Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-53

When you have a 9 cross 9 filter of the kind that we had that I had shown, let me see if we have that here, something like this right, is what I had told you. The approximation to this, to the kind of true log, it looks like this, so the approximation it looks like this and okay, I think something got shifted anyway, it does not matter and then your dxy, it looks like this and I was asking you can you still use integral image concept to be able to compute dyy, dxx and dxy, that was the question. How do you solve this? Hitesh, how do you compute using integral image? Something like this. So it is like this, I will just give you an indication about how to kind of go about doing this, it just requires a shift of I σ, just requires I σ to be shifted up, down, left, right, that is all it takes. For example, think about this okay, let me take the same example, I am just going to drop hints now okay, because you guys should also do your part right. So 1, this is what we had, 1, 3, 2, 4, 5, 2, 1, 6, 4, 6, 2, 5 okay, this was I and then 1, 4, 1, 1 right, this was the example that we had taken and for this we got an I σ which looked like, because it had 0s appended to it and then 1, 4, 6, 10 I think, 6, 11, 24, 0, no, no, no, sorry 10, 21, 26, 41 and then 11, 26, 32, 48, 32, 48 and then we had appended 0s on the outside okay, this is what we had know.

How to get it from I σ? Okay so it is like this right, so imagine for example, suppose I had a filter like this 111, 111, – 1, – 1, – 1, it does not matter, the third row could have been – 2, – 2, – 2 for all I get right, but then it is like it is like get one sort of box, another box, another box and so on. So now what you can do is you know, you can actually think about having 2 box filters now, one is this guy, another is – 1 times this right, which is like 111 okay and the idea is that you want to be able to use I σ to be able to do everything, I do not want to go back and recompute anything right, then it is like your usual thing which we know how to do. So suppose I sit here right and suppose I apply this 111, 111, – 1, – 1, – 1, I should get what should I get? 6 + 2, 9, 15, 17, 22, – 6, I should get 16 right, if I convolve at this location with that kernel, my answer should be 16 okay. Now the point is right, if you actually keep this kernel here right, then with respect to the box filter 1 right, if you try to compute what your ABCDs are right, so if you try to compute what your ABCDs are right, so your filter will sit here right, if you just take the top box which is the box 1 right and your center is here right for that filter, so actually keep that here right, then your A okay will then be what? Your A will be 1 right, your B will be 10, C will be 24, no C will be 10, C will also be 10 right, by 11, you are sitting here no C is this guy right on top of 24, why C 11, I am sitting here, C is 10 and D is 41 right, so now if you do A + D that is 42, – 20 which is equal to C 22 okay, now if you convolve 111, 111 you should get actually 22, 6 + 2, 9 + 6, 15 + 20 + 2, 22 okay.

Now with respect to another one, – 1, – 1, – 1, one way to get of right you know think about it is to actually shift this, see because – 1, – 1, – 1 is going to act on 26, 30 to 48 right or in other words you can think of i σ being you see shifted, either way right, either you think of – 1, – 1, – 1 acting on but then that is not the way you want to do it right, you want to do everything with respect to i σ, show that right if you were to take – 1, – 1, – 1 and all that it requires is for you to get of you know kind of shift this i σ up and then do this A + D – B + C and then take this minus that and that will give you that 6, I mean you can write away C you know, see for example if you take – 1, – 1, – 1 that sits here, so if you look at your A for that right that is like 10 and then B for that is 41 right, C for that is 11 and then D is 48 right, so what do you get 48 + 10 is 58 – 52 right that is 6 which is what you would have got 4 + 2, 6 and now you do 22 – 6 that is 16 and that is the answer here. So all that it requires is for you to be able to shift i σ and simply do this A + B on the one for the for one box filter then this A + B what are A + D – B + C for another box filter subtract the 2 and you have the answer. Okay now let us kind of move on okay, now to come now let us kind of look at you know how the okay Ranish what you are not convinced or convinced. Yes sir I need to try it. Yeah try it out I mean it is not any great rocket science right okay, no.

So let us come to this key point key point courtesy detection, okay prior to that right there is something else okay which I have to talk about. Now similar to the shift right you need this what you call this kind of you know an octave thing right even here but then right I mean as I said right so what these people do is they can actually go for right any size right I mean so in that sense right they have that disadvantage but the way right this is done is you take the first octave, so this is actual implementation it goes like this. So you call this as octave 1 okay and there is a certain kind of say reason to start like that okay so for example this I said is you see 9 cross 9 okay so I will just write it you know 9. So the first size right I said that was corresponding to σ equal to 1.2 right that is the minimum at which you start, you start here as a approximation of the log or the dog with a box filter that is at minus 2, 1 that approximation right which you made then what they do is that the even next one is that actually 15.

So it is like you know in one octave when you go up the scale right there you went like σ k σ right k square σ that is how we went right now here we do not have a corresponding k but instead what you have is a jump of sorry 21 right 21 and then 27 okay now there is a certain reason right why basically right now this is done actually this is actually I mean in order to completely understand this we need about 3 to 4 classes okay if you actually go through the exact analysis as to what a true linear Gaussian scale looks like and what approximation it will kind of you know bring it best to a box filter and then what will actually you know what will actually preserve the energy in the Gaussian what was the energy there how will a box filter version preserve the energy so lot of calculations there okay one needs about 2 to 3 classes just for doing that we cannot we cannot afford that right. So what I am going to do is I am just going to tell you briefly as to where these things emerge from so that you get a rough idea as to how this happens so the second octave right I mean I will first write down okay and then I will tell you how this comes second octave starts from where does it start from 15 and then here it is a jump of 12 okay 27 then I am not drawing to scale okay 39 then 51 and then octave 3 right goes up like this starts from 27 and then here it is a jump of 24 so it is like 51 what is it 75 and then 99 and then that you can have you can have further octaves also. Now the idea is that right I mean you I mean since you see computer even at 99 cross 99 there is no big deal right to do the computation because you know everything can be done with just the integral image all of these can be done with integral images okay. Now they still want to use octave simply because if you keep right domain and you know expanding the expanding the support right then you can actually catch them early on. The spacing rate that you have here see one I said that right this is like σ is equal to 1.

2 approximately okay this 15 and whatever this 27 are all chosen such that the mid way σ at either extreme okay that means if you compute between 9 and 15 you got like 12 right and between 21 and 27 right you got like say 24 and this jump of 6 comes roughly for a k of root 2 okay this jump of 6 pixels in the first octave right. You can roughly think of for example if I compute the σ here okay this would be like this is like 1.2 into say 15 by 9 okay that σ if you calculated it will be roughly root 2 times that 1.2 that you had in 9 roughly okay it is not exact none of these calculations is exact okay these are all approximations. So one way to think about it is the σ here and the σ here right it is like almost covering 1 octave it is like you know you have like you know like you have covered a whole σ right in going from 12 to 24 because you had whatever σ here you will have say twice that value of that σ here.

And similarly right if you go here okay this one will be like what is this 15 to 27 means 21 so you got the mid one is 21 and here it will be 45 right. Now 21 times 2 is of course not 45 but it is roughly like that right so in every octave the mid σ value at both extremes right are such that you cover roughly 1 σ okay across that octave and when you actually go up the jump is by 2 because that is called a dyadic sampling I do not know whether I used the word dyadic right. So when you did that dog right we used to go up by let us say right 2 σ right so here the corresponding here because there is no notion of σ and all the notion is in terms of the spatial resolution. So you kind of go up by spacing of 12 then you go by say 24 that is to say replicate to some extent what you would like in a kind of you know dyadic sampling to happen okay and again right this will be like whatever is this enter thing right it will be roughly this and this value will be roughly for example roughly see twice this like I said it is not exact. See when you go up the next scale that you expect a 2 σ to enter right I mean it would not be exactly 2 but 12 times 2 is not 24 it is like 21 but that is see because it is all a discrete approximation as you would not get anything that is exact in terms of continuous case but you can roughly think of for example even here right if you go 27, 51 what is that 34 right no, no, no 27 to 50 what did I say that 39 right this will be 39 correct this mid value.

So 39 is not of course 21 2's are is not 39 it is 42 but 39 is still close so it is like so you go up by so this mid σ at both extremes are such that when you go up they go up by a factor of 2 and when you go along the octave they go back by a factor of 2 if you check the 2 extremes the midway extremes of the σ's okay this is roughly the kind of the scheme and within the scheme you have to now find out so right think of this is analogous to the way right you were doing you were doing you know a log approximation okay but now since you have that filter already right you have the dxx which I showed you already have what that what that filter should look like right dxx, dyy and then you had whatever dxy so you have these things at whatever size you want okay they are there and then right all that all that you need is wherever the wherever whatever size is needed you just have to apply it at that size and then similar to that similar to the earlier case right we have to now find out where the extrema are right I mean that is what we did there also but now the point is how do you know pick the extrema now right what should flag whether something is interesting or not and there what we will be doing I mean that wherever we are we used to check a 3 cross 3 around it and then we used to check a 3 cross 3 above it and the 3 cross 3 below it right around 27 26 neighbors right 26 neighbors so similar to that we have to do here also so it is like you know it is like check here and then check above check below. Now in order to do that right so here what they do is they so in surf right because anyway I have gone for a kind of a box approximation what is done is the following so you have like H okay I mean X, σ is what is written but of course you know typically it should be X, Y so this ideally right is this matrix LXX of X, σ actually it should be X, Y okay LXY of X, σ LXY of X, σ this is the Hessian LYY of X, σ okay this is the Hessian. So each of this is actually a second derivative this is the second derivative so second derivative of the image I which you can compute using this approximation that you have with you DX, X, DY, Y and DXY at every point right you need to actually compute you will have right if you do just one convolution you will have all those values it is not like you do it individually but whatever right if you are sitting at a particular location then you have the LXX value for that location you have the LXY you have the whatever right LYX which is we are assuming to be same as LXY and then LYY. And then what they do is to check the strength of that particular say response right I mean you need to know whether that point should be considered as important or not so what you do is the actual thing that is taken right to actually check the significance is actually the determinant of H and that is computed as DXX into DYY minus 0.9 DXY square okay now where this DXX is all an approximation of LXX I mean LXX is actually the I mean log the actual log which you would have ideally like to have there but you do not have that so you have an approximation of the log through a box filter which is DXX DYY okay right do not get unduly worried by this you see 0.

9 business because this 0.9 is coming out of like I said right so when they actually equate the you know this one right energy in the true sort of a Gaussian and they try to equate that energy with respect to that of a box filter they want to make sure that the energy remains the same so this 0.9 factor is coming from there okay like I said the derivation of that is you know is kind of beyond this course but why does this make sense you tell me let us see right you guys have an exam next week next week right so why does this make sense a determinant of H so if this is a high value right I take it to be right I take it to be I take it to be a strong response I mean I will get a value and of course you know I can do the same thing right which I would do like you know 3 cross 3 compare and all that but first of all why does this make sense forget about the 0.9 and all okay let us not worry too much about that the determinant of H see what he is saying is after all that determinant is simply the product of the eigenvalues okay and if you do a Taylor series expansion right about let us say right if you think that you are already sitting at the optimum and you are just trying to search around it right and you know that and you know that if you are actually sitting at the optimum then if let us say X transpose right if X transpose HX right which is right which will be actually you know which will be actually one of the one of the terms right in your or let us say H transpose whatever right I mean let us say if it is you know delta X okay if you are right if you are taking you know a delta X kind of a deviation about X and you actually believe that you know X is the optimum then the first derivative is anyway 0 right if it is an optimum now everything hangs on this H now. Now if it turns out that turns out that H is actually a PSD then what it will mean is right that this number has got to be greater than or equal to 0 for all right for all of this delta X right so it should be greater than or equal to 0 which means that means that all the

eigenvalues are actually positive right so all the eigenvalues are actually positive that you can show by taking delta X to be if you just take that to be the eigenvectors then you can show that H delta X simply lambda delta X and therefore delta X transpose delta X is greater than or equal to lambda greater than or equal to 0 right.

So it actually means that all your eigenvalues are actually greater than or equal to 0 which means that which means which means that which means that that actually you are actually sitting at at a at a this one you know local minimum right you are sitting at a local minimum because any deviation from there is actually increasing the function value correct whereas if on the other hand if you find out that that at some point at where you are sitting if delta X transpose H delta X turns out to be less than 0 let us say less than or equal to 0 right then we then we then we know that your eigenvalues are all are all actually less than or equal to 0 right. But then then if you take the product right the I mean I know the product I know will still be all but the product will still be a positive number right okay. So so it means that right either way you go right either way you go I mean you will actually end up end up with the fact that whether it is whether it is PSD or it is NSD right I mean you know a negative semi definite or a positive semi definite this number right will always will always be will always kind of indicate to you the strength. But if you had let us say any other if you are sitting anywhere else right then okay then right then of course you know then your H is not it is not either in which case in which case this this number right will not will not I know it will not kind of indicate indicate indicator get a local extremum. The idea is to be able to catch the extremum right a local extremum.

So the local extremum can be caught by using a determinant of H does it make sense okay. So then what you need to do is so in order to be able to search for it right so so in order to be able to search for it so in order to localize okay now the yeah that is let me just write this down okay in order to do non maxima suppression in a 3 cross okay in order to localize points. So you have the strength right at all the at all the points in order to localize interest points in the image and across scales right in the image that means at that octave and across scales I mean across scales as well as octave okay across scales actually typically means you know along the octave. A non maxima suppression and NMS is in a 3 cross 3 cross 3 neighborhood right that is same as what we have seen before neighborhood is applied is applied. The maximum of the is a determinant of the determinant of the I mean hessian is taken to be the taken to be this interest point or taken to be the key point.