha > 1$  that means along  $Y$  axis you have a higher density. So the final coordinate will be  $\alpha F Y_c$  by. Correct, correct.

So it will get scaled by  $\alpha$ , not added, scaled. Okay, now see there is one other thing, okay, which can also happen that is like, you know, one of the assumptions which we are making is that this is exactly, you know, each of these pixels is actually a square or actually, you know, a perfect sort of, you know, a rectangle, right. We are sort of assuming that this angle is actually  $90^\circ$ . Okay, that means it is like, it is like, you know, they are exactly orthogonal to each other. But then whether we saw some operation, right, earlier, what was that operation which could actually turn something which was actually perpendicular to something else shear, right.

So what can happen is sometimes this need not be exactly orthogonal, I mean you could have some little bit of angle between them, they may not be  $90^\circ$ , it could be like, you know, not be exactly  $90^\circ$ . So it accounts for that, right. So this parameter here, this is turned into shear  $S$ . Okay, so this is like your shear, I mean if you want to think about it, but then this is about non-orthogonality, what do you call, non-orthogonality, this is to account for non-orthogonality. This is for non-orthogonality, this is for pixel density, pixel density and this is for the offset.

So in that sense, right, you have got like 1, 2, 3, 4, 5 unknowns. Okay, and in a most general  $K$ , which is the camera intrinsic, this is called the camera intrinsic matrix, called the camera intrinsic matrix and the most general form that you can take is this. And one other thing that you observe is what, something else which is interesting about this camera. Upper triangular. Exactly, upper triangular and there is one more thing, if it is just upper triangular, maybe it is not so interesting, with all non-zero diagonal elements, which means what, which means that it is always, right, invertible.

You can always invert this guy. Okay, so this matrix is upper triangular, upper triangular non-zero diagonal elements, elements, okay, invertible. Okay, this is, this as far as the camera matrix  $K$  is concerned, so which is why you can freely use  $K^{-1}$  and all. You know how to go from, right, you want to do any invertible inverse operations, all you can

comfortably do that. Now, let us kind of talk about this extrinsics.

Sir, what is the name we give to alpha? What is the name we give to alpha? I do not think we give any name, just a factor there. No, sir, you wrote something. No, no, no, I just said pixel density. Oh, pixel density. I mean, I do not think they have a name for that.

Even for S, right, they do not explicitly say it is a shear, but that is what it is supposed to encapsulate. Camera extrinsics. So right, I mean all this, okay, right, we are actually doing because, right, eventually what we want to do is, for example, you know, when we, when we, when we move to stereo, right, we will have, let us say, one camera here and then we may have another camera here. It could be that it is the same camera that we have shifted or it could be that we have actually two cameras, right, on a rig and we have a point here, right. So unless we know, for example, how to, how to, let us say, represent the coordinate of this point in this camera frame and how to actually, right, represent this point in this camera frame, then when we want to search for a correspondence or we want to know, for example, if you know the correspondence, how the triangulation should be done and so on, then we should know, we should know where this point, point goes here, right.

So it is like saying that, it is like saying that, saying that right, I mean, you know, if this point is here and if that point goes there, then we should be able to express through a projection matrix what will happen to this point if it is, right, I mean, you know, if it falls on this image plane and what will happen to this point if it falls on this image plane and if the cameras are not the same, then how should you account for that because it need not be the same camera, right, it can be different camera, that means the  $k$  could change and in all such cases and also, and also, right, there could, there could be a, you know, a relative rotation and translation between those two cameras, I mean, that is when you get a stereo, right. Pure stereo is like pure translation on the, you know, on the, on the image plane, that is the purest stereo, but, but yeah, you can have a general stereo where you have an  $R$  and  $T$  and therefore, right, so this is the intrinsics and the extrinsics together come to come now in order to be able to tell what is happening and that will actually decide, so for example, right, so for example, if I ask you, right, suppose, suppose, suppose I am able to tell, right, where this point comes here, okay, now this point, this point, I mean, so it is like saying if I get a back project, right, I have the optical center here, I have an image point, if I do a back projection, right, it goes all the way, that ray. So, I can say that my 3D point is somewhere in that ray, but, but I should be, I should be able to, suppose I ask where is this point, right, on this, on this image plane, we should be able to tell that whether is that lying on a curve on this image plane, is that lying on a line on this image plane, is that, where is it going, right, so such things when we wanted to, so for example, if I want to say where does this line map to on this image plane, I should be able to tell where it is, so that then I can search on that line for my coordinate sort of, you know, a

correspondence. I mean, if I do not know the correspondence, if I know the correspondence, then you can triangulate and so on, but all of that requires this extrinsics and intrinsic, before the, without that we cannot do any of this, so we have to be able to express the scene coordinate in terms of the image coordinate, whatever, you may have n number of cameras and you will have n number of intrinsic and that many extrinsics, because each camera is kind of positioned somewhere and you want to know what is that R and T, okay, so we are doing this exactly for that reason, okay. Alright, so the extrinsics, right, it will be like this, so to go to the extrinsics, okay, so see until now, right, what we said was I have a coordinate system, right, that is actually centered at the optical center, okay, so  $Y_c$  and then this is my optical axis aligned with, no sorry, this is my  $Z_c$  aligned with the optical axis, right, so here is my, let us say the camera center C, okay.

Now, what could happen is somebody could kind of have a world reference which could be aligned in some other direction, let us say this is  $x_w$ , this is a world coordinate system centered here and this is  $y_w$  and that is some  $z_w$ , okay and this person gives you the coordinate of this point in 3D, okay, which is what we are interested in, but then our camera is somewhere, so it is like this, right, so if he gives you the coordinate of this point with respect to  $x_w$ ,  $y_w$ , let us say  $z_w$ , right, with respect to, okay, this coordinate system, whatever it let us call this as O, okay, this is or let us say  $C_w$  or something, right, if you want to, okay, let me call this  $C_w$ , so this is like a world coordinate system which is centered somewhere else, okay, this is a world coordinate system with respect to which I know this. Now, the point is right, see we know kind of how to write, I mean how to say where this point will map on this image camera plane if this coordinate was expressed in terms of the camera coordinate system, that we just now saw, right, if you tell me that it is  $x_c$ ,  $y_c$ ,  $z_c$ , I know where it will map, but now you are not telling me in terms of this coordinate system, you are actually giving me in terms of another coordinate system. Now, you might ask why would somebody do that, the whole point is this, right, like I said here, if you had another camera sitting somewhere else, right, with respect to, then you need to be able to relate, relate the same point with respect to this coordinate system and with respect to the other coordinate system, right, that is when you can actually tie things up, the same point seen from there, seen from here. So the world coordinate typically even when we do the calibration and all that we typically tend to use some coordinate system which is convenient and we will call it the world, that is the way, so this is still single view by the way, right, I am not even talking about two view, but I am just trying to motivate this two view from here, but this is all still single view, I have one camera and we are still doing only what we can do with respect to expressing the coordinates and all that we are saying is can I express image coordinates given that the world coordinate is with respect, is not with respect to the camera coordinate system, but with respect to a different coordinate system which could be a world coordinate system arbitrarily chosen, okay. So all that we are still doing is all single view, okay, single view geometry, this is all single

view, okay, but then you see remember that is exactly these ideas that will come back when we do a two view geometry or when we go to multi view geometry, okay.

Now for example, so the point is right, I have this point P here, right and okay let us, okay, now, so now right what we can, what we know is, okay, now this is the, okay, now the C is the coordinate, now I have written that as C, okay, let us say that it is actually a vector, okay, C is a coordinate of the camera center or the optical center with respect to Cw or with respect to the world coordinate system, okay. So with respect to this coordinate system, this guy is at some C bar, okay and this point is here, okay, so right, so kind of one way to kind of think about it is, okay, so right, so kind of one way to think about it is I mean if you call this as xw vector that is xw, yw, zw, then one way to think about it is you kind of account for this translation that exists between the two in the sense that you can do say  $xw - C$  bar, right, because actually right this guy is at C bar, it is not at Cw. So one way to think about it is you do an x bar w or this vector - C bar, okay, that will be like, that will be, but then that by itself will not solve the problem because if you do just that and if you try to plot that point, right, with respect to, okay, where this  $xw - C$  bar will go, right, it may not exactly give you the same point because it may go somewhere there unless you rotate because these two have a mutual rotation also, right, if it was a pure translation I could have simply done this, right and I am there, but then because this could be differently rotated, so unless you kind of do a rotation, account for the rotation you cannot kind of come back to this, come back, so this, so the coordinate of this point with respect to the camera coordinate system will be some xc, yc, zc, right, that =  $xw - C$  rotated, so it is like translate first and then, and then, and then get us a rotate, so translate first and then you say rotate, you can also rotate first and translate, you can do both, but in this case it is like translate, first translate and then, and then you say rotate, then a rotate, so what this means is if this point is plotted this will exactly hit here, it should give you the same point P, it should not give you something else, okay, so what this means is that let us call this as xc, right, a vector xc which represents xc, yc, zc, so we have xc, right, is equal to R times  $xw - C$ , okay. Now, let us again go to the homogeneous coordinates because finally, right, we want to go back to that equation, right, where we had this 3 X 4 matrices and all, therefore what we can do is, let us call x tilde, xc tilde to be the homogeneous representation of xc and x tilde w to be the homogeneous representation of xw, okay, so now this is 4 X 1, this is also 4 X 1, right and then what this means is that now, see earlier right what we had, we had like x, y, z, right = whatever that the K matrix which is the most, in its most general form, then we had 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0 and then I had here xc, yc, zc, right, this is what I had, no, no, zc1, yeah, correct, zc1, okay, now which means that this is my xc, okay, tilde that I have there. Now we know that, we know from the previous expression that xc tilde, okay, now, so now see if I have instead of x, no, where is that, so here right, if in the previous one, if I convert this into xc tilde and I want to write it in terms of xw tilde, see this is xc, yc relation, xc, by the way, xc, xw relation, right but



then I want to kind of express it in terms of, in terms of you know the homogeneous coordinates, then what you can do is you can actually express that as  $k$ , right into, okay, now this one, right, we typically partition it as identity and then partition as 0, where 0 is a  $3 \times 1$  vector,  $i$  is  $3 \times 3$ , okay, that is a short form instead of keeping on writing 1, 0, 0 and all, right, so we have  $i, i' 0$  and then, okay, this portion, right, that we want will look like this, so you have  $r, r$  and then  $-rc$ , okay and 0 which is a vector 1 and then what is this,  $xw$  tilde, you can actually show this, I mean you can easily show, okay, this is  $1 \times 3$  by the way,  $r$  is  $3 \times 3$  and this one is  $-rc$  and last one is a scalar, that is just 1, so this is actually  $4 \times 4$ , this is  $4 \times 1$ , this is  $3 \times 4$ , this is  $3 \times 3$ , this is 0 vector, okay, is  $1 \times 3$ , okay and then this can be further simplified, so you can actually multiply these 2 matrices and that will give you  $k$ , okay, now yeah, so one of the things that we will do here is that we will call this a projection matrix, okay, this guy, so it is like saying I have the world coordinate of that 3D point with respect to some world coordinate system that is my  $x$  tilde  $w$  and I want to find out where it maps in my image plane, when my camera is somewhere else, okay, which is what we said is this  $c$ , right,  $c$  is where the camera is with respect to that world coordinate system, so in order to get it there by taking into account the camera intrinsics and the camera extrinsics, right, so this is called the camera projection matrix or simply a projection matrix.

And always, always, right, remember that a projection matrix relates 3D coordinate to the image coordinate, homography is always between 2 image coordinates, okay, this is not a homography, this is a projection matrix, it tells you where does a 3D point map on the image plane, okay, so the projection matrix gives you where it goes in my image plane, a homography is between images, there is no 3D camera projection matrix, right, called  $P$ , so this is a standard, this is not that point  $P$ , okay, this is a matrix projection maybe I am just using  $P$  because that is the most common thing and then  $K$ , now if you multiply these 2, right, you will get  $R - R C$  and then this will be like  $3 \times 4$ , this will be  $3 \times 3$  and therefore  $P$  is always a  $3 \times 4$  matrix, a camera projection matrix, this is the world coordinate and all of this is in the homogeneous form, okay, so act on a  $4 \times 1$  vector which is your 3D world coordinate appended with 1, the projection matrix acts on it, gives you a  $3 \times 1$  vector, scale the first 2 by the last one, you get your, okay. The other way to write it is there are different forms for this, you can actually pull this  $R$  out because it is common and then you can write this as  $I$  then  $-C$ , okay, you can also write like this and this sometimes, right, this also can be kind of written as  $K$  and then sometimes this is written as  $R$  slash  $T$ , where  $T$  is a translational vector,  $3 \times 1$  and then the interpretation here is right, you see this  $P$  is going to act on your, this  $1 \times 3$  tilde  $W$ , right, so if you look at this what would you interpret this as? Yeah, exactly, right, so it is like rotation, rotate first then translate, if you look at this form it is like translate first then rotate, okay, so and both forms are okay, this is like rotate, first rotate then translate, this is like first translate then rotate, okay, whatever be it but in the most standard form is this, okay,  $R T$  because it is

very clear, so it has intrinsic camera matrix which is  $K$  and then this is your extrinsic camera matrix, extrinsic camera matrix, camera matrix, it has only extrinsic parameters, no, so it has how many unknowns now, this extrinsic has 6, 3 rotations, 3 translations, right, all in of course you see 3D. And  $K$  has 5 unknowns and how many elements does  $P$  have, actually it has 12,  $3 \times 4$ , right, it has 12 elements sitting in it but then we can only estimate it up to a scale factor, right, because of this homogenous coordinates, all can derive with respect to homogenous coordinates, therefore  $P$  itself can actually compute it only up to a scale factor and there are 11 entries there and actually you have actually 11 unknowns, 5 sitting in  $K$  and 6 sitting in the extrinsics, right. So  $P$  by the way can be found only up to a scale factor, up to a scale factor, so in that sense it has actually 11 degrees or you know, yeah, so basically right 11 unknowns, degree of freedom is slightly more tricky, I mean I won't use that and say 11 unknowns, because for example, at  $R$  has,  $R$  has actually 9 elements in it but do you have 9 unknowns, the degrees of freedom is only 3, similarly translation of course is 3, so that way the degree of freedom and unknowns is not, they are not the same, okay. So then what this means is that, so see, so you know, any time somebody gives you, gives you an, gives you an say, you know, 3D point, you should be able to simply apply the camera projection matrix to get your, in fact, right, sometime maybe, you know, I will show you maybe one example of, one of my students is working on this ball tracking and all, so we are doing something in this indoor, what is that, some room, okay, where we have set up cameras and all, so you can actually see, right, when, how you can track these, see 3D point of the ball using 2 cameras, right, all this goes in there, okay, so we have to use, so we have to do the camera calibration, we have to know  $k$ , right, we have to know everything in order to go to tell where the ball center is, right and so on.

So actually all of this, I mean one can, one can actually take it up as a lab thing and then do it, okay, I mean if you are interested, you could also do that.