Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-79

Today, we will move, shift gears and move to what is called mid-level vision. So, until now what we saw was deep networks in the first phase of this course, then that was followed by low-level vision where we saw about edge detection, feature detection and all that. And then we moved to a geometry where we saw single view, two view, multi view geometry. The next in line is mid-level vision. This mid-level and high-level are not as big as the other ones. And mid-level vision in which the tracking part that comes in mid-level vision, I will combine it with YOLO when I teach YOLO for high-level vision.

So right now the first topic that we will focus on under mid-level vision is what is called optical flow. And I will actually explain along the way what we mean by this optical flow is by definition is the apparent motion of brightness patterns in the image. Now this the use of apparent rate comes for actually a reason. So this optical flow right, I have to just pull out this slide.

The thing is right see what we actually mean by this optical flow is that right it comes in when there is motion right. When there is motion it could be just object motion, it could be camera motion, it could be a relative motion between the object and the camera both could be moving or whatever. But then motion is such a strong cue right that you cannot actually afford to see right neglect it. In fact you know many a time motion is the strongest cue in a way. I will show you some examples.

And therefore right when you see what you mean by optical flow is that right we are kind of see no longer talking about one frame now. We are talking about frames that are varying with time right. So you have something like a video for example right where you have a successive bunch of frames right which you are watching. And when you are watching this frame right so what we call as spatiotemporal information right. I mean normally when you have just image you are talking about spatial information alone right.

So we are talking about what is called as spatiotemporal data. So the spatiotemporal data right so what we are saying is if you are looking at this image and if you can analyze what is going on in terms of the variation of the brightness patterns that can tell you a lot about what is happening in the scene right. Till now we were not so focused on object motion and so on right. When we were kind of looking at more or less static situations and so on. But now the

next right escalation in a sense comes when you want to look at brightness patterns that are varying.

Now when these brightness patterns vary right I mean typically look at a scene right where there is an object right and the camera is imaging this object and the object is moving. Now the motion of the object itself right when it is moving that is called scene flow okay that is called I mean as opposed to this optical flow okay that is called scene flow. And this actually refers to or you can call it scene motion this refers to 3D velocities of scene points okay. So it is like the actual motion of motion of points in actually a 3D world okay that is what we mean by scene flow. Then from the scene flow right when a camera is imaging through a perspective right a projection equation then what you get is actually a motion field right.

I mean this is what gets transferred to the camera right because we know that we can calculate where the x y coordinates are and as the object is moving you get a motion field right which is like saying where are they where the x y points going. And the motion field right this is the 2D projection of the physical motion of points of the physical movement of points in the 3D world. So this is a 3D projection on the image plane of course on the image plane of the physical movement of 3D points in the world points in the world. Now unfortunately right this is what we want actually right this is the motion field I mean this is what will tell you what object is moving which direction and where is it going and so on. But you do not what you observe is not the motion field right it is not observable it is not observable we cannot observe it we do not observe it actually the camera what we observe is the actually right intensity right the brightness pattern as it is called.

What we see is just this intensity in the image right we do not get to see the motion field directly but what we see are actually intensities. The idea is that these intensities right in turn should help us compute the motion field so but then what we observe however right what we observe are actually intensities. And the idea is to analyze these intensities in order to be able to make us inferences about the motion field right which in turn can probably tell you something about the scene flow and so on right. And what we observe are actually right intensities. Then so to say so right so what this is effectively means is that that for example right so what you are sort of saying is that based upon the optical flow right I want to be able to say within an image right by how much has a point move when you go from one frame to another one frame to another right how much is right let us say at every point moved in the sense that you want a velocity component which will be one along x direction one along y direction in order to be able to tell where every point is going right.

I mean you are kind of looking at the velocity of every pixel now right that is what you mean by we mean by the kind of motion field and that you are trying to infer based upon the optical flow which is simply the illumination right that you actually see. And theoretically I mean ideally right not I mean theoretically I mean ideally right you would want the optical flow that means the movement of the of the patterns ideally right you would want optical flow or the motion of the brightness patterns you would want of the optical flow to correspond to to

correspond to the motion field right. But in reality what happens is there are cases when it is not true in the sense that you may have a motion field but then you do not see any optical flow and you can have optical flow but then right there may not be a motion field at all okay. But then in most cases in most cases right it does work in the sense that this sort of a correspondence is true thankfully right in most cases it is true but there are exceptions where let us say such a thing right may not be true that is why we say apparent motion okay I mean it does not always represent the motion field but in most cases it does okay. So in a sense right what I think right we can say is that right we can say that but in most cases most situations in most situations the optical flow corresponds to the motion field corresponds to the motion field okay so in a sense right it works in most cases and that is the reason why there is so much interest in that.

But to just give you an idea about what the optical flow is and so on right let us just kind of go to the slides and we will kind of look at you know what might be the might be the use of you know having what might be the use of analyzing a spatial temporal volume right so to say in order to make right inferences. So one is visual motion right so when you see this visual motion it can be due to various reasons it can be camera instability it can even be camera motion for that matter it can be a jitter or whatever this causes visual motion and visual motion actually you know so it could also be because right some object is moving there I mean right so what we have here is not the whole set of things that can actually cover some few are listed here but then you can think about you know right visual motion can occur right due to object motion that most of the times right that is the case and it kind of you know indicates a dynamics in this scene right. When you say a dynamic scene right you really mean that some things are moving independently of the others I mean that is when you mean it is a dynamic scene for example right just because an image is changing over time it does not mean that it is actually a dynamic scene for example right I could have a perfectly static scene but I just move my camera so if you look at all my images right they are not the same but then right that does not mean that my scene is actually dynamic scene is not dynamic I am the camera is moving right that does not make the scene dynamic. So when you say that a scene is dynamic you really mean that there is one or more objects there that and there is a background perhaps right and there is more than one or there is more than one object perhaps there which is causing which is going through some independent motion or which are going through some independent motion right. So you have like moving objects behavior and most of the time you see this right somebody is you know you could have a game right where let us say right the people are playing the camera could be static but then there is so much activity going on right and you want to make inferences trajectories right object tracking so much you can do if you knew what was going on right if you could actually predict I mean if you could analyze that movement of the of the of the brightness pattern and be able to tell certain things.

And you know it is also nice that you know this also includes parallax in the sense that something being in the front will move something you know let us say even if it is a static scene and if you are moving a camera right there is going to be there is going to be a motion

pattern right because like I said it is a relative motion and at that time right you will notice that things in the front would have moved faster than things in the back which is the parallax effect again which means that which means that which means that right on the on the image plane when you see this motion of patterns right you sort of tend to you tend to know that something is moved faster than others and therefore and because the camera translated uniformly for everybody right so that sort of tells that something is in the front and something is in the back that is the parallax cue. In fact in fact right going the going a little backwards right the the Thomas C. Kennedy factorization method right that we did is actually one kind of optical flow. If you if you if you kind of recollect right we had we had an observation matrix right what is that observation matrix that just had the had the u and v in it right across all the frames if you guys recollect just a few classes before right we had a we had a we had a we had a motion matrix that had only only only use in v is in it right we had like various feature points coming along the column and then across the frames right there was there was a correspondence. So what is that really right so it is like saying it is like an optical flow you are telling where this point went right where this point went where that point went that is exactly what you are trying to indicate and that led led us to compute the structure of the scene right we could see we could say what was the structure of the scene.

So in that sense right we have already utilized it even though we did not sort of explicitly talk about it as being optical flow or flow or anything but yeah I mean it is optical flow is something that is not new to you in a sense right we have already used it but then used it in a sort of an you know indirect way. Now think about this right so sometimes motion is a foremost cue so you actually right there are there are even even GIFs and all where you see a small motion and it is a motion only right that actually tells you that there is something sitting there otherwise you know the thing is so very camouflaged that you cannot make out that there is a different object sitting there. So which means that motion is sometimes a foremost cue and here is what you have right in a sense you have a spatio-temporal volume right and you sort of represent it as I x y , t right until now we are only talking about I of x , y now we introduced a third sort of you know dimension which is the time right like I said this is now spatio-temporal data this is like a video. So and optical flow is typically between kind of like 2 frames and then you can do it successively. So the motion field versus optical flow right so think of the situation do I have like that okay yeah I do not have the slide for that but then think about the sphere right which is there which is kind of uniform right in the sense that the sphere it has the same sort of reflectivity wherever you go the material is exactly the same okay this sphere.

Now if I had this sphere and let us say let us say this sphere and there is some kind of a lighting right from there which is why you see that there is on the one hand there is some shadow and on the other hand there is a more brightness right that is because because you can think of this as a light being somewhere coming from there right that is why that portion is more illuminated. Now if you actually rotate this sphere now which is like an actual scene flow there is actually a scene flow right that is object is actually rotating what will you observe in the right images if I capture such a video what will I see? You will see exactly the

same picture right all through you will get the same picture at all picture because even though the object is moving right so what this means is that there is a motion field actually there is a 3D motion but then on the image plane right you do not see a change in the intensity at all right. So in that sense this is the situation where there is a motion like I told you right it is not always true that one means the other and as I always as I said earlier right so what you can observe is only the optical flow. So does the optical flow right imply the motion field or not right need not always so in this example what have you got? You have got motion field but there is no optical flow but now suppose I get the object still and suppose I move the camera right suppose I take a video but kind of moving the camera not the camera moving the light source I had a light source here now suppose I move the light source okay now there is no motion really right because this guy is still my camera is still watching from here but I am kind of moving the light source then what will happen you will see that this brightness that you are seeing on one side of the sphere will gradually shift and then right it will come out in this side. Now if such a video if I capture will there be optical flow there will be optical flow because it will look like look like that there was a pattern of motion but then really right there was no motion right in the scene there is no scene flow the object ever moved right so in that sense there are these extreme situations where there is no motion field but then there is optical flow or there is no optical flow but there is motion field but then in most cases right it works well okay so just because you know there are some extreme cases where right things do not work out it does not mean that we should simply throw it away okay.

Then why is optical flow useful like I said at depth 3D reconstruction that is what I told you just now which is this one example could be this Lucas Canade then motion kind of a detection right that helps you not just track right I mean you can even do action recognition and all right people do based upon how well you can track different body parts you can say what somebody is doing what is action is then compression so many things right mosaics right again I mean mosaic and all is all about right I mean having an optical flow and more right a lot more that you can do okay maybe some examples will come along the way. And as I said right so if you see this see the small little things right I mean so at every sort of a pixel right that is actually a motion vector right so in the sense that right the optical flow what you mean is at every pixel you want to say how this pixel move from here on to this next frame which means that you have to see of course 2 frames right and you are trying to say how it moved and the length of this vector right will mean at not what sort of rate of velocity or what speed it moved and then the orientation right will tell you as to right which way it is moving. So in a sense right typically they would show it like this at a particular location right you will have a vector that looks like that so this length tells you at what speed is this guy moving and the orientation tells you right in which way is it going and you can actually do it at let us say at every pixel and that in a sense gives you a notion of what is moving and what is not and so on and that is a big cube by the way as you will see further. So here are some examples right that you see now the first example right I mean where you see that if you look at the brightest patterns right they all seem to be seem to be coming inside right moving inside what would that be it will so suppose it is just you are just doing a camera lens adjustment right so it will be like zoom out right because when you zoom out I mean all

these patterns will start to shrink and then right come towards the center opposite will be zoom in right when you kind of change the camera you know focal length and then you get a zoom in so that is like right expanding and that will create motion patterns to go out. Then if you have a camera that is spanning from right to left right you will get actually motion vectors like this and you if you notice right they all are of the same length right that actually tells you a lot about this scene okay we will see right down the line there is one more example coming that which will make it even more clear.

Then you can have a camera right instead of let us say this kind of what you call a zoom in right you could have a forward motion right which is by the way right I mean you know changing the focus is not really a depth cube okay moving is actually depth cube okay. So even though these patterns and all may look alike right out there also I am everything is going out here also things are going out so are the two equal and all right we do not make those conclusions but we know that these vectors right are if you kind of find how those vectors are moving so when you actually move forward right you will get actually the motion to be as shown there. Rotation if you happen to be rotating about a point right then you can see that you know if you are rotating about that point then there is no optical flow there because that guy just remains stationary but then as you go around right you see kind of more and more motion especially at the boundary for the same rotation right you will get a lot more motion at the boundary right which you can imagine right a point I mean it is like this right I mean it is like having a vector right like that and then if you see right for the same angle a point here right can only move so much whereas a point here has to move right that much more right that is why in a rotation you see that the optical vectors right will sort of will be longer in length as you go outwards to the boundary and of course you know they kind of you know they also indicate a direction right which sort of indicates that there is a kind of a rotational motion going on. Then the next case is that of a horizontal translation right so here you see that this is something similar to this but then right this one this one right this one the interesting thing is right I mean so in this case suppose I asked you right what to say yeah so is it like is this a camera motion or is this like multiple objects moving this situation how would you make that up just by looking at the optical flow right that is the optical flow what would you conclude? Is it like a dynamic scene the camera is static or is it like the camera is moving and there are probably multiple objects some of them moving and some of them not or what would you say actually right this is a case where the camera is static okay the camera is static and there is an object behind which is moving and there is an object in the front which is moving and the behind object is like it looks like that is one so you can imagine that you know that is like probably it is like one big object that is moving together in a sort of a cohesive way that is why all its optical flow vectors are having the same length and all that whereas you see that there is a larger optical vector in the front right which means that there is an independent object in the front of the camera which is moving. Now why I said this is not camera motion is because right if one minute what did I one minute let me just think about it so if you had wait a minute right so if you had no camera motion oh okay this is just the opposite way no okay what this actually means is that the camera is actually moving okay because the camera is moving right I mean you get you get a kind of motion in the background

and because there is an object in the front because there is an object at the front right it is moving faster if there was no camera motion then what would happen the background would have actually would have actually 0 this one velocity.

Well the first one that I told is also not incorrect I mean you can think of a situation where there is a where there is some big object that is moving in the back but the simplest thing that you can sort of relate it to is if the background right everything looks like 0 optical flow that actually means that the camera is static and therefore you do not see any motion in the background and probably there is one object that is moving in the front which will kind of create an optical flow okay because there is motion in the background right this can this can be actually you know it is not wrong what I said right what it means is that it could be that the camera is moving and therefore the background seems to have a motion or it could be that the camera is static but there is a background object which is actually moving right I mean all all 3 things can occur is that okay I mean right if you think about it I mean all 3 can occur okay they are all valid. So for example I mean you know here so here is a 0 velocity right so it means that you know nothing is moving in the scene and then if you had something well in this case right here is where you would conclude for sure something in this case you can conclude for sure that the that the camera is static because because the background is all having 0 velocity right I mean you can think of a more complex situation but I do not want to get a venture what is that what is that complex situation exactly yeah that also can happen right I mean what what he said is correct right it can also happen that the background and the camera are exactly moving in the same same with the same velocity in which case it looks like the background is always static but there is something else in the front which will have a different optical flow but yeah but then actually right that is why this optical thing is a big you know there is a big psychological study on optical illusions right you must have seen therefore it is not true that whatever you see can make you know very you know clean inferences but most of the time yes right you right you will be able to tell what is going on okay so here right it is like it is like it is like there is an object here which is actually right which is actually right moving to the right okay that is an that is an independent object that is moving to the right and right here is a case where where the where the object is moving directly right towards the camera right which is the reason why you have this outward vectors right going from that point whereas the background and all right is still having kind of a 0 velocity right. So, you can have all these situations now where can you use it right I mean it is a big queue now here right nothing is moving in the scene but then your camera let us say has moved and you already captured multiple frames right then the optical flow can help you segment this image right because what it will tell is this tree in the front will move more and therefore it will have a different optical vector and the optical flow and then and then right the kind of the house the background right it is some other depth these guys are at some other depth right therefore you can see that you know based upon the optical flow itself we can segment this image right which is which is like a big queue now. All all that all that right you did was just to take the camera and move and then right you can also segment right objects based upon their optical flow because you know because if it is a rigid object right then then the whole segment will have the same optical flow you can kind of

group them together you can group you know a different bunch of objects together you can group a different bunch of objects having similar optical flow together I mean that way already you also know that there is probably multiple moving objects right in the scene and so on. Then 3D structure we already saw right I mean you know you can do something like structure from motion you can actually do something like super resolution and all right I mean I do not want to actually enter into this then recognizing events right I told you already okay now let us let us kind of do the math now okay this is fine now how do you how do you even how do you even right go about go about right I mean right analyzing optical flow now the way right it works is as follows.