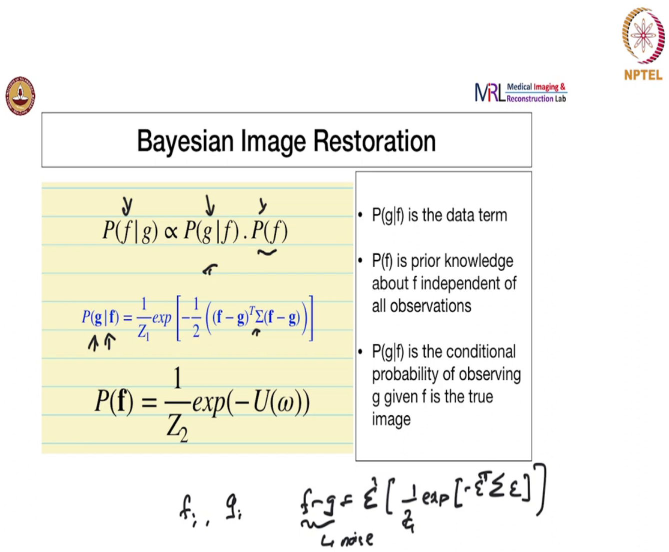
Medical Image Analysis
Professor Ganapathy Krishnamurthi
Department of Engineering Design
Indian Institute of Technology Madras
Lecture 12

Bayesian Image Restoration

(Refer Slide Time: 0:14)



Hello and welcome back. So, in this video we will talk about Bayesian Image Restoration. So, far we have looked at methods like linear filtering, we saw Boxcar averaging we also looked at bilateral filter ect, there we have assumed some prior knowledge about the noise in the image which is basically zero mean additive noise expectation goes zero and we have expected that especially in linear filtering we have some constraint on the image in the sense we assume that it is locally constant that is one way where we use the averaging.

So, other than that we have not used any constraints. So, the probabilistic framework which is the base and image restoration framework we are able to incorporate additional constraints and f sub for smoothness constraints on f. So, the idea here basically again for denoising is to determine an unknown function f from some observation g so in this case the observation g is nothing but the noise image itself. But it need not be in some problems for instance g can be just a bunch of features that are related to the picture or the problem at hand so for instance it can be used for segmentation classification etc.

So, for the purposes of this topic so we assume that g and f are vectors which is what is given in block face here block face letters g and f they are assumed to be vectors so we rasterize the image and so we have one big column vector we also refer to the individual elements of the vector as fi and gi.

So, the idea is to use bayes rule so we will write down bayes rule for this estimation problem remember we are trying to estimate the denoised image for the true ideal image f given the noisy observation g. So, the probability of that which is sometimes referred to as the posterior probability is proportional to the likelihood p of g given f times the prior p of f.

This p of g u and f is basically called the data term this is to make sure that that is that the image you estimate the f you estimate is not too far from g. So, it should be consistent with g so it is called the data term p of f is the prior knowledge about f independent of the observation g.

So, this is not something that g is giving this p of f is something that you are imposing for instance it has to be smooth etc. So, given this is the general framework for bayesian image restoration. So, people get clever about making mostly a by manipulating this p of f this p of f is the prior term that your knowledge of f that you want to impose is sometimes you can think of it as a regularization term.

It is called in the context of optimization you can think of this as the regularization term p of g given f is nothing but what is the what do you call the data fidelity term that whatever f you are estimating it should be consistent with whatever you have observed that is f is the denoised version of g that should be meaningful.

The conditional probability p of g given f which is the data term is characterized by some noise that is another thing so you are already incorporating noise here so you are saying that a g is a corrupted version of f and we are also now making an assumption about the type of noise that is the type of noise you are assuming is that it is a zero mean Gaussian noise.

So, why do you call this the noise distribution because if you look at this expression it is a Gaussian but it is cases g and f are vectors we have written it in this form with sigma being the covariant matrix but f - g you can say f-g=ϵ and so this becomes this expression p of g given f actually becomes $f - g = \epsilon[\frac{1}{z} exp[- \epsilon^T \Sigma \epsilon]]$.

So, this is just f minus g it is basically that is the noise that is a noise at every point that is the noise so you are just modeling the noise here so you are also and in the process you are also saying what is the probability of g given a certain f observing g given a certain s because g is the noisy version you are this probability distribution also defines what the noise is. The z1 is a normalizing factor for Gaussian distribution.

(Refer Slide Time: 4:58)

$$p(g|f) = \frac{1}{Z_1} \exp\left[-\frac{1}{2}\sum_{i=0}^{N-1}\frac{(f_i-g_i)^2}{2\sigma^2}\right]$$

$$p(A,B) = p(A).p(B)$$

$$\prod_{i=0}^{N-1} \exp\left[-\frac{1}{2}\frac{(f_i-g_i)^2}{\sigma^2}\right]$$

$$P(f) = \frac{1}{Z_2} \exp(-U(w))$$

$U(w) -$ Pot

Now, let us assume that if we assume that all the pixels in the image are independent of each other all the pixels image are independent of each other and that is you can draw each pixel from this distribution independently and then they all have the same variation and the equation we saw earlier simplifies to the following $p(g|f) = \frac{1}{z_1} exp[\frac{-1}{2}\Sigma\frac{(f-g)^2}{2\sigma^2}]$.

So, there is a summation in the exponent so how did this happen we are saying that all the pixels we treat all the pixels individual independently of each other so if you have an image the probability of getting that image that noisy image g is nothing but the product of the probabilities of the observing the individual pixels.

So, since this is an exponential distribution the product adds up in the exponent so if you want to rewrite this I will say $\Pi\, exp[\frac{-1}{2}\frac{(f-g)^2}{\sigma^2}]$. This is for a pixel i there are n pixels so for observing the entire image you multiply the probabilities assuming that all the pixels are independent.

So, given two per two pixel so if this rule p(A,B)=p(A).p(B) if a and b are independent events, so similarly we can write this for the entire image and this is once again the product adds up in the exponent this is the exponent so you can add up so from here you can get this expression. The a prior probability is modeled by another exponential distribution so 1 over z

2 x it is not Gaussian but is just an exponential minus u of omega. This u of omega is often referred to as a potential clique potential or I will just simply say potential.

(Refer Slide Time: 7:25)

$$p(g|f) = \frac{1}{Z_1} \exp\left[- \sum_{i=0}^{N-1} \frac{(f_i - g_i)^2}{2\sigma^2} \right.$$

$$p(f) = \frac{1}{Z_2} \exp(-U(\omega))$$

↳ clique potential

$$U(\omega) = \beta \sum_{k=0}^{k-1} u_{ik}$$

K , $u_k \rightarrow$ number of pixels in clique k

So, now if you let me write this down again so that then we can write this so we are using g and f as vectors so if we write it down in terms of individual pixel values what we can do we can write down is that the probability of observing g given f is the true image can be written as there is some normalizing constant the exponential and it is $p(g|f) = \frac{1}{z} exp[\frac{-1}{2} \Sigma \frac{(f-g)^2}{2\sigma^2}]$.

The probability p of f we saw that the probability $p(f) = \frac{1}{z_2} exp(-U(w))$. So, what is this u of w this is often referred to as a clique potential. So, clique potential is basically you are considering groups of pixels.

So, you can consider a group of pixels and each of them can be called a clique and for each group of pixels you have some you define some kind of spatial adjacency and while then each so and based on the spatial adjacency you also define these potential so for instance one potential that is often used is u of omega.

So, I will tell you what this is $U(\omega) = \beta \Sigma u_i k$. So, what is this where k is the number of clique so clique is given a neighborhood into how many ways can you split them into pairs three at a time four at a time etcetera like that is how we would call it and if you take every clique u k is the number of pixels, pixels in clique k which has the same intensity as center pixel i.

So, you can consider like a square neighborhood 3 by 3 neighborhood and count let us say I will use different color these are this is the center these are same maybe these are the same this let us consider this whole thing as a clique and maybe the others are slightly different it is different values I am just going to put a dot so you can count the number of pixels which have the same value as the center pixel.

Now, what this does is to enforce smoothness so you prefer images that have in this case these you prefer images which have very small neighborhoods with pixel values being almost the same. That is the smoothness that is what you are enforcing with this kind of potential function. So, this is in terms of probability so for doing this it is solving this problem is kind of hard because you see there are many arrangements of pixel values correct.

So, and that becomes much, much harder to do so one way you would solve this problem is you would actually formulate this a p(f|g) which is what we saw there.

(Refer Slide Time: 11:24)

$$\left[p(g|f) = \frac{1}{z_1} \exp\left(-\sum_{i=0}^{N-1} \frac{(f_i-g_i)^2}{\sigma^2}\right) \right]$$

$$\frac{1}{z_1} \exp\left(-\frac{1}{2}\frac{(f_1-g_1)^2}{\sigma^2}\right) \cdot \exp\left(-\frac{(f_2-g_2)^2}{\sigma^2}\right) \cdots \exp\left(-\frac{(f_{N-1}-g_{N-1})^2}{\sigma^2}\right)$$

$$\frac{1}{z_1} \prod_{i=0}^{N-1} \exp\left(-\frac{1}{2}\frac{(f_i-g_i)^2}{\sigma^2}\right) = p(g|f)$$

So, then beginning just try to write down how we are going to calculate f that is more important. So, calculation of f so we will write it in individual terms so $p(g|f) = \frac{1}{z_1} exp[\frac{-1}{2} \Sigma \frac{(f-g)^2}{2\sigma^2}]$.

So, in case you are wondering how we got here, how what is assumed is we what this assumes is that we treat each pixel as an independent random variable in the sense so if we choose a pixel it is chosen with a certain probability from this you can think of it that way. The other way of looking at this is that the noise is Gaussian distributed.

So, that is why there is a summation in the exponent so basically if you start from the exponential distribution and you are modeling the entire image with this so the probability of observing the first pixel will be times if you use just this distribution it will be f 1 minus g 1 squared by sigma square times so the probability of observing the entire image is nothing but the product of observing the entire product of observing each of the individual pixels.

So, this is for pixel 1 similarly you can write this for one half let us say again I am going to not consider the other factors f 2 minus d 2 squared by sigma squared and of course we assume sigma is the same for all of the pixels like that you can go all the way to exponential sorry this minus sign here minus f n minus g of n square in this case n minus 1 square by sigma square.

So, it is just the multiplication all of all each for each one for each pixel you have a term like this and when you multiply you sum in the exponent so that is why you get here so for this you can ignore this product from this product if you have if you want to write this particular expression in a compact form you write it in this solution so you would say $\frac{1}{z_1} \Pi \, exp[\frac{-1}{2} \Sigma \frac{(f-g)^2}{2\sigma^2} \ ]$

This is the probability that this is p of g given f if we go pixel by pixel so we take every pixel for every pixel i we estimate this probability and then the probability for all the pixels put together given that they are independently drawn is just a product for each of the individual probabilities. So, this is how we estimate this number and if this is our exponent the product sums up in the exponent so we get this expression I hope that much is clear.

So, the fundamental assumption here is that the probabilities of the individual pixels are independent or independent of each other and they all have the same variance sigma square. And if you think of it that way then this simplification can happen so this summation of the exponent we have one more here.

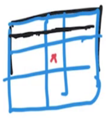
(Refer Slide Time: 15:13)

$$p(f) = \frac{1}{Z_2} exp(-U(\omega))$$

$$U(\omega) = \beta \sum_{k=0}^{k-1} u_{ik}$$

↳ Low

k – number of cliques



So, how do we define this $p(f) = \frac{1}{z_2} exp[-U(\omega)]$.

$U(\omega) = \beta\Sigma u_{ik}$. Where k is the number of cliques so cliques are basically how you in the context of this problem let me rewrite this clique.

So, in the context of this problem clique are basically the neighborhood of every pixel center pixel i if i is the center pixel you are looking at the neighborhood of the center pixel and you can choose different kinds of neighborhoods and each neighborhood you would call that as a clique. So, in this particular case u i k is the number of pixels in a clique k which has the same intensity value as the center pixel it is the same as center pixel.

So, if the neighborhood is the same then the potential is always increasing. So, consequently you will have very strong this this will lead to very strong edges because if you look at it if there is a neighborhood which straddles an edge so before that let me draw some pictures so that you can appreciate this better let me just use a different color so let us say a 3 by 3 neighborhood.

So, if there is an edge here, a very strong edge going through here has different pixel values. And this is the center pixel this this one is a central pixel now the edge will have a considerably different value than the center pixel correct so then u will be lower so this will be low if there is an edge around the center pixel u will be low so consequently p of f will be higher so think of it that way.

So, enforcing this constraint of I know having an increased potential whenever the cliques of central pixel have the same value as the center pixel leads to very strong edges. So, the other assumption is the Marconian assumption or Marconian property wherein you have for an image you will have there are n pixels so each of them will have a neighborhood so you will assume that each of those neighborhoods are independent of each other just as we assume that each pixel is independent of each other when we do the we did the data fidelity term or the likelihood term.

We assume for the prior term that each of these neighborhoods the way we treat them are independent of each other so then we can still when we consider each of the central pixels and we look at the entire probability prior probability we just multiply that as well.

(Refer Slide Time: 18:11)

$$p(f|g) = \exp\left(-\left[\sum_{i=0}^{N-1} \frac{(f_i - g_i)^2}{2\sigma^2} - \beta \sum_{k=0}^{k-1} u_{i,k}\right]\right)$$

$$\sum_{i=0}^{N-1} \left(\frac{(f_i - g_i)^2}{2\sigma^2} - \beta \sum_{k=0}^{k-1} u_{i,k} \right)$$

Simulated annealing, ICM, Mean-field annealing

So, taking that all into account we can rewrite we can rewrite this p of f given g as change color it is the write it in this form I will write down the last the final form so that you can work through the algebra itself $exp(-\ [\Sigma \frac{(f-g)^2}{2\sigma^2}\ -\ \beta\ \Sigma u_{i,k}])$.

so the maximize we want to maximize this probability so that is the probability of the true image f given the noisy image g that is here and we want to maximize this and in order to do that we wrote it down as a product of the data term or the likelihood term so and the prior which now acts as a regularizer.

So, now we have that add up the prior so that the probability of getting very strong edges are there so the images will have very strong edges the maximization of this probability actually means minimization of this exponent so minimization of the exponent $\Sigma \frac{(f-g)^2}{2\sigma^2} - \beta \Sigma u_{i,k}$ so this beta is just some hyper parameters that you will have to choose to make that work for the particular problem.

So, this is the last function that we are trying to optimize so we have now made this into an optimization problem. There are many techniques for doing this optimization; one is called simulated annealing. The other one is called ICM technique. There is also a mean field annealing.

The reason why there are these so many techniques is because this optimization problem is hard so for instance that is primarily comes from here and of course the size of the problem is also big because we are trying to estimate f i pixel by pixel and also this arrangement so we are looking at clique potentials where you are just counting the number of pixels with similar values.

So, but then the arrangement can be the deadly styles of so many arrangements and also depends on what the dynamic range of your image is so the number of images here increases exponentially so that way the problem is harder to solve we have to find the correct arrangement that satisfies this loss function so there are many of these strategies or so called greedy strategies to estimate this f i.

So, remember we want to estimate f i so that is the problem here and we have started off from the Bayesian framework figured out a data fertility term or the likelihood term we model this as a Gaussian of course it is the same as modeling the noise is Gaussian. And we then had a prior we impose the prior which is the same as regularization prior of course we can do lots of tricks there that is where people work on what kind of prior will lead to a certain type of image people have worked with.

So, you choose a prior and then you write down the expression of p of f given g which is basically trying to maximize this probability for a certain image f and when you do that in the process you will end up with this optimal problem that is what people typically solve in this line of research.

Once again I have presented this as some kind of pre-processing technique but it is not exactly pre-processing correct. There are so many people working on this as a research problem. It has been around for a long time, a difficult problem to solve. So, this kind of

Bayesian framework for solving these problems is generally difficult for them to solve and people have worked on this quite a bit this is a field of research by itself but then if you are going to denoise pictures before using them downstream for other processing maybe this is one method you can consider. So, this ends this week's lectures on Bayesian Image Restoration techniques. Thank you.