Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-52

So, this integral image itself is actually you know a nice concept and you know how it works right. So, the idea is that, idea is that you see for a shift right what we had to do was we had to have different octaves right within an octave we had to have different scales, we start with σ then we have k σ, k square σ and so on right. And our idea was that then we can go to the right next octave and so on that was way right we were implementing it and then the reason why we said we wanted to go you know from one octave to another was to simply reduce the size of the image so that you know we can do things faster right. So, searching in one resolution the highest resolution for all the scales is going to be a time consuming affair. So, we said we will down sample blur and down sample and then on that right you build another octave with different scale then go next octave and do like that right. Now, what you can show using this integral image sort of notion is that you know the if you try to do a convolution sum right which is what you do right I mean that is a convolution is independent of the size of the size of the kernel which is like which is like you know you know kind of a big thing.

For example, if I have a convolution that I have to do 1 cross 1 compare it with actually a 3 cross 3 compare it with a 5 cross 5, 11 cross 11 it looks like I have to do a lot more work right that is one of the main things. So, if your σ goes up then your base the spatial support goes up then your see computations go up. So, the claim is that using this integral image it is totally invariant whatever size you give me a 99 cross 99 kernel I will take the same amount of time that I would take for a you know a 3 cross 3 right that is the claim okay. And once you can make that happen right then you no longer have to kind of look at different octaves and all that right you can simply run it through because you really do not care now I mean I can have as big a size of filter that I want because that was the reason right we said that when there is a blob and you want to actually match it up right and the block is big right then you have to have a big σ right in order to actually fit that block right and whatever scale it matches that is when it will flare up or it will fire.

But then the problem was that you had to then create a kernel of that size you know do a dog approximation to that and do it. So, all that meant that you had to spend so much time in doing these calculations whereas in an integral image it looks like you can have you know a 5 cross 5 and then you can go to 40 cross 40 and then not bother in both cases the rate of computational requirement is the same. But how it happens rate we will see next okay so the idea of an integral image so what is so the whole thing rate revolves around this okay which is why I wanted to spend some time telling you know what this is. So the evaluation using integral images okay so the key point is this the evaluation using integral images requires the same number of the same number of lookups in fact it is more like a lookup calculation okay regardless of the of these image scale. Thus removing the need for a pyramid thus doing away or removing the need for a Gaussian pyramid and we see in those days right that was this led to a 5 X factor speed up but I am sure right it is much more you know maybe if you look up now right there must be more reason implementation which can which claim probably even more okay.

So this integral image comes from this Viola Jones paper 2001 okay but these of course you know but these people did a lot more work after that to show how it can be done okay let us not talk about what is an integral image right. So right given an image i the it is integral image let us call me suppose I call this i σ right that is the sum image or integral is like sum right so I call it summation so i σ okay now in at X, Y is = double sum i = 0 to i less than or = X, j = 0 to j less than or = Y, i of i, j what does that mean? Suppose I suppose right let us just take one example okay these things are best explained with an example okay first of all yeah okay yeah I think you know we will first take an example suppose some random example okay which I have taken here this is like 4 X 4 okay and entries are 1, 3, 2, 4, 5, 2, 1, 6, 4, 6, 2, 5, 1, 4, 1, 1 okay this is my i okay and I want to write up my i σ now I will tell you what is the utility of this i σ so let me just draw the same grid okay what will be the first entry now if you go by that i of X, i, i σ of X, Y is okay so 1 next 4 next 6 next 10 right then when you come to 5 you can only go up right because you have to have i less than or = X, j less than or = Y so you have to stop at 6 when you go to 2 no 2 + 3 you see when you go to 2 right you can go up you can go left right you cannot go spill to the right of 2 because i less than or j less than or = Y yeah this is just a top square yeah so it will be 3 + 5 what is that 3 + 2 5 11 yeah do you follow and then when you go to go to go to this 1 you can go like 1 + 2 you just add 3 more right so it will be 14 then you add another 10 that will be 24 for the for 6 when you come to 4 it will be 5 + 4 9 + 1 10 then the next when for 6 you add the whole column above 6 including 6 right less than or = okay it does not exclude your X, Y so 6 + 2 8 + 3 21 right then 21 + 23 26 then 31 41 then 1 11 then 21 23 26 then 27 29 30 32 then 32 33 38 48 something like this right so 1 4 6 10 6 11 14 24 10 21 26 41 11 26 32 48 okay now for the time being right suppose we just leave this off okay now if i were to conduct a grid okay where i can sum up the areas right in this manner and suppose i am interested in some shaded area here right at every point right i mean you know so for example imagine that you are actually computing the area just the same way right whatever is above right all the way to the top and all the way to the left for example if this is my point D then the area that i have computed when i am at D is all of this right so all of i do not know whatever color i do not want to show but all of that okay whereas if i say that my kind of point is A then what will be the area that i have that i have covered that area will be something like this right it will be this area that i covered then if i say that i have got a point B here i cannot even see that let us say B right then B will be all of this B will be all of that right is that correct and then say we have another thing that is C which is here what will be C? C will be all of this correct now if i was interested in the area and ABCD right then what would i write it as so if i am interested in the area of ABCD that is the shaded region right i mean let us say whatever this shaded region that i have here okay this guy the shaded region i want the area so what would you do you will say D right what do you mean you can do it in whichever way you want maybe right maybe you can say D - B but then if you knock off - C you would have also knocked away then right so you should add A back otherwise you would have knocked off A so you can say + A right or you can say A + D - of B + C why is this important why am i saying this looks like some you know like you see elementary fourth class and stuff right but why is this why is this irrelevant and do you see where i am going even though this looks yeah that is like they write area under area right within that rectangle that is the area within that rectangle. No, i am asking what is that rectangle is? No, they are the rectangle itself yeah exactly so which is why i read if you actually come back to your problem okay which is like i have an image okay i have an image and i have actually a box filter okay when i do a convolution with actually a box filter is it not it like is not it is not it like see for example right now if i keep my box filter somewhere okay let us take you know 3 cross 3 if i keep a box filter what am i doing i am simply summing up summing up all the elements under it right because all my weights are 1 i do not have to weight anything it is simply whatever is under it add it up right if it is a continuous case it is the area in this case it is just a sum of all the numbers under it. Now if i have a larger kernel let us say right this is like 15 cross 15 again right if i doing a convolution with the underlying patch is equal it is simply adding up all the right elements under it. Now can you kind of relate this to this i σ see we know that in i right if i wanted to do a convolution with i i can take a kernel like this which is our old fashion way of doing things right i can take a kernel move it all over but that is the same whether you whatever kernel you have Gaussian also you would have done this but then should we be doing this when we have a box filter with us right you know it looks like it is a very inefficient way to actually do it because after all moving it all around is simply equivalent to adding up the right elements under it and i σ is actually doing that right.

So now suppose I ask the following question right see suppose I have you know a 3 cross 3 kernel right so we have a 3 cross 3 kernel okay and I center it at I center it here okay now okay wait let me remove what was that number by the way 2. Let me take this example that I have here 216, 625 okay 4 okay fine alright so it turns out that it is the same example and I am actually centered here okay. So the point is right if I had done a convolution if I had taken this 3 cross it is all 1s right I mean for the time being ignore that factor outside it can be whatever okay I do not care it is just a uniform number I multiply once I get the area I multiply it. So if I just look at it right that is an arrow okay that is not 21 so if we look at it it will be just to what is it 9 + 6, 15, 17, 22, 26, 28 right that is what I should get if I do a convolution for that pixel which is at the center right so I should ideally get see 28. So I want to show that you can actually do precisely that if I from this is I σ image okay now the sort of convention adopted convention is this okay in order to be able to move wherever you want so what is normally done is you append one more row and column and put them as 0s okay I mean in this case you do not need but then if you are at the border right you might actually need it okay.

Now I again include these values 1, 4, 6, 10, 6, 11, 14, 24, 10, 21, 26, 41, 11, 26, 32, 48 okay now see I mean if I actually if I were to locate this 2 here in this I σ image which is appended which is you know an extended I σ image where is this 2 now? 2 should correspond to whatever is the area that I got that I would have gotten it to know that will be like 10 + 3, 15, 21, 26 right so it means that right it is here right that right 2 is right there and then of course you know and then you have these 3 numbers that for example for this 2 it would have been 11 for this 6 it would have been 24 and so on okay. Now I am going to indicate to find this convolution rate I am going to call certain things as you know A ‘ B ‘ so we are going to say this is A ‘ which is across this 11 and above this 24 I am calling going to call this at B ‘ left of 21 I am going to call this C ‘ and this 41 sorry this oh sorry not I mean left of 10 left of 11 left of 26 sorry A is the problem of these things can you tell me what was this number 6 and then 21 something had 10 + 10 okay. So A ‘ here C ‘ here B ‘ on top D ‘ is this guy 48 okay and now I am going to do this A ‘ + D ‘ right which will be 1 + 48 - B ‘ + C ‘ which is 24 + 11 I have like 49 - hey you know what happened I think I made a mistake somewhere wait a minute 48 + 1 or 10 yeah you should have known it is not 24 right this is 10 no B ‘ + C ‘ so this should be 10 + 11 which is C ‘ so 49 - what is this 21 right so that is = 28 which is this value okay. Now it is easy for you to show that if you take an integral image and if you actually if you know call this as A ‘ B ‘ C ‘ D ‘ then the convolution for that pixel which you are examining at 2 can be obtained by simply doing this A where your C convention is A ‘ B ‘ see you can also do it by using the whole image and so on I mean there are multiple ways to do it but then one of the most common conventions is this right mark your A ‘ B ‘ C ‘ D ‘ like that. Now why do you think right this is nice I hope that you need not do the looping every time not just looping suppose I gave you 7 cross 7 you still have only these 4 numbers to compute right if I give you 11 cross 11 I am just taking a small example yeah and once just that one time once you have computed σ I can run any size of filter on it now right I can use my scale whatever I want I mean if I want a 51 cross 51 I can run it because again it because my I σ I just have to compute once right and then whatever I want as a sort of convolution whatever filter size I want I can simply arrive at it using a simple whatever it what is this is like what 2 additions and 1 subtraction right that is all I need and with that you can actually compute a convolution of whatever size you want right that is the strength of this you know integral image and of course you know assuming that you have a box filter of course you know okay.

So if that has sunk in and this you can extend to anything okay by the way it does not it is not restricted to a 3 cross 3 and all I mean as long as you actually you know you have you know a definition for example you can have a row filter you can have a column filter every time you follow the convention it will all work you can have a 2 cross 2 filter it does not have to be you know exactly a 3 cross 3 you know it took a 3 cross 3 because it is you know easy to understand but it does not mean that this will work only for 3 cross 3 and all okay you can take even even size filters even for that it will work but anyway right I mean you know for the time being right let us not actually worry about I mean if you want it you can just check out one of these you know let us say let us say that if you just wanted to just for example right suppose we take you know 2 cross 2 okay let us let us take this okay and suppose I say that right I am actually centered here okay I mean again we will see the only problem with this even is that you will have to you will have to see tell where this enter is if you take that right then a convolution rate will be what 6 + 4 10 + 1 11 right 11 is what you should get now where is this 2 right you have got to locate this 2 here so 2 is sitting at 11 right that is what it is no so 2 is sitting at 11 and then 14 will be this 1 and then you know 21 this right so what will be your a ‘ a ‘ remains as what was that 1 right a ‘ is 26 no no 1 1 a ‘ is 1 right and then b ‘ will be 10 c ‘ will be 6 and then d ‘ will be 14 no no d ‘ is 26 right d ‘ is 26 so now you do know a ‘ + c ‘ sorry a ‘ + d + d ‘ which is 27 - b ‘ + c ‘ which is 16 right that is = 11 what should it be 4 + 7 11 right so whatever it is right as long as you follow the convention correctly you can do right any size filter any even one dimensional anything will work okay now having read once we know this right that right this is the power of the integral limit now this by itself is not enough right for us okay. So what they do is you know they go ahead and here is where of course you know they have done they have done a lot of work right in trying to find out what would be an approximation what would be a box filter approximation to your to a log that you need next right now this is this involves you know a lot of I mean so this involves a lot of analysis we will not go into that right we will simply say that in fact they have to compare a linear sort of you know a Gaussian scale space with respect to a box filter scale space and find out it by the approximation happens the best and so on. The studies show that σ = so for example if you have a σ = 1.2 okay which is where they start in fact I mean you know that is the smallest σ they actually start okay they are there this one box scale space and for that right you get actually actually a 9 cross 9 filter okay so let me say 1, 2 I do not know 3, 4, 5 so 3, 6, 7 and then let me put 2 more here okay 1, 2, 3, 4, 5, 6 so okay well maybe right why do not I just show that instead of instead of right doing this I think I think oh I do not think I copied the slides today okay. So the way so the way right this looks like is as follows so in the middle so here and already it is 0s okay these 2 columns in between it is got like 4 1s here and then again these 2 are 0s and then this repeats till this and then you have like - 2, - 2, - 2 okay then - 2, - 2 then again - 2, - 2 I think there is only 3 it looks like I missed a row okay there should be one more row.

Then this is like anyway I mean I will show you the slide tomorrow where this is all available so and then these are like 1s okay and then again okay I think I am kind of missing some rows and columns okay it is okay as long as you understand what is going on but then right the kind of key is going to be next right what I am going to ask you and this whole whole thing is 0 okay all 0s on the left okay something like this is what is what it looks like I mean okay this is what they show to be to be a good approximation to a log at $\sigma = 1.2$ this is actually a 9 cross 9 filter this is what is like your dy y and then a transpose of this will be your dxx and so on and similarly right they have a 9 cross 9 kernel approximation for your dxy. Now see the earlier case was that we had all 1s it was easy right we saw that we could do it but then finally right I want an approximation to the log I do not have box filters right box filters cannot approximate a log so when I go into approximating a log this is the filter I get which is a 9 cross 9 filter. Now can I still use that use that use an integral image idea even though I have a filter like this how would you do that? No but it is space variant. Where? Yes well not space actually that is not strictly correct space variant you mean that the numbers are not uniform that is what you are saying it is not uniformly just one number all over that is what you mean right it is a it is a space invariant filter this is a filter get applied everywhere it is not a space variant filter it is just 9 cross 9 right the entries are changing.

The concept that we use that the I σ was we had some special orientation. No orientation all 1s. Since everything was 1s they were able to probably subtract. But I am saying that I can still kind of write no I can still do that here what should I have to do? A combination of both. I do not want to do that I want to I want to use my integral well no I want I do not want to prompt that answer go ahead do it.

Ok so what you are saying is if I have - 2 I can actually remove that - 2 out and I can operate it later is that what you are trying to say? See I want to do all this operation using just the integral image I do not want write anything else. Yeah just that ok that is the only thing I know ok it is like it is like somebody who is trained only to do that right I know to compute use I σ I know to do that is all I know ok go back and think about it we ran out of time I did not realize ok.