Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-76

I think, you know, something is not clear about this plane sweep stereo, okay, which I had already explained, but just to clarify that the homography, right, that you have, what is the homography that we had for that? K L, I think what I had written was H pi P L, right, this is what I have written as a homography, that is K L R L + minus T L N transpose by d P the whole K ref inverse, right, this is what it was. So, one of the doubts was that do you actually, you know, match corresponding points and arrive at the homography or do you kind of directly use this equation, the idea is to use this equation because there is not really a real plane out there, right, I mean you have a 3D scene, okay, which is being watched by these cameras, right. And we are only, we are only trying to slice this guy, right, into let us say, okay, if you take, if you take, write this to be the, to be the, to be the camera, reference camera, just slicing it into, into various planes at whatever, at some equidistant locations and you are just, and these planes are virtual, right, they are not real. So, what you are trying to say is if, if this, if this 3D scene cuts, right, wherever it cuts, I mean, I am just trying to give some sort of an approximate figure but then, right, if you kind of look at the contour of points, right, that might cut this plane in the front, so what you are saying is that plane which is a d P, let us say, right, if this one is d 1 and then you have got like d 2 all the way up to whatever, right, N number of planes and you have got M number of views, so what you are sort of saying is that for one distance d 1, you can warp this, this reference frame using this equation because the poses are known, right, we know R, T, T, the normal as I said is 0 0 1, right, that is because it is kind of fronto parallel, right, so with respect to the reference frame, this is fronto parallel and the k L and k R are the intrinsic matrices. And so the idea is that you kind of warp in a sort of virtual, you know, in a virtual way, right, it is not like, it is not like, right, you take images from these 2 views and find correspondences to find the homography because there is no single homography by the way, right, because this is a 3D scene and there is no homography that will take you from this image to this image or to this image, there is no homography that can do that, if it was a case then it would be a planar scene, right, the idea is that this is 3D, there is a translation across the cameras and therefore as we had said right before that there is no homography that can bring you, bring one image and align it to the other. But here we are not talking about one homography, right, we are talking about a bunch of homographies that are coming from various depths, virtually created and you are trying to see as to where this image, how this image will look if it was warped by this homography and you are trying to compare it with this view and the hope is that those points, right, that actually lie at that depth which we do not know but then there will be points that will lie, that will actually, that the surface actually cuts, right, this plane and those points are at depth D or D1, right, with respect to the first plane and with respect to a reference camera and therefore, right, when you actually do this warping, so those points will come and if you try to match this warped image with this view then it is not true that it will match everywhere, right, it will match at some points where this, where this homography is obeyed by points that obey this homography, right and, and, and you know that is how, but then since you have multiple views I mean you can do it for, do it, do it across the cameras and as long as, and because the scene is the same all these points are supposed to appear at their respective locations, right, and when you warp them you should be able to compare them and you should be able to tell whether, whether the match is good or not for those points and then you sort of do it for D2, then you do it for D3 and that is how you do it, okay.

So, so there is, there is nothing like you know comparing feature points to get the homography, right, that is the way we did it in the beginning if you remember we said that you know if I have 4 point correspondences then I, that gives me 8 equations and that is all I need, but this is not like that, right, this, these are virtual planes, these are not real planes, these are virtual planes through which you are trying to warp and, and you are trying to get a sense for, for the, for this structure of the scene, okay, so I think this is one point that I just wanted to clarify. Then today, right, what I am going to do is I am going to show you a few examples, okay, of which I had said last time itself that I would do, so let me just go and, go and just show you a few examples, right, we do not, we would not spend too much time, but just to, just to illustrate a few of the things that we have already done, okay. So, so for example, in a typical kind of a stereo head, right, looks like this, so you have one camera here, you have another camera here, these are, so right, this is like a stereo pair and you can actually construct your own camera if you wish, right and you can actually take it around to capture images and this is what, this is what we had for a parallel camera, right, so the kind of epipolar line we know that they, they lie, they lie along the same line, I mean, okay, I should draw it correctly, but you know, so for example, right, so yeah, so maybe write this for some other point, so whatever it is, right, so, so you have these parallel lines and you know that the epipoles are at infinity, right, so the E at infinity, E dash at infinity, this is a case for a non-converging or a parallel camera case, right, where, where it is easy to search, right, so this is like, this is like, you know, the epipolar lines are, right, are simply the rows of the, rows of the images, so it is easy to go and search for them. When you have converging cameras, right, then your, then your epipolar lines are no longer, you know, right, they are not in general, they are parallel and, and it is also not true that the, that the epipoles should lie in the image, for example, in this case, you can clearly see that the epipoles are not in the image, right, they can lie outside the image, so you can ask where is the epipole, right, in this case and you can show that the probably epipole is somewhere outside the image, right, so this is, this is certainly possible, I think one of the students was asking during the, when we were actually talking about epipole, right, where they can exist and so on and I had mentioned at the time that epipole, right, does not have to be within the image, it can lie outside and, and what you do, so you basically, right, take a, take a coordinate from, from the left image and then you act your fundamental matrix on that after you have computed F, then you find the epipolar line and this is the line along which you have to conduct the search and the search could be using whatever, right, you can take a small patch and then, and then use it to compare across, that would be a traditional way of doing it but that is still over simplified because in reality, right, there are issues such as, you know, such as variations in terms of some illumination variations could kick in or, or you know, you could have something like, you know, whatever shadows or you could have noise or you could have a colour variation, we could have things like that happening and therefore, right, it is not such an easy thing to just take a patch and kind of do, do a cross correlation and sometimes there could be occlusions and sometimes there could be regions of low texture, where it is, where it is very difficult to match, right, suppose you give you two, suppose, suppose, right, I mean, think of a region like this and think of a region like that, I mean, there is, there is really nothing unique out there, right, for me to be able to match, I mean, because when I, when I kind of take a patch here, then maybe, maybe there is a whole bunch of things where it can match, right, on that, say, people or of course, I mean, if it is, if it is on that row or wherever it is, right, when you match, it is not true that, like, it is such an easy task, right, especially when there is a, when there is regions of low texture and therefore, right, people of course, have various other ways of, you know, handling that, what are called energy, energy minimization methods and so on, where you kind of believe that, you know, believe that locally, right, the, your depth should be somewhat smooth, right, which is a, which is a reasonable assumption to make.

The smoothness is something that I, that I keep telling is such a sort of a generic constraint and it is so very handy, right, and just a matter of formulating it and then usually it is a kind of a, kind of a Gaussian prior that is, that is put on the, put on the, put on the, say, depth map, right, simply means that things are locally, locally uniform or things are locally, what do you say, they, they look locally similar and which is normally true, even in image, if you look at a pixel around it, things look very similar, right. And again, right, if people are aligned, you can have a situation like that and anyway, right, I mean, so I think this is the chirality thing but this I think I had already drawn, so for example, right, so, so here is a case where the point is in the front of the camera, so right, and here is a case where, where, where the, where the point is, point is, so in this case, yeah, so in this case, which one, which one is it, so right, with respect to this, it is actually to the back of the camera, right, and so on. So, you can see that right, and then there could be a situation where this we discussed, right, where you can have both cam, the point coming, coming to the rear of both the cameras which is not possible, right, that means you cannot image that. Then this is that similarity ambiguity, right, that we had mentioned, for example, you could have a situation where, where the, where the reconstruction, right, could be up to a similarity, what it means is, you can only reconstruct up to, up to a scale factor, up to a rotation and so on, right, so you can see that, see this kind of, kind of this one, right, I mean, this kind of a construction, reconstruction where you can see that you know, the ambiguity is up to similarity, I mean, you can, you can have, you can have a projective ambiguity, right, which basically means that something like this could look like something like this and yet that is also equally valid, right, I mean, you cannot complain, I mean, because that is the, that is the ambiguity that, that, that you have in the problem fundamentally. The previous one, the previous one what it is saying is that, for example, right, I mean, you know, if you had, if you had an original structure like this or whatever was the original structure, then all these are, all these are sort of equivalent reconstructions.

I mean, you could have, you could have, you know, you could have a rotation of this, right, which is what this is, you could have a scaling, it may look smaller, so similarity is like what, you can have scale translation and this rotation, right. So, all three ambiguities could be there and still it is okay, it is perfectly okay, I mean, that is the, that is the level of ambiguity that is kind of, right, inherent in the problem and therefore all these are, all these are okay, I mean, and so we do not have a problem and we accept all these solutions. Then the multi baseline, baseline stereo results, right, typically would look like this, I mean, if you were to do this, do this kind of a plane sweep stereo, right, this one, right, this is at, this is at left, this is at plane sweep stereo, right, so what this means is that, so by left neighbour, right, what they mean is you have taken, you know, a reference image to be somewhere in the middle, I mean, I kept taking one onto the left, but you can take, right, take in the middle that to be the reference and then you can check what is happening to the images on the right when you warp them, what happens to the images on the left when you actually warp them or you take the reference and warp whichever way, in one case, it is just the inverse of the original homography, right. If you do that, so what really happens is, so for example, right, when you actually, so when you actually warp, right, with respect to, with respect to various, so you see, so you see this kind of, kind of it, a blurring effect, right, so what this really means is that, that is what it means, so you cannot align images through a simple homography because it is actually a 3D scene, right, so when you actually bring them up, so surely, right, there are points here which I think, you know, you are not able to immediately spot, but there are points where the overlap is perfect and there are points where it is not, that means, that means, right, when you kind of look at it, it looks like, looks like something is blurred out, right and that is because they are not, they are not, they are not being aligned and they cannot get aligned. It is a, if it is a 3D scene, right, they cannot get aligned, it is not possible to align them, that is, that is a fundamental,

fundamental thing which we know, using a single homography you cannot, you cannot align them, right.

Therefore, if you actually, if you actually bring these guys together, you will find that, you will find that only some points which are lying at that depth will match across views, right and that is what you in fact use. One simple way is to sort of look at this whole volume, right, which is called a kind of, you know, a cost volume, right, it is called a kind of a cost volume, that means, you warp and then, and then you kind of build a volume and through this, right, you try to, you try to, you try to kind of see, measure what kind of similarity you have. So for example, one way to look at it is if I compute the variance, right, along the, along the frames, then wherever I get a low variance, that means, that those guys, those, those guys are all okay, that means, that this is like one volume is for, is for, so one depth. All views warped with respect to one depth will be like one volume, right. You guys follow that, right.

I am just saying I have these, these very different views, I choose one depth and I warp it, all of them, then what will you get, I mean, you will get a, you will get a stack of frames, right. If you look at the stack of frames, they will not be aligned because it is not like all points are going to get aligned but there will be points, right, in this volume wherein if you compute the variance along the stack, then what will you find? You will find that there are certain points that had the same, that were lying at that, at that depth of the plane. So whichever view you take, because your r and t and all are already estimated, so what it will mean is, they will align, they will align, they will kind of line up, right, in the, in the, in the, you know, right, I mean, in that place because this is after the warping, right. This is not a reference view, right. This is a reference view warped, right and, and when you, when you kind of, when you kind of say compare it, compare it with, with your, with the image that you already have, you will find that, you will find that, right, that there are, that there are these pixels, right, that are actually aligned, right and therefore for them the variance is going to be very low, that is for the other pixels that are not aligning, right, you will have a variance, right, which will show up to be high because they are not, they are not being, right, they do not get aligned and therefore reducing that you can tell as to what value to assign, right, to that particular, particular this one, the pixel location.

So you know that, that this pixel is at, is at, is at depth, depth is d1 and so on, right and, and maybe for another plane, right, another set of points will come and then they will line up and therefore you will get that or different depth and so on. This is again, you know, this is again I think, you know, another, another case of, case of the same thing, okay. I just thought I will just show you these examples just because I have been telling that I will show, but.