Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-50

So, he goes for actually a pyramidal structure. Why a pyramid right will become obvious, a pyramidal structure. And as I said right I mean why so earlier right when we talked about a pyramid right I said that a pyramid is useful under certain conditions. What was that? When did we say that a pyramid is useful? The pyramid you understand right. So, what we are saying is you have at a certain this one okay resolution then you down sample and down sample then you down sample. If you do that I said that there are certain advantages to kind of having a pyramid like that.

Can somebody recollect? One of the advantages that I had mentioned at that time is if you wanted to sort of what you call if you wanted to align two images and if you wanted to find out what is the underlying transformation. Then it is easy to sort of work at a lower resolution get a good initial estimate use that to start the if you are doing it iteratively use that to start the next thing and then it will help you to kind of say converge fast. In a more stable way than sort of searching everything at a high resolution that is typically the reason why let us say people go for you know this one a pyramid okay. Here the idea is that in a pyramid by which he calls each one is an octave down.

Octave is like what I mean either you go up by 2 or you know or you go down by half right that is when you say that you are moving by octaves. So, what he does is the following right. So, he says so what he does is so for example if you still go back to what we had. So, we had σ k σ right all the way up to what is it we had something like k power 4 σ we had this right. Now this call this is octave 1 okay which is at the which is at the highest spatial resolution.

Now see somewhere here right we would have had a k square σ you know k square σ then we have a k cube σ. So, the k square σ is actually 2 σ right. So, what he does is he creates another octave which is called as octave 2 and in octave 2 right what he will do is this you see 2 σ that you had know this is actually an image here which is actually blurred by you know blurred by 2 σ. So, this 2 σ blurred image is actually down sampled. Down sampled means just decrease its size by half along both dimensions and then at this at this decreased size you gain you gain again right I mean you know.

So, for example you have something like you see k square σ here right. So, k square σ is like you know 2 σ right. So, you start with 2 σ then the then the next one is again 2 root 2 σ goes up like that. So, you see you can think of the next guy here down sampled down sampled and then you have another bunch of images sitting in that octave. The I mean the reason right why why something like this is done is because instead of scanning structures at at one resolution and trying to find out right where I where I can get my structure instead of that if you have a 2 σ blur and if I down sampled the image right then it means that then it right effectively means that a large structure right could still be captured even at a even at a even at a lower σ right at 2 σ because you have down sampled they made a bigger structure would now become smaller and even with instead of going up to 4 σ even at 2 σ right you might be able to able to you know trap that trap that structure right you might be able to kind of catch that structure right.

So, that is that is that is the reason why. So, instead of working at one resolution and going through all the scales it makes a lot more sense to actually down sample an image and of course aliasing you do not have to worry because this blurring and down sampling is actually a good thing to do right like I said it is directly down sampling is not a is not a good idea that is I think we discussed right at one point of time. So, that is anyway happening here because if we blurred it with with some Gaussian which is a 2 σ and we are down sampling it and then. So, there is a next octave starts from the 2 σ from the earlier octave, but then that 2 σ down sample and then it goes on then if you want one more octave you can add and so on the the the. So, he shows that reducing this I mean you do not you do not miss things otherwise you would have to computationally do a lot more in order to ensure that ensure that you get you get most of the features and there is no way to no way to write that there is no way to verify this.

There is no way that somebody there is nothing like a ground truth right where we can say that right these many features I have to get and then and then which algorithm gives me that right we can think of that no no in an image right it just like I said it just fires wherever it feels right things are there is a there is a good activity half a foreground with respect to its with respect to its background I mean that is the way you have to see right an activity that is that is significant enough at a region as compared to what is around it. Ok whenever that happens it is interesting ok that is what we that is that is loosely the interest point and each one could could have it in a you know right in their own ways and this this this paper builds it up like this ok. Now now comes the point as to the oh oh ok ok oh you are saying how do I come back and and locate where it is well because you sampled it by that factor no so you know so you just have to come back. Yeah so sorry this is also the reason why why for example right when you are coming back right you still have to do some sub pixel accuracy alignment and all you should do otherwise it should not happen that that when you expand it right you lose out in terms of the spatial accuracy that is why

that is why they do not go too many octaves it is not like you have octave 3, 4, 5 and all they do not go too many because when you come back right your location could be in trouble. Yeah so so I think they just go like you know I think it is just octave 1 and octave 2 I do not think there is there is too many octaves there max I think it is only octave 3 or something not more than that yeah that is a valid question.

Then ok so one of the things right so this so this is a key point right as he calls it or we might want to call it an interest point. So the key point right it defines the key point through something ok what he calls through or through a dominant orientation. So instead of octave 2 in octave 1 itself we can go to higher σ but because of the dominant orientation that is the only reason yeah that is the only reason. So this dominant orientation it is also it is like this right so what we want to do is if I have an image and if I find my if I find my if I if I apply my shift right this algorithm then I want to be able to tell that that I have a feature here I have a feature here I have a feature here whatever right I have bunch of features each is a shift feature but I but in addition to that right I also kind of want to be able to say something more about it ok when I even show that feature ok. And that that is that is an that is called an attribute of that feature a descriptor is something else ok.

A descriptor we will we will write a descriptor is something that will that we will use eventually for actually matching and so on. But even to do to be able to define a define a key point just by the extreme right he does not define a key point ok. Now, because all of this is kind of spread out right I mean it is not easy to catch them all under one roof so I will just I will just indicate it indicate what those steps are. So first of all a thresholding is done on the on the on the extreme arid so that whether it is something which is just a just an incidental extreme arid is not taken into account a thresholding is done on the thresholding is done on the even log or dog extreme arid ok. That means only the things that cross a certain value right will even be kind of you know will even be considered.

And then a unique orientation is assigned and it turns out that you know you can have ok I think it is assigned to each extrema or key point or to each interest point. So, how this is done is right so around the local extrema so if you have found a local extrema right around the local extrema what you do is it is I will write down ok. So around the local extrema a suitable window is chosen this can be 16 cross 16 something like that is chosen. Again there are some hyper parameters and in fact this strictly speaking this window in fact depends on the σ the scale at which you are and so on. But anyway let us not let us not let us not let us not go no go too fine into the implementation issues just want to understand that what is going on.

And the histogram yeah this is what is important of 36 bins each bin of each bin of 10 degrees each bin of 10 degrees is computed using the points around the extrema. What

does this mean the points and first write this then I will tell you what it means of course, it is fairly self self explanatory, but anyway. So, what this means is that like for example, you have you have a kind of right this an orientation. So, you have a theta going from 0, 10, 20 degrees all the way up to say 360 degrees. So, you got you know that many bins and the and the and the idea is that is that right when you actually when you actually look around this local extrema this is your local extrema.

No, actually right this is for knowing the dominant orientation of that point. So, that right when it appears in another image right you should know how that orientation is. You should know the kind of relative orientation between the two and that helps you to solve for these rotational ambiguities. That is what I am saying right these are like you know you could have considered 72 for all you care right yeah. Well some in fact, some papers I have seen they write like you see 24 or something you know which could be a 15 degree interval right.

Some people say 15 that is an implementation issue that is what I am saying let us not worry too much about why 36. Yeah how ok if you are asking why 36 that is for computational purpose and maybe it will still work if you get if you have see 24 just a matter of coarser theta that is all. But I am saying more important than that is the idea behind when what this bin is trying to capture. What is bin is trying to capture is for example, right I mean you know at each point you know to compute the gradient magnitude and the orientation right. If you have an h pixel we know how to calculate using a local neighborhood around it we know how to find out I mean you know that right you know you just have to do tan inverse gy by gx that will give you the theta and then magnitude is like gx square by gy square whatever right.

So, you know both you can do. So, what is done is so, if you if you if you find a certain angle right and if you find a certain gradient strength right you kind of drop it into that bin ok. And then and then another point that you might you might find you know different orientation different strength, but what you what you are dropping in each of those bins is the strength of that this one right. So, it is the it is the gradient strength of the gradient right.

So, you are plotting. So, what you are plotting is the strength of the gradient with respect to the orientation. And what you do is the one the one bin that gets the maximum orient maximum strength right that that is what you kind of tell is the kind of dominant orientation of that extrema. So, it is what means that. So, it means that around it right many of them are kind of aligned in a certain way right. So, that you want to say that it is like you know what to say I mean you know it is like saying that if you had a local structure and if you had an arc or something right then maybe it has a certain orientation right and that and that

or it is a it is a line segment or something which whose orientation you try to capture.

So, when you say that that extrema has a dominant see the extrema itself will have its own orientation right, but the dominant orientation is coming from examining what is around it and trying to find out what are the strength of the gradients of those pixels, but what orientation are they and then once you kind of put them all into this into this sort of histogram then you get one bin right that will be that will have the maximum this one. Are you guys following this? So, the gradient right. So, maximum gradient will appear in some bin and that we declare to be the dominant orientation of that key point. This is ok right. So, you have something like a like a see dominant orientation.

So, just to finish this let us say the bin with highest peak. So, the bin with highest peak the bin with highest peak or the highest gradient norm or whatever or the highest gradient strength. Is chosen as the as you say reference orientation, as a reference orientation for that key point, for that for that key point that means for that extrema and in fact, one more thing that they do is any other bin any bin with these are all again that these are all actually very fine implementation issue, but I thought this is a little this is a little more interesting. So, I thought I will anyway write greater than 0.8 peak is also considered a possible orientation.

So, what this means is that that at that point right when you show this. So, for example, right so, so, ok. So, if I come back to this figure. So, what will happen is right. So, when they when they want to show a shift know sometimes right you find actually you may find actually two arrows even one like this and then another like that ok.

If you are wondering what those two arrows mean what it means is that there are actually two possible orientations perhaps right. And one is like then the strength of the of that norm the gradient norm is what is the length of that arrow and the orientation is whatever is this orientation right. So, anything that is above 0.8 peak right they kind of suspect that you know the thing is that when you are trying to match it with another ok, when the same image appears under let us say you know a different orientation and also at that time when you try to create a match. So, if there are if there is multiple match possible right like for example, a lot of them even it is one of them with respect to whether one of these key points if you are able to match it is like at the key point you have got you know two two options right when you know either of them could actually work when you do the matching.

So, this is to make sure that you do not miss things because you are right eventually when you when you go to stereo and all right this all this will come very handy. Why are we doing this because when we when we go to go to a geometry right geometry is all about

picking points because if you have plane image there is no geometry that you can make any sense of right. I mean if I give you a plane image what geometry can you even talk about in there right. So, geometry comes because there is a lot of activity in the image and there whenever you want to find out structure when you want to find shape of an object right all these points you have to find right what those points are where did it go in the other image where it should I look for it. So, if so at those all those places that this will come and the more robust matches you have the better you have a sense of the scene right.

I mean if I have a 1000 points that that that I can match that gives me a much better sense of what I am looking at than let us say having 100 points right. So, so, so, so, so, the idea is that these are all kind of what do you say you know this one enablers. So, these are like enablers to get the match going. Okay. So, which is also the reason why you have why you could have.

So, so, so, the so, the key point right so, you will typically find it as some x y what is that σ and then actually a theta and then theta can also be theta 1 theta 2. Okay then comes a description right this is this is this is only to say that I have these many key points and these are all to be matched right and then some of them could surface in the other image some of them may not surface also okay. Sometimes because of occlusions right you do not even see them. So, all of that can happen. My idea is that I have got an image right here and I have got I have got to say another image of that scene okay and I want to be able to able to right suppose I have got these you know shift points that I have got that I have got here and I have got some shift points here and then I want to be able to able to find a match right I should be able to say right that this shift point is this I should be able to say right that that shift point is that and so on.

And in general right and we do not want to make you can of course you know make this make this match itself elegant if you knew a little bit more about the underlying scene. For example, if you knew by at what angle this was taken these two images were taken and then maybe that that simplifies that that will sort of automatically constrain where this guy could have gone right. But in general right we do not if you do not want to assume anything and we if you still want to want to do a match within within a reasonable search region right and not really constraining it to a line or something like that which you can if you knew a little bit more about what is going on. So if you just wanted to search within a local space right you want to be able to have a descriptor right. So what you what you need is actually a feature this one a descriptor.

And again the sub pixel accuracy and all that is there right I do not want to repeat all that whatever we did in Harris current detector right something similar also happens here you have to get down to the sub pixel accuracy of the extrema and all. I am not going to repeat

all that we have already done that so you can just you can just read it up you will easily follow that now. Having done that once I will be able to keep repeating. Now, the shift this one a descriptor is a 128 cross 1 vector dimensional vector dimensional vector is a one terminal and which is derived as follows. Now, we have I think I can finish this now what do you do? So you take a 16 cross 16 area or grid area is chosen.

So it is like this right. So now you have you have the interest point right and that another say reference orientation is something which you want to which you want to which you want to store in any case ok the dominant orientation which we found. Now what we are saying is we have this interest point and around that right we will so these steps are very similar to what we did just now, but there is but there is some changes. This chosen and the gradient norm and orientation of course, you know some of these things if you have already done for when you actually defined the k-part you can use use that information you do not have to recompute are found for for each pixel for each pixel. Then their angles are recomputed. So what this means is that so what this means is that you have you have these angles right.

Now these angles are recomputed I mean when you do this right you will get the the orientation right. So these angles are recomputed with respect to the to this dominant orientation. That means relative to the to the extremum point where these are with respect to this dominant orientation or what is called using the say reference angle I think this is a different different wording using a reference angle using the reference angle for that key point for the shift using the reference angle of the shift feature point. Of the of the no say say again difference between the gradient angle and orientation not of the not of the key not of the key point right. So what we are saying is we were trying to find out around the key point with respect to the key point where are the others.

This is like see you are finding the gradient norm and orientation are found for each pixel right that means around the extrema and and for the extrema itself right you have like a kind of a dominant orientation which you have assigned it. So now now what you are saying is with reference so for example if you have an absolute angle see typically when you compute an angle it will be with respect to the x axis right. Instead of computing with respect to the x axis if you have a dominant orientation with respect to the dominant orientation where are these where are these other points where their angle is with respect to that. Sir, so are you computing the angle of the point or.

Point of these points. Ok. So how will you store it. We will see that. So you are calculating the angle between the orientation and the point where. We are trying to see for example what you are saying is you are taking a neighborhood right around the extrema you have already found a dominant orientation for the extrema that means this whole sort

of local surface right has a certain orientation and what you are saying is the points when you compute orientation for all the other points that will have a reference direction which is which is your x axis. Instead of that you are trying to find out with respect to this this key point where are they oriented the reason being that if let us say if in the second image right if the whole thing.

So I think we have run out of time right we will we will just continue in the next class.