Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-66

Let us today right look at, look at this is special cases, ok, 2 special cases which will throw some light on what is happening, special cases. One is what is called the parallel camera case, the parallel cameras. So right what we want to know is if you have parallel camera, so that effectively means that means that you have 2 cameras, you are left and right and then your optical center is here, optical center is here and these 2 are actually separated of course by some Tx or something and that is all, ok. So in the sense that that the separation is exactly along, along right on the this one image plane, right. You have the x axis of the image plane, this is a translating along that, ok. You are not rotating, you are not moving forward, you are not doing anything and you are just doing a baseline, that is called a baseline, right.

So this is the best sort of stereo situation which you can have. This is the one that most people will use or will want to use where you have 2 parallel cameras, where we have kind of these parallel cameras. So this is the non, this is the type which is actually non-converging. The other way to call them is non-converging cameras and the idea is that right we want to know in such a case where should we be searching, right.

Really the whole idea is that you have a point, so we saw what actually fundamental matrix, oh yeah right, prior to that just one sort of a quick, this one on that view synthesis thing because somebody later got confused with what I said. I said the fundamental matrix you said will relate a line to a point but now you are relating a point to a point. So the view synthesis rate all that I was saying was that if you have a view of a scene right from a certain viewpoint with a depth map right, that is key. That means you have some stereo setup with which you have been able to estimate a depth map okay with respect to one of the views, let us say with respect to the left camera you have a view, you have a depth map. And that means depth map means you know that you know how far away right, no, where it is each point okay.

So along this z axis you know how far away is each point, that is your depth. That is all assuming that your optical axis aligns with your z and so on right, which is what we have been assuming all through. Now when you have that now the point is if I virtually place a camera somewhere else and I want to see what the scene looks like, of course I cannot go too far away okay because there should be a lot of overlap between the two right, otherwise you will see things that you cannot see in the first frame right. So it is actually a virtual camera right, so it is not even capturing anything, it is a virtual camera. So when you actually keep it there right, you are now trying to ask what will the scene look like if I saw it from there right.

And you can actually do it through the image coordinates right in the following, I mean one thing is you can go through 3D but you can also show very simply that if for this left camera right, if you had x ~ is equal to what is it, some k left and then i0 and then x ~ right, where this x ~ is actually xyz but now the z is known okay, that is a big difference now, the z is known now. That means the depth map is given to me, I have somehow calculated that I have another camera which is a stereo pair, I have calculated, we still have not talked about how to compute depth but let us say we can right, given the disparity let us say we can compute depth. Now that z is known okay, that is a key difference between synthesizing a view versus whatever else right we are actually doing. I mentioned about this view synthesis because I felt that you know that is something that you people must be wondering you know, people talk about novel views, they say you know that the scene looks like this when seen from there and all, I was just thinking that after all right with stereo with so much that you have learned you should be able to, that is one step away. So having come this far I thought it should be easy to just let you know that, I am not going to ask you anything on this in the exam but just let you know that such a thing is just one step away.

So now for example right, now if you actually, now the way you do view synthesis is that you have an image right with respect to the left camera and then let us say okay, this need not be right there, let us say this is kept somewhere else and this is actually a virtual camera right. So we are kind of right keeping it there at some r and d right from with respect to the first camera center, optical center. Now when you are saying that right, so that means you specify r, you specify d right and you are saying that how will the scene look like. So it basically means that I need to know that right given this 3D scene that I am seeing right, I need to know that right if this guy back projects right through my whatever right I mean here is my camera center let us say, when I get a back project array from a coordinate in the image right, it will go back project means it will go all the way but now I know the z, so I know where to stop right because the z is given to me for every point right, that is the depth map. That means for every point right, I know how far away is that particular 3D point from this camera.

So which means that when I back project I know where to stop on that ray because I know where is z and now right all that I need to do is from that point onwards right whatever is this camera center oops whatever is this camera center right, I need to simply come there okay let us say right that is my c ‘ and then wherever it hits this ray, this projected ray hits the image plane there that is where this point is and then the intensity should be exactly whatever is the intensity here let us say whatever right, some point here in the image plane that intensity should not be copied there and then that is what that is the intensity of this point right in 3D, its image is with us and all that we are doing is transferring that intensity to another grid where it is a kind of a virtual view right and you can think of you know a corresponding point as okay now what typically is done is since you are projecting from an image coordinate here, so you scale this by z. So you get like 1 by z x ~ is equal to 1 by z kl and then this can be simply write you can rewrite this as xyz, x ~ is 4 dimensional, so it becomes xyz and this is your actual image coordinate, I mean that is the coordinate that I am going to pick from here and you are going to say back project and therefore it you can call that as let us say x put sort of red you know a double ‘ below just to indicate that it is actually image coordinates like xy1 and then this becomes like 1 by z kl xyz sorry capital xyz and then when you want to go to the other point right that will be like you know whatever x ~ ‘ which is a corresponding point for that 3D right. So x ~ ‘ is now what some k virtual you can assume some intrinsic camera matrix if you wish then you have an r and t which you specify and then this goes through the same you say x ~ right because that is at rate that is coming back and now this right this is called the camera pose together it is called the camera pose r and t right will tell you the camera pose that is what people will tell and they say that the camera is at a different place there is they will say that it has a different pose they would not say it has a rotation and translation and all the word used is pose right. So there is a pose for that camera and all that you have is kv and then you have a pose and then this x ~ vector is but xyz1 right the same xyz but that same xyz is nothing but from here right I can write it as kl inverse z x ~ which is the image coordinate and then always you see remember that I know the z here it is not a case of unknown z okay therefore I can do all this. So then this becomes z kl inverse or kl inverse x ~ and then 1 and then this is your x ~ ‘ okay.

So you have pose you have the depth this z is a depth you have the camera intrinsic and then and then you have an image coordinate. So you can directly go from the left image coordinate to the you can see virtual image plane and you can synthesize the view okay but always remember that right even though you are even though somewhere the fundamental matrix right you may be able to see here at all but then it is not that okay fundamental matrix again maps lines to points this one is actually a point to point mapping now because z is known because z is given right this is a point to point mapping therefore this should not be confused with what I said the other day about the fundamental because somebody had this doubt they said oh you said fundamental matrix line and point now you are saying point to point this is not a fundamental matrix anymore right this z and all is coming in now the actual value of z and therefore this is how you actually synthesize a new view right if you read any paper right where they do view synthesis this equation would be central there and if you are wondering where it comes from this is where it comes from right follows in a very simple way. So, what the fundamental matrix has something like the cross. Yeah so yeah so all that I am saying is right see the fundamental matrix is also something like yeah I mean I know right it is actually you know a cross product and so on but all that I am saying is one should not think that the r and t right which are actually sitting here right they all kind of I think okay right I think probably the student did not put forward the confusion correctly confusion was that somewhere I said line to point correspondence and somewhere I said it is a point to point. So somewhere right that so the fundamental matrix confusion that there was line to point suddenly I am talking about point to point I am not saying this is fundamental matrix they are saying that that should not be confused with this this is a different thing altogether okay that is all.

Which one which inverse mapping first image plane is yeah x ~. Yeah because you are stopping at some z no you know the z right you can only go so far and you have to stop there I mean you cannot go further you cannot go the some there are some complications that can happen which I do not want to get a go into okay when you change the view right for example right it is like saying that they are saying that when a point right I might be able to see in this image plane from this view but when I change my view right there could be some other point sitting in front of it exactly in front of it. So whenever a back project I do not see that point in fact I see the point in the front that is called occlusion such things can let us not get into all that right then we will go nowhere okay. So the parallel camera case right is like this okay so we want to know as to what kind of a search will then happen and where will be the epi pole and so on right so in this case. So the way to kind of right do it is as follows so yeah so let us go back and write down the fundamental matrix so we had so this is the right camera this is the left camera what was the fundamental matrix equation something like this right k r minus t and what is it this one the t cross r and I think kl inverse right this is what it was okay.

Now when it is a parallel camera case right that means r is identity you did not say rotate what about t the vector this is vector t, t is simply tx00 right because you have only a base you have only shifted translated along the plane of the camera along the x axis so there is only a tx shift there is no 0 there is no ty there is no say tz and therefore it what will be this say t cross when you kind of write it down as a matrix what it 000 I think it is 00 minus 1 and 010 something like that right something like that should be your tx okay. Now if you now right I am going to take a simple case here I am going to say that k r is equal to kl right I am going to take a typical serious situation where you just translate the camera but then whatever I am saying will still apply even if you take a full blown k k not equal to k kr not equal to kl at this portion I am going to leave it to you to find okay because everything if you derive here it will take a long time. I am just assuming that it is equal to k I mean in fact taking a very very simple situation like f000 f00001 okay but in reality this is not a condition this is not a necessary condition for what follows okay I am just taking it as a simple case because then we can do this kr inverse and all very quickly and show something otherwise if you take 5 parameter unknown in that k matrix then that inverse and all right will be kind of clumsy but whatever I am saying does not restrict it so it does not mean that stereo means that same camera should be translated you cannot change the camera at all okay you can have 2 non identical cameras perfectly okay you can have all 5 unknowns thrown into k and k ' no problem okay but I am just taking it as a simple case here. So then what you can show is this f right I think you know this should be very straight forward to show I just had the form here so it is like 000 00 t x by f and 0 minus t x by f and then what is it 0 yeah correct minus t x yes you are right minus t x and t x oh then I think even here what is it like minus t x and the f yeah I think even here I think it is minus t x by f here and then and then a t x by f anyway it is it is equivalent up to any scale factor but I think if you multiply this what you will get minus t x by f there and then t x by f in the bottom. Now what you what you want to know is okay now we know that we know that right with this fundamental matrix which is relating p to p what is the p l to p r right so now if I take a point okay let us now now right I can actually scale this okay because like I said right the fundamental matrix is only known up to scale factor therefore I can equivalently write this as 000 00 minus 1 010 because it is only known only up to scale factor so I can just knock off this t x by x and simply keep it in a simple form.

If I take this and now right what I want to do is I want to see as to where this takes a point right so I want to take an say x ~ which is in my where is it left or right left right from a left camera I take a point and now I have what is it 000 00 minus 1 right 00 minus 1 010 and then I apply it on this x y 1 this will give me an epipolar line will give me l 1 l 2 l 3 for the for the epipolar line corresponding to that point right and therefore what will you have you will have 0 minus 1 y okay this is my this is my l ' right corresponding epipolar line for that point and now if I now take some say take some what should I take now I should take some x ~ ' right and find out what condition should it satisfy for it to be on this epipolar line and x ~ ' is coming from where from the right image plane right and I want to know what condition should it satisfy for it to lie on this epipolar line and what is that condition that should be that it should lie on this line right which is like x ~ ' yeah dot product right with this guy should be 0 okay whatever dot product or transpose or whatever so then let us call that as x ' y ' 1 and then you have like 0 minus 1 y is equal to 0 for some point whose coordinates are x y in the first plane and this means that what does this mean minus y ' + y is equal to 0 or it means y is equal to y ' so what does it mean or y ' is equal to y in fact right this is the other way y ' is equal to y so it actually means that means that only points that satisfy this condition that their y ' is equal to y right can lie on the bipolar line. So what this means is that your search space is now is now exactly this so if you pick any x comma y sorry not a line if you pick a point here at some x comma y now the corresponding point right lies exactly at the same y on this on this epipolar which is a horizontal line right this is a special case because you have a parallel camera situation therefore right we just pick a point which is at x comma y you do not have to go up or down or you do not have to go some angle and all right you just have to go look at the same row in the other image and sort of look for a match and match you can do using cross correlation whatever it is that you want to some patch intensity something you can do right which is simple but the point is right so this epipolar line right will go like that and therefore and therefore right I mean all these all these epipolar lines right I mean you know will actually turn out to be turn out to be horizontal and the point is right so if you want to find out what is you see epi pole where is the epi pole right for this arrangement because epi pole you know is a place where if you drop c to the other camera plane wherever it intersects the image plane that is your that is your that is your e ‘ and if you and if you do the other way around that gives you your e right. Now you know that right epi pole e is such that right Fe should be equal to 0 and it is straight forward to show that if you just use in SVD or something then this null of f right turns out to be 1 0 0 you can show that f times this is also 0 and Fe SVD if you do right this is what you will get as kind of your null so which basically means that the means that the epi pole is that actually is that you know is that you know infinity okay that means you can see that no it is a parallel camera case right it is a parallel camera case and therefore right epi pole will go both this one and that one right they will both so the epi pole will be at infinity okay and that is one way to know actually that is one way to know somebody right gives you a camera arrangement and if you are able to compute the fundamental matrix and if you find the null right that actually tells you what is the camera arrangement there that is a converging camera situation or it is not right the epi pole actually is very kind of powerful that way and the other thing yeah right so yeah epi pole at infinity so yeah so I think that this is a special case this is a parallel camera case. Let us also look at one more thing that is of interest typically okay there is one other situation at which you can look at which is forward translating camera forward translating camera what is it like forward translating means you take a camera and walk along the optical axis like this right you do not rotate you do not go this way just walk along the optical axis so it is like this I mean I am not kind of right good at drawing and all so imagine something like this so your c can be here and your c ‘ can be there right so this is your first camera I will not call left right now I will call first and second because there is nothing like left right now or front and back if you wish right front and back again right you can ask what kind of a where do you think will be the say epi pole for this. This is a converging camera case converging case converging cameras anyway we will see it I do not think it is obvious right fine so let us see right let us go the same way so what do you do you first compute the fundamental matrix kr minus t T cross r kl inverse okay and again use the same simplification again it does not matter okay you can take a general case but I will just take this simplified case where now T x right will take the form what will so now okay now yeah now T itself right T itself is now what 0 0 T z because you are moving only along z right so 0 0 T z now so it turns out that you will get 0 yeah so what you get you will get like 0 minus T z 0 0 0 0 0 you can actually verify that this is correct and now r is identity of course because you have not done any kind of a rotation r is identity then this f takes up a form which would be like 0 minus what is that T z by f or something f square I believe whatever that you can actually verify all this I do not know whether it is actually a square or f but I have written it as f square I think it is f square 0 0 0 0 okay and again you can scale this is again only up to a scale therefore you can say that equivalently this is 0 minus 1 0 1 0 0 0 0 0 okay that is my fundamental matrix corresponding to the forward translational case.

Now I can ask right what is this in null of this and you can show that the null of f right again if you do an SVD right you can show that this is 0 0 1 by the way right I mean if you try to okay first let us find out where does the point map right so where is the epipolar line so let us take 0 minus 1 0 1 0 0 0 0 0 I apply it on a point xy1 in my left image or not left image now it is the first image right and this gives me minus y x 0 this is my this is my say epipolar line and now we want to look at x ‘ y ‘ 1 for it to lie on this epipolar line right this should be equal to 0 right we should satisfy this condition if it has to lie on this which then means that minus x ‘ y + xy ‘ is equal to 0 which then means that your x ‘ is x and y ‘ is y right. Now what does that mean that means that see any I mean you need actually 2 points to tell where the epipolar line is right one point is this if xy so on my right plane image plane right I am saying that if on the left I start at xy then my x ‘ and y ‘ are both again mapping to the same xy the other thing is the right epipolar itself the epipolar is this e vector is 0 0 1 that means it is the origin right so the epipolar lines look like these pass through the origin so all the epipolar will pass and that is what you actually see you know if you walk with a camera you get the zoom in and zoom out effect right that is because all the points are either moving towards the epipolar or they are sort of moving away I mean that is I do not know how many of us pay attention to all this but really right that is what will happen and the nice thing is this is even the scene is 3D and everything right does not really matter okay. So yeah so these 2 cases are actually right good to know just to have a feel for right I mean you know what is where the epipolar might be or how the right how this is right epipolar lines could turn out to be or where you should search and so on okay. Now let us again let us again write get back to this okay yeah there is one topic right which we will not cover okay that is I think I clicked on something okay now what is called actually rectification what is called rectification camera rectification it is called it is not already a rectifier right that you learn in electronics. So there is something called camera rectification right I will not actually talk about this the rectification is actually a nice operation it is actually consists of just homography so you might wonder if you are interested read it okay so this rectification operation is such that if you give me non-converging cameras right then I can actually then I can rotate them in a manner that they actually come back to a parallel configuration that is actually rectification and you know it does not involve any translations it does not involve anything it involves rotational homography.

The moment you say that translation is involved then it means that right the 3D and all right we will come into the picture right it is seen independent it has got to do with the cameras alone and this is fairly involved okay we cannot we do not have time to go through the rectification process but for those of you who are interested you can read about it I will not ask write anything on this in the exam but I just wanted you people to know that this is precisely done for this reason that when you have a parallel camera set up the search is so easy right you just have to go along the same row whereas the moment you have any other configuration of converging cameras and you know right every time I have to wonder where the epipolar line and so on. Therefore rectification is actually a standard procedure you have a camera pair stereo camera pair because nobody knows right whether they are actually whether they were set up in the correct way there could be a mutual rotation or something so rectification they will anyway do it is a standard procedure but the interesting thing is that it can be done as a rotational homography. So if you remember homography if you do rotation right 3D should not matter right the scene depth does not matter for rotations okay and yeah so it is an entire topic in itself okay so just remember that right that basically typically when let us say people do stereo right what they really mean is this okay that is they have a rectified pair that is called a rectified stereo pair because then that makes the search and everything very sort of easy right it is a kind of a convenient thing to do. The next thing right that I wanted to talk about was yeah right estimating F itself now all the time right we have been assuming that we know F right I mean we wrote down F as some K something and all right now F so computing the fundamental matrix. See this whole thing is a lot more involved by the way okay I am only telling things that are most relevant okay actually this is the geometry is not an easy thing it is very interesting but at the same time there are several things right that one has to be aware of I mean I will try to talk about something like what is called ambiguities you know like for example let us see it is not true that if you take this take a camera setup I should not say this now because you will get more confused but the point is right when you when somebody okay one of the things that you should realize is when somebody shows a depth map just as you have an image right when somebody says I have a depth map right what they show is again an image right they do not show you know a 3D scene right they show an image what does that mean that actually means that and then if you look at that image right it will have like you know like simply gray scale values it would not have this color and all sometimes people can show it in color but normally it will be a gray scale image where let us say bright point will typically mean that something is near and dark point will mean that something is far.

And the way you have to interpret that is a depth map is that is that right from the camera center right through that image coordinate that you have if a ray had passed and kind of hit a point somewhere in the 3D scene then the z component of that particular guy with respect to this camera is at depth right that is the way to actually interpret that. Now let me ask you a question so in the parallel camera case right I have a depth map right let us say I take the left camera I see a scene right in front of me there is actually you know a 3D scene right and assume that somehow if I can get the correspondence that I can compute depth as we know right once you know a disparity that means if the disparity is high then we know that the point is probably close to the camera the disparity is low then we know that the point is farther to the camera that we know right that comes from that is Fx by z except that we have to explicitly show that now which we will do today.