Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-67

But we know that right, so if you have disparity large that means the point is closer to the camera, disparity small point is farther off, this is very clear. So that means I have a sense for depth. Now if I take my left camera and I give you a depth map, I say that oh this is the depth map you get from this left camera, what can you say about the depth map for the right camera? What should be same? That is the question. I mean suppose I say that the same image you can use as a depth map for the right, will that be correct? Because depth map means you have to give me an image now. See when you say same right, what do you mean? No, no, let us say the entire thing overlaps. Both cameras see exactly the same scene.

I have a depth map here. Yeah, no but I do not understand what do you mean by same? That means I will use the same image. Sagneet, what should you do? Do you really have to do so much I mean right? So warp in the sense that why, see what when you say same, same what you are referring to is the z right because the cameras are parallel right. Therefore z but then where it maps will change you know.

You kind of see project it and project it back. You have an intrinsic matrix sitting here right, you have a T sitting here, all that will change where this point goes because it is all about where does that ray hit. I mean if it hits at the same location that means you have a disparity which is 0. I mean how can it hit at the same point right? If I give you a depth value and say that at x, y it is let us say 50 and if you say that in the right also at x, y it is 50 then it means that how can they be at the same location? Yeah, it should be that is what he means by warping. So what he means is exactly that but I am saying that so when you say same right I think what you in your mind what you have is the z is the same but when you show it as a depth map you cannot say use this for the right guy that will be wrong right.

So all the subtleties I think I mean if you just think little and little more about it you will realize that there is you know there is nice structure to all this right. You cannot go wrong right as long as you think correctly you cannot go wrong okay. Computing the fundamental matrix right. So one thing that we know is the matrix is actually governed by x ~ ‘ fx right is =. So we have this right so x ~ now we know that this is actually a correspondence right.

We know this is not arbitrary points right x ~ and x ∼ ‘ are corresponding points in this. Suppose I know z okay now the point is we know this equation is valid right. Now what we want to do is if I had now if I wanted to estimate f I do not have my f with me right. If I had to find my f I could actually do it I mean no I could actually do it by simply right getting enough number of these point correspondences across these images okay. And in this case need not be a parallel camera and all right whatever be the situation I have actually write 2 images with me and I want to find out what is that fundamental matrix that actually okay you know what is f and what is of course f transpose that relates to P L and P R and P R and P L right.

Now given that given these point correspondences now right if you think of x and x ~ ‘ is point correspondences now and f has a matrix right which is okay this is simple right I mean I do not have to spend too much time on this all of you know this. So this is like f21 this is similar to that homography thing that we did except that there is a small difference okay which I will highlight so f33 is 3 right. So you have f and therefore if you actually write this equation down then for one point correspondence I will just write down okay what you get and then the rest right you can check x1, x1 ‘, x1, y1 ‘, x1, y1, x1 ‘, y1, y1 ‘, y1, x1 ‘, y1 ‘, y1 ‘, y1 ‘, y1 ‘, 1 okay that is how 1 row will look and then this is multiplying a vector f okay which is like this guy stacked up okay as a kind of a 9 cross 1 that is the way you do not typically right you just convert it into something as an unknown and known. So the known is this matrix on the left which is a data thing right where you know that x ~ prime and x ~ are basically this is a correspondence that somebody gives you so that correspondence will have to come from something like sift or surf because we do not know the epi-polar we do not go and search now because we do not know where the epi-polar is because we do not know f right. So these correspondences have to come from something like what you already learnt sift or search or whatever right surf or whatever.

So these point correspondences have to actually come from that. Now one point correspondence is correspondence gives me actually one equation right of this kind and I have actually 9 unknowns but I know that my f actually has only 7 unknowns in it right we said yesterday right it is rank 2 and then it is also it can also be found only up to a scale factor right. So it has actually only say 7 unknowns okay. So actually right there is a very kind of famous algorithm what is called the 8 point algorithm okay and what is done there you can of course extend it to the case when you have let us see typically you do not start a stop at 8 points you typically take because of noise right you will typically construct a matrix that will be like say m cross 9 where m is much much larger than 8 actually but then there is an 8 point algorithm which if you assume that you have exact point correspondence then somebody gives you the most ideal case no noise or everything right then an 8 point algorithm works as follows. Suppose you call this matrix as A then what it will do is it says that read Af = 0 and then write this is actually 8 cross 9 and this is 9 cross 9 and because of the fact that you know A is A has A can only have a maximum rank of 8 right so we know that you know there is a non-zero F right which it maps to 0 and that can be found as kind of say right null of A.

So right so this null of A will give you this F okay and then it will exactly satisfy Af = 0 because A has only rank 8 right and that F right is only known up to of course you know up to a scale up to a factor but then it need not have rank 2 this F right has found out right there is no we have not there is no we have not automatically enforced rank 2 on it whereas we know that the fundamental matrix should have a rank 2. So as compared to the homography and all that we did right where we simply accepted the matrix as it is and we said up to a scale factor here we have to do a little bit more right we have to kind of get an F hat right which is as close as possible to F and should have a rank 2 right. So what this means is that so the F if you so you can reshape F as this matrix right F11 F12 whatever F that you get from the SVD right whatever you do here SVD you get your let us say F whatever right whichever way you get your F now that F you can reshape as a 3 cross 3 matrix. And now what you are expecting is a matrix F hat which is as close as possible to F but should be rank 2 how do you do that? There is you know a theorem right which you if you have done SVD you would be aware of that right but I am saying that why do you do that is because right there is actually right you know a theorem that says that if you want if you have a matrix A whatever right in this case F right you have this matrix A and then if you want let us say right a certain rank approximation of this right which is of course of the same size right the other matrix will also be of the same size but then in terms of a rank approximation right you want the best approximating matrix so you specify a rank right you say that I want this new matrix to have a rank whatever 2 or something in this case but then I want it to be as close as possible to F then the only thing right the only thing that can actually give you another matrix F hat which has the lowest Frobenius norm between F and F hat right that is what you mean when you say closest you mean in the Frobenius norm sense okay. So in the Frobenius norm sense SVD only does that I mean that is what he is referring to.

So then what you should do is you should do a decomposition of F as U σ let us say V transpose and the σ will have typically all you see non-zero singular values so it will have like 0 σ 2 0 0 0 σ 3 and if you order them right if I am writing this in a particular order right I am assuming that right σ 1 is greater than or = whatever right I mean greater than equal to σ 2 and then greater than or = σ 3 then what you should do is you should force σ 3 to 0 right. So what you do is you make this as σ 1 0 0 0 σ 2 0 0 0 0 this becomes and if you multiply U σ transpose then this F hat right you pick any right F hat in the world for that matter you can try it using MATLAB pick any F hat for that matter of 3 cross 3 in the world and try to find out the Frobenius norm between the actual F and this F hat the only thing that will give you the smallest Frobenius norm is this nothing else. I mean you can randomly keep on choosing rank 2 of course you have to choose rank 2 right rank 2 F hat as many matrices of size 3 cross 3 you can choose and find the Frobenius norm between F hat and F hat every time keep computing that number the only number that will be smallest is this okay. So this so right that is how you compute the compute the fundamental matrix now there is this one other thing right which I wanted to tell that is the actual depth itself right we still have not seen how to compute depth right that is actually pretty straight forward. So depth from this one at disparity right and we will talk about the parallel camera case.

You can also do something similar using a triangulation idea even if you have non-converging sorry even if you have you know if you do not have a parallel camera case that means you have converging case also you can do but then we will not enter into that right we will just talk about the case where let us say we have actually rectified so that we have parallel cameras okay. Now the way to see it is if I have x ~ right with respect to the left camera I have kl and then i0 right and then x ~ which is my 3D point whose depth I want to find out then a corresponding match right once I have found out that means I have the fundamental matrix I have searched I found this guy x ~ ‘ to be the best match for x ~ right so this is a correspondence okay that through search. See I mean right there is a whole literature right that even tells how to get afraid do this searching right one way is like I said I take a patch here and then try to match it there but then that could be sensitive to noise and other things. So there are very many robust ways there are actually energy functional that you can evolve that will also take care of smoothness in the depth because it should not happen that one guy has one value depth the next guy suddenly shoots up in nature it does not normally happen unless you have you know this one a discontinuity otherwise things are typically locally smooth right so all those constraints you can put in and make it a far more complex problem that is the way it is actually solved okay. But for this course we do not want to get into all that we just want to understand what is the basics what are the basics.

So say x ~ ‘ will then be k right and I will have identity and then a t and then this is a right x ~ right and where this t is simply a baseline b00 and this baseline we know okay this is a known baseline. In the next class right when I talk about structure for motion that is like what happens if some of these things are not known okay I know that it is a parallel camera setup but I do not know the actual spacing I do not know the baseline value but I know that it is a parallel camera setup that I can find out through the fundamental matrix I know that the epi pole is at infinity so I know that it is a parallel case but I do not know the exact baseline then what do you do okay that is what is structure for motion where the camera poses are not known right this is a slightly simpler case where you are assuming that if somebody gives me everything right so b00. So then what will happen so you will have like kr and then x + b yz right I mean assuming that this is some xyz so x + b kr right so now if I take my kl = kr that is like the same camera right I just translate by some baseline right and suppose I put this as f0000 okay all this is not needed that is what I am trying to tell you I am just making it simple for to just make it easy for us to do on a board. So then what do you have x ~ is = okay what is this so you will have this fx and then fy and z right and then x ~ ‘ and you will have fx + fb fy fz right and therefore right if you compute the image coordinate because that is what you have right a disparity. So if you compute the image coordinate you have for the x for the left guy like xy1 or you can simply write x equal well if you wish whichever way fx by z fy by z no comma and then 1 and then on the right hand side you have x ‘ y ‘ 1 this is again a coordinate right that you have found out this will be = fx + fb by z fy by z 1.

No sorry yeah no this is there is no f there that is just 1 okay now a disparity right I mean as you can see right y1 is = y ‘ which is expected right we cannot have a change in the y coordinate right that has to be in the same row the only thing that will change is where does x map to right therefore if you compute you know disparity as let us say ∂ that is x ‘ minus x right then you can show that it is simply = fb by z right. So it is only a function of z the focal length and then the baseline or in turn you can say that this is = fb by this one ∂ right. So if you want to know the depth all that you need to do is like you know the focal length know the baseline know the know this one a disparity and then right and then okay you have your z and this is exactly the z that we are talking about when we say depth map depth map and all for that point what is the z value right that you can plot and. As a fx + fb know because you will you will do x I mean this is x + b right because there is only translation so x becomes x + b and therefore fx + fb by z. So the z right so as you can see if you have a disparity that is small that means ∂ becomes small then your z will be large that is to say that the point is farther away if you say ∂ is very large sorry wait a minute right yeah exactly right and if you have to say ∂ is large then it means that your z is small that means you have a point that is actually that is actually close to the camera right.

So this disparity and z will go right will go kind of you know inverse to each other. So right here is where issues like these and all come up right in the sense that okay now if you I mean if you run an open CV or something right this is what it will do actually there are there are 2 things right which I did not talk about there is one thing called a camera calibration okay calibration means that you should know your k in this case we are assuming f is known and all right suppose let us say it was a 5 parameter matrix you would have had to know everything right in order to do this you can do this but then you should have you should know the know the this one right interjects of the camera. So that involves what is called a camera calibration which is a one time effort that is also fairly involved again right one cannot teach all that in a course of in a course that has so many things to cover but in but in some advanced course you know they do talk about that camera calibration right one can do that okay it is not so difficult and all but just that right people talk about then distortions and all that you can incorporate non-linear distortions all that it comes as a nice package the calibration part okay. So because there could also be lens you know sort of you know distortions and all sometimes you would have seen you know you will have this what is called you know a fisheye effect and all right so all that you can counter in that. So calibration is something that we that we did not talk about and that is something that you need to do and open CV and all right if you run that is what it will do it will do the calibration it will do it will do a rectification then it will do this calculation of depth and give you.

But now I think right having found all this right you should now have the confidence that if somebody says I have a depth map and all right you should at least be able to able to able to relate it to geometry, epipolar all that right you should know fundamental matrix all of this you should at least be able to speak a few words about all that right so that compared to what you knew right prior to this course and at least right after you do the course right there should be a difference in terms of what you can understand about actually right a 3D geometry. So I will stop here so next class right we will do what is called structure from motion.