Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-61

Now that we know what is a projection matrix, a camera projection matrix right, since we have understood what it is. So now it is time for us to venture into what is called stereo, which is a very interesting sort of a concept and it works as follows right. Imagine that you have 2 cameras, it could also be the same camera that you have translated, it could be 2 cameras that are on a stereo rig okay. Stereo rig is like a hardware right on which you can mount 2 cameras, it is called a stereo rig and what that typically means is that right this is actually held you know steadfast right in a sense on that rig, so that the cameras do not move and the idea is that and they are kind of they are also put in a way that it happens to be a pure translation, I mean that is ideal, that is the ideal situation for a stereo rig. But then we will talk about what is called, we will actually discuss both, this is what is called converging cameras and there is what is called the parallel cameras okay. The parallel situation actually makes things easier for us to solve, converging cameras are probably more general because it may be difficult for us to make sure that the cameras are exactly parallel and so on.

Therefore our sort of discussion will in general take into account that there could be a mutual rotation and a translation right. Translation is a must in a stereo but rotation ideally can be avoided but then right you may not always be able to avoid it. Therefore we have what are called converging cameras as well as parallel cameras and stereo itself right consists of 2 things, there is something called a dense stereo and typically right this is what is common in the sense that for every when you say dense what you mean is for every this one pixel in your image for every pixel right you want to be able to associate a depth map, a depth value. So that eventually end up with what is called a depth map and somebody says depth map this is what they mean.

For every pixel in the image you get a corresponding value for where the scene point is right, what is its depth from either camera 1 or camera 2 right depending upon the depending upon which camera you are going to say referring to. And there is something called sparse stereo which is also possible in the sense that you know you may have the correspondences for a bunch of points and right and then okay you might ask us to what kind of what kind of a depth do I get for those points alone and then if need be right one can I mean there are other other advanced methods that can do what is called region

growing and all and start to fill in okay. But we will actually look at both because I mean right sparsely also is interesting so we will also look at that but for the time being right our focus is on what is called dense stereo that means we want to estimate a value of or what do you say assign a value depth value to each of the pixels in the image. So if you have n by n then you are looking at like n by n to be your depth map okay. And a stereo rate has been seen for a long time it is like having a camera set up like this and let us say that right I have a 3D point in the world okay let us call this as x and x is some xyz okay this is a vector xyz okay and all our alignments and all would be okay now with respect to some coordinate system right we have to talk about.

So let us assume that without any loss of generality we can assume that the first camera is wherever is where the world reference is located you can always change it if you want but for simplicity let us assume that the world reference is here that is camera C and that is the what is it that is the optical center of the first camera C ‘ is the optical so right this is the optical center of camera 1 or what is called a left camera this is the right camera or whatever camera 1, camera 2, left light does not matter but typically we refer to stereo as a left right pair. So the optical center is here for the second camera and that as I said could be the same camera translated or it could actually be 2 cameras that have been permanently fixed on a stereo rig okay and you are watching a 3D world okay and the idea is that you so for example now if I join the ray right that goes from here to this 3D point then similarly right the same point is also observed by this guy by this other camera right. So what that means is that okay that is a straight line by the way okay so you have a ray right from the from camera center C for the first camera to X and then also right you are also observing it from C ‘ and let us say that let us let me let me just draw the image plane okay. So as I said that we want to draw it for a general situation okay now let me just move it a little bit up okay and let me well I have a figure in the slide but I thought it is better to sort of do it like this and then you have something like this okay and let us say that this is my camera C ‘ center okay and of course you know I mean I have not drawn it very correctly but then right I mean you can imagine the optical axis going through this and so on right that will be through the center of the plane and so on but this ray is from oops but this ray is from the camera center to the C 3D point right in the world. And the idea is this right so okay now we know that this is the image plane 2 or the right image plane this is the left image plane okay and the line that actually connects the 2 okay is called the base line all the base line that connects the 2 optical centers of the camera right that is called the base line okay.

Now certain things right now okay what does it really involve I mean why are we interested in stereo because we want to estimate the right depth of this point okay with one camera we cannot do because if we just use this one view right till now what we have seen is what is called single view geometry right where we saw that if you gave me a 3D point then I could write down the image coordinates for that point right given the camera

intrinsic and whatever right with respect to some world coordinate if there is an extrinsic we could also include that. Now the world coordinate system is centered here okay this is where the world coordinate system is for simplicity okay with respect to the center of the first camera which then means that c ‘ is some r and t is that some r and t with respect to the first okay. Now yeah so okay now what do we want right so in order to be able to solve this problem right it involves actually 2 things okay one is what is called a correspondence search a correspondence search and the other thing that it involves is what is called a triangulation okay these 2 are basically right essential for us in order to be able to compute the depth of this point because for example right imagine that that this cuts this image plane this ray of that 3D point cuts it at x where x is some x, y right some x, y coordinate on the image plane and then this is x ‘ this is the image of x in the second image plane of the same 3D point in the second image plane. But then when you see 2 images right we do not know automatically where is x ‘ but if you knew x ‘ right then so that is the problem of searching for a correspondence right that is the problem of searching for a correspondence which means that I need to be able to locate x ‘ in this other image before I even start triangulating because without that how do I triangulate right because by triangulation I mean that I have to kind of see you know do a back projection I have to back project the ray and hopefully there is no noise or something in a completely noiseless situation the 2 rays will intersect in the 3D space and that in some sense should tell me where that point is right. Now but then the thing is we do not know what is x ‘ is and especially if you are looking at a kind of dense stereo right we do not have shift and all okay so we will use shift or something else but not for really doing a dense correspondence search okay.

So what this means is that I want to be able to locate x in the other place and how we do typically locate I mean now it looks like a full blown sort of a 2D search in the second image looking for where is this x ‘ and even look for where is this x ‘ would typically entail that you know maybe take a small crop a small region around x and then try to compare it where is it in the other image and then maybe there is a good correlation or you can do what is called a normalized gray scale correlation or whatever right you may want to do any of those things that will sort of photometrically match to the extent possible and then you will decide that well maybe if the match is really decent right between this patch and that patch then you will say that probably right that the center of that patch is where my x ‘ lies. And once you have that x ‘ then you can go back and you see triangulate okay of course all this we have to do show kind of mathematically that we can do all of this which is possible but at the sort of a conceptual level right this is what we are aiming to do. And okay now the first question is right I mean you know when you have a situation like this is it really true that we have to do a full blown 2D search right it is quite possible that you know we may not have to do that on the you know at first glance it looks like I have to look for x ‘ everywhere in the image but now right we will actually look at what is called

what is called as epipolar geometry okay. And this epipolar geometry will involve a few things of interest and then once you bring in the epipolar geometry then that by itself imposes certain sort of say constraints on what when on you know where the image can be formed I mean it cannot be arbitrarily anywhere on the image plane and so on it has to lie along actually you can reduce your correspondence search to 1D search turns out that if you enforce the geometry epipolar geometry right then you can actually restrict the search instead of going for 2D you can restrict to 1D but then whether the 1D search rate is going to be along some line at some inclination and all right that we have to see but at least right what can be shown is that with the right epipolar geometry constraints thrown in the search can be reduced to a 1D plane okay in this in the second image plane okay or kind of vice versa. I mean you can be in the first image I mean you can ask the same question from right to left also right you can have a point x ‘ in the right image plane and then you can ask where should I search for x in the left image plane and that again can be shown is not a full blown 2D search it is just a 1D search okay all right.

Now let us kind of look at with that background right let us kind of look at what do we mean by this epipolar geometry.