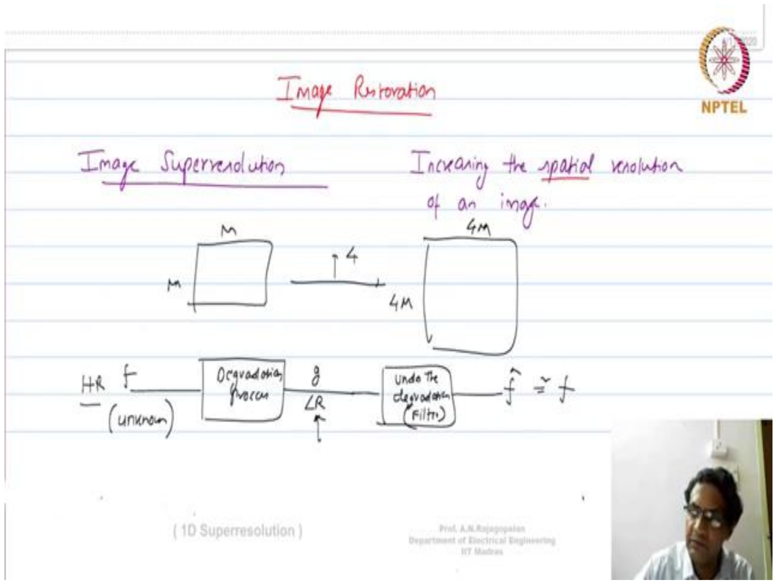
Image Signal Processing
Professor A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture 82
1D Super resolution

(Refer Slide Time: 0:21)



Let us look at one more example and their image restoration. In case our main topic is still image restoration, which as I said is an inverse problem and we will look at what is called image super resolution. Earlier, what we saw was a D Blurring problem which was again ill-posed and as you could see, a blurring problem was so very straightforward; even if you had to do a space variant blur you could do it so very easily, right in the early part of the course.

But then D blurring as you saw later, it was an inverse problem and then which was far more involved and there were so many conditions that had to kind of go through in order to be able to get a clean D blurred image. Similarly, super-resolution is again another sort of an inverse problem that comes under the umbrella of restoration. And super-resolution refers to actually increasing the spatial resolution of an image. What this means is you would like to increase the spatial resolution of an image.

Now, I mentioned spatial because there is also something called temporal super resolution and we are not interested in that, increasing the spatial resolution. So, this deals with the task of improving the spatial resolution of an image. There is another task what is called temporal super resolution, which means that if you got like 30 frames per second video can you increase it to 60 frames per second. Hey, that is like a temporal factor of 2. You would like to increase it by it a temporal factor of 2.

Here we are not interested in that problem. That is a different problem. We are interested in spatial resolution. As I said under this restoration umbrella you can have so many problems, right from deraining, dehazing, high dynamic range imaging, deblurring, super resolution, denoising, so many of them. And the idea is to just look at a couple of problems. Deblurring, we have already seen. I thought I will also tell you; give you a window into the image super resolution problem.

And then once you understand there are two such problems then you would know or so what it entails to solve any of these problems, like a restoration. Now, increasing the spatial resolution, so what it actually means is that, I have let us say I have a lousy camera, assume that I have a cheap camera. I do not want to pay much like this take up cell phone and then it is a kind of a low-resolution camera or low-resolution image and I would like to know, bring it up to a higher resolution.

So, let us say it was M cross M, then maybe I want it to be 4 M cross 4 M, which means I want to go up by a factor of four. You are going up by a factor of 4. Something like that. It could also be 2, it could be you know more of a 3, 8, whatever it is. So, you would say why to go up by a certain factor and how do you do that? Is this issue about super resolution? There is not anything super about super resolution. What it involves is just as, there is let us say a high-resolution image, which you are not able to see.

This goes through a degradation process, which is your camera and the camera yields a G which is actually a low-resolution image. Now you want to put up, so basically this is degradation process, which we will have to model, of course. Like I said the whole idea is to undo a degradation. So, first of all we should know what is a degradation process? Then you have to undo the degradation here through some kind of restoration filter.

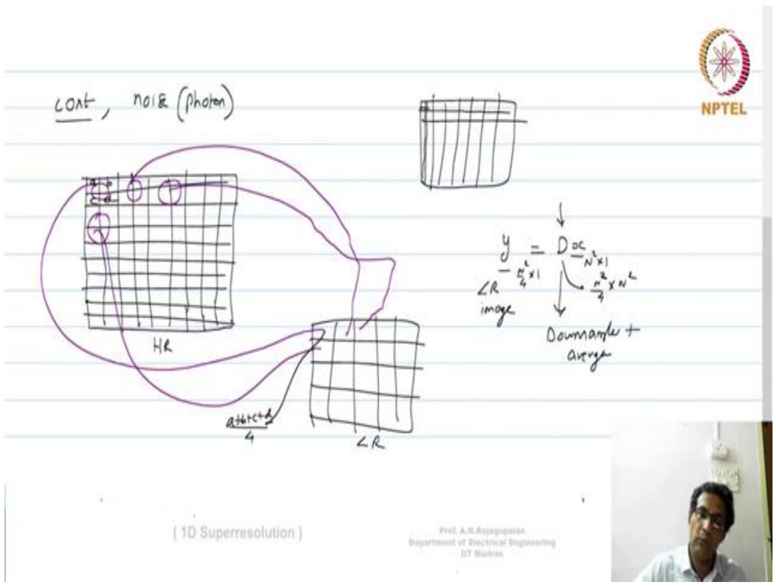
Degradation and then in order to be able to get some f hat, which we think is approximately close to f, where f is your input high-resolution image. Now, this is unknown. We do not get to see this, so if you have a poor camera you only get to see this LR image which is of a much lower resolution and you are saying that have some kind of filter here, which will be a restoration filter, whatever it is. But some algorithm here which will take this LR image and then blow it up by a factor of whatever 2 or 3 or whatever in order to give you back a high-resolution image.

So, now by what factor can you go up and so on, it depends upon several things. We will see each one of them one by one. Now, the need for spatial resolution, of course, obvious, I mean, every one of us would like to, even if you give me a high-resolution image you have a fantastic camera, I might want to ask can I get a blow it up further by a factor of 2 or 4, simply because I simply like to see even more fine details.

And this is something or you could have a very low solution cheap camera and you could ask can I can I get a, build a high solution picture out of that so that it matches the quality of my friends camera.

Because he has paid more, I paid less. Now man I cannot afford something very costly, I take a low-resolution camera, can I use some signal processing algorithms or image processing algorithms in order to be able to build a high-resolution picture from this low-resolution image. And the one that we are talking about is of course, are we talking about spatial resolution all through. Now the one of the need arises because of cost.

(Refer Slide Time: 5:28)



I mean, cost is one thing, but then cost is not the only thing why you would want to create or do super resolution because one thing is you want to buy a cheap camera and do super resolution, other is noise. especially photon noise, because them as we know, even if let us say a manufacturing, even if the manufacturing industry can really bring down the spatial resolution right off your sensor, let us say right here is where we are today and suppose, let us say, suppose they say we can manufacture something that it will be finer than that and then later they say, oh we even go finer than that.

Now, your manufacturing process may allow it, but somewhere right the photon noise will start to limit your ability to kind of go down, not because you have a manufacturing problem, but then because of the fundamental physics of the problem, because we know that each of these is like is like a bucket. Maybe, each sensor is like a bucket; which try to take in photons and more the area of the bucket, then the more likely that we will get enough number of photons.

And if we try to reduce the sensor size too much then what will happen is availability to kind of capture enough number of photons in a given period of time, because the exposure time has got to be small. We cannot make the exposure time arbitrary large for reasons that we have already studied. Earlier, the illumination could change if you keep along the exposure time or maybe there could be motion blur and so on. This (())(6:49) could may not be static.

And therefore you cannot expose for too long in time, and then if you have short exposure time, but then your sensor size is very small, then your photo noise rate is going to be very high, because the uncertainty in terms of the number of photons that it can collect is going to be higher, and therefore noise could then put a fundamental limitation to what you can do you. Now, this is how it all started by the way. So, the people that started super resolution started with these things in mind.

But now down the line I will tell you what is going on in super resolution as of today. People are looking at for more difficult problems, especially within the deep learning framework, but I think this introduction to super resolution that we are going to do which is kind of simplistic, but then it will at least throw some light into what really awaits, if you have to do a full-grown super resolution problem. And the other thing that I have also wanted to say was interesting.

I mean, one of the big guys in super resolution, initially he was a person that actually said that why, let us say why kind of one cannot do of super resolution and after a couple of years they wrote a paper on how to do super resolution. And so, such funny things in research do happen, kind of researchers can sometimes guess things wrongly and then they come back in and they sort of realize that, okay that is not the case, well actually I can do it.

Now, so the way you get to think about the super resolution problem is something like this. It assumed that ideally if you had a high-resolution kind of a camera, that means you have got like, you are able to sample the scene very fine. So, it means that you are able to sample the scene in a very fine manner. So, your spatial sampling is very good, it is at a very high rate that means your sensor size is really very-very small and you are able to sample.

Now this is ideal rate, think of this is what you would get from a high-resolution camera. Now, when you capture an image, the same scene when you capture it with a low-resolution camera, because of the fact that your sensor size is large then a low-resolution camera, what really happens this, four of these pixels, so these will map to one pixel here, that is what really happens. And then similarly the next four would to the next pixel and then similarly the next four will go to the third pixel.

This will go here and so on. So, you have, suppose you are going down by a factor of two. So, suppose the spatial resolution between these two images differ by a factor of two that will mean that four pixels here get averaged to give you one because see, sensor after all is like an area. So, if you have, let us say the individual intensity is here, now what is the super resolution problem, if the individual intensity is here where ABCD, what you see here is simply A plus B plus C plus D by four.

Assuming that they are being weighted equally after all the sensor area, we can assume that it leads to an equal weighting of all the same points and therefore what are you seeing simply an average value. And you say given this average value, give me ABCD. So you can already see that it is a very-very ill posed problem. So this maps to something like, so if you want to write down a mathematical model for this, a simplest model that you can write as Y, which will be our low resolution image.

So, LR means low resolution image, HR means high resolution image or SR means high resolution image, now what we, or the super resolve image, SR means super resolve and you can simply write this as D times X plus, maybe if you have some noise and so on but let us say if you have no noise, it is like D exists, X where this D will be would be a down sampler come average. So, this has to be a down sampler plus averagor. So, what it really means is, this matrix should have entries such that it would average.

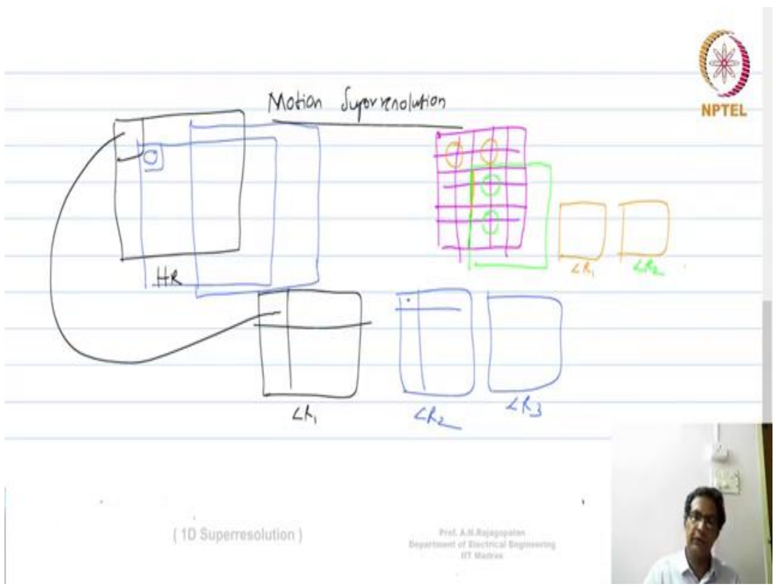
Four of these values and then kind of map them to one. So, it is going to lose, go to normally down sample, which is the reason why between 1D and 2D there is a difference. In 2D it is a sensor, it is a physical thing, it is a physical area therefore it will average intensities. Therefore, when you talk about low resolution you are also talking about a sensor size that is bigger and therefore there is an averaging effect. So that averaging effect needs to be captured by this D. So, it is not only a down sampler, it is also an averagor. That has to do both.

And as you can clearly see, if you are going for something like let us say a factor of 2, then maybe, if this guy is N square by 1 and let us I write Y will then be N square by 4 cross 1, because it will have like N by 2 by N by 2. There will be the image size Y, if X is N by N, Y will be N by 2 by N by 2 therefore if you stack it up as a vector, become N square by 4 and therefore d will then be N square by 4 by N square. That will be the matrix that you will have.

Now, in such a situation, one option, one of the simplest things that you can, so you can imagine that there can be many XS that can lead to the same LR. So, this is your LR, this is your original image, HR which you would have liked to see. Now as you can see right ABCD, why I can even put B here, A here, d here, and C here, and then I will still see the same average. Therefore, you can see that there are multiple solutions for this problem. There are multiple higher resolution images that can map and give you the same LR because so in this case, clearly a solution exists, but it may not be unique.

This is not unique in fact. And therefore one of the things that you could do is what we did earlier, something like a (())(12:27) solution, but then people have realized that (())(12:30) it is not the one that really gives you what you would like to see. You need to do kind of more work. Now the kind of progress rate that has happened this in terms of not just doing super resolution one image, but then solution if you had multiple images. Now what is it okay?

(Refer Slide Time: 12:53)



That is why what we are going to talk about is what is called motion super resolution. There is also something called a single image super resolution and so on. And especially, and so for example in such a case (())(13:05) solution could be one but that not be so great. Therefore, one can do some kind of an optimization with some L1 norm of the gradient of the image and then have this cost just as we did for he may deblurring and solve for the original image.

Now, that can give you something better than what simply a (())(13:22) solution can give or one could go in for something like a deep network, where you can train and then train for super resolving a single image. But what we are going to be focusing on is motion super resolution because given the background that you have, this is the one that you will understand the best, because it is really not a kind of a deblurring course and therefore I do not want bring in deblurring into to this course.

So, motion super resolution. So, what is the idea behind motion? Now clearly if I take an image and if I try to interpolate it, if I take an LR, now let me first draw the HR image. So, I have the HR, now let me just draw it like. Now let me first wipe this off. There is the reason. So, let us say that I have a high-resolution image and corresponding to that I have one low resolution picture. This is LR 1. This is my original high resolution. This is what I would have liked to see, I do not this, I see this.

Now, if you would allow me to move my camera, suppose I need to move my camera that is why it is called motion super resolution. So, if I can take my camera and move around, at the simplest thing which you can do is an in-plane translation. So, if you were to do something like an in-plane, you could also rotate, you could also have an affine motion, you could have homographic, I mean, complete 6D motion, whatever it is that you ideally can, I mean, theoretically you can have any of that.

But just to understand the basic of super resolution, the best thing to do is to take the simplest of cases. Suppose you move in whatever manner, move just fine amount, move by a fine amount, you are not going to move a lot, you are going to move by a fine amount, then what might happen is, you might actually look at this part of the scene and for this right again, if you try to capture an image corresponding to this, you will have another image here which will be LR 2.

And as you can see the average of the pixel that is going to happen here in order to get you a pixel intensity here need not be the same as what might happen here. For example, with this guy, this might give you some average here or average of some pixels in order to give you the first pixel here. And therefore if you can capture multiple such observation, then you take one more, perhaps, that you went there and then you took one more high-resolution picture, one more image corresponding to which you will have one more LR3.

All that you capture LR1, LR2, LR3, because you cannot, you do not see any of these high-resolution pictures, just moving the camera, your are kind of sensing different averages of the scene, you are sensing all of the matter kind of a low resolution because you do not have a high resolution camera, but the idea is that if with one image, low resolution picture, you can only sample the scene at certain points, but now that you are moving right the hope is that you are sampling multiple points.

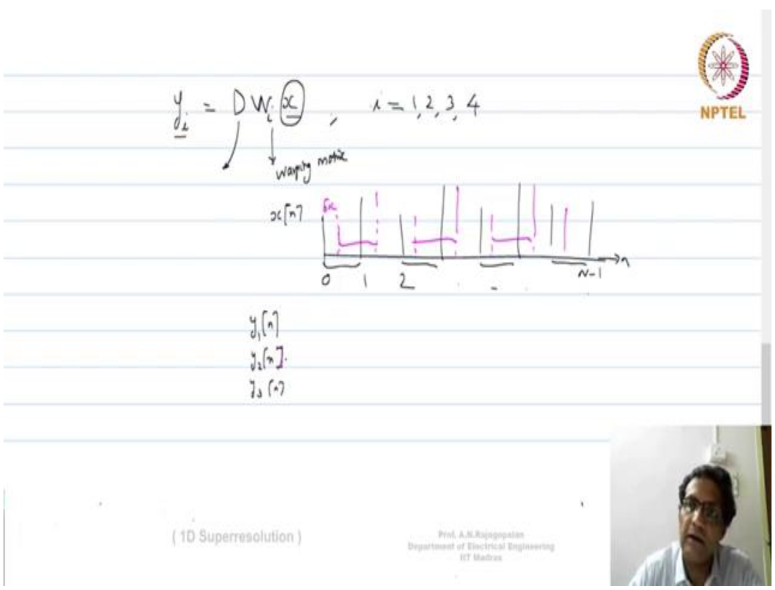
You are sampling the in between points that you ideally like to sample. And so the idea is that if you can move like this, now can you kind of aggregate all this information through your LR images in order to be able to build up a high resolution picture. Because idea is that under what conditions can you get us do it. Now, clearly there is one condition. It does not always mean that super resolution is possible. For example, one condition under which super resolution is not possible despite camera motion can be explained like this.

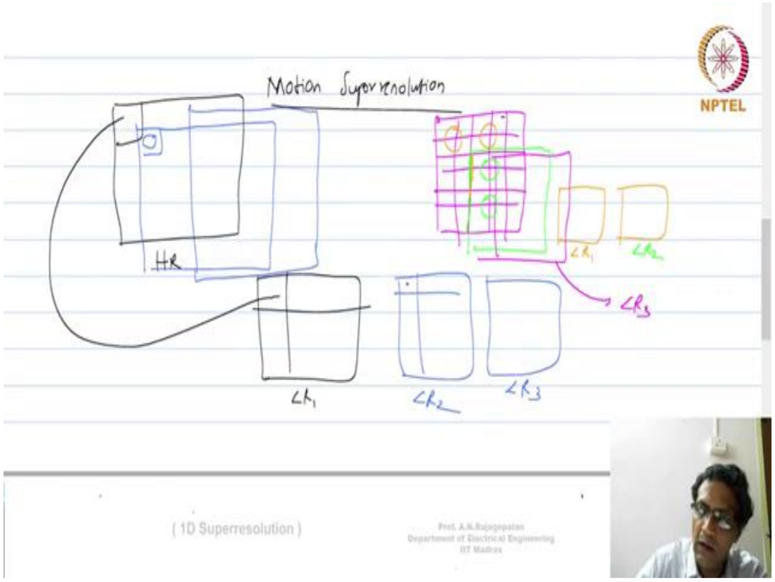
Suppose you are translating, let me say that your original picture was here and let us say this is your high-resolution grid, and you captured an image and you know that these four pixels are mapped to one, then these four mapped to the next and so on. Now, corresponding to this, let us say that you have an LR 1. Now, if your motion in the high-resolution grid is such that your next image unfortunately for one of the crazy reason, it gets formed here. Let us say if it gets formed here.

Now if the next image, if you capture here, then what it will mean is, you will be capturing this average but this average is already captured by LR1, because that is how LR1 was getting formed. And therefore, and this was already there in LR1, this was already there in LR, no sorry this is not that. Now this is already there. And then whatever subsequent is going to come is already being captured by LR 1, therefore your LR2 may carry nothing new.

The idea is that you should sample some in between points and not the same set of grid pixels or grid points that the first, that let us say first sort of camera pose gave you. So as a camp pose, the idea is that you are capturing different averages and therefore aggregate all this information in order to be able to come up with a high-resolution picture. That is the idea behind doing super resolution.

(Refer Slide Time: 18:08)





So, the way to express this is to actually write it down as an expression where you have Yi which is the Ith observation to be equal to D. Now Wi, I am going to introduce a matrix Wi. Let us now, we should understand all of these vectors plus some let us say noise Ni. Now, let us for the time being, let us just the noise out, let us assume you have an ideal model. Because once you bring in non-idealities, things will change. But for the moment let us just assume that everything is nice.

Now, what is the role of? So, this we know already, this is a down sampler and averagor, this is the warping matrix. Now, what this means is that, so what this means is here when you took an image here, corresponding to that you had LR 1, so when you say that you are going to super resolve X, this is what, you are going to get X, which is a super resolve version of Y1 in fact. Because Y2 is coming from something else which is not the original X. So, the idea is as follows.

So, what this warping will do is that it will tell you, so what you are effectively asking is a following. If I give you multiple observations, let us say for a factor of two I give you four observations, what you are asking is, get me that X such that when I do not know at all, that means when I take my first picture Wi as identity and I get Yi to be down sampled and average values of X, then I move my camera therefore there should be some warping that I should apply on X.

So, that I come here to this green area and average the pixels there in order to get my Y2 let us say. And then similarly, I should apply some of the matrix W let us say three on this original grid because I may have gone somewhere else, after that I could have gone somewhere here let us say. And for which I may have an LR3 and so on. So, you need to, so each case you will have a certain warping matrix which is acting on the high-resolution image, X.

So, this is acting on the high-resolution grid and after you have that then you can down sample average in order to show the observation, in order to see the observation. So, what you are effectively asking is give me that one X, we are not asking for each of those warp versions, we are just asking for the one X corresponding to LR1 that X which is a high resolution version of LR1 which when warped will give me, which with no warp and down sampled and averaged will give me Y1, which when warped and down sampled and average will give me Y2, warp and down sampled and average will give me Y3 and so on.

That is what is your optimization problem, which you are trying to solve and on top of that you can, of course, add other constraints on X and so on, which is your high-resolution picture. Now in the sense that in a regularization framework, the observation term that you will have this norm of YI minus let us say DWi X norm square, in a sense that is what you look at. Now, we will kind of look at a very simple situation. We do not want to solve the super resolution problem in its sort of generality in this course.

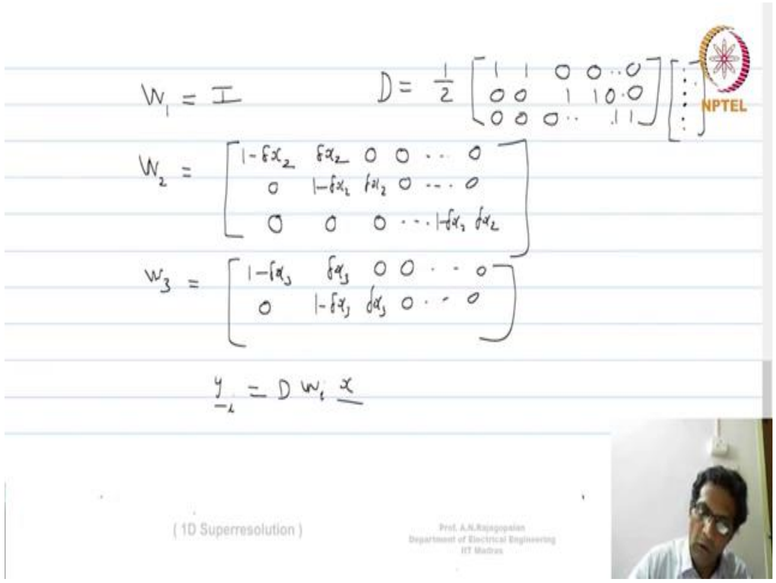
But I just wanted to show you a very simple situation. Instead of 2d look at look at the 1D equivalent of this example of the super resolution problem, what will be the 1D equivalent. Think that I have a sequence Xn which is a high-resolution sequence, 0, 1, 2 whatever. Now, what I have is basically a low-resolution observation corresponding to this sequence X, this is your Xn versus N. Now, let us say whatever, it is of some length N, capital N.

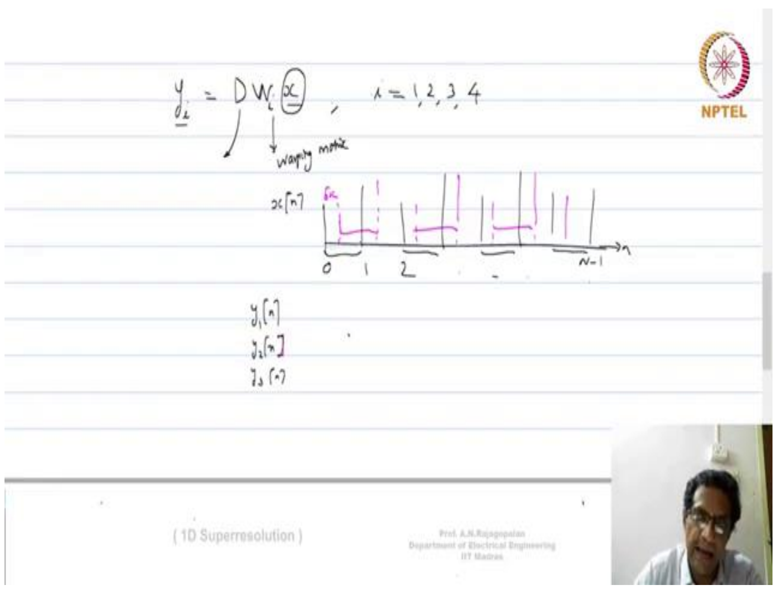
Now, my first observation Y1 of N is coming as average of these two, average of these, average of these two values and then go on and finally average of those two values. Then Y2 of n, this is like because of motion, I could have moved somewhere here, in the sense that, it could be… Now wait a minute, so the idea is that I could be here. And then right here I could be and then here and then I could be here, here, so I moved by a global motion which is simply a constant.

So, I moved by some constant amount, let us say that this is delta X or something. So, if that is my high-resolution motion just as we said, for images you could have a 2D translation. Imagine that this is just delta X in the time domain. Then what you are effectively saying is that Y2 of n will then be the average of these two, average of these two, average of these two and so on and that gives you Y2. Similarly, you can get Y3 of n and the Y4 of n and so on. Now, you can get all of them in the same manner.

Y2 of n, then let me put this is Y3 of n, so you say I am like Y3 of n, Y4 of n and so. Now, the idea is that given these observations what we would like to do is we like to get our X out, X is your high-resolution sequence. We are kind of looking at the 1D case. And we would like to like to get our Xn back.

(Refer Slide Time: 23:17)





Now before we can do that, so we need to understand as to how this model is going to look like. And now look at your W. Now let us say W1 is simply identity because there you have not moved at all and it is simply an average of the original X values. What about let us say now W2? W2 will have, let us say if it is Delta X 2 is the motion. Delta X2 is the motion, then this would have 1 minus W x2 because now, because we are saying that that we would like to arrive at these values by doing a lenient interpolation.

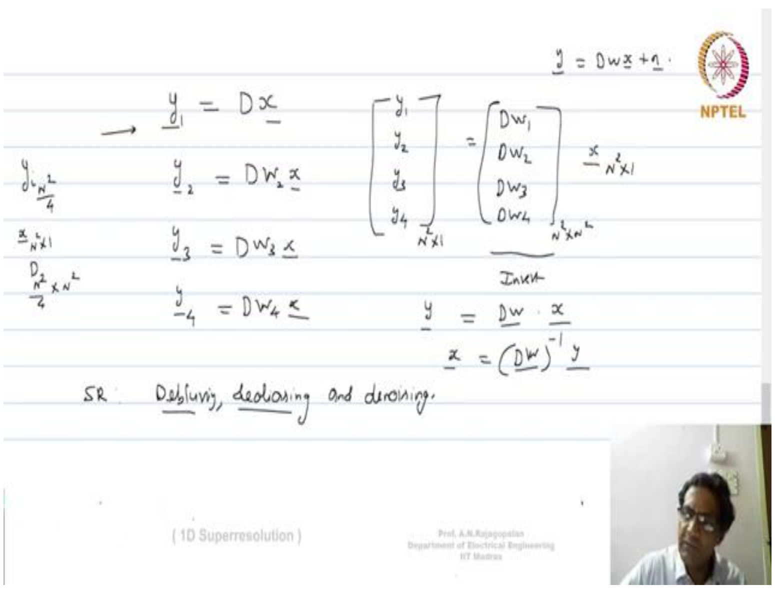
If it is 2D then you will do a bilinear interpolation, but in this case it is going to be linear interpolation one minus delta X2, delta X2 then like 0, 0, 0, 0, then you will have 0, 1 minus delta X2, delta X2, 0, 0, 0 and then all the way up to 0, 0, 0, all the way up to 1 minus delta X2, delta X2. This will be the form of W2.

Similarly, W3 will have 1 minus delta X3. So, in general if you have any delta Xi, so you will have delta X3, delta X3, 0, 0, 0, then 0, 1 minus delta X3 and so on.

So, we know how to do linear interpolation. What about your D itself? Now D is this special matrix so if it is a down sampler and averagor, and therefore, for the 1D case will be 1, 1, 0, 0, all the way up to 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, all the way up to 1, 1 at the final values. So when you look at Y is equal to D or Yi is equal to DWi X, this is what you mean, so you are going to warp X by the warping matrix, so which means that you can get these in between values and those will have to be averaged and down sample which is what this will do.

See, for example, so when this acts on an X, after of course, the warped X, what it is going to do, it is going to average the first two values. In order to, I mean it give you the first entry, then it is going to average the next two value, so if 1 comma 1 will give you the second entry, then the third entry and then the another final two entries. So that is how it is going to take care of down sampling as well as average and that is why we call this down sampler come averagor kind of matric.

(Refer Slide Time: 25:39)



And now what is super resolution? That entails in something like this. So, let us say I have Y1, I have captured one low resolution observation that is simply DX. And now Y2 is DW 2x because Y2 will come from a warp version of X. Then you got Y3 is equal to DW 3x and then we got Y4 which is equal to D W4x. Therefore, I mean if you think about it, if X is let us say, so if you can write this up plus Y1, Y2, Y3, Y4 is the start of all the LR observations, so if you are going up by a factor of 2, then your DWn.

Let write this down as DW1, DW2, DW3, DW4, so each of these DWs and as you can see is of a size, so if this is, let us say if you are down by a factor of two then your Yi has a size which is n by 2 by n by 2 or n by 2 into n by 2, in fact, so which is like lexicographic ordered, so it will be n square by 4. X lexicographic ordered is n square by 1, and D is n square by 4 by n square.

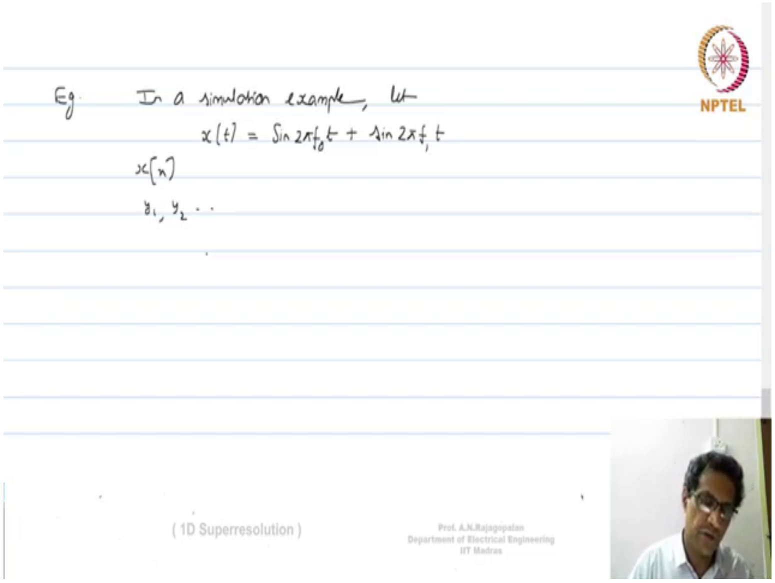
And therefore clearly what this will mean is when you stack them up this will become n square by n square because you have got like four of them, n square by 4, n square before, n square by 4, so this becomes n square by n square, this again because you have stacked up 4 observations, this together become n square by 1 and this is X which is n square by one.

And now you can actually invert this matrix provided, of course, you do not have a pathological condition like the one that I mentioned where LR1 and LR2 are exactly shifted versions of each other. Unless that happens, you will be able to invert this matrix and suppose you call this as y is equal to DW into X, then you can compute X as inverse of DW. This matrix inverse times y and that will be X. So, in a sense like the super-resolution problem we can define it in three words it is deblurring, dealiasing, and denoising.

Now you will wonder why did I bring in deblurring suddenly. Well, inherently there is a blurring. Whenever you do super-resolution because you are down sampling, they are averaging, they are averaging four pixels, any averaging is like blurry. Therefore, you try to deblur, try to dealias, because you are trying to go from an alias image to an unalias image because your Xn is unalias in this case.

They want a sequence where is Y1, Y2 are all alias so dealiasing and denoising because of the fact that typically your model will be something like Y is equal to DWx plus n, and therefore you might also want (())(28:33) noise. So, somebody asked you what is a special super-resolution in three words, it is simply deblurring, dealiasing and denoising.

(Refer Slide Time: 29:00)



Now, let me just give you a 1D example, where let us say, which will kind of throw some light on this in order to kind of look at how you can kind of perform this kind of dealiasing. Let us just take this example of mine. Now, let us just look at this example. Suppose, let us say in a simulation example, what you have. Suppose you have a simulation example, supposing a simulation example, we have let X of T is equal to sin 2 pi F naught T plus sine 2 pi F 1 T.

And the whole idea is if we had samples of X of T, let us call this as X n. If I give you down sample then average versions of Xn which will be Y1, Y2 and so on, can you get a reconstruct X from Y1, Y2. If you just interpolate Y1 and Y2 to the size of X, the hope is that you will not be able to retrieve your X back. The idea is can you combine multiple low-resolution observations like Y1, Y2, Y3, Y4 and so on in order to be able to reconstruct X.