Bayesian Denoising/Deblurring of Poisson-Gaussian Corrupted Data Using Complex Wavelets

Timothy D. Roberts and Nick Kingsbury
Sigpro Laboratory, Cambridge University Engineering Department (CUED), Cambridge, U.K.

Abstract—In this work we summarize our algorithm used to denoise/deblur observations which result from a mixture of Poisson and Gaussian noise sources. Our algorithm, MSIST-P, is an extension of the MSIST algorithm, in that it accounts for a spatially varying statistical distribution. Using our own simulated ground truth data, we’ve obtained better reconstruction results than the Richardson Lucy algorithm.

I. NOTATIONS

In this document we use the following matrix/vector notations:

- $\mathbf{x}$ is a vector corresponding to the ground-truth image;
- $\mathbf{A}$ is a matrix representing the forward blurring operator;
- $\mathbf{A}^T$ is a matrix representing the inverse blurring operator;
- $\mathbf{b}$ is a constant vector representing the background signal;
- $\mathbf{y}$ is a random vector modeling the measurements;
- $\mathbf{y}_b$ is simply $\mathbf{y} - \mathbf{b}$;
- $\mathbf{x}_n$ is an estimate of $\mathbf{x}$ obtained after $n$ iterations of the algorithm;
- $w$ is a $j$-indexed with real and imaginary parts in alternate locations, comprising the coefficients of a complex-valued wavelet representation;
- $w_n$ is the wavelet transform of $\mathbf{x}_n$ obtained after $n$ iterations of the algorithm;
- $w_y$ is the wavelet transform of $\mathbf{y}$;
- $\mathbf{W}$ is a real matrix representing a forward wavelet transformation;
- $\mathbf{W}^T$ is a real matrix representing an inverse wavelet transformation;
- $\nu^2$ is a regularization parameter representing the assumed variance of the additive Gaussian noise;
- $\varphi^2$ is a regularization parameter which controls the shape of the re-weighted L2 sparse penalty $\mathbf{w}^T \mathbf{Sw}$;
- $\lambda^2$ is a regularization parameter which controls the relative strength of the re-weighted L2 sparse penalty $\mathbf{w}^T \mathbf{Sw}$;
- $\Lambda_\alpha$ is a diagonal matrix of wavelet subband weights, which approximates $\mathbf{W}^T \mathbf{A} \mathbf{W}^T$;
- $\Phi$ is $\mathbf{W}^T$;
- $\sigma_{w,j}^2$ is the variance of the $j$-th coefficient in $w$;
- $\mathbf{S}$ is a $j$-indexed diagonal matrix of wavelet coefficient inverse variance estimates;
- $s_j$ is the $j$-element of $\mathbf{S}$;
- $\Psi_y$ is a $j$-indexed real diagonal matrix of scaling coefficients corresponding to $w_y$;
- $\Sigma_w$ is a diagonal matrix whose elements are the elements of column vector $\nu^2 \mathbf{1} + \Psi_y$;
- $\mathcal{T}\{\mathbf{q}, \mathbf{b}\}$ is a thresholding operator which sets elements of vector $\mathbf{q}$ less than $\mathbf{b}$ to 0, this represents the noise variance of $w_y$;
- $\mathbf{g} \sim N(0, \sigma_q^2)$ is an additive white Gaussian noise vector (detector noise);
- $\mathcal{Q}\{\mathbf{q}\}$ denotes the quantization of $\mathbf{q}$ to 16-bit unsigned integer precision;
- $\mathcal{P}(\lambda)$ denotes a Poisson random vector with mean $\lambda$, which models the shot noise of our imaging system.

II. ALGORITHM

The MSIST-P algorithm follows an extension of the basic MSIST algorithm [1] and applies the Gaussian scale mixture model (GSM) [3] to a combination of Gaussian and Poisson noise sources. It involves simple matrix-vector multiplies and indexing operations for each iteration:

\[
\begin{align*}
\mathbf{w}_{n+1} &= (\Lambda_\alpha + \Sigma_w \mathbf{S})^{-1}[(\Lambda_\alpha - \Phi^T \Phi)\mathbf{z}_n + \Phi^T \mathbf{y}_b] \\
\mathbf{S}_{2j,n+1} &= \mathcal{T}\{\mathbf{S}_{2j,n+1}, 0\} \\
\mathbf{S}_{2j,n+1} &= \mathbf{S}_{2j,n+1} + \frac{1}{\sigma_{w,j}^2}(\mathbf{w}_{2j,n}^2 + \mathbf{w}_{2j-1,n}^2) + \epsilon^2
\end{align*}
\]

This represents a threshold Landweber iteration in the wavelet domain, followed by a hard-thresholding in the spatial domain, at each iteration. In this algorithm, we have chosen to use the 3D version of the DT-CWT for $\mathbf{W}$ because of it’s directional selectivity and approximate shift-invariance properties.

III. VARIATIONAL INTERPRETATION

The MSIST-P algorithm assumes that the following observation model can be simplified by assuming a single Gaussian noise source with a spatially-varying variance.

\[
y = \mathcal{Q}\{\mathcal{P}(\Lambda \mathbf{x} + \mathbf{b}) + \mathbf{g}\}
\]

We develop a cost function and algorithm here in order to recover $\mathbf{x}$ from these blurred, noisy, quantized measurements. Neglecting $\mathcal{Q}\{\}$, we can rewrite eq. (5) using our spatially-varying Gaussian assumption:

\[
y \sim N(\lambda, \Sigma)
\]
where $\lambda = Ax + b$ is the local variance due to the Poisson noise, and $\Sigma$ is a diagonal matrix comprised of the elements of $\sigma^2 = \lambda + \sigma^2_j$.

Thus, eq. (6) results in the following posterior:

$$p(w|y) \propto p(y|w)p(w|s)p(s)$$

(7)

where we assume that $\lambda = Ax + b = AW^T w + b$ is uniquely determined by $w$, $p(w|s) \propto \exp(-\frac{1}{2}w^T Sw)$ captures our Gaussian scale mixture prior, and we assume an independent prior for each $s_j$. We've made a correction to the estimate of $b$ (+1.5) for all cases, which produced more accurate estimates using when experimenting with our ground truth data. We've tuned $K$ using information about the relative sparsity of the volumes, as some data sets have more zero-blocks than others. The parameter $\nu^2$ is varied in the same way but from a value approximately 10 times the largest observation in $y$, with a stopping value of 20% above $\sigma^2_j$ to a stopping value just below $\sigma^2_j$. This ensures a sparse solution while mitigating the risk of over-regularization. The parameter $\epsilon^2$ is chosen to decrease geometrically by a factor of 0.64 on each iteration from a value 20% above $\sigma^2_j$ to a stopping value just below $\sigma^2_j$. This ensures that the penalty function converges to a sparse solution, while keeping a near quadratic penalty $J(w, z)$ in early iterations so as not to become trapped near undesirable local minima in early iterations.

Convergence is assumed when the relative error $\frac{\|x_{n+1} - x_n\|}{\|x_n\|} \leq 0.1\%$, and the volumes look reasonable.

**REFERENCES**

[1] Y. Zhang and N. Kingsbury “Fast L0-based Image Deconvolution with Variational Bayesian Inference and Maximization-Minimization”.
