Digital Image Processing

Prof. P. K. Biswas

Department of Electronics and Electrical Communication Engineering

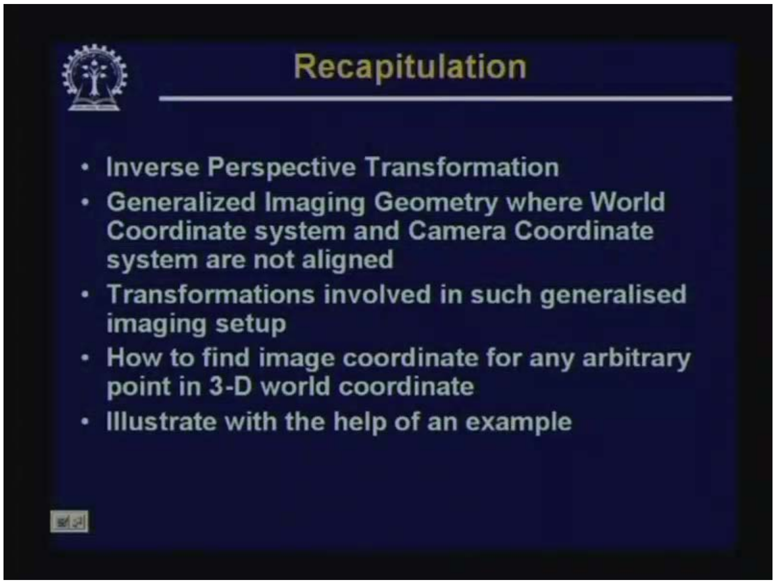
Indian Institute of Technology, Kharagpur

Lecture - 8

Camera Calibration and Stereo Imaging

Hello, welcome to the video lecture series on digital image processing. Till the last class, we have seen that given a particular type of imaging geometry where a camera is placed in a 3D, in a 3D coordinate system where the coordinate system and the camera coordinate system and not perfectly aligned; in that case, what are the set of transformations which are to be applied to the points in the 3D world coordinate system which will be transformed in the form as seen by a camera. Then, followed by that if we apply the perspective transformation; then we get the image coordinate for different points in the 3D world coordinate system.

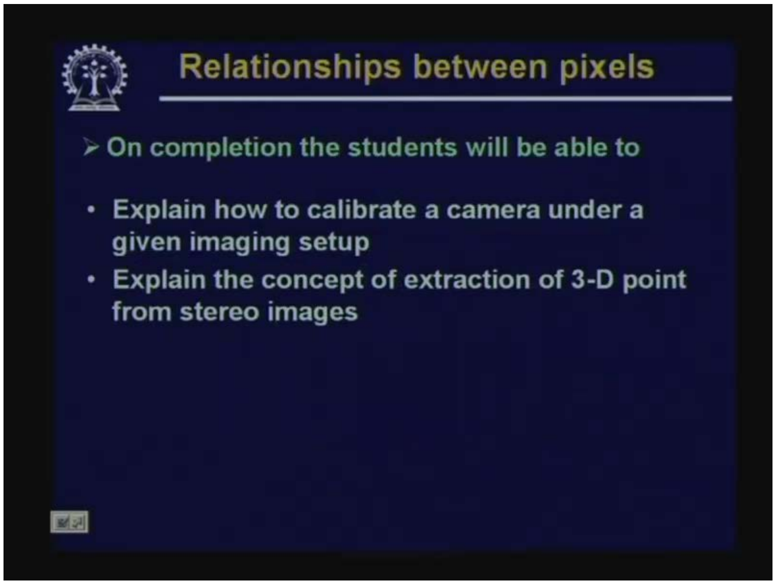
(Refer Slide Time: 1:52)



So, what we have seen in the last class is once we have the image points in an image plane; how to applied the inverse perspective transformation to get the equation of the straight line so that the points on that straight line map to a particular image point on the imagining plane.

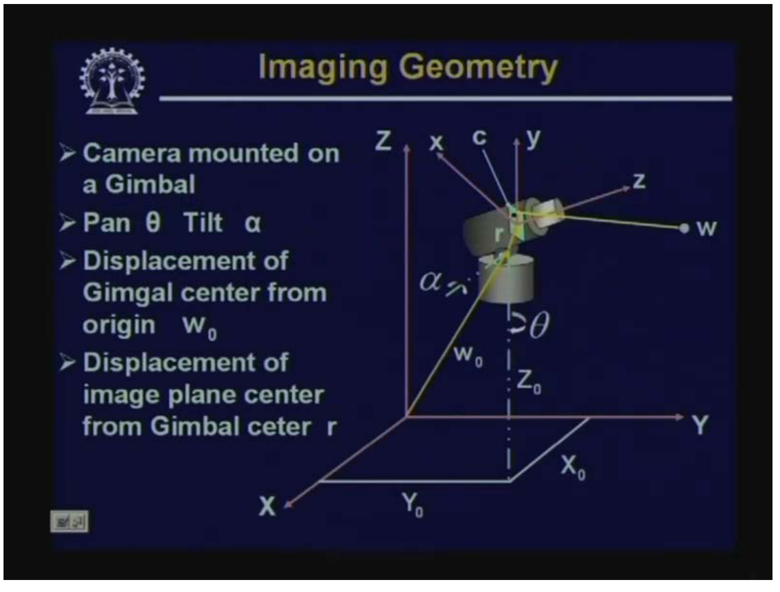
Then we have seen a generalized imaging geometry where the world coordinate system and the camera coordinate system are not aligned and we have also discussed the set of transformations which are involved in such generalized imaging setup and then we have also seen how to find the image coordinate for any arbitrary point in the 3 world coordinate system in such a generalized imaging setup and the concept, we have illustrated with the help of an example.

(Refer Slide Time: 2:48)



In today's lecture we will see that given an imaging setup; how to calibrate the camera and then we will also explain the concept of how to extract the 3D point from 2 images which is also known as stereo images.

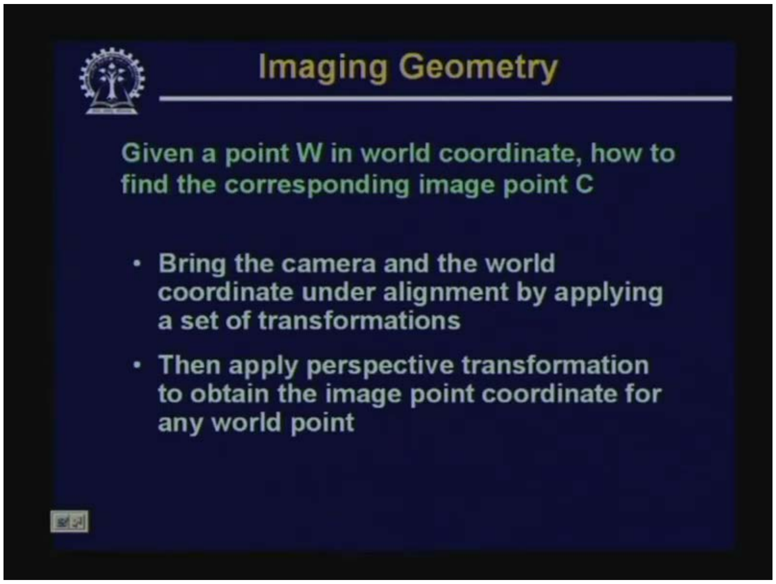
(Refer Slide Time: 3:09)



So, in the last class that we have done is we have given an imaging setup like this where the 3D world coordinate system is given by capital X, capital Y and capital Z. In this world coordinate system, we had placed the camera where that camera coordinate system is given by small x, small y and small z and we have assumed that the camera is placed is mounted on a Gimbal where the Gimbal is displaced from the origin of the world coordinate system by a vector W_0 and the center of the camera is displaced from the Gimbal by a vector r.

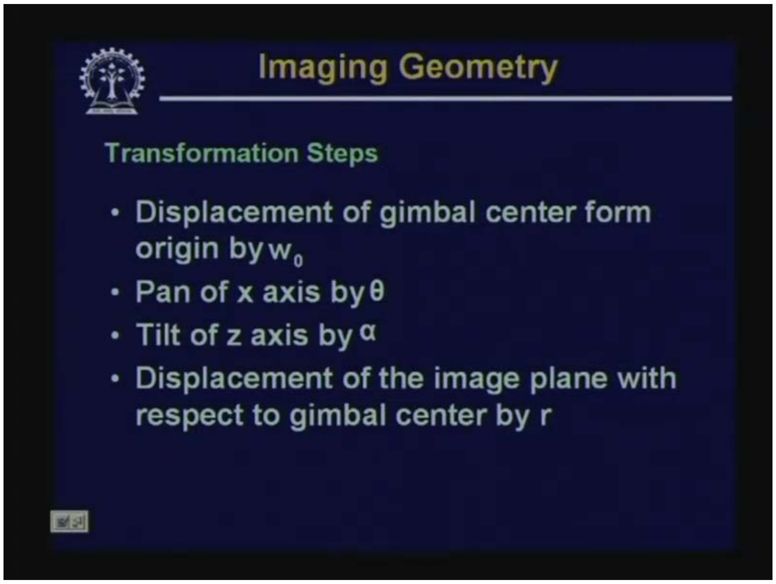
The camera is given a pan of angle theta and it is also given a tilt of angle alpha and in such a situation if W is a point in the 3D world coordinate system, we have seen that how to find out the corresponding image point corresponding to point W in the image plane of the camera.

(Refer Slide Time: 4:20)



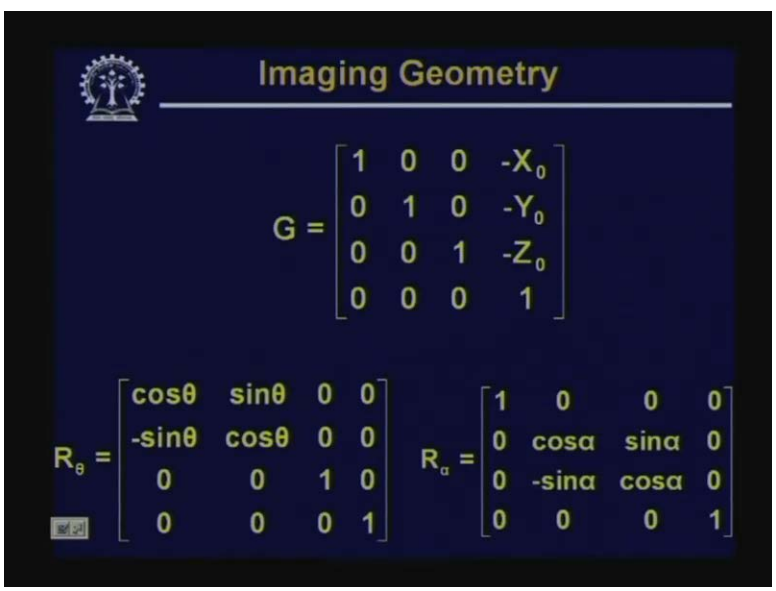
So, for that we have done a set of transformations and with the help of the set of transformations, what we have done is we have brought the 3D world coordinate system and the camera coordinate system in alignment and after the 3D world coordinate system and the camera coordinate system are perfectly aligned with that set of transformations, then we have seen that we can find out the image point corresponding to any 3D world point by applying the perspective transformation.

(Refer Slide Time: 4:55)



So, the type of transformations that we have to apply is first we have to apply a transformation for the displacement of the Gimbal center from the origin of the 3D world coordinate system by vector W_0 followed by a transformation corresponding to the pan of X axis of the camera coordinate system by theta which is to be followed by a transformation corresponding to tilt of the Z axis of the camera coordinate system by angle alpha and finally, the displacement of the camera image plane with respect to Gimbal center by vector r.

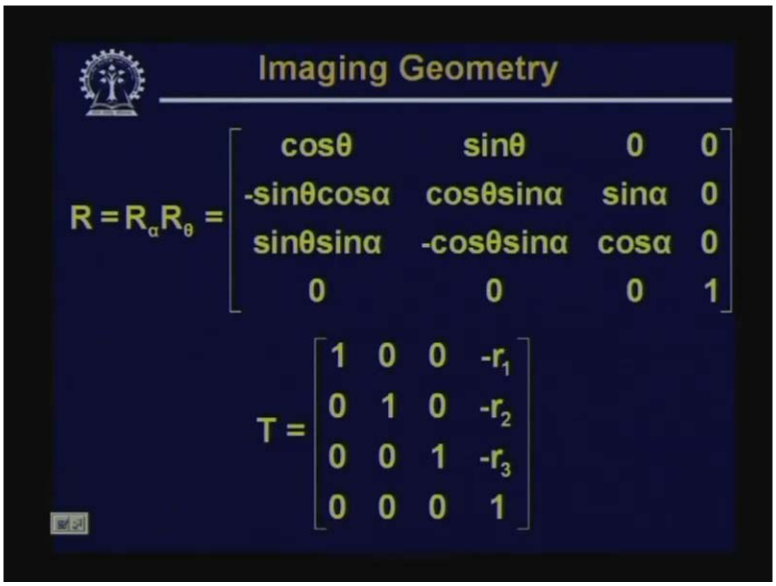
(Refer Slide Time: 5:37)



So the transformations, the first transformation which translates the Gimbal center from the origin of the world coordinate system by vector W_0 is given by the transformation matrix G which is in this case (1 0 0 minus X_0) (0 1 0 minus Y_0) (0 0 1 minus Z_0) and (0 0 0 1). The pan of the X axis of the camera coordinate system by an angle theta is given by the transformation matrix R theta where R theta in this case is (cosine theta sin theta 0 0) then (minus sin theta cosine theta 0 0) then (0 0 1 0) and (0 0 0 1).

Similarly the tilt, the transformation matrix corresponding to the tilt by an angle alpha is given by the other transformation matrix R alpha which in this case is (1 0 0 0), (0 cosine alpha sin alpha 0) then (0 minus sin alpha cosine alpha 0) and then (0 0 0 1).

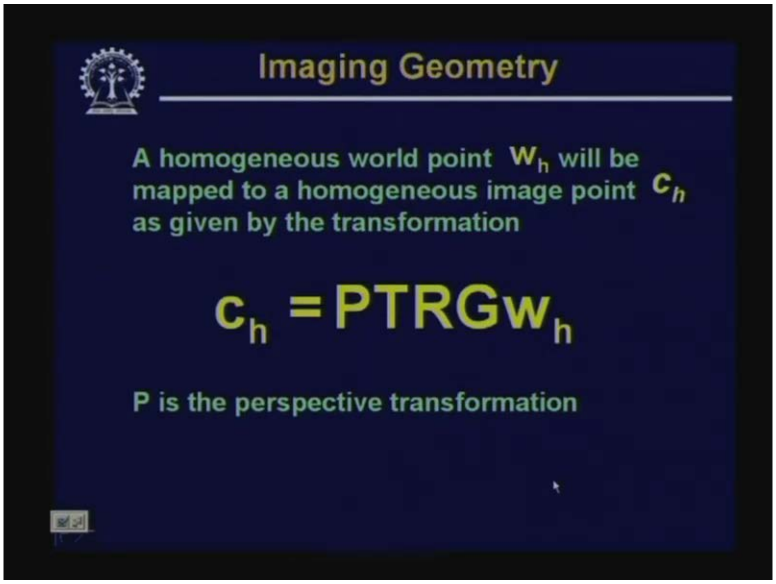
(Refer Slide Time: 6:58)



Then the next transformation we have to apply is the transformation of the center of the image plane with respect to Gimbal center by the vector R. So, if we assume that R has the components, the vector R has components - r_1 r_2 and r_3 in XY along the X direction, Y direction and Z direction of the 3D world coordinate system; then corresponding transformation matrix with respect to this translation is given by T equal to (1 0 0 minus r_1) (0 1 0 minus r_2) (0 0 1 minus r_3) and then (0 0 0 1).

We have also seen in the last class that the rotation matrices R theta and R alpha can be combined together to give a single transformation matrix R which is nothing but the product of R alpha and R theta.

(Refer Slide Time: 8:06)



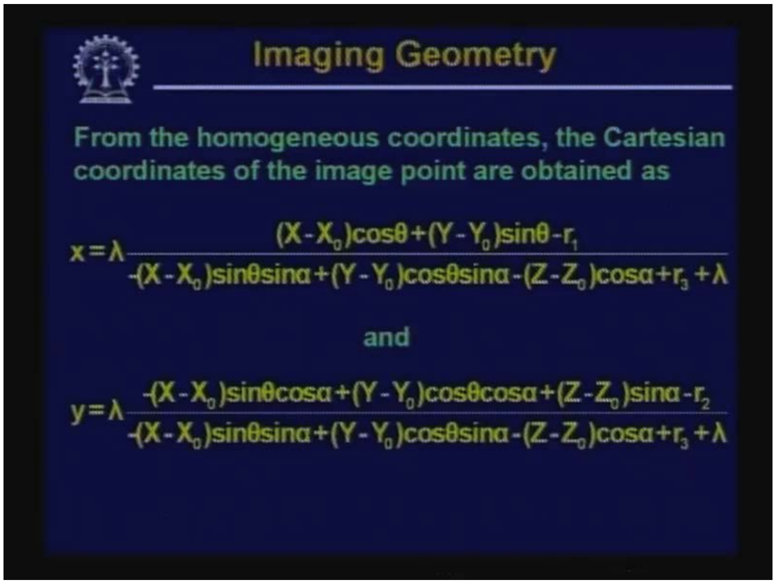
Now, once we get this transformation matrices, then after the transformation first by the translation matrix G, then by the rotation matrix R, followed by the second translation matrix T; what we do is we aligned the coordinate system of the camera with the 3D world coordinate system that means now every point in the 3D world will have a transformed coordinate as seen by the camera coordinate system.

So, once we do this, then finally applying the perspective transformation to these 3 dimensional world coordinate systems gives us the coordinate of the point in the image plane for any point in the 3D world coordinate system. So, here you have find the final form of the expression is like this that both of the world coordinate system and the camera coordinate system in this case, they are represented in the homogeneous form.

So, w_h is the homogenous coordinate corresponding to world coordinate W and c_h is the homogenous form of the image coordinate C. So, for a world point W whose homogenous coordinate is represented by w_h, here you find that I can find out the image coordinate of the point W again in the homogenous form which is given by this matrix equation that c_h is equal to PTRG and w_h.

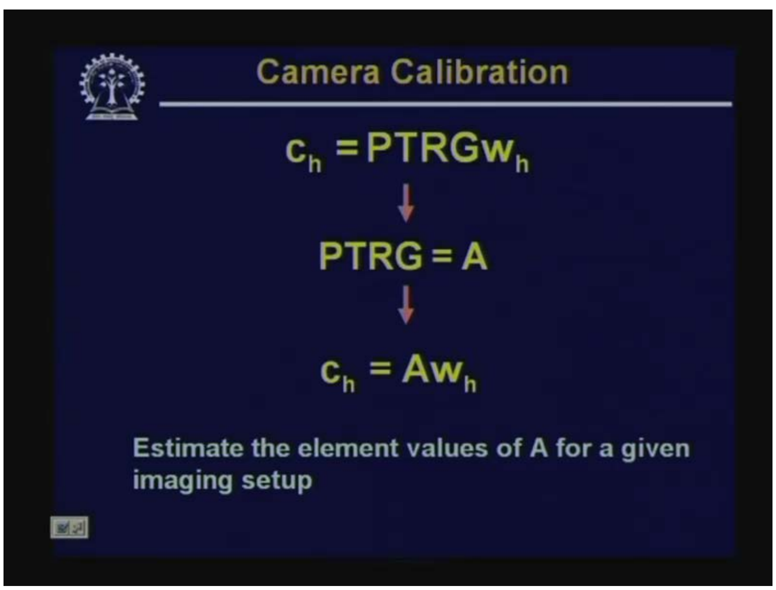
And, here you note that each of these transformation matrices that is P, T, R and G all of them are of dimension 4 by 4. So, when I multiply all these matrices together to give a single transformation matrix, then the dimension of that transformation matrix will also be 4 by 4.

(Refer Slide Time: 10:13)



So, what we have now is of this form. So, after doing this transformation, I can find out the image coordinate of the corresponding point W where X and Y coordinates will be given by these expressions.

(Refer Slide Time: 10:28)



So, after doing this what I have is I have an equation, transformation equation or a matrix equation so that for any world point W, I can find out what is the homogenous coordinate of homogenous coordinate of the image point corresponding to that particular point W.

Now, the transformation which is involved that is P, T, R and G as we said that each of these transformation matrices are of dimension 4 by 4; so the combined transformation matrix if I represent this by a matrix A, then this matrix A will also be a 4 by 4 matrix and now the inter transformation equation in matrix form will be c_h is equal to A into w_h.

Now, find that given a particular setup, the transformations T, R and G, they depend up on the imaging setup. So, G corresponds to translation of the Gimbal center from the origin of the 3D world coordinate system, R corresponds to pan angle and the tilt angle and T corresponds to translation of the image plane center from the Gimbal center.

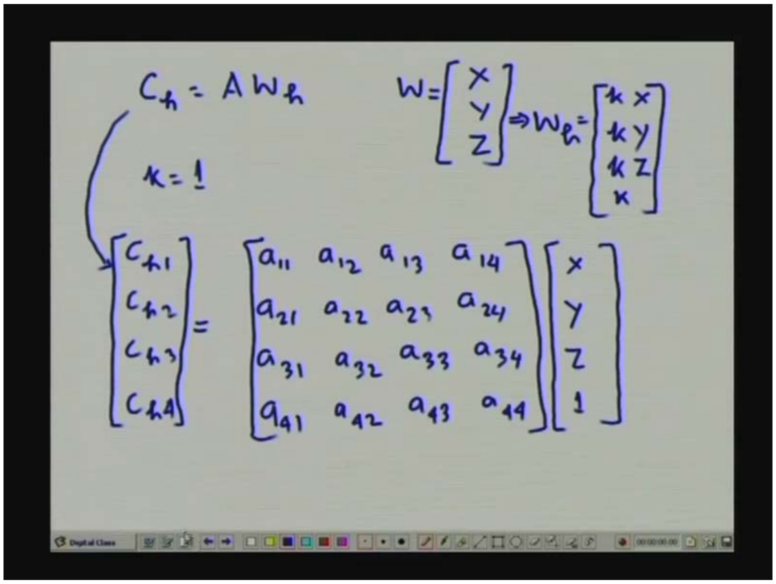
So, these 3 transformation matrices depend up on the geometry of the imaging system. Whereas, the other transformation matrix that is P or perspective transformation matrix, this is entirely a property of the camera because we will find that the components of this transformation matrix P has a term lambda which is equal to the wave length the focal length of the camera.

So, it is possible that for a given camera for which the focal length lambda is known, I can find out what is the corresponding perspective transformation matrix P. Whereas, to find out the other transformation matrices like T, R and G, I have to do the measurement physically that what is the translation of the Gimble center from the origin of the 3D world coordinate system, what is the pan angle, what is the tilt angle. I also have to measure physically that what is the displacement of the image center, image plane center from the Gimbal center.

And in many cases, measuring these quantities is not very easy and it is more difficult if the imaging setup is changed quite frequently. So, in such cases, it is always better that you first have an imaging setup and then try to calibrate the imaging setup with the help of the images of some known points of 3D objects that will be obtained with the help of the same imaging setup.
So by calibration, what I mean is as we said that now I have a combined transformation matrix for the given imaging setup which is A which is nothing but the product of PTR and G. So, these being a 4 by 4 matrix, what I have to do is I have to estimate the different element values of this matrix A. So, if I can estimate the different element values of the total transformation matrix A from some known images, then given any other point in the 3D, I can find out what will be the corresponding image point.

Not only that if I have an image point, a point in the image by applying the inverse transformation, I can find out what will be the equation of the straight line on which the corresponding world point will be lying. So, this calibration means that we have to estimate the different values of this matrix A. Now, let us see how we can estimate these values of the matrix A.

(Refer Slide Time: 14:41)



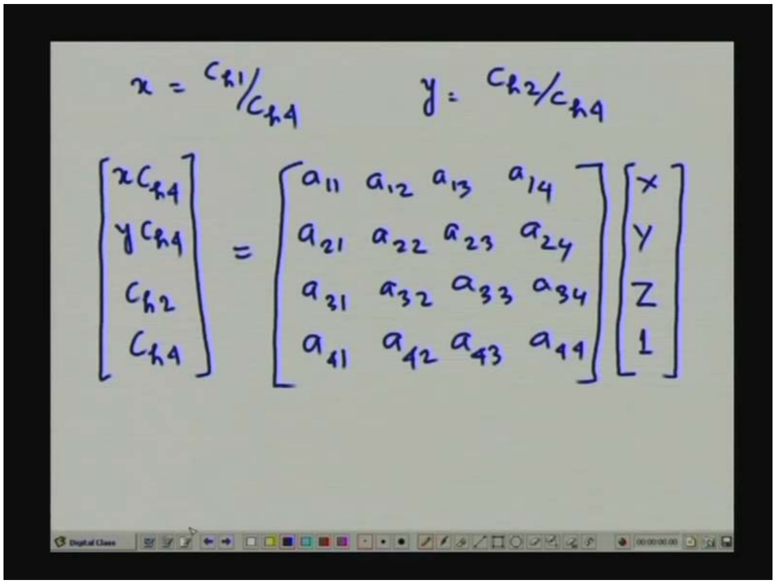
So, here you find that we have this matrix equation which is of this form. That is c_h is equal to A into w_h where we have said, the w_h is the world coordinate of the 3D point put in homogenous form and c_h is the image point on the image plane again in the homogenous form and A is the total transformation matrix.

So here, if the world point W has the coordinate say, X Y and Z, the corresponding homogenous coordinate system will be given by w_h is equal to some constant k times X, some constant k times Y, some constant k times Z and the 4th element will be K. So, this will be the homogenous coordinate w_h corresponding to the point W.

Now, without any loss of generality, I can assume the value of k equal to 1. So, if I take k equal to 1 and if I expand this matrix equation; then what I get is I get the component say c_{h1} c_{h2} c_{h3} c_{h4}. This will be, now I expand the matrix A also. So, A will have component a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22} a_{23}, a_{24}, a_{31}, a_{32}, a_{33}, a_{34} then a_{41}, a_{42}, a_{43}, a_{44} into the homogenous coordinate of the point in the 3D space which is now XYZ and 1.

So, you remember that we have assumed the value of k to be equal to 1. So, I get a matrix equation like this. Now, from this matrix equation, I have to find out or I have to estimate the component values a_{11}, a_{12}, a_{13} and so on.

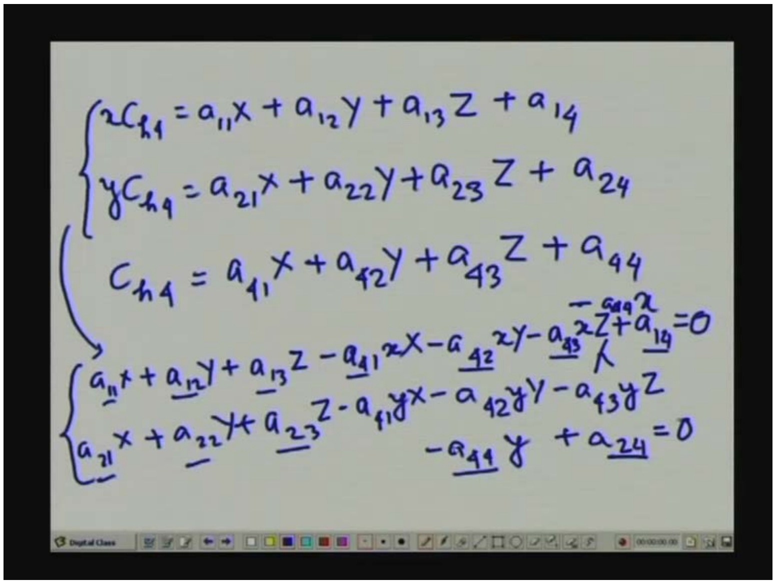
(Refer Slide Time: 17:40)



Now here, once I have the homogenous image coordinates as c $_{h1}$ c $_{h2}$ ch 3 and c $_{h4}$; then we have already discussed that the corresponding Cartesian coordinate in the image plane is given by x equal to c_{h1} divided by c_{h4} and y is given by c_{h2} divided by c_{h4}.

So, this is a simply conversion from the homogenous coordinate system to the Cartesian coordinate system. Now here, if I replace the values of c_{h1} and c_{h2} by X time c_{h4} and Y time c_{h4} in our matrix equation; then the matrix equation will look like xc_{h4}, yc_{h4} then c_{h2} let it remain as it is, then finally we have c $_{h4}$ this will be equal to a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22}, a_{23}, a_{24}, a_{31}, a_{32}, a_{33}, a_{34} a_{41}, a_{42}, a_{43}, a_{44} multiplied by the 3D point coordinate in homogenous form which is (X Y Z 1).

(Refer Slide Time: 19:34)



So, if I expand this matrix equation, what I get is xc_{h4} will be given by a_{11} X plus a_{12} Y plus a_{13} Z plus a_{14}, then yc_{h4} will be equal to a_{21} X plus a_{22} Y plus a_{23} Z plus a_{24} and c_{h4} is given by a_{41} X plus a_{42} Y plus a_{43} Z plus a_{44}. Now, find that while doing this matrix equation or while trying to solve these matrix equations, we have ignored the third component in the image the image point. That is because the third component corresponds to the Z value and we have said that for this kind of calculation, the Z value is not important to us.

Now, from these given 3 equations, what we can do is we can find out what is the value of c_{h4} in terms of XYZ and if I replace this value of c_{h4} in the earlier 2 equations, then these 2 equations will simply be converted in the form a_{11} X plus a_{12} Y plus a_{13} Z minus a_{41} x small x capital X minus a_{42} small x capital Y minus a_{43} small x capital Z plus a_{14}, this is equal to 0 and a_{21} capital X plus a_{22} capital Y plus a_{23} capital Z minus a_{41} small x small y capital X minus a_{42} small y capital Y minus a_{43} small y capital Z plus a_{24}, this is equal to 0.

So, these 2 equations are now converted in this particular form. Now, if you study these 2 equations, you will find that x and y, small x and small y at the coordinates in the image plane of a point in the 3D world coordinate system whose coordinates are given by capital X capital Y and capital Z. So, if I take a set of images for which the point in the 3D world coordinated system that is capital X, capital Y and capital Z are known and I also find out what is the corresponding image point image coordinate in the image plane; then for every such pair of readings I get 2 equations. One is the first equation, other one is the second equation.

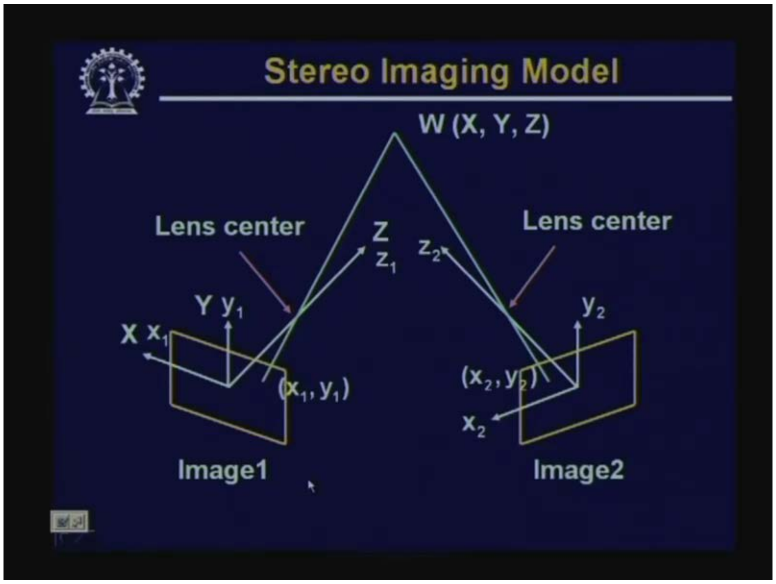
Now, if you study this particular these 2 equations, you will find that there are 6 unknowns. The unknowns are; one is a_{11}, a_{12}, a_{13}, a_{41}, a_{42}, a_{43}, a_{14}, a_{21}, a_{22}, a_{23} then you have a_{24}. So, the number of unknowns we have in these equations are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11; so, 11 or 12? 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, I have missed something?

Sorry, there should be 1 more term, minus here there should be 1 more term - minus a_{44} x and here should be one more term - minus a_{44} y. So, this a_{44}, this is another term. So, there are 12 unknowns. So for solving these 12 unknowns, we need 12 different equations and for every known point in the 3D world, I get 2 equations. So, if I take such images for 6 known points, then I can find out 12 equations and using those 12 equations, I can solve for these 12 unknowns using some numerical techniques.

So once I have the values of these 12 unknowns; so, what I have is the transformation matrix A, the total transformation matrix A using which I can find out what will be the image point of any point in the 3 dimensional coordinate system. And, for given any point in the image plane, I can also find out what will be the equation of a straight line on which the corresponding 3 dimensional world point will exist.

So, this camera calibration using this procedure can be done for any given imaging setup. But the problem still exists that given an imaging point an image point, I cannot uniquely identify what is the location of the 3D world point. So, for identification of the 3D world point or finding out all the 3 X, Y and Z coordinates of a 3D world point, I can make use of another camera.

(Refer Slide Time: 26:25)



So, let us look at a setup like this where on the left side I have image 1 and on the right side I have image 2. The image 1 is taken is captured with help of 1 camera, image 2 is taken with the help of another camera.

So, image 1 has the coordinate system say X_1 Y_1 Z_1, image 2 has the coordinate system X_2 Y_2 Z_2 and we can assume that the 3D world coordinate system that is capital X, capital Y and capital Z is aligned with the left camera. That means the left image coordinate system is same as the 3D world coordinate system, whereas the right image coordinate system is different.

Now, once I have this, given a point W in the 3 dimensional world in the 3 dimensional space, you will find that the corresponding image point in image 1 is given by X_1 Y_1 and the image point for the same point W in image 2 is given by X_2 Y_2.

I assume that both the cameras are identical that means they have the same value of the wave length lambda. So, they will have the same perspective transformation as well as inverse perspective transformation. Now, once I know that in image 1, the image point corresponding to point W is point X_1 Y_1; then by applying the inverse perspective transformation, I can find out the equation of the straight line on which the 3 dimension the point W will exist.

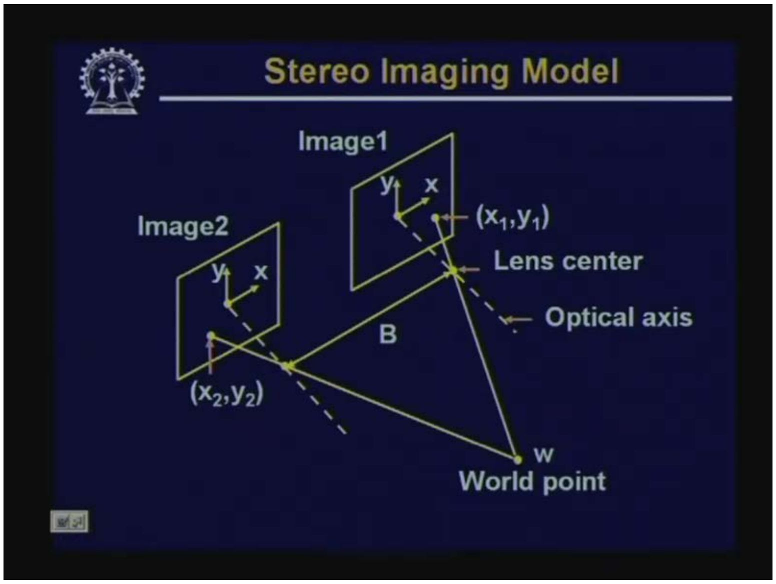
Similarly, from image 2 where I know the location X_2 Y_2 of the image point, if I apply the inverse perspective transformation, I also get equation of another straight line on which this point W will exist. So, now you find that by using these 2 images, I got equations of 2 straight lines. So, if I solve these 2 equations, then the point of intersection of these 2 straight lines gives me the X Y Z coordinate of point W.

But here you find that we have taken a general stereo imaging setup where there is no alignment between the left camera and the right camera or between the first camera and the second camera. So, for doing all the mathematical operations, what we have to do is we have to again apply a set of transformations to one of the camera coordinate systems so that both the camera coordinate systems are aligned.

So, these transformations will again involve may be a transformation for some translation, the transformation for some rotation and possibly it will also employ some transformation for scaling if the image resolution of the image resolution of both the cameras are not same. So, there will be a set of transformations, a number of transformations and the corresponding mathematical operations to align the 2 camera systems.

But here you find the positioning of the camera is in our control. So, why do we consider such a generalized setup? Instead, we can arrange the camera in such a way that we can put the imaging plane of both the cameras to be coplanar and we use the coordinate system in such a way that the x coordinate system and the x of one camera and the x coordinate system of the other camera are perfectly aligned. There will be a displacement in the Y axis and the displacement in the Z axis.

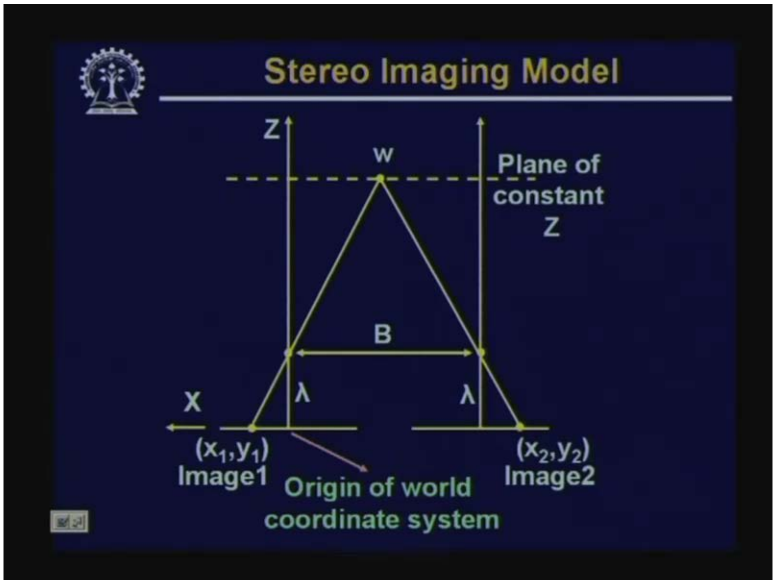
(Refer Slide Time: 30:22)



So effectively, the camera setup that we will be having is something like this. Here you find that for the 2 cameras, the image plane 1 and the image plane 2, they are in the same plane. The X axis of both the cameras, the camera coordinate systems, they are collinear. The Y axis and the Z axis, they have a shift of value B. So, this shift B, this value B is called the camera displacement.

We assume that both the cameras are identical otherwise. That is they have the same resolution, they have the same focal length W. Again, here in the given 3D point W, we have in image 1 the corresponding image point as x_1 y_1 and in image 2, we have the corresponding the mage point as x_2 y_2.

(Refer Slide Time: 31:21)



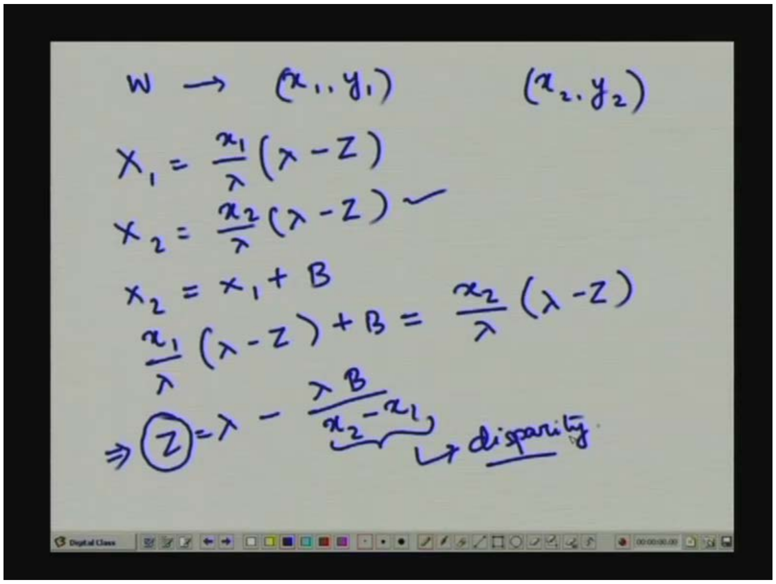
Now, this imaging setup can be seen as a section where you find that these XY planes of both the cameras are now perpendicular to the plane. So, I have X axis which is horizontal, the Z axis which is vertical and the Y axis which is perpendicular to this plane.

So, in this figure, I assume that the camera coordinate system of one of the cameras, in this case the camera 1 which is also called the left camera is aligned with the 3D world coordinate system capital X, capital Y capital Z. The coordinate system of the left camera is assumed to be X_1 Y_1 Z_1, the coordinate of the right camera is assumed to be X_2 Y_2 Z_2.

Now, given this particular imaging setup, you will find that for any particular image point say W with respect to the cameras camera 1 and camera 2; this point W will have the same value of the Z coordinate, it will have the same value of the Y coordinate but it will have different values of the Z coordinate because the cameras are shifted or displaced only in the Z axis not in the X axis or Y axis.

So, origin of this world coordinate system and origin of the left camera system, they are perfectly aligned. Now, taking this particular imaging setup, now I can develop a set of equations.

(Refer Slide Time: 30:03)



So, the set of equations will be something like this. We have seen that for image 1, for point W, the corresponding image point is at location (x_1, y_1); for the same point W in the right image, the image point is at location (x_2, y_2). So, these are the image coordinates in the left camera and the right camera.

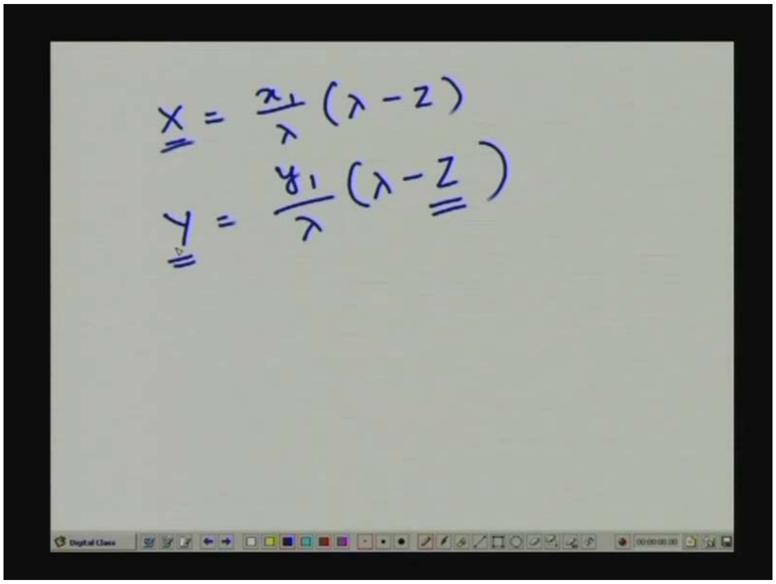
Now, by applying inverse perspective transformation, we find that the equation of straight line with respect to left camera on which point W will lie is given by an equation say X_1 is equal to x_1 by lambda into lambda minus Z. Similarly, with respect to the right camera, the equation of the straight line on which the same point W will exist is given by X_2 is equal to x_2 by lambda into lambda minus Z where this X_1 is the X coordinate of the point W with respect to the camera coordinate of camera 1 and capital X_2 is the X coordinate of the 3D point W with respects to the camera of the second camera.

Now, recollect the figure that we have shown that is the arrangement of the camera where the cameras are displaced by a displacement B. So, with respect to that camera arrangement, we can easily find out that the value of X_2 will be simply X_1 plus the displacement B. Now, if I replace if I replace this value of X_2 which is X_1 equal to B in this particular equation, then I get a set of equations which gives X_1 by lambda into lambda minus capital Z plus B which is equal to x_2 by lambda into lambda minus Z. And from this, I get an equation of the form Z equal to lambda minus lambda times B divided by x_2 minus x_1.

So, you find that this Z is the Z coordinate of the 3D point W with respect to the coordinate system of the first camera, it is same as the coordinate system, the Z value with respect to coordinate system of the second camera, it is also the Z value with respect to the 3D world coordinate system. So, that means it gives me that what is the Z value of the 3D point for which the left image point was X_2 Y_2 X_1 Y_1 and the right image point was X_2 Y_2 and I can estimate this value of Z from the knowledge of the wave length lambda, from the knowledge of the displacement between the 2 camera which is B and the from the knowledge of the difference of the X coordinate that is X_2 minus X_1 in the left camera, in the left image and right image.

So, this x_2 minus x_1 - this term, this particular quantity is also known as disparity. So, if I know this disparity for a particular point in the left image and the right image, I know the lambda that is focal length of the camera and I know the displacement between the 2 cameras, I can find out what is the corresponding depth value that is Z.

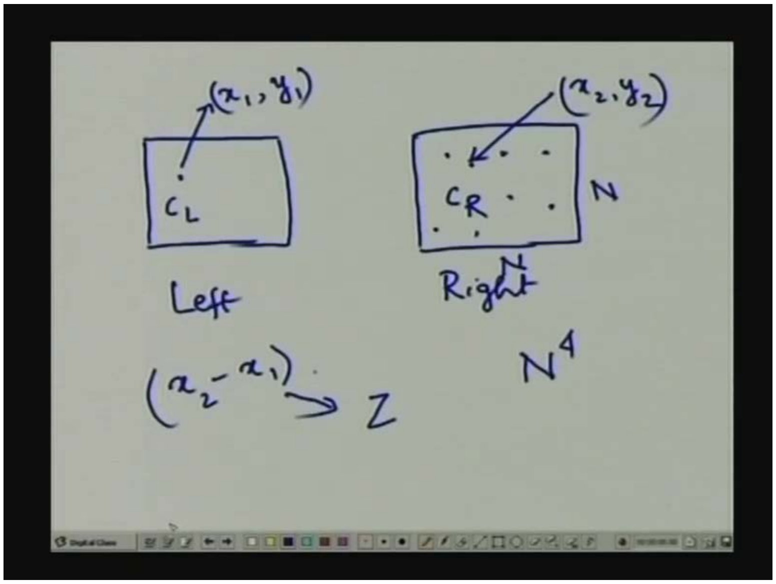
(Refer Slide Time: 37:25)



And once I know this depth value, I can also find out the x coordinate and the y coordinate of the 3D point W with respect to the 3D world coordinate system for which we have already seen. The equations are given by X is equal to x_1 by lambda into lambda minus Z and Y equal to y_1 by lambda into lambda minus Z.

So, first we have computed the value of Z from the knowledge of disparity, camera focal length and the displacement between the cameras and then from this value of Z and the image coordinates in say the left image that is the (x_1 y_1), I can find out what is the X value, X coordinate value and Y coordinate value of that particular 3D point. Now in this, you find that the very very important computation is that given a point in the left image, what will be the corresponding point in the right image.

(Refer Slide Time: 38:33)



So, this is the problem which is called as stereo correspondence problem. So, in today's lecture, we are not going to deal with the details of the stereo correspondence problem that is how do you find out a point in the left image and the corresponding point in the right image. But today what we will discuss is about the complexity of this correspondence operation.

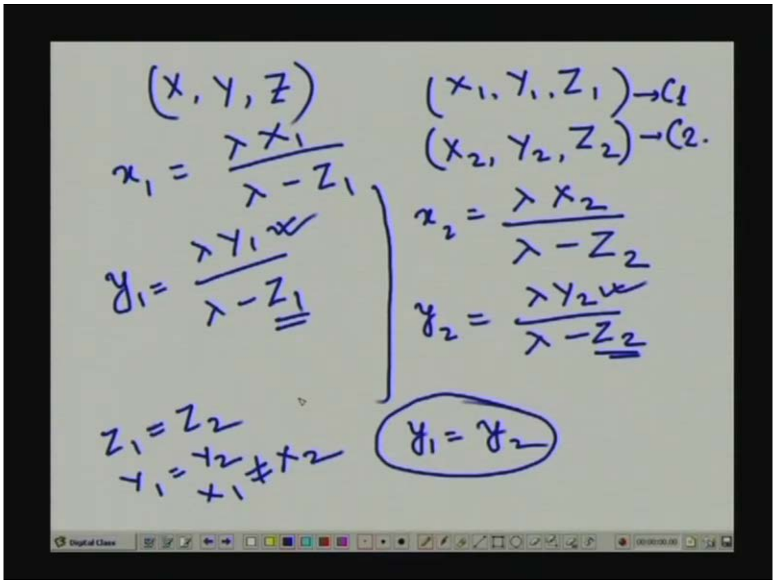
So, our problem is like this; we have a left image and we have a right image. So, this is the left image and this is the right image. So, if I have a point say C_L in the left image, I have to find out a point C_R in the right image which corresponds to C_L and once I do this, here I find out what is the coordinate image coordinate of this point C_L which is $(x_1\ y_1)$ and what is the image coordinate of this point C_R which is $(x_2\ y_2)$.

So, once I know this image coordinates, I can compute x_2 minus x_1 which is the disparity and then these x_2 minus x_1 is used for computation of the Z value Z. Now, what about the complexity of this search operation? Say, I identify a particular point C_L in the left image, then a corresponding point C_R in the right image may appear anywhere in the right image.

So, if I have the images whose dimensions are of order N by N that means I have N number of rows and N number of columns; then you find that for every point in the left image, I have to search N square number of points in the right image and because there are N square number of points in the left image, in the worse case, I have to search for N to the power 4 number of points to find out the corresponding correspondence for every point in the left image and the corresponding point in the right image.

So, this is a massive computation. So, how reduce this computation? Fortunately, the imaging geometry that you have used, that helps us in reducing the amount of computation that will be doing.

(Refer Slide Time: 41:14)



So, you find that for the point X, Y, Z in the 3D space, the corresponding left images is given by x_1 is equal to lambda times capital X_1 divided by lambda minus capital Z_1. So, I assume that capital X_1, capital Y_1 and capital Z_1, they are the image coordinate system of the first camera and i also assume that capital X_2, capital Y_2 and capital Z_2, they are the coordinate system of the second camera. So, this is for camera 1 and this is for camera 2.

So, with respective camera 1, the value of x_1 image point x_1 is given by lambda times capital X_1 divided by lambda minus capital Z_1. Similarly y_1, the y coordinate in the first image is given by lambda time capital Y_1 divided by lambda minus Z_1.

Now, with respective to the second image, the image coordinate x_2 is given by lambda times capital X_2 divided by lambda minus capital Z_2. Similarly, y_2 is also given by lambda times Y_2 divided by lambda minus capital Z_2.

Now, you find that the imaging system or imaging setup that we have used in that we have said we have seen that is Z_1 capital Z_1 is equal to capital Z_2, capital Y_1 is equal to capital Y_2 but capital X_1 is not equal to capital X_2. This is because the 2 cameras are displaced only in the X direction, they do not have any displacement in the Y direction neither they have any displacement in the Z direction.

So, for both the camera coordinate systems, the x coordinate sorry the z coordinate and the y coordinate value for both the cameras will be same whereas, the x coordinate will be different. So, by applying that since Z_1 is equal to Z_2 and Y_1 is also equal to Y_2; so you find that among the image coordinates on the 2 images image 1 and image 2, y_1 will be equal to y_2.

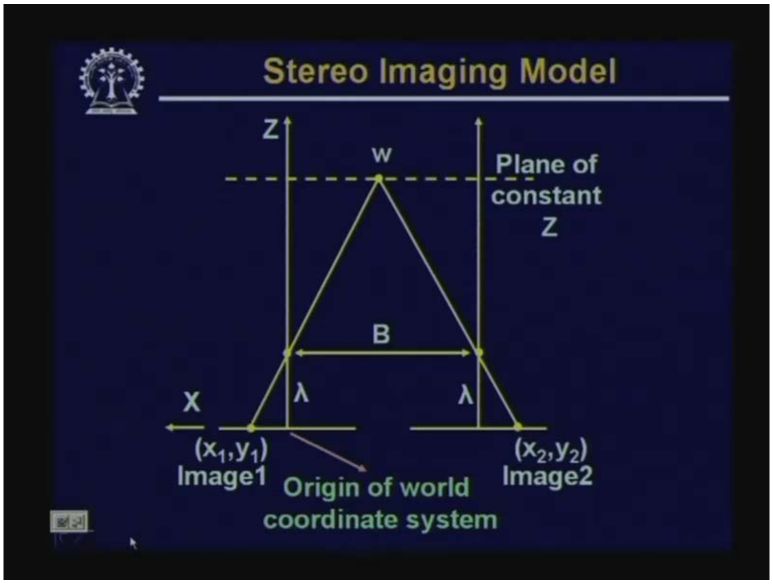
So, what does this mean? This means that whatever is the (x_1 y_1) value of point C_L in the left image, the corresponding right image point C_R will have a different X value X coordinate value but it will have the same y coordinate value. That means 2 corresponding points image points must lie on the same row.

So, if I pick up a C_L belonging to row I in the left image, the corresponding point C_R in the right image will also belong to the same row I. So by this, for a given point I do not have to search the entire right image to find out the correspondents but I will simply search that particular row to which C_L belong that particular row in the right image to find out a correspondents.

So, this saves a lot of time while searching for correspondence between a point in the left image and the corresponding point in the right image. So, till now we have discussed that how using 2 different cameras and having a stereo imaging setup, we can find out the 3D world coordinates of the points which have a point, an image point in the left image and a corresponding point in the right image.

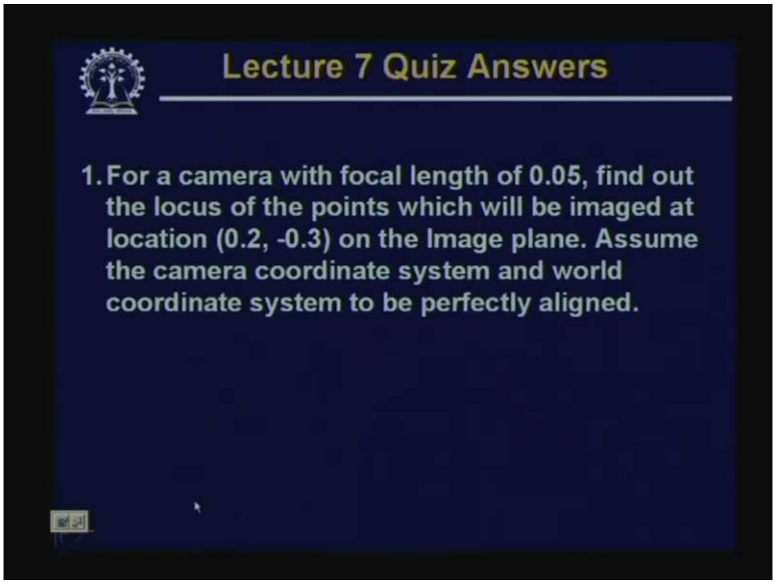
But by studying this stereo imaging setup, you can find out that it is it may not always be possible to find out a point in the right image for every possible point in the left image. So, there will be a certain region, there will be a certain region in the 3 dimensional space where for which spacer for all the points in that space, I will have image points both in the left image and the right image. But for any point outside that region, I will have points only in 1 of the images; either in the left image or in the right image but I cannot have points in both the images and unless I have points in both the images, I cannot estimate the 3 dimensional XYZ coordinate of those points.

(Refer Slide Time: 46:34)



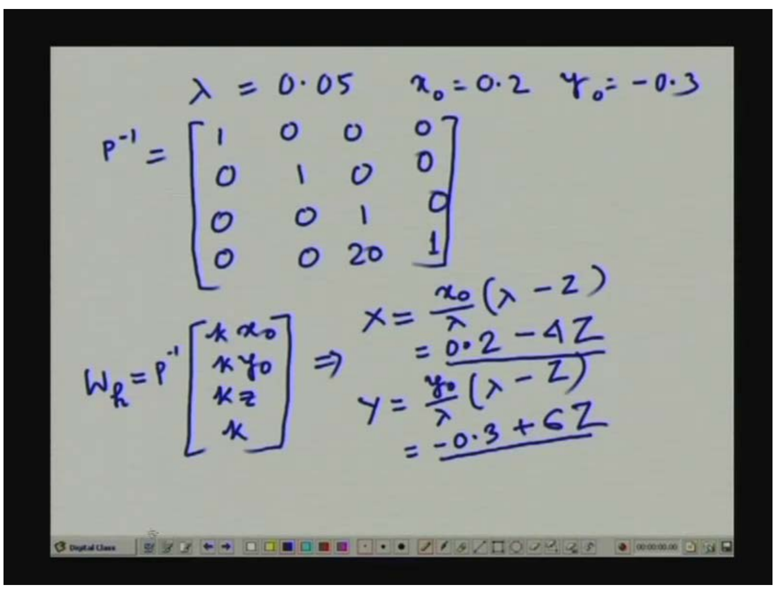
So, till now we have seen that using a single camera, I cannot estimate the depth value of a point in the 3D but if I have 2 cameras and using stereo setup, I can estimate the depth value of the 3 dimensional points.

(Refer Slide Time: 46:54)



So, now let us try to answer some of the questions that we had asked in the last class. So, in the last class we had asked a question that for a camera with focal length of 0.05, find out the locus of the points which will be imaged at location 0.2 and minus 0.3 on the image plane. Assume the camera coordinate system and the world coordinate system to be perfectly aligned. Now, this particular question can be answered very easily.

(Refer Slide Time: 47:33)



You find that in this particular case, the value of lambda is equal to 0.05, x_0 is equal to 0.2 and y_0 is equal to minus 0.3. So once I know the value of lambda, I can find out what is the inverse perspective transformation matrix. So, inverse perspective transformation matrixes is given by P inverse which will (1 0 0 0) (0 1 0 0) (0 0 1 0) then 0 0 1 upon lambda; so in this case, 1 upon lambda will be 20, then 1. This is the inverse perspective transformation matrix.

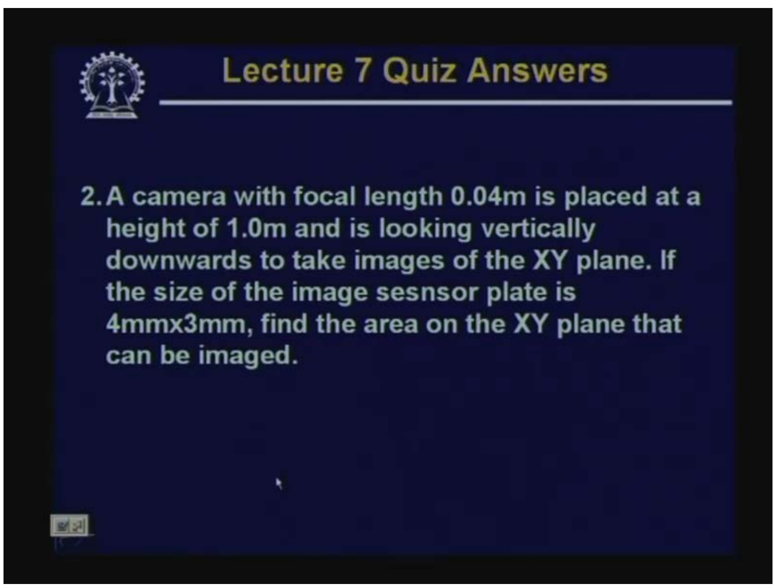
And because x_0 and y_0 are the image coordinate points; so, I can find out the corresponding homogenous coordinate point of the 3 homogenous coordinate of the 3D point as P inverse k times x_0 k times y_0 k times z and K.

So, here we have find that when we discussed about the inverse perspective transformation, we have said that this Z is taken as a free variable which helps us to get the equations of the straight lines and by solving this, you will get the equations in the form the derivations we have already done in the previous class.

The equations of the straight lines are given as X equal to x_0 by lambda into lambda minus Z and if you compute this, in this particular case, it will come out to you come out to be 0.2 minus 4Z and the Y coordinates is given by Y equal to y_0 by lambda into lambda minus Z and if you put the values of lambda again, this will come in the form minus 0.3 plus 6Z.

So, these 2 equations together that is X equal to 0.2 minus 4Z and Y equal to minus 0.3 plus 6Z, these 3 equations together gives you the equation of the straight line on which the point W will belong. So, this is how the first problem can be solved.

(Refer Slide Time: 50:15)

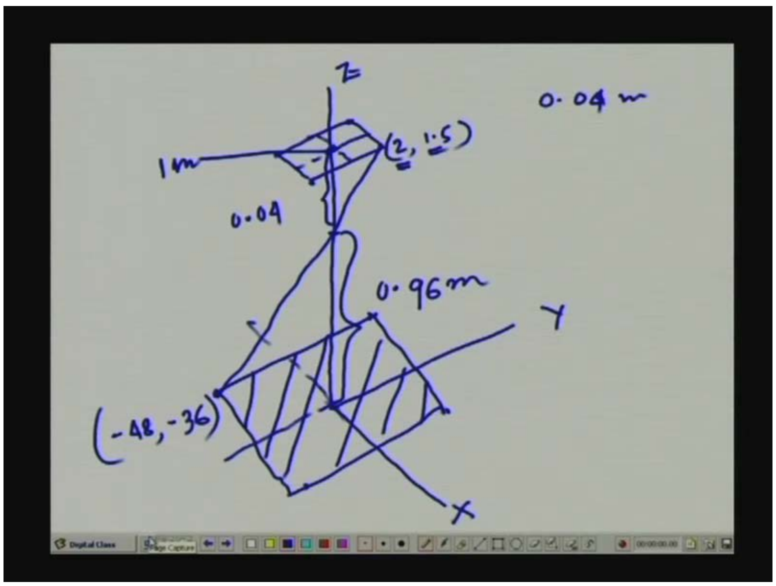


Now, in the second problem, we had seen that a camera with focal length of 0.04 is placed at a height of 1.0 meter and is looking vertically downwards to take images of the XY plane. If the size of the image sensor plate is 4 millimeter by 3 millimeter, find the area on the XY plane that can be imaged. So, here again you find that we can apply a set of transformations to bring the image coordinate system, to bring the camera coordinate system and the 3D world coordinate system under alignment.

So, the set of transformations that you have to apply is firstly, the camera has to be translated by a vector (0 0 1) after that because the camera looks vertically downwards, we do not have to apply any pan of theta but we have to apply a tilt of 180 degree. So, just by these 2 transformations that is first a translation by (0 0 1) followed by a tilt of 180 degree; I can get the transformation vector, the transformation matrix which will align the 3D world points with the camera coordinate system. After doing this transformation, if I apply the inverse perspective transformation; so here if you find that it has been shown that the sensor plate size is 4 mm by 3 mm. So, if I take the corners of the sensor image plane as extreme image points, then using this extreme image point, if I apply the inverse perspective transformations on these extreme image points, I get the extreme points on the XY plane which gives us the bound of the region on which the imaging can be done.

Now, this problem can be solve can also be solved very easily without going for all these mathematical transformations.

(Refer Slide Time: 52:21)



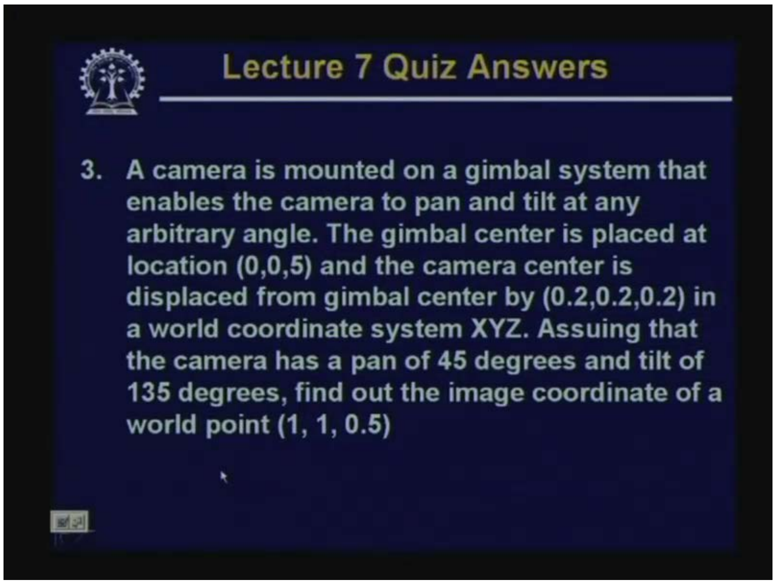
So, the solution is can be obtained like this. Say, I have this imaging point, so imaging plate is something like this. Let us assume that this is the Z axis, this is the X axis and this is Y axis of the 3D world point. So, I assume this is X axis, this is Y axis, this is Z axis and this is the center of the imaging plane. The size of the image plate is given by 4 mm by 3 mm, so here it will be 2, then 1.5 and similarly I can find out what are the coordinates of the other extreme point of this image plane.

Now, if I take a line through this, line from this point passing through the focal point center of the camera; now in this case, the center of the camera, the focal length of the camera has been given as 0.04 meter. So, from this point to this point, the length of this line segment is 0.04 meter. The height of this, this is at 0.1 meter, at a height of 1 meter. So, height of this is 0.96 meter.

So now, by applying the concept of similar triangles, I can find out that the x coordinate and y coordinate here will be 4 times the X coordinate sorry 24 times the X coordinate and Y coordinate here. So here, if X coordinate is having a value 2; this point, this will be in the

negative side, so X coordinate will have a value of minus 48 and here the Y coordinate is having value 1.5. So, in this case, the Y coordinate will have a value minus 1.5 into 24 that is 36.
So, in the same manner, I can find out 4 other extreme points on this XY plane. So, the region bounded by this 4 points gives me the region on the XY plane or area on the XY plane which can be imaged by using this particular imaging setup.

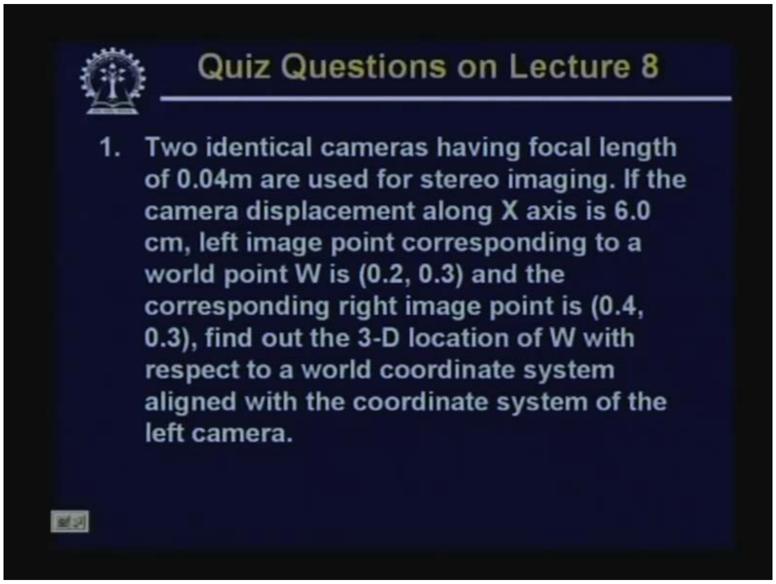
(Refer Slide Time: 55:15)



Now, coming to the third one; a camera is mounted on a Gimbal system that enables the camera to pan and tilt at any arbitrary angle. The Gimbal center is placed at location 0.05 and the camera center is displaced from Gimbal center by 0.02 (0.2, 0.2, 0.2) in a world coordinate system X, Y, Z. Assuming the camera has a pan of 45 degrees and tilt of 135 degrees, find out the image coordinate of a world point (1, 1, 0.5).

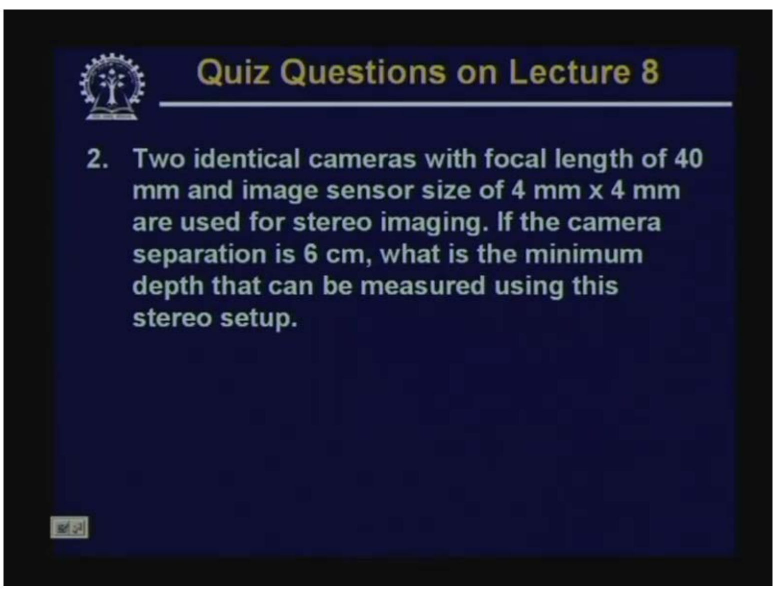
Now, this is a direct procedure which we have done when we have discussed about the generalized imaging setup. So, you simply replace the transformation coordinates in that equation by the translate transformation matrices of this particular problem, you directly get the solution to this particular problem.

(Refer Slide Time: 56:13)



Now, I have some question for today's lecture. First problem is; 2 identical cameras having a focal length of 0.04 meter are used for stereo imaging. If the camera displacement along X axis is 6 centimeter, left image point corresponding to a world point W is (0.2, 0.3) and the corresponding right image point is (0.4, 0.3); find out the 3D location of W with respect to a world coordinate system aligned with the coordinate system of the left camera.

(Refer Slide Time: 56:57)



Then, second problem 2 - identical cameras with focal length of 40 millimeter and image sensor size of 4 millimeter by 4 millimeter are used for stereo imaging. If the camera separation is 6 centimeter, what is the minimum depth that can be measured using this stereo setup? Thank you.