Modern Computer Vision

Prof. A.N. Rajagopalan

Department of Electrical Engineering

IIT Madras

Lecture-34

And, this goes exactly the same way that you would do in a 1D case. So, all that you need is linearity and instead of time invariance, you need what is called shift or space invariance. That is what we, that is the word that is actually coined for an operation like this. So, you can, so a discrete case rate you can think of some is it ∂ m n, which is, which if you input to a, let us say, you know, to an imaging system. And, suppose you see an output which is let us say h of m , n and then I simply write this as, you know, 0 , 0 just to indicate that it is an impulse applied at, you know, m equal to 0, n equal to 0. Now, when I shift this impulse, right, let us say, I do m - n prime and then n - n prime.

So, this is a Kronecker ∂, this is not a Dirac, this is like a Kronecker ∂. So, it is like 1 at m equal to 0, n equal to 0, 0 elsewhere. So, if you, if you do this, right, then maybe, right, you can write this as m n and then you can say that this probably depends upon, depends upon where you apply. Of course, for a shift invariant system, it should not matter, but in general it might matter, right.

If you, it is not, it is not always true that if you shifted there. So, that is why there is a difference between impulse response and, and, you know, response to an impulse. It responds to an impulse can, can vary. So, when you say impulse response, that means, you are actually indicating a convolution kind of or an LTI system. And therefore, right, what we, what and then this, this follows that, you know, if I have any and why do you choose a ∂? That is because any f m n, right, which is any image, you can simply write it in terms of, you know, this, this one in terms of the impulses as f of, you know, m prime , n prime and then ∂ m - n prime, right, n - n prime, right.

I mean, that is how, so you can express any, any, I mean, that is how you do any, any signals, right. If I gave you, how do you express, you know, a continuous signal, right. I mean, you will have an integral and you will have impulse, right, I mean, shifted and scaled, scaled in terms of, of course, the scaling is now coming in terms of the intensity f of m prime , n prime. So, you see that, you know, if you, if you substitute m prime is equal to m and then n prime is equal to m, right, that is the only time when this guy is non-zero and that time you will get f of m , n. So, this is fine.

But then what, what will happen is that when this kind of, when such a thing is being input and if you assume linearity, right, if you say that my, my kind of system is linear, then, then it is like saying that, you know, I have one ∂ getting applied at m prime , n prime and, and, you know, that is getting scaled and then I have, I have another, I have another. So, what do we do in terms of, in terms of a superposition kind of the principle? We say that, you know, if you have, say, f 1 giving g 1, f 2 giving g 2, then f 1 + f 2 give, you should give g 1 + g 2 and alpha m 1 f 1 should give you alpha g 1, right, where alpha can be real or complex or whatever. So, same thing, right. If you just apply that, then, then, you know, this will simply turn out to be f of m prime , n prime and then this, this ∂, right, will get replaced by h of m , m , n and then you have m prime , n prime. Now, on top of this, if you now assume, assume space invariance or shift invariance, right, which is, which is, is LSI, right, so, this is called LSI.

So, if you assume, assume shift invariance, right, then, then what happens is, then this guy, right, you should know, then what it means is, if ∂ m n went, went in and this came out, then the ∂ m - prime m - m prime n - m prime should simply lead you to f of m prime , n prime and then h, h of m - m prime, you know, n - n prime and then simply 0 , 0, because that should be exactly the same as what you had when you had ∂ m n. And therefore, right, this 0 , 0 is no more information really, right, it just, it makes no sense to, you know, carry it along. So, you simply drop it and you, and you kind of, and you are back to this expression, right. So, so, in a sense, what you are sort of saying is that when I kind of perform a filtering operation, the, the filter remains the same wherever I go, and that is one of the things that you are saying and then the other thing is this kind of a linearity. That means, if you have, right, if you add, you know, if you give multiple inputs, you know, at the same time, that is simply, you know, the, the output is just the sum and the outputs of what you would, what you would have gotten for, you know, for each of those, you know, individual inputs.

So, this all follows exactly the same way, right, that, that you would do with respect to 1D. This is all you would have done in your 1D case, just that. Which notation? Oh, semicolon, that just a semicolon, that is a semicolon, oh, ok. This is just a semicolon to say that, right, there is still a functional dependence on m prime , n prime and when you assume shift invariance, you know, whether you have m , n or m - m prime, you still get 0 , 0 and therefore, there is no point taking it along because, I mean, it does not, it does not depend on that 0 , 0 anymore. If it was dependent on m prime , n prime, then you would have actually taken it forward and no longer it has, it has no significance.

So, you just throw it off and that is why you get this H of m - m prime , n prime. There is no point carrying it. I mean, we carried it initially because it could have been a function of m prime , n prime. So, it is like saying that, you know, if I applied an impulse today, right, and I see something and then, and then if I come back and apply an, apply an impulse, right, later, tomorrow, I mean, ideally I should get the same output that I got, that I got the previous day, but then shifted. But if that does not happen, then it means that it is a, it is a function of time, right, when you are applying.

It is like your m prime , n prime, but if it is not, then why carry that? Then, now, now one interesting feature, right, that, that you will find, right, in terms, and of course, you know, and this convolution is again, again the same thing, right, that I said. So, if you had to flip, right, you would actually flip. Of course, now that we are, we are into traditional filtering, right. So, so, so, right, we should, we should, we should know what we do. For example, if I had something like f, right, and you know, if my, let us say, origin was here, and now, if, if let us say, right, this happens to be my filter or the image, and if I had to flip it around, right, then I could either, right, like I said the other day, you could either, you know, flip it, you know, about the y axis initially.

So, in which case, right, it will, it will come like this, and then, right, followed by, followed by, by, by flipping with respect to the, with respect to the, right, x axis value, right, something like that is what will happen. I mean, you will first flip it this way, and then, you will kind of flip it this way, or you can flip it first itself like this, in which case, right, this guy would have become like this. If you had flipped it, you would have gotten like this, and then, and then, right, if you had flipped it this way, right, then, you would again be back to this, right. So, whichever way you flip, it should not matter, but, right, that is what you need to do. Just as in 1D, you flip about the origin, here you have to flip about the two axis in order to get that m - m prime n - n prime effect, and then, you translate.

Then, right, all of that is exactly identical, right, slide this window all over the image, do a, do a kind of weighted average. Now, one of the things, right, that is actually, you know, interesting is what is called separability. One of the notions, you know, that we have when you have a 2D filter is one of separability. So, it is like saying that, you know, if I had a filter h of x y, well, I mean, I am going to be a little abusive here. I mean, sometimes I will write in terms of, you know, discrete, sometimes continuous.

So, so, please, please bear with me because sometimes examples are easier to give in continuous when appropriate. So, what it means is, right, if I can split this up into h 1 of x into, say, h 2 of y, right, that is, that is, that is when I say, when I say that I have a separable filter, right, and this is the separability, especially, right, when you go to higher domains, you know, because it has, it has various other implications, which, which I do not have time to discuss about here. But in the spatial domain itself, right, there are, there are certain implications in terms of, in terms of, you know, the time taken to kind of, you know, perform the operation and so on, if it is convolution especially in this case. So, can you get, tell me something, you know, can you, can you, can you give, can you give me an example where, where you have a continuous function, which is separable? x y x y, you know, I mean, give me something more interesting, I mean, you are just saying x into y, you know, this guy is like, you know, give me something more interesting, no? e power e power e power, well, no, but, yeah, e power, but then something else that you have seen, yeah, e power something that you have seen. Yeah, ok, yeah, so, so, what I am saying, there are so many of them, which you have already seen.

For example, a Gaussian is what I thought, I mean, right, that would be the first thing I thought you would say. What is, what is, what is a Gaussian? So, so, if I, if I write h of x y, I mean, if, if you take the same sigma, right, then it is 1 by 2 pi sigma square e raise to power - x square + y square by 2 by 2 sigma square. This is what you are indicating, but, I mean, better to say that, you know, it is a kind of a Gaussian. I mean, why I am saying is that these are the kind of things that are more common, I mean, right. For example, e raise to x square + y square also separable, but then that is not something that you often use, right.

I mean, a Gaussian you have all seen, right. I mean, Gaussian is such an important function. Of course, you know, see, if it turns out that, that, you know, your sigma's are not the same, then you will get like 1 by root 2 pi sigma 1 e raise to - x square by 2 sigma 1 square into, what, 1 by root 2 pi sigma 2 e raise to power - y square by 2 sigma 2 square, right. But if it is, if it is the same sigma, then, right, this is what you get and, and of course, it does automatically splits as you can see. And then, what are their functions? Gaussian then e raise to power, I mean, if you, if you have this case, right, 2 pi by n, right.

If you have, if you have a DFT, right, what will you have? 2 pi by n m k +, + let us say, right, n l. I mean, if you had, if you had a 1 DFT, it will be 2 pi by n k n, right. This is what you have. If you wanted a separable 2 D, right, that is what it will be, e raise to power j 2 pi by n. Similarly, a continuous case, right, you can.

So, so, right, these are actually things that you have already seen and there is something else, which is special about a Gaussian, which is not so special about the other thing that we have written. What is that? Let us just, let us just, let us just see, right, how much, how much attention you guys must have paid, right, during the filtering or signal processing course. A Gaussian is something even more special, of course separable and all that is ok, that the other guy is also there. What else is there very special about a Gaussian? Its Fourier, Fourier transform is also Gaussian, right. What is the, what is the, what is the Fourier transform of e j omega naught t? Some shifted version.

Some shifted version? Dirac. Yeah, it should be a ∂, Dirac ∂, right. I thought you said shifted version of itself. No, no, no, no, no, no, no, ∂. Oh, shifted version of ∂, ok.

I heard it as shifted version of itself, that is what I was wondering, where does that come from? So, but then, so, so, in that sense, right, I mean, you do not, you do not, the functional form suddenly changes, right, and normally that is the case, but the special signals like, like a Gaussian, right, where when you, when you take the Fourier transform, it still kind of retains its shape. Anything else that, that retains its shape? We talked about impulses, no, can you, can think of something with respect to an impulse that retains its shape? Retains its shape. In the sense that you take the Fourier domain, retains the shape. Gaussian retains. What is the Fourier transform of this Gaussian? Gaussian is Gaussian.

No, no, I know, but what is the form? Erase to power - omega square + nu square by 2 into sigma square. Sigma square goes on top, right, that is, that is a 2D Gaussian Fourier transform, Fourier transform of this guy with this normalization factor thrown in. What is the importance of this 1 by 2 by sigma square? Normalization. Yeah, exactly, if I take the area under, under this curve, it is 1, right, and therefore, g of 0 should be 1 here, right, because g of 0 is like the area under the curve, right.

So, that should be 1. And so, so what was, so Aniruddh, give me one signal, other than a Gaussian, which retains its shape. Impulse strain. Exactly, right, impulse strain, an impulse strain is again an impulse strain, right, the Fourier domain. And then, then now already you can talk about combinations of Gaussian with impulse and all that, right. You can get so many signals like that whose Fourier transform on this.

Now, the point, right, to actually notice that when you have, when you have a separability like that, right, what it means is, you know, it actually makes life, life a little simpler. See, for example, I mean, if I, if I had an image, right, and suppose I had, I had a, you know, let us say, you know, m cross m filter, I have an m cross m filter and, and this image is actually n cross n, right. And now, when I want to do a convolution, I will have to, of course, take this filter, apply it, slide it everywhere, right. So, if I, if I just remain at, even, even if I stay at, stay at one place, right, there is at least m square, m square multiplications that I have to do. Then I have to do maybe m square - 1 additions and so on.

This is, let us look at the multiplications, right. Now, at one pixel I have to do, I have to do m square multiplications and there are n square locations like that, right. So, I am looking at even, even in the, in the simplest case, right, if I just look at the multiplications, looks like I need to do m square into n square multiplication. Is that correct, right, m square into n square multiplications. Now, if you had a separable filter, right, then what happens is, equivalently, right, what you can do is, you can take this by, by a separable filter, right, what you really mean is something like this, a one-dimensional filter multiplying another, another, you see, one D filter.

That is, that is, that is the, that is the, right, you know, you know, this is an equivalence of this. So, for example, I mean, let me give you an example, right. If I had, let me, let me ask you, right, suppose I had 1, 2, 1, 0, 0, 0, - 1, - 2, - 1, this is, this is, this is a separable filter. How would you separate it? 1, 2, 1, 1, 0, - 1.

1, 2, 1, 1, 0, - 1, right. So, you see that, so you see that, you know, this is a separable filter and yeah, 2, 0, - 2, yeah, right. So, 1, 2, 1, 1, 0, - 1, right. Now, the, the, the, the good thing about this is that you can get the, get the equivalent effect, whatever you got through the 2 D filter, which you did here, you will get exactly the same effect. If you were to first do a convolution let us say, I mean, right, I mean, if you, if you try to do, do a vertical filter, 1 D filter, if you, if you, if you applied on this image vertical, vertical, I mean, the, so the first filter, if you took it, if you were to apply, if you had to apply the convolution, finish it and then you apply, apply a horizontal filter, right, you will get the same effect. Whatever you got with the, with this filter that you applied like overall, but then in terms of computations, now if you take this filter, right, this has how many entries, what is its size, whatever, yeah, m cross 1, right.

So, so, the size is m cross 1 and therefore, in terms of the multiplication, how many multiplications do you need now? m into n square. m into n square and so, it is like 2 m n square, right. So, that is all you need, where there you needed n square by m square n square. So, you are looking like m square n square by, see, 2 m n square. So, the gain is, what is this, m by 2, right and if m is, m is, you know, reasonably large, so you can already see that, you know, that, that makes a lot more sense.

So, so, separable filters, right, people could have look out for and even in the, oh, you will have more, more, more additions. Why do you have more additions? I mean, there also you had like m square - 1 additions. Here you will have like 2 m - 1 additions, right. It would not be more. So, so, so, but the separability, of course, you know, this is like one way to, one way to look at separability, but it has lot of implications in terms of, if you do Fourier analysis at all, which we are not kind of, which we are not going to see going into, ok.

But just, right, I just wanted to say that even this convolution operation we are doing, you should, you know, remember that if you had a separable filter, then, then, right, things can be easier. Then, among the filters, right, that are most commonly used, when I will, I will just, I will just talk about two of them. One is called a box filter. These are all handcrafted, right, I mean, unlike the ones that you saw with respect to CNN and so on.

These are all handcrafted guys. So, in a box filter, it is like, or we call it a, call it a uniform filter. So, so, so, it looks like, I mean, if you had, you know, 3 cross 3, then you will have like 1 by 9, then, right, all ones, 3 cross 3. I think I had to draw it properly. 1 by 9, 1, all ones inside, so that it sums to 1. And there is actually a reason, right, why, why, I mean, there is a physical reason for this, why this average and all the, so, it should sum to 1.

In the simplest way to kind of think about it is, you do not want, so, an average when you apply, right, nowhere, it will, what to say, I mean, I mean, this has to do with energy conservation, conservation, so on. I mean, there is a, I mean, there is something to do with optics and all that. When you have a lens, a lens cannot, you know, cannot, cannot, cannot inject energy and so on. So, the box filter, we will just use it as an, as an average, right.

We just want to smooth things. If you want, whenever you want to smooth, right, we will use an average. And similarly, if you wanted, you know, a difference operation, right, then, then, then, right, I mean, you know, you can kind of, you know, think of filters of whatever kind, the simplest, right, could be, could be, could be whatever, right, 1, 1, 1 and then, you see, right, - 1, - 1, - 1 or the 1 and then 0, 0, 0, right, in the middle. So, so, you can think of and of course, you know, then depending upon the orientation, right, I mean, then you may, so, for example, right, if you had an image, where, let us say, well, say, you have 255 intensities here and then you have got like 0 here, that is a very sharp edge, right, that you have here, here is where the transition happens. Then, if you have to pick up, pick up, you know, a vertical edge like that, right, then, see, the other filter that you could have is for a, is for horizontal edge, right, which you could maybe write like this. So, if you wanted, wanted a vertical edge, right, you would probably, right, go for that filter.

If you wanted a horizontal edge, something like this, right, you would have to go for a, go for a filter like that, let us say, this is 255 and this is 0. And the other thing, right, that you will typically find is that, sometimes, right, sometimes, this will have like, you know, 1, the other one that I showed, right, this is actually a standard filter 1, 2, 1, this is called Sobel 0, 0, 0, 1, - 2, sorry, - 1, - 2, - 1. So, you see that, right, there is an, there is an extra 2 that is, that comes in there, that is to actually wait, I mean, you know, so, for example, right, that is to, that is to give more waiting to the central pixel. We can see that, when we actually do it, I mean, I will, I will show that. But the other thing that you actually notice is that, you know, why do you, why do you think they do this 1, 2, 1 and - 1, - 2, why do not they, why do not they, they just do this, 2, 0, - 2, what is the, or 1, 0, -, if you want some weightage, let us say, right, why do they, why do they do that, why do they have like, why, why, why cannot they be happy with just a, just a, just a, of course, you know, in this case, right, we showed that, right, doing this is, is equivalent to, you know, multiplying with those two filters.

But I am saying, generally, right, when you have like 1, 2, 1 and then - 1, -, what do you think would be the reason? So, so, what exactly are you doing to counter noise? You are averaging, right, you are averaging in the, in a sort of a direction along the edge, right, you see this, no. So, so, so, right, what you are doing is, you are sort of adding up intensities, right, along the, along the, so, if you are looking at a vertical edge, it is ok, it is ok to average along the edge, I mean, you should never, never average orthogonal to the edge, I mean, then you will, then you will end up smearing the edge. But it is, but it is actually perfectly ok to go, go along the orientation of the edge and do an averaging because that will counter noise, right. I mean, if you had noise, then we all know that, you know, averaging actually reduces, provided the noise is independent and so on, it reduces the standard deviation of the noise, right, depending upon how many you average. So, that is the reason why you have this, you know, typically you will have like, you know, some averaging going along in the, in this direction along the edge and then a differencing that goes on in the, in the, you know, in the horizontal direction, which is, which is across the edge, right, orthogonal to the edge.

So, it is the same thing, right, we could also have for a horizontal filter. Now, so, so, in a sense, right, so, difference is like, you know, a gradient filter. So, in a sense, right, we will call these as gradient filters and this we will call as an, you say, average, and I will just show you a few examples just to, just to, right, pick your mind and, you know, write what, what those things may, may look like. If you just, so, it is like, this is like a small little, this and all, we will skip, this is how you know what to do.

Now, and now, think about this, right. So, I have, I have a bunch of, bunch of filters, right, which I am, which I am going to apply on this image on the left, right, and then you have to tell me what is the kind of output I will get. So, if I, if I actually convolve with, with a 2D filter like that, what will I get as output? Identical, right, I will get the same. So, it is unchanged. What about this? So, what do you see here, I mean, what kind of, what kind of edges have emerged? Edges have come, but what are those edges more like? They are more prominent vertical or more prominent, right.

Here, horizontals are more prominent. Now, we can see the eyebrow, this one, right, what is that? Forehead, lines, streaks, mustache. What about this? Sharpen. Sharpen. No, the right guy is a blurrer, but then now, you are subtracting the, the, the, the original from the blur, right.

So, it should sharpen. Oh, was the answer already there? No. Good. So, now, the point is, right, so, I think, I think there are few more examples. Now, the one of the, one of the other filters, right, one minute, I think there are few more examples that could be interesting or else we will first, we will first finish this and then we will come to that.