Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-75

And the last thing that I want to talk about is actually what is called a dense sort of reconstruction, 3D reconstruction. If you see, I mean so the job of the bundle adjustment is to only return to you the camera poses which are actually more accurate, okay that is all it does. It gives you more, so the bundle adjustment rate gives you accurate camera poses, accurate to the extent possible, it does not mean this is all zero error and all. So you have the camera poses with you and you have, if you have n camera poses you have n views, right. Now you want to do a kind of a dense reconstruction because you saw that in bundler you had only a few feature correspondence, you cannot use that as actually a 3D depth map now. Now there is actually a smart idea what is called plane sweep stereo and I like this because this is one thing that actually fuses homography which we did earlier which was for actually 2D scenes and so on and actually a 3D, okay.

So this is nice and then, so the idea is as follows, right I will just tell you what the idea is, the equations are easy to follow. So idea is like this, I mean I have my views, right this is coming from let us say you know a bundle adjustment, I have these poses and let us say I fix my camera to be reference, reference camera to be this, then this is my first camera, then let us say 3 and so on, okay those are my other cameras which means that I know my R1, T1, right I know my R2, T2 everything of course, right with respect to my say reference. Now what is done is as follows, right now you have a 3D scene here, right whose kind of a you know a dense reconstruction you want that means what dense reconstruction actually means that for every pixel in this image I want to assign a depth value that is what it means or I need X, Y, Z for every point and not for just some sparse points and all right that is what we mean by a dense 3D reconstruction that means I want to find out what exactly is a 3D point in the scene for every point on my image plane, okay. Now the way this works is right, think about, so for example right so if you did a bundle adjustment and other thing that you also get is the 3D points, right you get that information also for those parts, I mean from that right you can get a rough estimate about what is the maximum depth and what is the near and far points, right some rough idea you will have, you can of course you know you can maybe make it a little larger provide a little bit of relevance but you will have something like you see D far and D near, right what that means is the farthest point and then probably right a nearest point I mean in your 3D scene you have let us say some sort of an estimate about that.

Then what you can do is you can actually divide this guy into n number of planes, okay these are the parallel planes, imagine that you have got parallel planes and these are all parallel to this plane, to this image plane with respect to that means with respect to the first camera, okay. So the first camera is looking this way then I am looking at slicing the I mean 3D scene right with let us say whatever depending upon how much accuracy you want, right I can have let us say right 20 planes coming in between D far and D near or I can have 50 planes coming in between D far and D near but then you can figure out that you cannot have it too fine because you would not actually gain anything if you make it too fine, why? What will happen if you make it too fine? Anyway I will leave it to you, okay so you cannot make it too fine either, okay so let us say right you have we have kind of quantized your see 3D scene right and here is a camera, my first camera, okay and then right it looks at the first and then these are all kind of virtual planes right they do not exist in the real in reality that is virtual planes that I am actually looking at. Now the homography right as we know homography is what so for example right I mean if I know that if I know that I have a planar scene in front of me, a homography will allow me to transform okay the image that I have here okay which I am assuming is of this plane which in reality is not it is of the whole 3D but for the time being if I assume that whatever I am seeing here if I were to use the use a plane equation right of this plane and of course I have my camera pose and if I apply a homography so as to be able to go to my next camera so that I synthesize a view as to what this image would look like if it was seeing a plane right at that you see depth whatever right if you start with the nearest it will be like D near. So if a D near if you think of a plane a fronto parallel plane and if you actually warp this view such that it generates or synthesizes the view with respect to camera 1 then I can do synthesize with respect to camera 2 what is that equation that is the homography I think I wrote it down once but I did not show it actually but that is your homography right also if you let me just write down the standard notation right for that homography we write it as H okay H pi p l okay in this case that is why let us be let us be clear about what we mean so k l r l + minus T l N p transpose by d p the whole into k reference inverse okay. Now what is the l N p r l is the l th pose and p is the p th plane okay so what this means is if you fix a plane let us say that you fix the plane then given the reference view you can synthesize the scene as seen by let us say a camera which is l which is the l th camera which is the l th pose so the l th camera has its own intrinsic let us say typically it may be the same camera and then you have the pose with you which is r l and then you have a T l and this N p right because it is a fronto parallel plane with respect to the first camera N p is easy N p is simply 0 0 1 it is a fronto parallel camera.

So the idea is like this right so it is interesting to see right what will happen if you actually do this now if you if actually if in the 3D scene if there were points right at that depth if there were points at that depth then what will happen when you warp it right and then you see it in the second scene so if you stack up all these so see it is like this right I have a bunch of observations that I already have with me these are not synthesized these are my original images these are my views that I have. Now I am taking my say reference view okay of this scene and I am actually applying see homography right whatever H pi what is this H pi p 1 with respect to one plane then I have H pi p 2 all the way and each one will synthesize a view okay all the way up to whatever right I mean so I have what is this you know whatever right I mean m number of views right so l going from so I have like H pi p m so I have m number of views so I have synthesized now what do you think will happen if I compare these two images okay so if you compare these two images then if a point was actually at that depth right then when you warp it right when you warp the reference view this will exactly match at that location with the observation do you see this if it actually obeys the homography that means you have something at that depth d p okay yeah right d p is d p is the depth of the plane which means a perpendicular distance from the center of the camera this is the perpendicular distance or in this case right it is a frontal parallel camera so really there is nothing perpendicular distance of the plane from the camera from the camera center which is also our depth. So what it is saying is if there was a point there then because of the fact that it is actually at that point and because you are you are actually because that point is lying on that plane a planar homography applied on that point will take you exactly where it ought to be as seen from the second view as seen from the third view and so on therefore it if you check for a photo consistency of this right then what will happen is what was the point so what did I do so I mean no wait a minute right okay I changed I am so I changed only the view right so what will happen is in the same view because right what the point is because it is a planar homography you will find consistency across all the frames literally because everywhere you have to map like that unless of course you know you cannot see it in which case you will not be able to map it that is why it is called a winner take all strategy that means that if you compare this intensity with this typically you will take a small region not just one value you will find that wherever there is photo consistency right you will get a lot of lot of close intensities there and that basically means that that point is that actually depth dp that is why it is able to obey this any other point it I mean so for example if my plane was here and this point was here think about a point here okay which will also be acted upon by this right but that is at a kind of a different depth so that should ideally have been applied this homography which is at let us say dp whatever at d1 or dt d3 or something but I am applying this planar homography on that that has a depth which is not the same as this depth therefore when you apply the homography it would not match so some other point here which is this point which is at actually a different depth if you if you if you warp it okay some other point here when you warp it let us say it came here if you try to compare the location here at the same location these intensities would not match because that is not coming from that depth that is coming from a different depth that means this dp should have been different there for the same r and t for the same u see I am seeing it from here my r1 and t1 are fixed but the depth is not right and I do not know the depth so the only way to check is I will warp it and after warping I will compare right if there is a consistency I know that okay that is a correct depth all others will not match that is okay but then for that point I found the depth then I can repeat it repeat it for all the planes so what will happen when I actually come to this plane and I warp that that guy will come and correctly fall that means it would not come here it will go somewhere else that is where this point will be is this clear this is called a plane sweep right because you have a you have you are sweeping a bunch of planes across and and you are actually right try to check for a photo consistency loss and based on the photo consistency loss now for every pixel now we can assign a depth value right because you know that d1 is for this guy d3 is for that guy d5 is for that guy so so you can so you can imagine that imagine that right for the whole grid you have values now and and and you know and the point is even if something is not visible it is okay what is that is why we say winner takes all we do not we do not expect it to match in every view we look at at least you know a good number of views it can also happen that right something may have let us say you know 5 sort of photo consistent views okay for one depth and then maybe right you know 7 consistent views for another depth right in which case in which case in which case we will take the one which is probably 7 whichever is higher it can also happen right because if the depths are very close right so the so if your sectioning is very close then it can happen that right you get kind of you know 2 depths but whichever gives you the maximum consistency you go with that and this is how you do actually a dense 3D this one reconstruction okay and this does not require require a rectified arrangement at all right I did not try to rectify I have RNT I have got cameras arbitrarily placed it does not matter right I do not need rectification I can directly you can rectify and do it if you want but it does not require that so that is why plane sweep studio is actually a nice way to just a homography it is all just a homography principle that is used to sort of to sort of figure out you know the you know right depth value for every point okay I will stop here okay so next class right I mean I want to so with this right the traditional you know this one okay structure SFM and all is over okay geometry just that I want to say discuss a few deep network works okay next class and then and then we go to mid level and high level those are smaller topics related relatively okay.