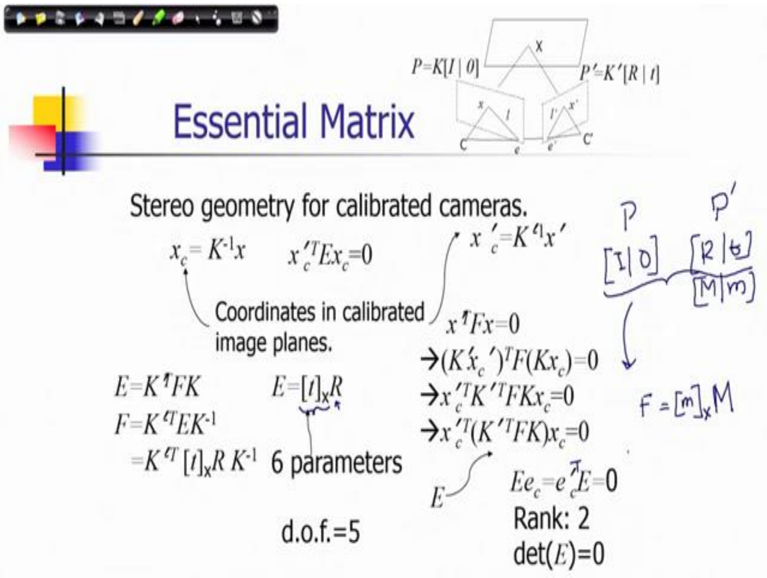
Computer Vision
Prof. Jayanta Mukhopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture -18
Stereo Geometry Part- III

We will continue the discussion on Stereo Geometry and we discussed the concepts related in this geometry like epipoles, epipolar lines, epipolar planes, then the fundamental matrix and how fundamental matrix converts an image point to an epipolar line in the other plane where the corresponding image point lies. We also discussed about the fundamental matrix of calibrated camera setup when this matrix is called essential matrix. Now in this lecture, we will further elaborate the parameters or elaborate the properties of this essential matrix.

(Refer Slide Time: 01:08).



So, in brief, it can be called as stereo geometry for calibrated cameras. So, in this case, we represent say the projection matrix P=K[I|0]. So, since the calibration matrix I can convert it into a canonical form of [I|0] and also the other matrix I can also convert it in its canonical form, because just by applying rotation and translation parameter itself all the image coordinates are represented in normalized image coordinate system.

So, I can also convert it as [R|t]. So, this is how in the canonical form the coordinates can be converted just by multiplying with the inverse of the calibration matrix. So, you see that in the first camera x is converted to xc which is the coordinate in the canonical setup optimal camera. Similarly x' is converted to xc' by applying there inverses of the respective calibration matrix. So, this is what is the expression.

So, now, if I apply the relationships of fundamental matrix in these setup, then that fundamental matrix is called essential matrix or E, but the relationship remains the same.

$$x_c'^T E x_c = 0$$

You note that in general it is $x'^T E x \quad = 0$ when we do not have any idea of K and K' from the projection matrix. So, how an uncalibrated fundamental matrix; that means, fundamental matrix for uncalibrated stereo setup is related to a calibrated stereo setup? So, how they are related? So, their relationships can be expressed using those calibration matrices itself.

As you can see with this simple derivation that

$$x'^T F x \quad = 0,$$

then convert x'^T into its canonical form as $K' x_c^T$. Similarly Kxc and then you consider the middle part of this composite matrix. So, that is the 3 X 3 matrix. So, this matrix is nothing, but this is the essential matrix E. So, this is what is E. So, the relationship between fundamental matrix and essential matrix can be expressed with using these calibration matrices as

$$E = K'^T F K$$

K' is the calibration matrix of the second camera and K is the calibration matrix of the first camera.

Similarly you can also get fundamental matrix from essential matrix by using those calibration matrix. So, the calibration matrix, it reduces into this from because if I consider once again you note that relationship; if you remember it. Suppose I have [I|0] and then [R|t] see these are the 2 camera matrices, this is P and this is P'. So, if you remember that this structure was [M | m] in general and the fundamental matrix was given as this $[m]_X$ matrix, which is a skew symmetric matrix into M. So, that is how the fundamental matrix can be derived.

So, in this particular setup using the same relationship I can write

$$E = [t]_X R$$

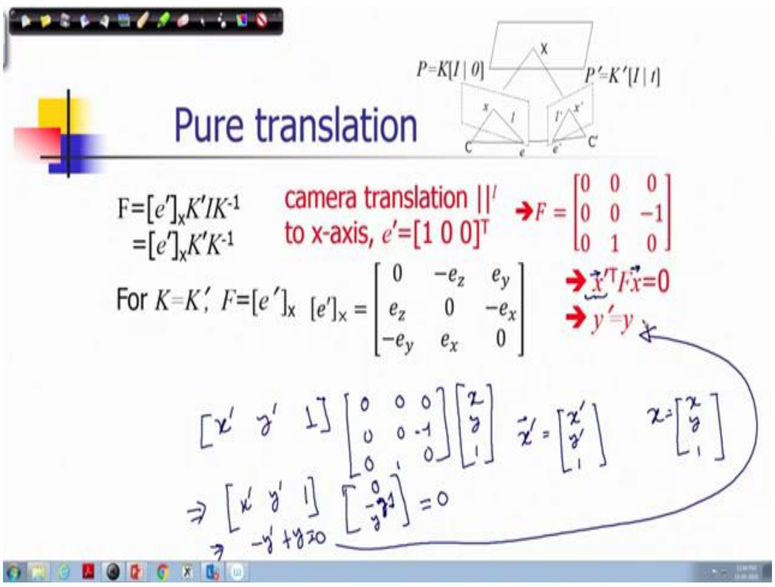
So, this is how the essential matrix can be represented or can we derive. So, I will replace that value. So, here you note that there are only 6 parameters involved in this particular operations like 3 parameters related to the translation of camera center. There is a translation parameter and rotation of the orientations rotation of the axis, so that is also 3.

So, there are 6 parameters which are involved in constructing the essential matrix and since scaled is involved; that means, if I multiply essential matrix with a scale value K still all this relation holds. So, out of them one again 5 are independent and also the epipoles in its canonical representation, so it also has those relationships holds; that means,

$$Ee_c = e_c'^T E = 0$$

There should be a transpose here, because it is a right zero, but we have to take it a transpose operation and its rank is also 2 like F. So, det(E)=0 and as I mention d.o.f of E=5, because a scale factor is associated with E and it has 6 parameter. So, it should be 5.

(Refer Slide Time: 06:44).



Now, consider a situation of special kind of stereo set up where you have pure translation of camera centers, pure in the sense there is no rotation of axis. So, in that case we can very easily express the projection matrices having a simple structure like if you have pure translation, then projection matrices can be expressed as say K[I|0]. I am representing again in the uncalibrated form then P'= K'[I|t]. So, simple translation; so, K't is that translation of the camera.

So, your fundamental matrix is expressed in this particular setup is in this form

$$F = [e']_X K' I K^{-1}$$

You remembered that if it is in general case it should be rotation matrix R. So, in this case rotation matrix itself is an identity matrix because there is no rotation and then this can be reduced as

$$F = [e']_X K' K^{-1}$$

That is a structure of fundamental matrix and if I consider both the calibration matrix is a same which means they are the same camera. They are using the same camera for taking the images or it has been fabricated in such way that all other intrinsic parameters they remain the same. So, then fundamental matrix takes a very simple form of $[e']_X$

So, just from the epipole you can compute the fundamental matrix itself. Whereas, $[e']_X$ is given in this form;

$$[e']_X = \begin{matrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{matrix}$$

This is the definition of $[e']_X$ and a since you have if you consider a special kind of translation that it is a parallel to x axis then, e' is the vanishing point of x axis and which is given by [1 0 0]. So, only to nonzero elements and the structure is very simple. So, if I apply this relationship of corresponding points, then we can show that the y coordinates actually they remain the same.

So, in this representation there we need to highlight few things that x' here is a vector. So, x' is a vector and in my representation we consider for example, this points are say

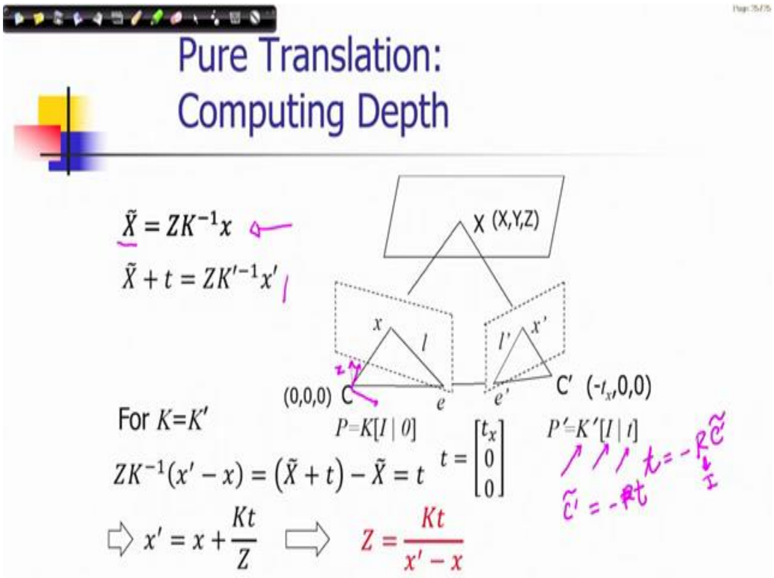
$$x' = \begin{matrix} x' \\ y' \\ 1 \end{matrix} \; and \; x = \begin{matrix} x \\ y \\ 1 \end{matrix}$$

$$F = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{matrix}$$

$$x'^T F x = 0$$

$$[x' y' 1] \begin{bmatrix} 0 \\ -y' \\ y \end{bmatrix} = 0 \rightarrow -y' + y = 0$$

(Refer Slide Time: 11:20).



Let us discuss how we can compute depth under pure translation. Consider these two cameras in the stereo setup as we have already mentioned that we have the corresponding camera matrices P and P' and where P is given by the K[I|0] which means the center of the first camera would be at in the origin and also this translation between these two cameras, it is given by $t = \begin{matrix} t_x \\ 0 \\ 0 \end{matrix}$

So, you are considering that there is also no rotation. So, it means the horizontal direction it is a pure translation and there is no rotation matrix R. As we can see in the case of second camera matrix here, we have same identity sub matrix here in this part instead of a rotation matrix of or we can say rotation equal to the identity matrix which means no rotation and this is a translation part and this is a second camera whose calibration matrix is also different which is K' and our objective is that to compute the depth here.

So, in this case since it is a canonical form so, your principal plane is also the xy plane of the coordinate system in the first cameras camera centric coordinate system. So, this is your canonical form. So, this is the z directions if I consider. So, your depth will be the z coordinate in this case. So, let us see how we can compute this z considering the corresponding points x and x' in the stereo view. Now the center camera center of the second camera, it should be $(-t_x, 0,0)$. As you can see that translation t in this form we will give this particular form because as we know that the camera center c'. I am using tilde to denote that it should be expressed in word coordinate.

$$t = -R\tilde{C}$$

R = I. So,$\tilde{C} = -t$. So, which will give you (-tx, 0, 0) that is the camera center in this form in this kind of configuration and then we can compute the corresponding world coordinate of the same point. So, here same point is X and in our convention when I am using the world coordinate in nonhomogeneous coordinate convention I will be using tilde on top of it.

So, as you can see that what we have done here.

$$\tilde{X} = ZK^{-1}x \rightarrow \tilde{X} + t = ZK'^{-1}x'$$

Similarly if I consider the other camera, there also we can use the same convention, but since there is a translation involved here.

It is because here also there is no rotation. So, your depth your principal plane is also the same xy plane and you can use Z coordinate as the corresponding depth. So, now, if I consider K=K'; that means, two camera have the same calibration matrix then this relationship can be further simplified. It takes a very simple form and as we can see from here itself if we subtract this equation, if you subtract see this equation from this one then you will get this relationship; that means, you get

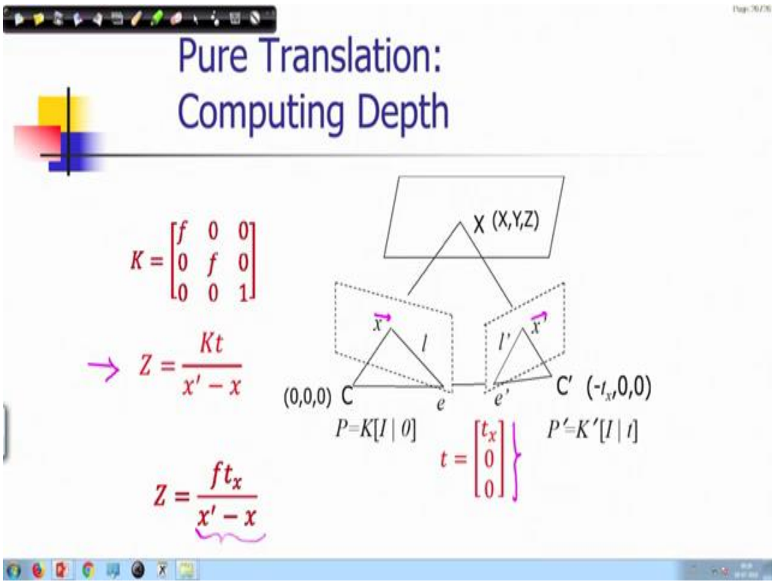
$$ZK^{-1}(x' - x) = \left(\tilde{X} + t\right) - \tilde{X} = t \rightarrow x' = x + \frac{Kt}{Z}$$

That is how we will get and from here we can get x'; that means, the second image coordinate of the since sin point in the second camera. So, this is a relationship between x and x’. So, there is a shift in the horizontal direction and the amount of shift is determined given by this Kt/Z and from this shift you can determine the depth

$$Z = \frac{Kt}{x' - x}$$

So, this is how we can compute depth and this relationship can be further simplified once we take the take a simple form of K.

(Refer Slide Time: 17:55).



So, in the calibration matrix we consider only focal length is a parameter and all other parameters they are initialized to 0, which means principal point is also the center of the image coordinates there is no skew and under this situation you have a very simple calibration matrix. So, this relationships; that means,

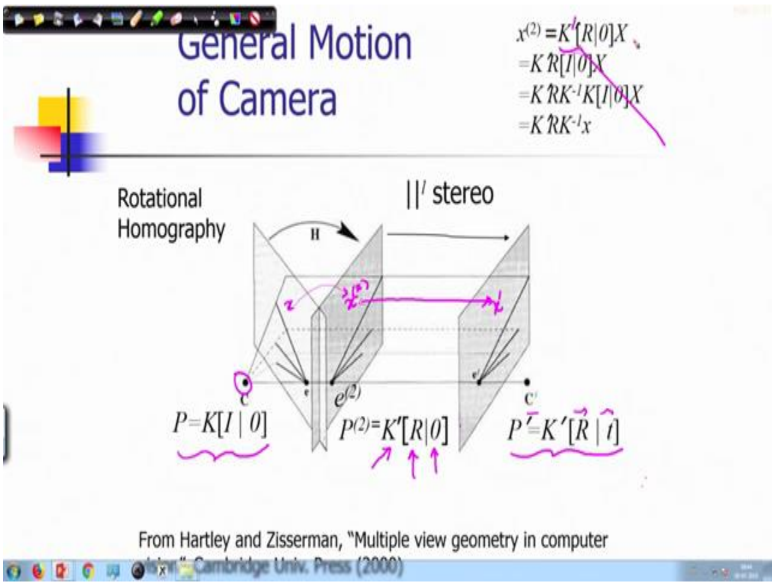
$$Z = \frac{Kt}{x' - x} \text{ } and \text{ } t = \begin{matrix} t_x \\ 0 \\ 0 \end{matrix}$$

$$Z = \frac{f t_x}{x' - x}$$

So, this is one of the familiar equations of computing depth in a stereo setup where optical axis is parallel between two cameras and also both the camera have the same focal length. They have same identical calibration matrix and the shift between these 2 camera is tx then we compute depth incidentally this value x’-x which is a shift of the corresponding point along horizontal direction or this is the x coordinates. These are all scalared x coordinates. You should note here once again that here this x and this x' it denotes the vector whereas, finally, in this computation when you have considered x'-x this is the scalar x coordinate of those two points.

So, this is also called disparity. So, in a stereo setup we can compute the shift or disparity and f and tx could be the parameter of the stereo imaging system. So, only using image, computations over those stereo images you need to compute disparity at every point and that is how you can obtain the depth.

(Refer Slide Time: 20:09).



This could be further generalized when we have a general configuration or arbitrary configuration of two cameras instead of having a parallel stereo configuration. Let us consider that you have a general projection matrix for the second camera, whereas the first camera is a reference camera and there we follow the same canonical structure with the calibration matrix K and in the second camera we have these general structure where you have the calibration matrix K prime which is a different camera. It could have different focal length etcetera and it has rotation and translation parameters as well, which means you know they are not separated by simple horizontal motions C' or the center of the camera is rotated and translated with respect to C. It has a general configurations.

So, under this situation, how we can compute this depth that we would like to discuss. Now there is a simple way by which you can map this problem to a parallel stereo problem and use the previous relationships what we derived here itself. So, what we can do? We can consider that let there be another imaginary camera which has the same center C, but it is related by only the rotation with the same rotation matrix R and then we can compute the homography between these two image planes. We know that it establishes a homography we have already discussed in two dimensional persecutory projection during projective transformation also and we will also see how this homography also could be, could be octant we can also find out here itself.

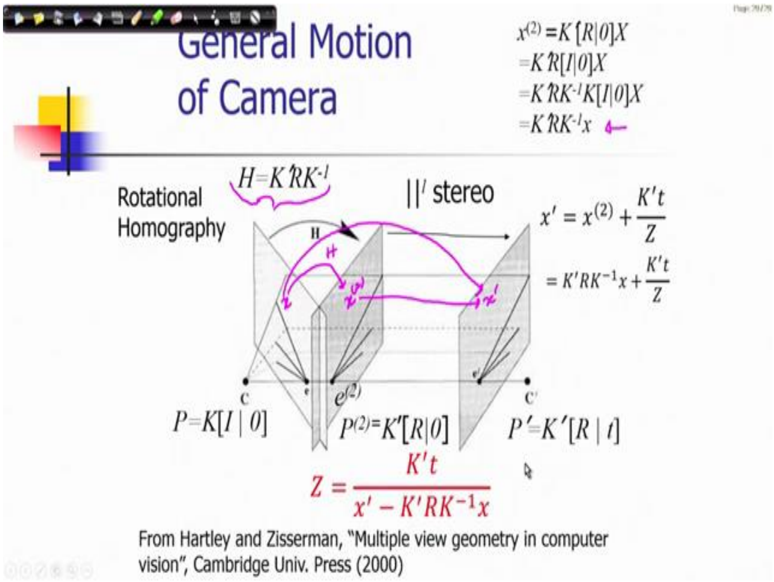
So, once we have these things for example, he you consider originally you have two point say x this is the reference camera. So, we will be considering this is x and another point say x'. So, these are the say 2 corresponding point of the same. Now due to rotational homography we bring x to the new imaginary camera setup where will be we will be considering this point is a x 2. Now these 2 since we have exactly applied the same rotational matrix R,

Now this image plane and this image plane they have they are they have that parallel stereo configuration. So, now, you can apply the parallel stereo results under the same; that means, what we need to do, we need to consider now the corresponding points under this 2 camera setups and x' and use them use this corresponding pair of corresponding points to compute the depth. So, let us elaborate this computations further. So, let us first compute the rotational homography in this case.

So assume that you have the second camera let me rub this thing just to make it more clear. So, consider that the second camera matrix as P 2 that is a imaginary camera as I mentioned which we considered whose image plane should be parallel to the image plane P prime after performing this rotational transformations on the image planes. So, P 2 since there is it is related with the original reference camera by the rotational matrix R and also. I would like we would like to have the same calibration matrix of the second camera K prime.

So, the projection matrix of P 2 is given in this way that it that it has a calibration matrix K' and in parameter of rotational matrix R and the center is still at origin 0 and that is the same as the reference camera. So, under this configuration what should be the homography between this 2 sins. So, here we can see that x2 can be written as in this case it is the image of the same scene point x. So, this could be written as K'. So, this is K' which is not very invisible from this part. So, K'[R|0]. So, this projection matrix P2x that is what is x 2 further. Further simplifying it so, this is x 2 and once again we can simplify it in this form which means I can take R outside.

(Refer Slide Time: 25:40).



So, it becomes K'R[I|0]X. Now this part can be expanded into this way. So, our objective is that out of this will be taking this part here and then K[I|0]X can be considered as x that is the original scene point. So, once again this is x2 which is mapped by homography H and then x'. So, originally x and x' is the corresponding points and now it has become x2 and x' is the corresponding point. So, now, the relationship between x and x2 and x is given in this form.

$$x^{(2)} = K'RK^{-1}x$$

So, the rotational homography under this situation it will become $K'RK^{-1}$. So, this is a rotational homography. So, these are rotational homography. So, once you have obtained this now you can apply the stereo results between x2 and x', which means that I can write

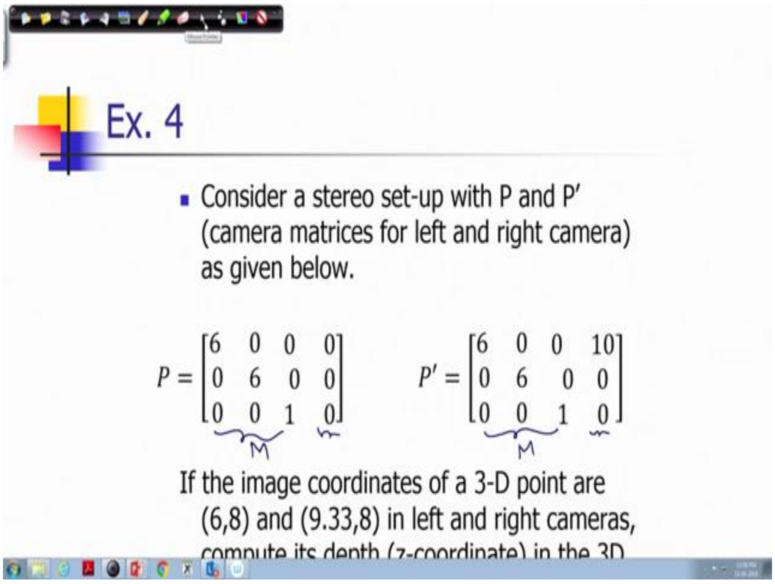
$$x' = x^{(2)} + \frac{K't}{Z} = K'RK^{-1}x + \frac{K't}{Z}$$

So, that is how x' and x is related. Now it is a simple mathematical manipulations algebraic manipulations.

$$Z = \frac{K't}{x' - K'RK^{-1}x}$$

So, from here we can get this relationship. So, in this way we can obtain the depth or Z coordinate under this situation of the same of this kind of arbitrary general motion of camera.

(Refer Slide Time: 28:06).



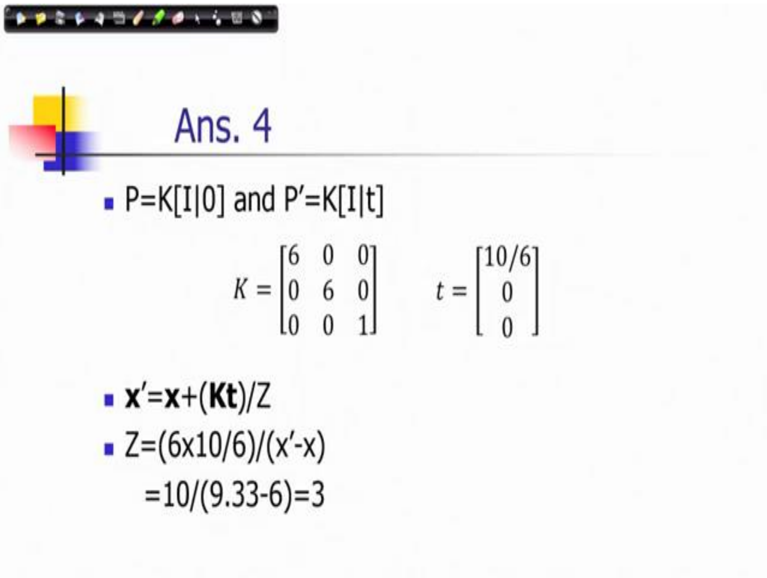
So, let us work out a problem to understand this relationships that how the depth could be computed or Z coordinate would be computed. You consider a stereo setup with these projection matrices P and P' which are given here as it is shown and now what you need to find out that you have been given 2 image coordinates 3-D points are there (6, 8) and (9.33, 8) in left and right cameras.

So, you should note that we have already satisfied that constraint of keeping the y coordinate same because that is what we discussed as a property of this particular scenario you can find out that the component of sub matrix m, they remain the same and only these column vector they are different. So, which means it is a case of camera cameras which are related with only translations of origins because you know that this is given by minus m inverse this column vector this is given by m inverse this column vector and that is a only thing because others are either rotations or rotations except there is no rotation relative rotation between this two system.

So, this is a case of your translation. So, in the under this case you we already seen that your y coordinates of the corresponding point they should remain the same because the there is a special case of its a special case of you know stereo and the fundamental matrix is given by simply by E prime cross as we have seen earlier also.

So now let us compute the depth of this particular you know in the particular what should be the depth of sin point?

(Refer Slide Time: 30:14).



So, this is the scenario your P = K[I|0] and P’ = K [I|t] where

$$K = \begin{matrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{matrix} \; and \; t = \begin{matrix} 10/6 \\ 0 \\ 0 \end{matrix}$$

So, this is what is K and this is this is what is t. So, you note that t needs to be scaled by the value 6. Because in the previous case it was Kt. So, you will get if I multiply this one then only you will get the previous matrix. So,

$$x' = x + \frac{Kt}{Z} \rightarrow Z = \frac{\frac{6 * 10}{6}}{x' - x} = \frac{10}{9.33 - 6} = 3$$

So, with this let me stop my lecture here and we will continue this discussion in my next lecture. Keywords: Essential matrix, Fundamental matrix, Projection matrix, computation of depth