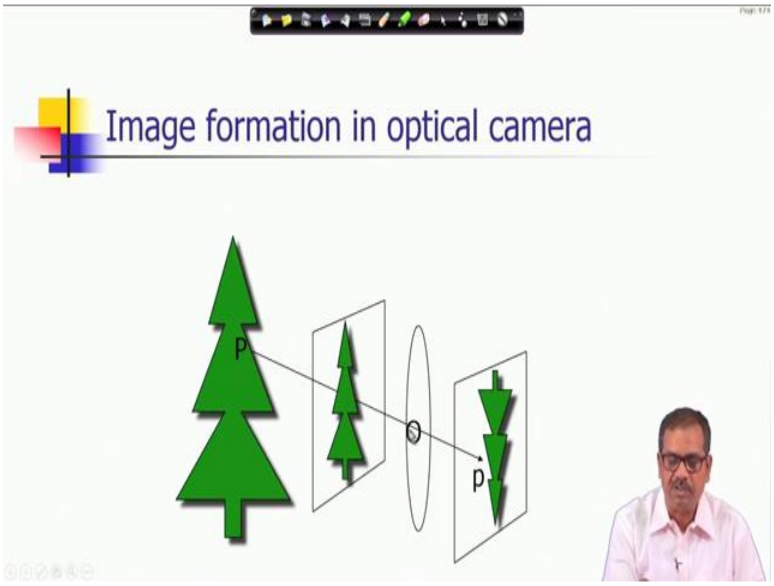
Computer Vision
Prof. Jayanta Mukhopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 11
Camera Geometry Part - I

(Refer Slide Time: 00:20)

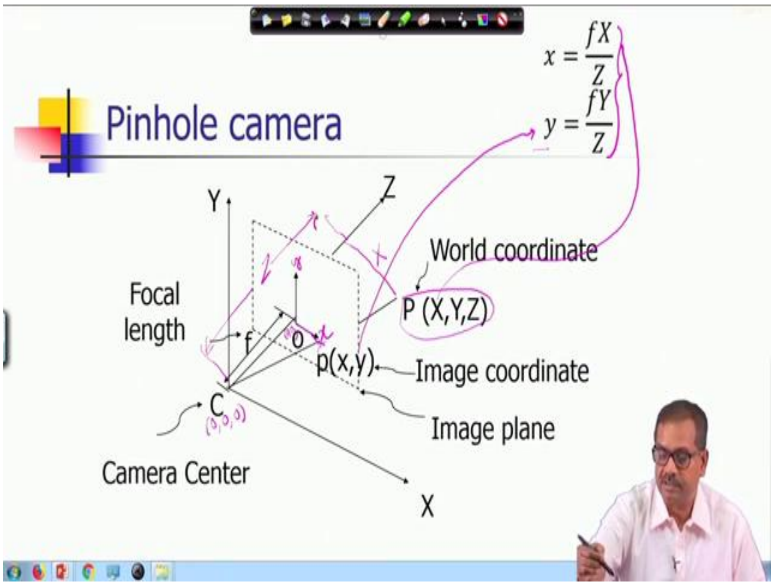


In this lecture we will discuss about Camera Geometry. We have already discussed how a 3 dimensional same point is mapped to a 2 dimensional image point. Here as we know that if I consider a 3 dimensional scene point P, then from there if I draw a straight line which passes through the centre of lens and which intersects image plane at another point some small p that forms the corresponding image point.

Now, logically we can consider also in the same way of image has been formed, in the front of the camera where the corresponding image plane is placed at the same distance where the usually if the sensors were placed behind the lens, the distance between that plane of sensors that distance if I keep it same then you will get the images of the same size and there is a logical transformation of coordinates from this point to this point that we can consider.

So, this is a convenient way of representing the coordinate system and we will use this principle while explaining the mapping of 3 dimensional scene point to 2 dimensional image point using this kind of projective geometry.

(Refer Slide Time: 01:40)



So, let us considered the principle here, how these are related. So, this is the configuration, what is usually shown for pinhole camera mapping and here in the pinhole camera once again center of projection is taken as the point from which all the rays are emanating and which is projected on the image plane, so the point of intersection that is the image of the corresponding 3 dimensional points.

So, only thing here that you should notice here, we are interested to know also the mapping of physical 3 dimensional coordinates in the scene to the corresponding 2 dimensional image point in the image plane. Though the complete ray, it is represented by this point in the sense of projective geometry, but physically images are formed by the phenomena of reflection from where energy of the reflected point those are focused at this point.

So, we are interested to build up that relationship and mathematically if I know this point, then it is very easy to establish the relationship between the coordinates of the image point and also the coordinates of the 3 dimensional point where now, we have to considered some convenient representation of this coordinate system. So, let us considered the center of projection or the camera centered at the origin of this coordinate system and also the axes of the 3 dimensional coordinate system those axes, the X axis and Y axis of that 3 dimensional coordinate system that is parallel to the coordinate convention what is followed in the image plane.

The origin of the image planes coordinate that is formed by the intersection of the Z axis with that image plane and this is considered as the origin of the image plane and as I mentioned this X and Y axis they are axis parallel with respect to the 3 dimensional coordinate system. So, with this convention then it is very easy to compute the coordinate x and y from the 3 dimensional coordinate which is represented here by upper cases of X, Y, Z or capital X, Y, Z. We can apply the law of similar triangle, we assume this plane is at a distance f because f is the focal length of a camera.

So, in that case by applying this law of similar triangle, we can find out that so, these are the points and this is the image coordinate and as I mentioned this is the focal length and then these are the expressions that we get the coordinate as

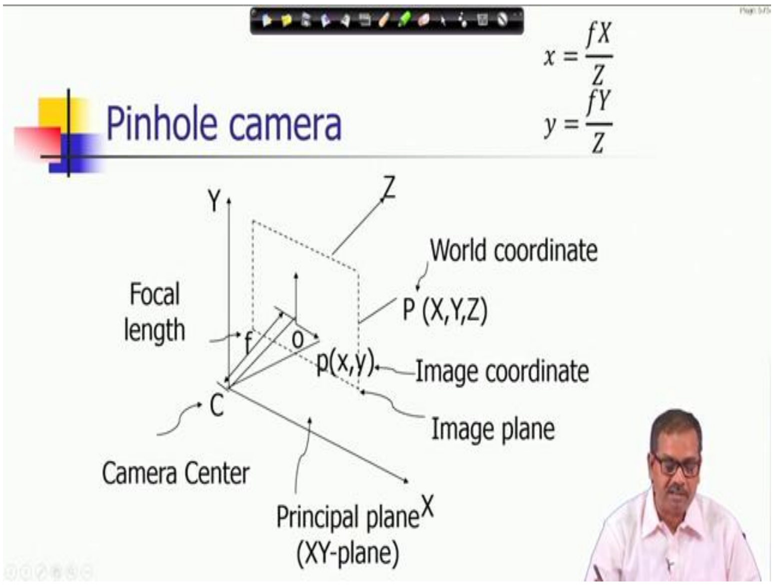
$$x = \frac{fX}{Z} \quad y = \frac{fY}{Z}$$

So, you are using this coordinates of the scene point to get this competition, you can apply the law of similar triangles if you consider the triangle for finding out x coordinate.

So, this is the X from the Y Z plane and so the corresponding x value at this point and this is a similar triangle so, that will be. So, this length is f and this length here in the coordinate system is called Z, so it is $\frac{fX}{Z}$. Similarly, y is also considered as $\frac{fY}{Z}$

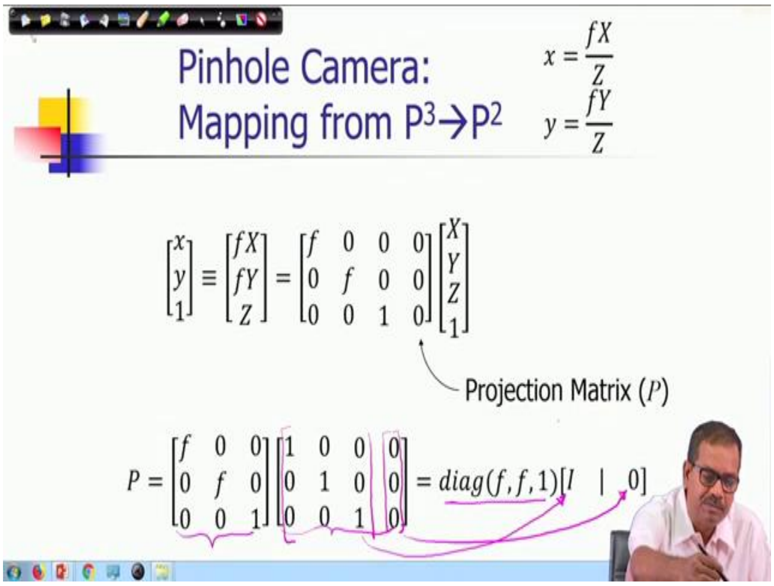
So, this is how we map a 3 dimensional coordinate points to its corresponding 2 dimensional image point, you see that we are using actual scene coordinates for which this image has been formed. So, if I give you X, Y, Z, it is very easy to get this point x and y, but inverse is not true. If I give you the image point, naturally it is represented by the whole ray and it is unless you specify some other constraint, you cannot say that it is image of which point. So, let us see how you are we can resolve all this things.

(Refer Slide Time: 06:31)



And another thing I should mention another nomenclature will be as follows, we call the XY plane of the original 3 dimensional coordinate system as the principal plane and the Z axis of that coordinate system is called principal axis.

(Refer Slide Time: 06:51)



So, mathematically we have seen already, we have written in the form of an equation which is given by this $x = \frac{fX}{Z}$ and $y = \frac{fY}{Z}$ and so, this equation can be represented by a matrix form if I use a homogeneous coordinate system for representing the points, how you can do it. let us see. So, this is the representation as you can see that the coordinate point x y z, it becomes in the homogenous coordinate system you have an additional dimension.

So, it is a 3 dimensional homogenous coordinate system and or you can say also this is a 3 dimensional projective space that is why it is a point in the projective space P^3 or P^q that is how we can you know call it and this point is mapped a point in the 2 dimensional projective space which is given by $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$, 1 is a scale factor. You can see that the trick here in this representation, since this fX and fY if I divided by Z then we get the actual image coordinates.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} fx \\ fy \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

So, the above matrix multiplication will provide you this representation of fX, fY, Z in the projective space in a 2 dimensional projective space which is equivalent to the image point. That is why, this mapping can be expressed as a linear mapping in the projective space and it is a mapping between a 3 dimensional projective space to a 2 dimensional projective space.

So, that is the difference between homography, what we discussed in earlier lectures, where a homography was defined in our context as a mapping from a 2 dimensional projective space to another 2 dimensional projective space that is one difference, here it is from 3 dimensional projective space to a 2 dimensional projective space.

The other difference was that, a homography was invertible. So, mapping was unique 1 to 1, but in this case as you know that a 3 dimensional point now, it is mapped to a unique image point, but it is not true for the image point when it is again projected back to the 3 dimensional space, then it becomes a projected ray and the any point in that ray is represent represented by this image point. That is why, this mapping is not invertible and you can see also from the dimension of this matrix, it is not square matrix that is one of the conditions where you should have an invertible transformation for a linear transformation of course, it should be non-singular also.

But in any case, since it is 3×4 matrix; so we can say that this mapping is not an invertible mapping. So, this matrix is called projection matrix and this matrix can be also represented in this form, you can see that it is formed by composition of two matrices,

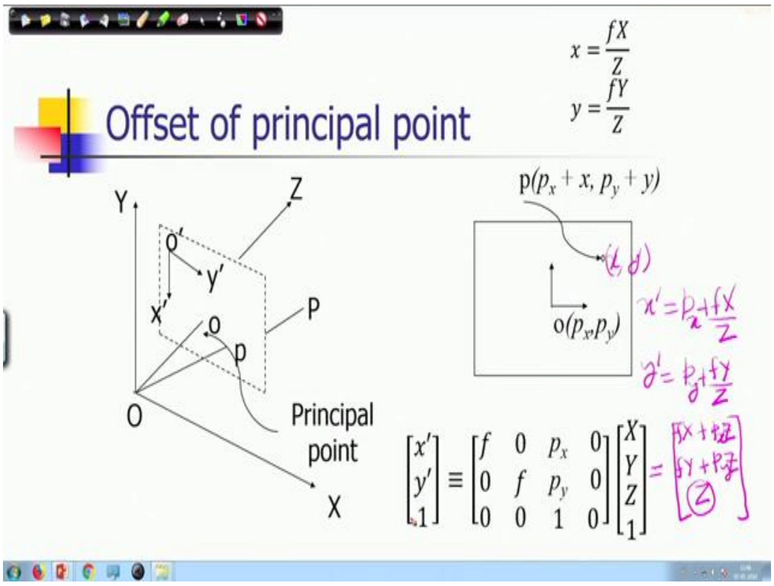
$$P = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = diag(f, f, 1)[I \mid 0]$$

Where one is a 3×3 matrix. The other one is 3×4 matrix and the corresponding the structure of the matrix are very simple and you note the notation what you are using for representing this matrix, this particular matrix which contains information of the focal length that is represented in a short form, we will be using this notation because it is a diagonal matrix we will be simply listing its corresponding diagonal elements.

Whereas, for representing the other part of this you know the right hand side matrix in this component. So, there are two sub matrixes, one sub matrix is an identity matrix. So, this is a 3×3 identity matrix which is shown here in this form and the other sub matrix is a column vector which is a 0 column vector which is also shown here in this representation.

So, $diag(f, f, 1)[I \mid 0]$ is a short form of representation of a projective projection matrix, there are motivations for representing in this form that would be clear when we perform the corresponding different mathematical operations on this matrices. So, let us proceed and we will understand why we are taking this kind of representations.

(Refer Slide Time: 11:50)



So, we will try to generalize this coordinate mapping considering different other scenarios. So, in the canonical form as we discussed earlier, we have considered the projection plane or the image plane the axes in the image plane they should be parallel to the axes what are used for the 3 dimensional coordinate systems. So, there are two coordinate systems in this configuration, one coordinate system is the 3 dimensional coordinate system which is also called world coordinate system. The other coordinate system which is called image coordinate system. So, this image coordinate system also has its own coordinate convention and that is the representation of the corresponding 2 dimensional image plane and you can also considered that is represented by a 2 dimensional projective space.

So, it has its origin here and in this convention in the 2 dimensional real 2 dimensional space. So, it has its origin and this origin is formed as I mentioned as an intersection of the optical axis or principal axis which is a Z axis with the image plane. So, that is how this origin is formed and then axis are parallel I told you.

So, this is the canonical configuration and in this case we are assuming of course, image plane is at a distance of focal length f in the further canonical form, even we can normalize it at a distance 1, we will discuss in the previous representation itself. We have seen that you have taken out the focal length outside of the projection matrix and the other right hand side component which contains only identity matrix and the 0 column vector.

So, if I use only that projection matrix representation, then that is the minimal representation of this minimal configuration of this imaging where projection plane is also taken as a unit distance from the center of projection. let us consider, we can have a different coordinate system in the image plane, we have origin; we can have origin at different positions say. So, in the canonical form what I discussed here when the image plane is at a distance of f from the principle plane, then we consider that x and y mapping of a 3 dimensional point X, Y, Z to a image points x, y are shown by this equation. So, that we have discussed already these are the two equations.

Now, so, when you consider a coordinate axis when it is a different coordinate system; that means, origin maybe shifted at some other point, but we may consider our principal, our coordinate axis, they still remain parallel to X axis and Y axis. So, when the origin has been shifted to a different point so, the principal point will have a coordinate which is other than 0, of the image plane. Suppose, this is p_x and p_y so, any point in a image plane say; any point in image plane, it will be now represented as ($p_x + x$) and ($p_y + y$).So, in this case, if I consider this as the new coordinate system say this is $x^{'}$ and $y^{'}$ under this transformation of the origin or translation of the origin, then

$$x^{'} = p_x + \frac{fX}{Z} \qquad y^{'} = p_y + \frac{fY}{Z}$$

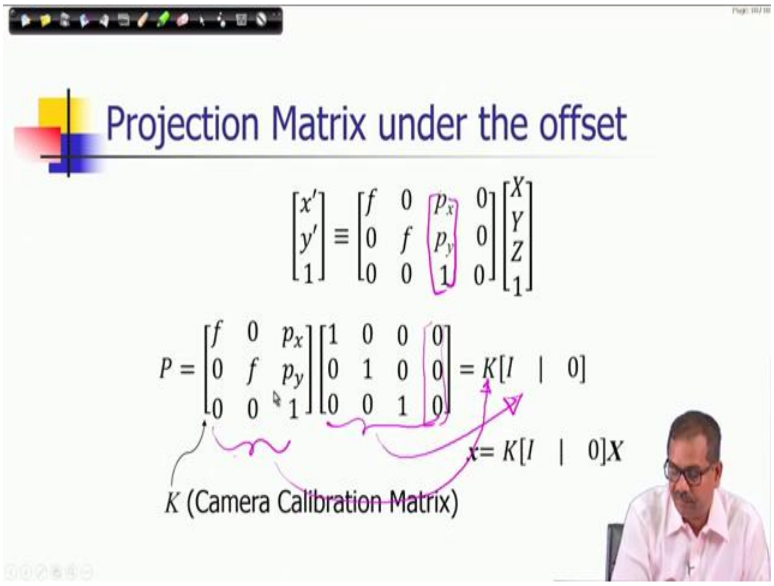
So, this is how this coordinates should be interpreted and in the matrix form if I transfer this relation, then we can write this relation simply in this form; that means, the projection matrix as you can see the third column of the projection matrix will have will be change from 0, 0, 1 to p x, p y. 1, you can you can verify the corresponding expressions we will see that the x prime is equal to fX.

So, this these operation will give you fX plus p x so, if I simplify this matrix multiplication, we will find out these operation will give me

$$\begin{bmatrix} f_x + p_x Z \\ f_y + p_y Z \\ Z \end{bmatrix}$$

. So, this is the 2 dimensional point of the 2 dimensional projective space and Z is the scale factor. So, finally, when I divide this by Z, it would be $p_x + \frac{fX}{Z}$ and similarly for y coordinate, $p_y + \frac{fY}{Z}$. So, this is equivalent to this coordinate system. So, this is how the shift of the origin in the image plane affects the projection matrix.

(Refer Slide Time: 18:04)



So, this is the form of the projection matrix when there is an offset in the origin of the image plane and this offset is given by p_x, p_y and which forms the third column of this projection matrix so, this particular information is contained in the third column of the projection matrix. So, once again, you can decompose this matrix into the form

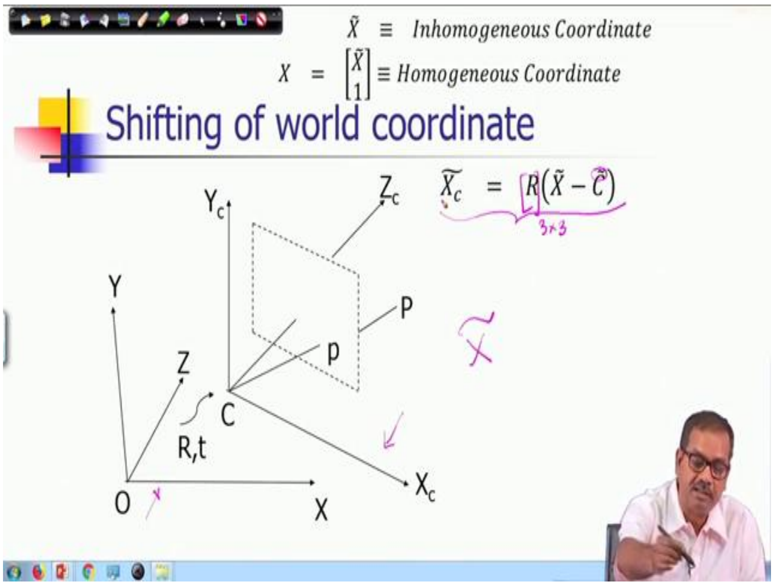
$$P = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = K[I \mid 0]$$

what I discussed earlier, you can see that $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ remains the same that is canonical form of projection matrix when your image plane is at a distance of unity in front of the camera center.

So, this mapping is very simple, it is an identity matrix and you know this column is this 0 vector column and once again this is represented by this particular matrix and K is representing $\begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$. So, this 3×3 matrix is represented by the symbol K here which is called incidentally calibration matrix, we can see that this is camera calibration matrix and you can find out in this matrix, we have some of the parameters which are defining the mapping of the image coordinates, canonical form of image coordinate to its corresponding coordinate what has been observed in what is given to us.

So, what are the elements there? You can see that once again the diagonal elements particularly, it is found by the corresponding focal length it is diagonal [f f 1] that is these are the offset principal. So, finally, the relationship between the image point and the 3 dimensional scene point that can be summarized in the following form: $x = K[I \mid 0]X$ as this is the image point, small this represented by this convention I will be using the small letter to denote any image point and sometimes I will be using the bold font just to denote that it is a column vector. Similarly, the upper cases, they are the world coordinate system and they are all represented in homogenous coordinate system.

(Refer Slide Time: 20:48)



Now, let us consider that your world coordinate system also is different from the camera centric coordinate system; that means, your origin of the world coordinate system is different and also the axis of the world coordinate system there at different orientations which are not parallel to the axes of image plane. So, this is the canonical configuration what we discussed earlier so, this is the canonical form we discussed. So, this is a canonical form, but this is the actual coordinate system, world coordinate system what we will be considering.

So, that first we need to consider the transformation between this coordinate system and that and the corresponding canonical coordinate system. So, there are operations like rotations of axes and translation of origin those are involved in transforming any coordinate to the corresponding canonical world canonical camera centric coordinate system. So, this could be related by $\tilde{X}_c = R(\tilde{X} - \tilde{C})$ where R is given by a rotation matrix which is a 3×3 rotation matrix.

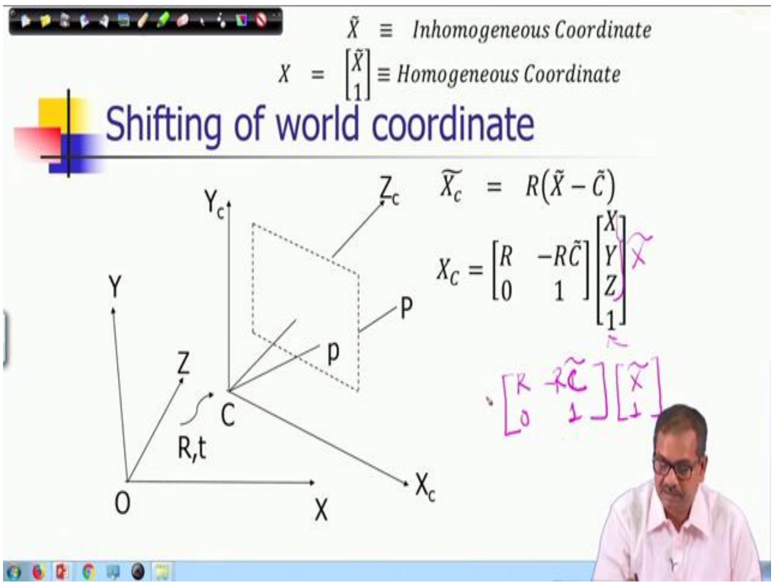
So, first we have to shift the coordinates of the 3 dimensional point. Incidentally, you note that there are two type of representation of coordinate representations in our discussion; one is that representation in the non-homogenous coordinate space or in homogenous coordinate space. So, that is the normal 3 dimensional coordinates what we use in or analytical geometrics and the other representation of the same coordinate point in a 3 dimensional projective space or we call it homogenous coordinate system.

So, in this case this relations, it involves only the coordinates in the usual convention of non-homogenous coordinate system, even the centre is also represented. when I will be using symbol tilde over the point symbol of that point, then I will consider the coordinate representation is a non-homogeneous coordinate representation. Otherwise, if this tilde is not there at the top of that symbol then usually unless I mention that is representating homogenous coordinate system.

So, you can see that the first you have to translate the point from camera center. So, translate the origin by this operation. So, this is the corresponding location of the center. So, first you translate the point to this center and then you apply the corresponding rotation at this point and then you will get the corresponding transformed camera coordinate system as in the camera coordinate system. So, this is the transformation which is well explained by any analytical geometry book .you can t refer to it.

So, R is in the structure of it is a 3×3 matrix which is which is called rotation matrix and . So, any coordinate X , X' , it can be expressed in the coordinate system camera centric coordinate system by this particular operation. So, now, what we need to do? We need to apply this transformation of X to this canonical coordinates camera centric coordinates system and then we can apply the same projection matrix.

(Refer Slide Time: 24:50)



So, which can be expressed in this form so, you can see that in this expression, I can express this transformation of a coordinate system in the world coordinate the camera centric coordinate system, all are expressed in the form of a 3 dimensional or homogenous coordinate part.

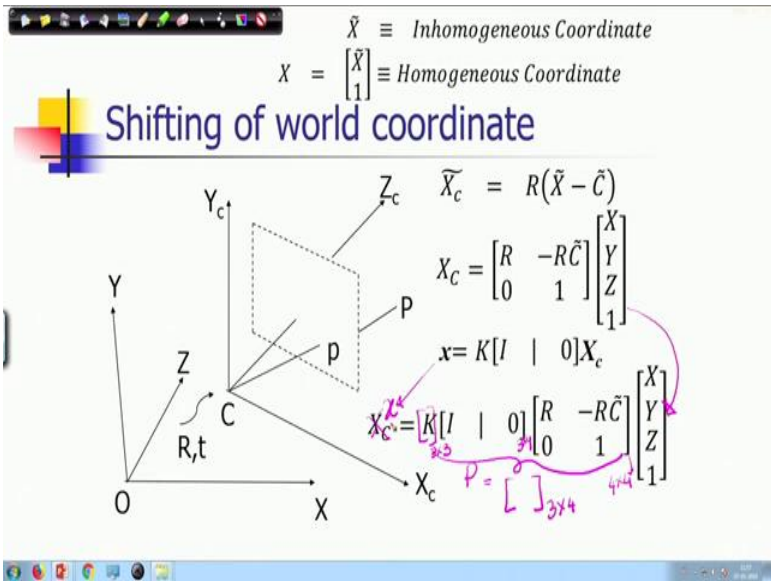
$$X_c = \begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

You notice that this is the $\tilde{X}$ part that is the non-homogenous coordinate representation if I add another additional dimension, then it becomes homogenous representation and this operation so, this $R(\tilde{X} - \tilde{C})$ which is explained by this sub matrix multiplication.

$$\begin{bmatrix} R & -R\tilde{C} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{X} \\ 1 \end{bmatrix}$$

Where $\tilde{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$

(Refer Slide Time: 25:59)



We can apply the relationship with the projection matrix, with the camera centric system. So, you know already that when your coordinate convention is the camera centric coordinate convention that is a canonical form once again we discussed while deriving projection matrix.

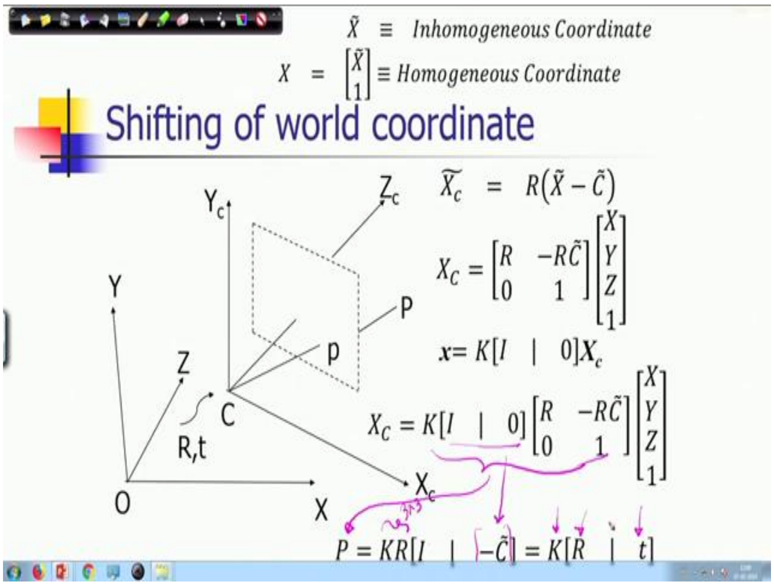
The relationship between the image point and the corresponding coordinate in the camera centric coordinate system is given by $x = K[I \mid 0]X_c$. So, once again this K is the camera calibration matrix which is a 3×3 matrix and I is an identity matrix and 0 is a corresponding column vector and this can be further expanded. So, this X_c again can be related with the 3 dimensional coordinate system which is observed 3 dimensional coordinate system in this form.

$$x = K[I \mid 0]\begin{bmatrix} R & -R\widetilde{C} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

So, finally, this is a mapping of point ,So, finally, you see that the whole expression $K[I \mid 0]\begin{bmatrix} R & -R\widetilde{C} \\ 0 & 1 \end{bmatrix}$, it provides you a corresponding cascade of matrices and if I multiply all these matrices, this is a matrix which will be of dimension 3×4 because K is of dimension 3×3, this is of dimension 3×4 and this is of dimension 4×4. So, if I perform matrix multiplication and this matrix, we represent by this a symbol P which is called projection matrix.

So, this projection matrix of 3×4 so, if I multiply once again this projection matrix with the 3 dimensional coordinates in the projective space of a same point, then I will get its corresponding image point once again in a 2 dimensional projective space. So, that is how, we derive the relationship in a more general case.

(Refer Slide Time: 28:02)



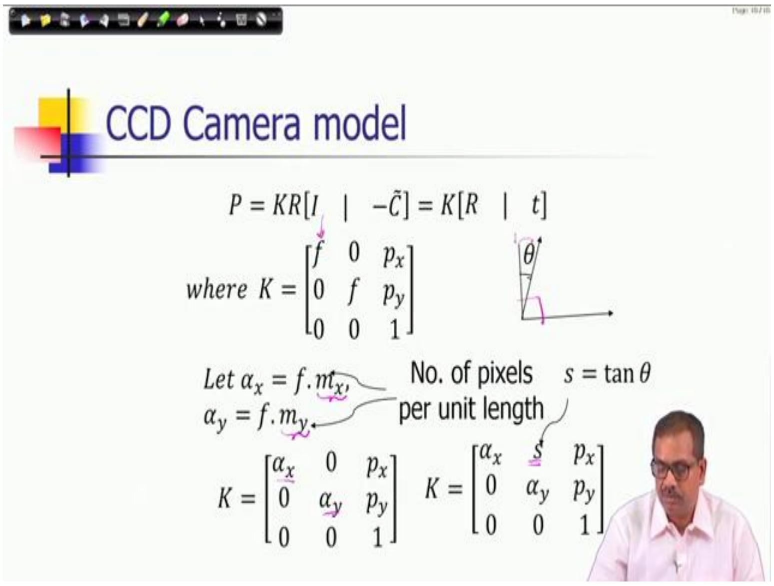
So, to summarize this representation once again in a brief form, we are showing here. So, as I mentioned so, this matrix is consider, this is as their projection matrix which is shown here in the symbol P and it and the whole you know composition can be shown in different way.

Suppose, I am taking out the rotation matrix out of this operation so, I can also equivalently represent this matrix as $P = KR[I \mid -\widetilde{C}]$. So, in this case, you will see K R, this is a 3×3 matrix because K is a 3×3 matrix and R is also 3×3 matrix. So, it is a 3×3 matrix and once again this is the identity matrix and here instead of a 0 column vector of this form, what we get here? We get negative of the coordinates of the camera center that is very interesting.

So, we can from the projection matrix elements itself we can determine the position of the camera center in the world coordinate system there l we will come to that part in my next slides that computations will be elaborated. The other form of representations once again you see that if I only take K out of this box and put R inside the this part 3×4 matrix, then it has two components one is the rotation matrix and the other one is a translation.

So, this gives you the corresponding coordinate transformation parameters of the world coordinate to the camera centric coordinate system; that means, you can get the rotation matrix, you can get the translation of the coordinate system origin translation of the origins. If I can find out what is the matrix K so, from there we can get all this components. So, this forms are interesting because this can give you different information from the projection matrix itself.

(Refer Slide Time: 30:21)



So, when we observe a digital image, actually we observe the pixel coordinates and how a pixel is defined? A pixel is defined by an element by a sensor and which is usually a CCD sensor for a digital camera, it could be also CMOS sensor there are different sensors, but each pixel is information of each pixel is captured by a sensor. So, we usually called this arrangement of pixels as a CCD camera arrangement.

So, in the CCD camera arrangement if I consider that configuration, then we need to also consider the parameters involved in digitization of the points in that corresponding you know sensor plane. So, there is a resolution involved there so, let us see that we have already shown that how a projection matrix is component of projection matrix we have shown without considering the parameters of a CCD camera, but let us consider that per unit length there are m_x number of sensors.

So, m_x is the sensors number of sensors in the horizontal directions and m_y is the number of sensors in the vertical directions. So, this is how a unit length has been covered by that many number of sensors. So, if I consider a length f in the calibration matrix so, because this is f is a multiplication factor of that length. So, in our case, we will consider the f can be replaced by this transformation as $\alpha_x = fm_x$ and $\alpha_y = fm_y$ in their respective places.

So, which means that our camera calibration matrix when we are considering this number of pixels per unit length along X and Y axis then it will look like

$$K = \begin{bmatrix} \alpha_x & 0 & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

So, these are the diagonal elements instead of f, we have now α_x, α_y which takes care of the resolutions of the pixel resolutions of the camera. And then there is also another issue that the alignent of the sensors in the horizontal direction and vertical directions may not be always rectilinear.

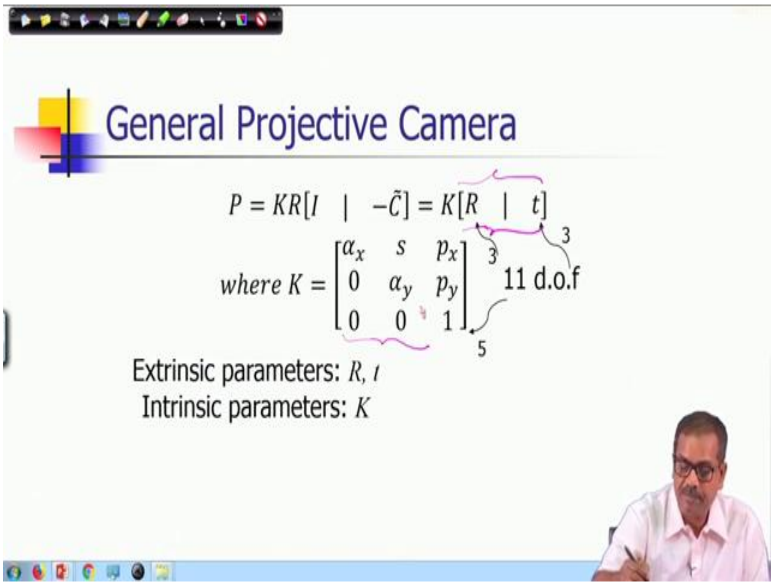
There would be small deviations because of the manufacturing, very small though still that should be accounted for. So, let us considered that instead of the angle it forms 90 degree, there is a small deviation. So, it forms an angle θ from this perpendicular direction.

So, this is providing a kind of a skewed coordinate system. So, we need to take care of that, we need to correct that skew and this is taken care of by another introducing another parameter into the calibration matrix which is shown here,

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

which is given by s which can be derived from the corresponding coordinate transformation, you can also verify that. s would be $\tan\theta$ or even you can say $\sin\theta$; θ is very small. So, you can write it as $s = \theta$ also in the radian unit.

(Refer Slide Time: 34:01)



So, this is the final form of the calibration matrix and then a general projective camera matrix can be can be described when I consider all sorts of parameters, all sorts of transformation in this form that it consists of the elements these are the corresponding elements like K has 5 elements like α_x, α_y, s, p_x, p_y and also the other component s which contains the other parameters rotation and translation.

So, these parameters are grouped into different forms. So, there are 11 degrees of freedom in this particular projection matrix because it is associated with a scale factor and so, there are 12 elements. So, considering the scale factor so, there are 11 parameters and there are 11 independent parameters.

So, 3 of them will denote a rotation matrix which is a rotations along 3 axis, then 3 will denote the translations of a 3 dimensional coordinate of the origin to the camera center and 5 parameters are related to the calibration matrix which are given by this elements α_x, α_y s , p_x , p_y . So, in total there are 11 independent parameters and out of them, the extrinsic parameters which are related to the transformation of world coordinate to the camera centric coordinate system, those are given by this matrix that is the right side matrix of 3×4 matrix and the intrinsic parameters are inside the camera calibration matrix. So, these 5 parameters are called intrinsic parameters. So, with this let me stop here and we will continue our discussion in the next lecture.

Thank you for your listening.