Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-49

Scale invariant, so you can see that there are there are a bunch of them in terms of the half, then we saw blob detection, then we saw corner detection, we saw edge detection. So, we are looking at something like scale invariant feature transform, again the goal is the same. Goal is again the same that we want to be able to match features across images and then you know based upon the draw some inferences, right, either about the geometry or about the about the mutual you know transformations between the images and so on. So, this is a very kind of popular among the among the ones right that the ones that exist and this is the most popular / the way and this is this is due to David Loewe and it came as a conference paper in 99 and then came in in a journalist in 2004. And of course, again draws a lot you know heavily from what we have seen till now into the log right what we saw in the last class right. So, draws heavily from what log can do, but then but then right this this person he proposes a bunch of things along the way some of which are intuitive and some of which may not be that intuitive also, but yeah the point is it works right and it is not like deep learning okay.

It works because I know there is a lot of you know kind of a systematic thinking that is gone into it. I mean sometimes right some things they assume which I think we just have to we just had accept right. I mean there is just that is just a compromise between computation and accuracy and so on. Okay now the first thing right that I wanted that we want to do is the difference of Gaussian which is dog right this I think at some point of time I had said that right this is something that will come up again a difference of Gaussians.

So, we want to sort of you know establish a relation between difference of Gaussians and the scale normalized log scale normalized log. The scale normal log of course right we know what it is and it turns out that you know actually you can you can of course the advantages of that dog is far more easier to implement just request to Gaussians with different σs lot more easy to implement than actually log then that turns out that there is actually a good. So there is an approximation there, but still then in practice it works very well okay and we will see what that is. So in order to do that let us first look at what is this the first thing which is let us say g of x we will take a Gaussian you know 1d again okay and then we will I mean whatever happens here happens there in 2d - of x square / 2 σ square. Okay now let us first look at ∂ g / ∂ x okay in this case okay let us say ∂ g / ∂ x which turns out to be 1 / √ 2 π σ into e raised to power – x square / 2 σ square because it is with respect to x we will have get – 2x then maybe 1 / 2 σ square.

So this we can write as – what is it 2, 2 will cancel off and then – 1 / √ 2 π σ cube x into e raised to power - x square / 2 σ square okay I do not think we need this bracket - 6 okay. Then let us look at second derivative so this is - 1 / √ 2 π σ cube x into e raised to - x square / 2 σ square into - 2x 1 / 2 σ square + e raised to - x square / 2 σ square and something that I had said last time right I think if you are if you are attentive enough right you will see that something that I had asked you to prove comes out of this / the way I would not tell what it is but I leave it to you to figure that out. So then you get - 1 / √ 2 π σ cube and let us pull out this e raised to power - x square / 2 σ square and then we will get 1 - what is this x square this 2 and 2 will cancel right x square / σ square. Now let us kind of look at this one right so till now we have not examined what is ∂ g / ∂ σ see till now whatever this one derivative we have taken right we have always taken with respect to x or y right. Suppose we examine this Gaussian with respect to σ okay then what will you get you will get 1 / √ 2 π and then you have like what is this 1 / σ into e raised to power - x square / 2 σ square.

Now what will it be - x square / 2 and then what is this - 2 / σ cube - 2 / σ cube and then what is this + we need this e raised to power - x square / 2 σ square into - 1 / σ square right. This is ∂ g / this guy right this equation 1 if you try to write find it is a derivative with respect to σ - x square / 2 - 2 / σ cube + e raised to - σ - 1 / σ okay. Now and if you do 1 / √ 2 π let us just simplify this okay e raised to power let us first pull this out just to clear the mess then what happens 2 and 2 will cancel we get x square / σ power 3 right x square / σ cube σ 4 σ power 4 in fact and then we get - 1

/ σ square or if I pull out – then I get – 1 / √ 2 π and then I can pull 1 σ square extra because this guy also has σ square and then e raised to power – x square / 2 σ square and then you will get 1 – x square / σ square which is confirm that everything is going on okay. So, you get – 1 / √ 2 π σ square 1 – x square / σ square yeah okay. Now it is interesting right if you compare let us say they call this expression number 2 and call this expression number 3 then you see that the second derivative right ∂ square g x / ∂ x square if you see that almost has a similar form right except that except that you need a unit a σ right.

So, you need a σ to multiply ∂ square g x / ∂ x square right. So, you can write ∂ square g x / ∂ x square is equal to this guy ∂ g / ∂ σ right where here it is a function of σ. So, this actually called the heat equation or something where this is nothing to do with image processing okay but anyway turns out that this is also useful for us okay. Now what you can do is now here is where a certain amount of approximation goes and till now it is okay right. Now what I do is and also this ∂ g σ / ∂ σ okay.

So, so what you have is σ ∂ square g / ∂ x square. In fact there from now on it I am going to simply write this as you know this one del square f which is a which is a Laplacian the same thing will will hold good okay ∂ square f that is also the reason why I said that you know see a Laplacian will be will then involve an x square + y square / c / σ square right and which will also come here if you do this with respect to σ. So, the 2D is just straight forward then this right. So, the way right this is written is you write this is your g of σ not ∂ g / ∂ σ. So, this you write as g of x , y , k σ – g of x , y , σ / k σ – σ limit what do you think k attending to 1 right.

This is what it will it will be actually or in other words right you can actually transfer the σ on to the left and you will get σ square δ square f is equal to limit k attending to 1 g of x , y , k σ – g of x , y , σ / k – 1. Now it turns out that see on the I mean on the left is your scale normalized log right this is what we saw last time this is what gives you a scale normalized log and we saw what is the use of that right. I mean it it allows you to catch the extreme and all very well right without without going to damπng the damπng the magnitude of the strength. On the right hand side right now this what what basically people have found is that this right they directly approximate it as g of. So, so what they do is so this they directly approximate as simply a dog which is a difference of Gaussian and k strictly speaking where k should be very close to 1 for this for this equality to be kind of meaningful.

But in reality what is found is the range if approximately equal to simply you know a difference of Gaussian which means that you can simply say g of x x y x , y , k σ - g of x , y , σ and it is approximately equal to a difference of Gaussian which is what is what dog is right. You just take one Gaussian subtract it from the other for for a reasonable range of values of k. Now this is where the this is where you know this is kind of empirical. You know strictly speaking that is what should hold, but in practice it turns out that k all the way. So, a tyπcal value of k that people use is value of k is actually is actually k is equal to √ 2, which is like 1 point what is that 4 or something right.

Now strictly speaking there should be 1, but but it holds very well. In fact, you can plot this and see yourself ok. I mean I would encourage you to to plot this for different values of k and compare with a log on the left. Turns out that you know it is in fact, people go all the way up to 1.6 and all and this is a range of values where it works very well ok.

And this was known ok, this was this was known this is not related to the paper really. This is not like you know lower founded or something whether this was known, but then what he did was he exploited this this you know in a very nice way. So, last time right when we saw log right we had sort of a scale space right where he said that you know we will convolve with with right different σs. And and then it depending upon whether the structure is coarse or fine right for smaller σs the finer structures right will will will fire. And then for coarser σs right the the whatever the the coarser structures right will fire for larger σ right.

Now what he did was the following right. So, because computationally rate I mean after all right all this had to be done in a very fast manner right. For example, open c will have a shift implementation that can be run in in you know in one hundredth of a second or something like milliseconds in fact. And all that is happening because of these approximations and all that have been made and to show that right it still remains reasonably robust and so on. And and the way right he does this is as follow I mean he does a bunch of things right I will go through one / one.

So, first thing right that he actually does is that that he builds up what is called what is called you know a dark scale, a dark scale space right instead of a log instead of the instead of in the log domain. So, then what would you do and of course he chooses k equal to √ 2. So, what that means is the first one image right that you have is actually blurred with σ ok. And then the next one because you want this approximation will be actually k σ. And then the next one will be actually k square σ.

Then you have k cube σ and then let us say k power 4 σ. I mean like this again how many how many measures should you this is called an octave I mean I will tell you what that octave mean, but for the time being let us keep that aside. But let us say right if you did not have a you know a pyramid or something and you simply wanted to search through all σs right at a sort of at a what you call you know at a at a particular this one spatial spatial resolution right a spatial resolution that is which is the highest kind of resolution that you have which is the image size right. It is all at the image size / the way right. So, when I say I apply σ that means that I take the image which is apply at that at that spatial resolution.

So, it is like the highest resolution right that is what you have an image is captured at whatever highest resolution you have. When you have a pyramid you will down sample it right, but here as of now there is no down sampling these are just you know the same image taken and then blurred successively. And what would you next what would you do next I mean the obvious thing to do would be to actually catch the extrema right because after all that is what you want to know I mean right. So, the course structures and the finer structures you want to be you want to catch them. So, then what would you do you would take these two because now is where they where the dog comes into play because you are no longer using a log right you want to approximate it using a difference of Gaussian.

So, what you will do is you will actually subtract these two right that is what that is what we said here right. And you will subtract these two and let us say let us call this is dog 1 and then you will take these two subtract them you will get dog 2. And then you can take these three these two you can get dog 3. In fact, in his paper it he goes up to dog 4 ok. / the time / the time that you are a dog 4 how much is your σ up / now? 4th σ right.

So, you are already here like here you are talking about if k is √ 2 then you are you are looking at already like 4 σ there. And here is where you got like k square. So, it is actually 2 σ here ok. And and each of these being an approximation to the log at that scale right each of these is actually approximation now now what can you do? Now what you can do is you know you can look at let us say suppose I look at dog 2 right and I want to catch extrema now. So, what I can do is just the same thing that what I said last time.

So, I can I can have I can look down and up right and I and I πck something here I take a 3 cross 3 neighborhood I take a 3 cross 3 neighborhood up the scale down the scale and in that scale right / scale I mean in that σ ok. That is why it is called scale space scale has to do with σ space has to do with the with the resolution of the image. So, up the scale down the scale and at that scale you compare with how many neighbors? 26 right 26 neighbors 9 above 9 below and 8 around you. And then if you find that right you are the you are the highest right in terms of strength again right. If you find that the magnitude of the or the or the strength of your of the of this dog response right.

If the strength of the response we would not call it log response, but it is actually an approximation of the log response. But whatever is the dog response that you get if that strength is the highest there then you would actually think of that as an extrema. So, it could be a minima it could be a maxima right. Again like I said could have a blob that is that is dark in the middle and white outside or could be white in the middle and dark outside both are blobs for us right. And here the interpretation is not strictly in terms of a blob or something just that it you know the insπration is drawn from log, but nowhere it is he it is not about you know it is not specifically meant.

So, for example, if you were to ask what is the key point then here right the interest point that you are trying to find out. The interest point is whatever fires up right under this scheme in a sense right. And yeah you can think of it as blob it is again not strictly blob I mean it does it does fire you know whenever whenever there is an there is a reasonable activity in this sense ok. But because we have seen log you can relate to that ok, but then it is not like it is a blob detector or something ok. And ok so so what you can do so when you can compare with up and up and below.

So, you have got right you know you got to compare with 26 neighbors ok. With this itself right you could have been happy, but actually in this paper right he also takes up doc 3 ok. Just to be sure that right if there is something right if you know if you you should not miss something. So, he also compares that doc 3. So, doc 3 where doc 4 and doc 2 and then and then he finds the extremum, but that is an implementation issue right how many how many you how many images you take here ok.

Now it does not end here ok this is of course, on one way to see see how you can actually, but of course, we still have not talked about I mean you know there is still a lot more ok to the sift than this I mean it is not simply that you know that that that becomes the extremum. You want to declare something as a key point right. So, the key point he actually declared that defines a certain orientation for it which is a dominant orientation and I mean actually it is a smart way of doing things right we will we will we will see them one / one, but it is not over yet I mean this is just the extrema right how to find an extrema right we are only yet you know we are only there as yet. The other thing that he does is this this the kind of speed up. For example, right I mean it is like you see if you had if you if you stay at the at the same spatial resolution which is the highest resolution at which somebody is given you the image and you do this right.

So, what would you do suppose I wanted to I wanted to examine I mean it could have structures right from small to large right. Now, if I wanted to go to the larger ones then I will have to increase my σ right because that is when that is when that is when it will actually match up with the with the underlying structure. So, but then that means that you are actually operating at a spatial resolution that is the highest and you still have to create you know many many σs for that and you have to scan through it is almost like scanning through different different σs to find out to kind of locate all the interesting structures in the image. And as I said right you do not just just show the way you still deπct it is of course, if you know there is a deπction which is even which is even more interesting. But it is like saying that you know I found something interesting there at the center, but then I still tell that at what scale I found it that is very important at what scale right ok did I find it because when you match it right I mean these things could could still be important and not only that you want to know whether that structure is a small one or is it a big one what kind of structure did we identify.

Now, if you had to speed this up right what we do what would you do I mean if I if this is time consuming then what would be the next thing that would occur to you based upon some things that we that we can as we discussed you know of course, not not not in the immediate past, but some somewhere sometime in one of the earlier lectures how would you speed up convolution well convolutions are going on here I mean that and all we are anyway doing right. So, all that is there ok you will do convolve in one shot that that anything else that is there anyway what else can we do. In fact, I said I dropped a hint in the beginning that is a pyramid right a pyramid. So, what he goes for is actually a pyramid a Gaussian pyramid.