Bayesian image analysis in Fourier space (BIFS)

John Kornak*

Abstract

Bayesian image analysis provides a solution for improving image quality relative to deterministic methods, such as linear filtering, by balancing a priori expectations of image characteristics with a model for the noise process. A reformulation of the conventional Bayesian image analysis paradigm in Fourier space is given, i.e., such that the prior and likelihood are defined in terms of probability densities across spatial frequencies. Spatially correlated priors, that are relatively difficult to model and compute in conventional image space, can often be more efficiently modeled as a set of independent processes across Fourier space. The originally inter-correlated and high-dimensional problem in image space is thereby broken down into a series of independent one-dimensional problems; using 'parameter functions' to capture variation in the model's prior parameters over Fourier space. The Fourier space independence definition leads to easy model specification and relatively fast and direct computation that is on the order of that for deterministic filtering methods. Specific examples of applications are given and contrasted with Markov random field based model solutions.

Key Words: Bayesian image analysis, Fourier space, Image priors, Markov random fields

1. Introduction

Bayesian image analysis models provide a solution for improving image quality in image reconstruction/enhancement by incorporating *a priori* expectations of image characteristics along with a model for image noise. However, conventional Bayesian image analysis models, defined in the image space, can be limited in practice because they can be difficult to specify and implement (requiring problem-specific implementation) and slow to compute. Furthermore, commonly used Markov random field model priors in Bayesian image analysis (as conventionally used for the type of problem discussed here) are not invariant to changes in resolution and not naturally isotropic (directionally invariant).

Our approach to overcoming the difficulties of conventional Bayesian image analysis is to reformulate Bayesian Image analysis in Fourier Space (BIFS). Spatially correlated prior distributions (priors) that are difficult to model and compute in conventional image space can be successfully modeled as independent across locations in Fourier space. The original high-dimensional problem in image space is thereby broken down into a series of one-dimensional problems, leading to easier specification and implementation, and faster computation.

2. BIFS modeling framework

Consider x to be a true or idealized image (e.g., noise-free or with enhanced features) that we wish to recover from a sub-optimal image dataset y. (We use the common shorthand notation of not explicitly distinguishing the random variables and the corresponding image realizations (Besag, 1989), i.e., we use lower case x and y throughout.) The Bayesian image analysis paradigm incorporates *a priori* desired spatial characteristics via a prior distribution for the true image x and the noise degradation process via the likelihood.

Instead of the conventional Bayesian image analysis approach of generating prior and likelihood models for the true image x based on image data y directly, we formulate them

*University of California, San Francisco, USA

via their discrete Fourier transforms representations, $\mathcal{F}x$ and $\mathcal{F}y$, respectively. To indicate that we are working in Fourier space, we denote distributions with ϕ rather than the usual π notation. The prior distribution in Fourier space is denoted as $\phi_{\text{pr}}(\mathcal{F}x)$; the likelihood as $\phi_{\text{lk}}(\mathcal{F}y|\mathcal{F}x)$; and the posterior is $\phi_{\text{pst}}(\mathcal{F}x|\mathcal{F}y) \propto \phi_{\text{pr}}(\mathcal{F}x)\phi_{\text{lk}}(\mathcal{F}y|\mathcal{F}x)$.

The key component of the BIFS formulation that leads to the modeling and computational gains, is that we define both the prior and likelihood to consist of a set of independent distributions across Fourier space locations. Spatial correlation in image space is induced by allowing the parameters of the distributions to change over Fourier space. This independence definition can be contrasted with conventional Bayesian image analysis using Markov random field (MRF) priors, where Markovian neighborhood structures are used to define correlation structures across pixels via joint or conditional distribution specifications (Besag, 1974, 1989; Geman and Geman, 1984).

When defining a spatially correlated prior in image space via a set of independent processes across Fourier space, the full conditional posterior at a Fourier space location $k = (k_x, k_y)$, or for volumetric data (k_x, k_y, k_z), now only depends on the prior at k, i.e., $\phi_{\text{pst}}(\mathcal{F}x_k|\mathcal{F}y) \propto \phi_{\text{pr}}(\mathcal{F}x_k)\phi_{\text{lk}}(\mathcal{F}y_k|\mathcal{F}x_k)$, where we use $\mathcal{F}x_k$ as shorthand for $(\mathcal{F}x)_k$. The joint posterior density for the image is then $\phi(\mathcal{F}x|\mathcal{F}y) \propto \prod_{k \in K} \phi_{\text{pr}}(\mathcal{F}x_k)\phi_{\text{lk}}(\mathcal{F}y_k|\mathcal{F}x_k)$ where K is the set of all Fourier space point locations. It is precisely this independence property of the BIFS formulation and the corresponding reduction of the posterior distribution to a product of independent distributions that provides for the simple definition and computational benefits of the BIFS formulation. (Note that we index Fourier space along direction $v \in \{x, y, z\}$ by $\{-N_v/2, \ldots, 0, 1, \ldots, N_v/2 - 1\}$, rather than the common alternative of, $0, \ldots, N_v - 1$, as it leads to a more convenient formulation for BIFS prior models that are naturally centered around the zero frequency position of Fourier space.)

In order to account for the fact that medical images are generally real-valued, the Fourier transform must be conjugate (Hermitian) symmetric on the plane (or volume if 3D). A real image output is ensured by considering a realization of the posterior distribution to be determined by the half-plane (half volume), the half being conjugate symmetric to the first (see (Liang and Lauterbur, 2000), pp. 31 and 322). Therefore, for real-valued images, the BIFS posterior is only evaluated over half of Fourier space and the remainder is obtained by conjugate reflection.

Where does the independence property over Fourier space come from? That the independence assumption is a reasonable one to make can be heuristically gleaned from the fact that spatially correlated Gaussian processes in image space, lead to exactly uncorrelated Gaussian variables in Fourier space (where the Gaussian variables are mutually independent but with heterogeneous variances) (Zeger, 1985). Further, the Fourier transform of any spatially correlated stationary process in image space is asymptotically independent Gaussian in Fourier space (Peligrad and Utev, 2006), allowing a wide range of spatial correlation structures to be modeled by BIFS priors.

2.1 The BIFS prior and its parameter function

The process of defining the BIFS prior distribution capitalizes on the independence definition of the prior over Fourier space. First the distributional form for the prior at each Fourier space location is specified, $\phi_{\text{pr}}(\mathcal{F}x_k)$. Then the parameters of each of the priors are specified. In order to specify the parameter values across all Fourier space location, we define a *parameter function* over Fourier space that specifies the value of each parameter at each Fourier space location. Specifically, for some parameter α_k of $\phi_{\text{pr}}(\mathcal{F}x_k)$ we define the parameter function f_α such that the $\alpha_k = f_\alpha(k)$. For many problems it is desirable to define a spatially isotropic prior, which can be induced by defining $\alpha_k = f_\alpha(|k|)$, where

$|k| = \sqrt{k_x^2 + k_y^2}$ in 2D or $\sqrt{k_x^2 + k_y^2 + k_z^2}$ in 3D, i.e., such that f only depends on the distance from the origin.

It is convenient to define separate (and independent) priors and associated parameter functions for each of the modulus and argument of the complex value at each Fourier space location. Working with the modulus and argument provides a more natural framework than working with the real and imaginary components, because contextual information (e.g., smoothness, edges, or features of interest) most strongly relates to the magnitude of the process involved rather than the phase information.

2.2 BIFS likelihood

Because we model based on independence across Fourier space points, a range of different noise structures (defined in Fourier space) can readily be incorporated into the likelihood $\phi_{\text{lk}}(\mathcal{F}y_k|\mathcal{F}x_k)$. For example, the combination of Rayleigh noise/Rician likelihood Rice (1945) for the modulus with uniform argument on the circle corresponds to the likelihood model of Gaussian noise in image space.

2.3 Posterior estimation

It is at the posterior estimation stage that the computational gains of the independent BIFS formulation are ultimately realized. Posterior estimation in conventional Bayesian image analysis tends to focus on *maximum a posteriori* (MAP) estimation (minimizing a 0-1 loss function), because it is the most computationally tractable. In the BIFS formulation the MAP estimate can be efficiently obtained by maximizing the posterior conditional at each Fourier space location, i.e, $x_{\text{MAP}} = \mathcal{F}^{-1}(\mathcal{F}x_{\text{MAP}})$ where $\mathcal{F}x_{\text{MAP}} = \{\mathcal{F}x_{k,\text{MAP}}, k = 1, \ldots, K\}$ and $\mathcal{F}x_{k,\text{MAP}} = \max_{\mathcal{F}x_k} \{\phi_{\text{pst}}(\mathcal{F}x_k|\mathcal{F}y_k)\}$. This contrasts with conventional Bayesian image analysis, where even the most computationally convenient MAP estimate requires iterative computation methods such as conjugate gradients or expectation-maximization to obtain it.

2.4 Implementation

Implementation of BIFS modeling requires the following steps: 1) Fast Fourier transform (FFT) the original data from image space into Fourier space; 2) Define the prior in Fourier space; 3) Define the likelihood in Fourier space; 4) Combine the prior and likelihood via Bayes' Theorem to generate the Fourier space posterior; 5) Generate posterior estimates/summaries/simulations (in Fourier space); 6) Inverse FFT posterior estimates/summaries/simulations back to image space.

2.5 BIFS properties

We here discuss two important properties of the BIFS formulation that lead to meaningful advantages over MRF-based models beyond those already discussed.

a) Invariance Property: Defining the prior in Fourier space leads to a natural invariance property of the prior form with respect to changes in image resolution. When increasing resolution in image space the central region of the Fourier transform at high resolution corresponds to that of the complete Fourier transform at lower resolution (the lower resolution image is a band-limited version of the image at higher resolution). Therefore, only the extension of the prior over the higher frequencies is required in the higher resolution case and the spatial properties for the lower frequencies remain unchanged. The BIFS invariance property contrasts with MRF models, for which in order to retain the spatial properties of the prior at lower frequencies, an increase in resolution would require careful manipulation of neighborhood structure and prior parameters to match spatial auto-covariance (Rue and Tjelmeland, 2001).

b) Isotropy Property: In order to define an isotropic BIFS prior all that is required is for the prior to depend solely on the distance from the center of Fourier space, i.e., $\phi_{\text{pr}}(\mathcal{F}x_k) = g(|k|)$. The relative ease with which isotropy is defined can be contrasted with MRF-based priors where local neighborhood characteristics need to be carefully manipulated by increasing neighborhood size and adjusting parameter values to induce approximate spatial isotropy (Rue and Tjelmeland, 2001).

Note that isotropy is not a requirement of the BIFS prior formulation. Anisotropy can be induced by allowing the prior function to behave differently in different directions from the origin through appropriately defining the parameter functions, i.e., such that α_k cannot be fully defined as $f(|k|)$.

3. Examples

In the two examples below we develop BIFS processing for MAP estimation for the following simple BIFS model structure at each Fourier space location:

- Gaussian (normal) prior for the modulus: $\text{Mod}(\mathcal{F}x_k) \sim N(\mu_k, \tau_k^2)$
- Uniform prior on the circle for the argument: $\text{Arg}(\mathcal{F}x_k) \sim U(0, 2\pi)$
- Gaussian noise model for the modulus $\text{Mod}(\epsilon_k) \sim N(0, \sigma^2)$
- Uniform noise model for the argument $\text{Arg}(\epsilon_k) \sim U(0, 2\pi)$

where ϵ_k is the complex noise at Fourier space location k. (Note this model is not Gaussian noise in image space, for which we use a Rayleigh noise model/Rician likelihood, for the modulus.)

The global posterior mode can then be obtained by generating the posterior mode at each Fourier space location based on conjugate Bayes for the Gaussian/normal distribution Gelman et al. (2003) with $x_{k,\text{MAP}} = (\frac{\mu_k}{\tau_k^2} + \frac{y_k}{\sigma^2})/(\frac{1}{\tau_k^2} + \frac{1}{\sigma^2})$. The remaining component is to define the parameter functions for the BIFS priors, which we choose separately for each example.

3.1 Example 1 - breast MRI reconstructions

Figure 1 shows the results of applying a BIFS prior for the reconstruction of a 2D slice of contrast-enhanced breast MRI data. The left image shows the original breast MRI image slice and the second shows the same image with added Gaussian noise (approximating the effect of background enhancement commonly seen in contrast enhanced breast MRI as a confounding factor to detecting breast cancer). The third image shows the reconstruction from the noisy data based on a prior $\phi_{\text{pr}}(\mathcal{F}x_k)$ that has an inverse square root (inv-sqrt) parameter function for the modulus: $f_\mu(|k|) = a/\sqrt{|k|}$ with the standard deviation of the modulus modeled as proportional to the mean: $f_\tau(|k|) = cf_\mu(|k|)$. The right image shows a 1st-order intrinsic Gaussian MRF (IG-MRF) reconstruction (i.e., pairwise-difference) (Besag, 1989). Both priors reduce the noise and enhance the image, though the BIFS prior does a better job of preserving features (we tried a range of parameter values for the IG-MRF and the result shown was visually the best). The reconstruction ***took less than 0.3s to compute*** (including forward and back Fourier transforms) using Matlab R2012b on a Macbook Pro with a 2.3 GHz i5 processor (c.f. conjugate gradients optimization for a pairwise intrinsic

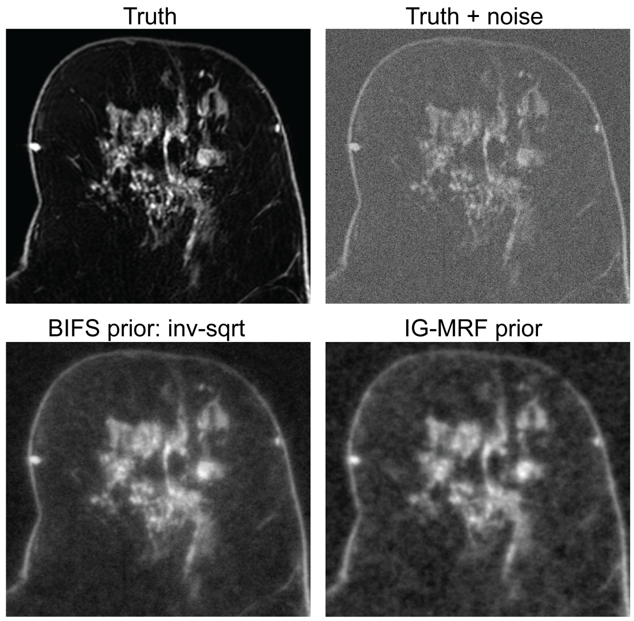

Figure 1: Comparison of BIFS inverse square root prior and 1st-order intrinsic Gaussian MRF prior for noise reduction of breast MRI data.

Gaussian MRF prior, which needed 18s for 100 iterations – there was no noticeable change with additional iterations).

3.2 Example 2 - lesion enhancement

In this example we used simulations to drive the parameter functions. That is, the parameter functions were generated from the empirical estimates of the mean and standard deviation at each Fourier space location over the simulated datasets. We simulated 1000 256×256 images representing lesions/tumor patterns. The number of lesions was modeled as a Poisson process; the lesions were simulated as randomly positioned truncated Gaussian probability density functions (resembling bumps) with random intensity, standard deviation (sd) on each axis, and correlation, distributed uniformly between -1 and 0, so that the process was not isotropic. Figure 2 shows a single new realization of the process on the left, with added noise in the center, and the BIFS MAP reconstruction on the right. The BIFS reconstruction clearly enhances the simulated lesions, especially note the upper left one which is all but lost in the Truth + Noise image. We forego showing a 1st-order intrinsic Gaussian MRF prior reconstruction because it would be a straw man comparison given the lack of capacity to incorporate the diagonal directional effects; incorporating directional preferences could be done for the Gaussian MRF, but would require careful engineering of parameters in larger neighborhood structures (Rue and Tjelmeland, 2001).

4. Conclusion

The BIFS modeling framework provides a new family of Bayesian image analysis solutions with the capacity to a) enhance images beyond conventional standards; b) allow straightforward specification and implementation across a wide range of imaging research and applications; and c) enable fast and high-throughput processing. These benefits along with the inherent properties of resolution invariance and isotropy, make BIFS a powerful tool for the image analysis practitioner.

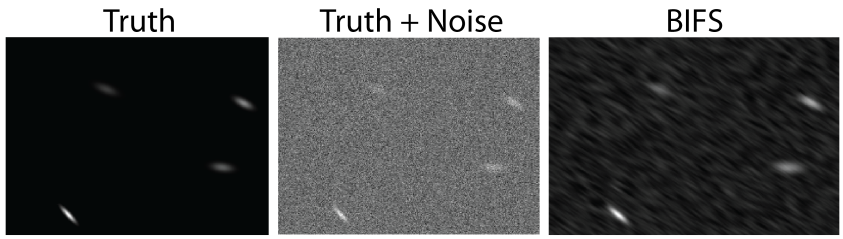

Figure 2: Simulation study and reconstruction of lesion/tumor patterns.

References

J. E. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 36(2): 192–236, 1974.

J. E. Besag. Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407, 1989.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, 2nd Ed.* Chapman & Hall/CRC, 2003.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

Z. P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging: A signal processing perspective*. SPIE Optical Engineering Press New York, 2000.

M. Peligrad and S. Utev. Central limit theorem for stationary linear processes. *The Annals of Probability*, 34(4):1608–1622, 2006.

S. O. Rice. Mathematical analysis of random noise-conclusion. *Bell Systems Tech. J., Volume 24, p. 46-156*, 24:46–156, 1945.

H. Rue and H. Tjelmeland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29(1):31–49, 2001.

S. L. Zeger. Exploring an ozone spatial time series in the frequency domain. *Journal of the American Statistical Association*, 80(390):323–331, 1985.