

Denoising Three-Dimensional and Colored Images Using a Bayesian Multi-Scale Model for Photon Counts

John Thomas White^a, Subhashis Ghosal^{b,c,*}

^a*SAS Corporation, Cary, NC, U.S.A.*

^b*North Carolina State University, Raleigh, NC, U.S.A.*

^c*Research is partially supported by NSF grant DMS-1106570*

Abstract

X-ray images of distant stars and galaxies are typically registered by low photon counts at the pixel level, for which Poisson distribution is a sensible model description. Such a pixel by pixel photon count data can be represented in a multi-scale framework, where the likelihood function can be factorized in independent factors which are functions of relative intensity parameters corresponding to different levels from the whole frame down to the pixel level. In a Bayesian approach, a prior is assigned independently on these relative intensity parameters and the image is reconstructed using the posterior mean of intensity parameter of each pixel. A novel prior which allows ties in the values of relative intensity parameters of neighboring regions has been recently shown to be very successful in finding structures in images. We extend this idea to reconstruct colored images from noisy data. In the context of astronomical X-ray images, color represents the energy level of photons, which are also typically

*Corresponding author

Email addresses: John.Thomas.White@gmail.com (John Thomas White),
sghosal@stat.ncsu.edu (Subhashis Ghosal)

URL: <http://www4.stat.ncsu.edu/~sghosal> (Subhashis Ghosal)

recorded by telescopes. The additional data on energy level of photons can give valuable information about the celestial objects under study. The energy level can be considered as the third dimension of images. In a more general sense, the technique we develop applies to three dimensional images, and can be used to process medical images.

Keywords: Bayesian denoising, colored images, Chinese restaurant process, multi-scale model, photon counts

2010 MSC: 68U10, 62H35, 62F15, 62P35

1. Introduction

X-ray images of distant celestial bodies are typically very faint with low photon counts in each pixel of the screen of the detecting telescope. In such a low-photon count regime, a Poisson model for pixel-wise counts seems to be very appropriate. The likelihood function in the Poisson model nicely factorizes in a multi-scale framework; see Kolaczyk (1999); Crouse et al. (1998); Kolaczyk and Nowak (2004); Willett and Nowak (2003); Starck and Murtagh (2006). A multi-scale framework is particularly attractive for a Bayesian approach, where independent prior distributions may be assigned on the parameters that arise in different levels of hierarchy, leading to independent posterior distributions. The prior distributions assigned on the parameters at each level induce shrinkage effects that help denoise the image. Recently White and Ghosal (2011) introduced a novel prior distribution which is capable of tying the neighboring values of parameters with certain probabilities. This helps finding continuous structures and boundaries from noisy images.

In this paper, we consider images that can be represented by three dimensional photon counts. The third dimension can be another spatial dimension, a temporal dimension, or a spectral dimension consisting of the energy level of each photon observed. In high-energy astrophysics, when taking X-ray images using a detection device, discrete energy levels are also recorded along with photon counts. Splitting the counts over different energy levels, we obtain a three dimensional array of counts corresponding to each pixel and energy-level combination (to be called a voxel below) that can be modeled by independent Poisson variables and can be represented in a multi-scale fashion. We shall extend the method introduced by White and Ghosal (2011) in the three dimensional setting.

Only a few three dimensional image denoising methods such as Krishnamurthy et al. (2010) are available in the literature. Most methods apply image smoothing techniques separately on slices of the third dimension. The drawback of this approach is that it does not offer any smoothing over the third dimension, and hence will not be able to rectify measurement errors associated with the energy levels. In contrast, the proposed denoising method will simultaneously smooth out the whole data as one unit instead of smoothing each individual third-dimensional slice. In a three-dimensional multi-scale model, each block in a coarser level is split into $2 \times 2 \times 2 = 8$ smaller blocks in the next finer level, until we reach the finest (i.e., voxel) level. The large block is often called the parent and the 8 smaller subblocks are called its children. The resulting compartment is called a parent-child group.

The basic idea behind our method is a three dimensional analog of the procedure

in White and Ghosal (2011). In the multi-scale representation of voxel-wise photon count data, the likelihood can be factorized as functions of relative intensity parameters that control the distribution of photon counts from parent to children at each level of the multi-scale representation. Instead of putting a continuous prior distribution, we consider an auxiliary Chinese restaurant process (CRP) [cf. Pitman (1995)] to determine the extent of ties among the relative intensity parameters and which relative intensity parameters are to be tied together. This leads to be a probability vector in a reduced dimension, on which we assign a Dirichlet distribution. The same assignment of prior is followed in each parent-child group within and across levels, except possibly for a change in the parameter of the underlying CRP. By invoking a crucial conjugacy property, posterior parameter updating can be described analytically. This leads to a fairly closed form expression of the posterior mean, which is used as the denoised image. The method is completely data driven in that all smoothing parameters, such as the parameters of the underlying CRP, are obtained automatically within the method by maximizing the marginal likelihood function of the these smoothing parameters. The main difference between the two-dimensional and the three-dimensional cases is that in the former case a 4-person CRP is used, while in the latter case, an 8-person CRP is to be used. A CRP in a parent-child group produces different configurations as a result of these ties where these terms are defined in Subsection 2.1. A 4-person CRP gives a distribution over 15 possible configurations and an 8-person CRP gives a distribution over 4,140 possible configurations. The resulting complexity poses substantial computational challenges. Nevertheless, the analytic form of the posterior mean due to the conjugacy of the posterior distribution and the factorization due to

the multi-scale structure allow us to handle the computation efficiently.

The outline of the paper is as follows. In Section 2, the statistical model and the prior are formally defined and the smoothing parameters are estimated from the data. Simulations are performed in Section 3 on a three dimensional Shepp-Logan phantom image [cf. Schabel (2006)] to investigate error reduction of the proposed method. An astronomical image with a third dimension corresponding to the energy component of each photon is denoised in Section 4.

2. Model and methodology

In the context of low intensity image processing, a three-dimensional image can be viewed as a three dimensional array of count data. To begin with, we first assume that the array length in each direction is the same and is of the form 2^S . Let $X_{(l,j,k)}$ stand for the photon count at voxel (l, j, k) , $l, j, k = 1, \dots, S$, and let \mathbf{X} denote the entire array. In the low photon count regime we are considering, it is reasonable model the voxel-wise count data as independently Poisson distributed: $X_{(l,j,k)} \sim \text{Poisson}(\lambda_{(l,j,k)})$, independently, where $\lambda_{(l,j,k)}$ is the intensity at voxel (l, j, k) , $l, j, k = 1, \dots, S$. Let $\boldsymbol{\lambda}$ stand for the entire array of $\lambda_{(l,j,k)}$, $l, j, k = 1, \dots, 2^S$. At intermediate scales $s = 1, \dots, S - 1$, there are 8^s block-voxels that are denoted by (l, j, k) where $l, j, k = 1, \dots, 2^s$, whose photon counts are to be denoted by $X_{s,(l,j,k)}$. Clearly, these values are obtained by summing the photon counts of its 8 children block-voxels through the relations

$$X_{s,(l,j,k)} = \sum_{l'=2l-1}^{2l} \sum_{j'=2j-1}^{2j} \sum_{k'=2k-1}^{2k} X_{s+1,(l',j',k')} \quad (1)$$

for all $l, j, k = 1, \dots, 2^s$, $s = 0, \dots, S-1$. Therefore $X_{s,(l,j,k)}$ is Poisson distributed with parameter $\lambda_{s,(l,j,k)}$, say. These parameters are connected by the following relations

$$\lambda_{s,(l,j,k)} = \sum_{l'=2l-1}^{2l} \sum_{j'=2j-1}^{2j} \sum_{k'=2k-1}^{2k} \lambda_{s+1,(l',j',k')} \quad (2)$$

for all $l, j, k = 1, \dots, 2^s$, $s = 0, \dots, S-1$, with $\lambda_{S,(l,j,k)} = \lambda_{(l,j,k)}$. Let $\rho_{s+1,(l',j',k')} = \lambda_{s+1,(l',j',k')}/\lambda_{s,(l,j,k)}$, $l' = 2l-1, 2l$, $j' = 2j-1, 2j$, $k' = 2k-1, 2k$, $l, j, k = 1, \dots, 2^s$, the relative intensity parameters at level $(s+1)$,

$$\boldsymbol{\rho}_{s+1,(l,j,k)}^* = (\rho_{s+1,(l',j',k')} : l' = 2l-1, 2l, j' = 2j-1, 2j, k' = 2k-1, 2k),$$

the resulting vector of relative intensities and

$$\mathbf{X}_{s+1,(l,j,k)}^* = (X_{s+1,(l',j',k')} : l' = 2l-1, 2l, j' = 2j-1, 2j, k' = 2k-1, 2k),$$

the corresponding photon counts, $l, j, k = 1, \dots, 2^s$, $s = 0, \dots, S-1$. Note that given $X_{s,(l,j,k)}$, the vector $\mathbf{X}_{s+1,(l,j,k)}^*$ is multinomially distributed with parameters $X_{s,(l,j,k)}$ and $\boldsymbol{\rho}_{s+1,(l,j,k)}^*$. Thus the likelihood function has the multi-scale factorization given by

$$\mathcal{P}(X_{0,(1,1,1)} | \lambda_{0,(1,1,1)}) \times \prod_{s=0}^{S-1} \prod_{l=1}^{2^s} \prod_{j=1}^{2^s} \prod_{k=1}^{2^s} \mathcal{M}(\mathbf{X}_{s+1,(l,j,k)}^* | X_{s,(l,j,k)}, \boldsymbol{\rho}_{s+1,(l,j,k)}^*), \quad (3)$$

where \mathcal{P} and \mathcal{M} respectively denote the probability mass functions of the Poisson and multinomial distributions. Thus the multi-scale representation gives a very useful reparametrization of $\boldsymbol{\lambda}$ in terms of the total intensity parameter $\lambda_{0,(1,1,1)}$ and the

level-wise relative intensity parameters $\boldsymbol{\rho}_{s+1,(l,j,k)}^*$, $s = 0, \dots, S - 1$, which will allow separate smoothing in each level of the representation. In particular, in the Bayesian approach we shall be pursuing, it is natural to assign independent priors on parameters at different levels, leading to independent posterior distributions. The independence allows easy computation of the posterior means of the original λ -parameters: If $\lambda_{(l,j,k)}$ is written as $\lambda_{0,(1,1,1)}\rho_0\rho_1 \cdots \rho_{L-1}$, then its estimate $\hat{\lambda}_{(l,j,k)}$ is given by the posterior mean

$$\mathbb{E}(\lambda_{(l,j,k)}|\mathbf{X}) = \mathbb{E}(\lambda_{0,(1,1,1)}|\mathbf{X})\mathbb{E}(\rho_0|\mathbf{X})\mathbb{E}(\rho_1|\mathbf{X}) \cdots \mathbb{E}(\rho_{L-1}|\mathbf{X}), \quad (4)$$

which gives the denoising of the observation $X_{(l,j,k)}$ at the voxel (l, j, k) . The prior for the parameter $\lambda_{0,(1,1,1)}$ will be chosen to be degenerate at $X_{0,(1,1,1)}$, leading to the first factor on the right hand side $X_{0,(1,1,1)}$, the estimate of the total intensity.

In most image processing contexts, the actual intensity parameters are typically unimportant — only the relative intensity parameters $\bar{\lambda}_{(l,j,k)} = \lambda_{(l,j,k)}/\lambda_{0,(1,1,1)}$ matter, since they describe the shapes and sizes of the objects under study. Estimates of $\bar{\lambda}_{(l,j,k)}$ are simply given by $\hat{\bar{\lambda}}_{(l,j,k)} = \hat{\lambda}_{(l,j,k)}/X_{0,(1,1,1)}$. To judge the quality of the proposed denoising method, we shall compare the values of $\hat{\bar{\lambda}}_{(l,j,k)}$ with $\bar{\lambda}_{(l,j,k)}$.

In many situations, such as in the context of X-ray images of galaxies and supernovas, photons have a relatively few energy levels, much less than 2^S , the physical length and width of the photographic plate. Hence it becomes necessary to consider a reduced number of energy levels 2^{S^*} , $S^* < S$. Hence k in (3) will be restricted to 2^{S^*} , and in levels higher than S^* , the 8-dimensional multinomial distribution will be replaced by

a 4-dimensional multinomial:

$$(X_{s+1,(l',j',k)} : l' = 2l - 1, 2l, j' = 2j - 1, 2j) \\ \sim \mathcal{M}(\cdot; X_{s,(l,j,k)}, (\rho_{s+1,(l',j',k)} : l' = 2l - 1, 2l, j' = 2j - 1, 2j)).$$

Thus in the scales that do not split the third dimension, the likelihood is the same as the two-dimensional case where there are only 4 children in each parent-child group. Notice here that the third subscript is set to k since this is a two-dimensional split of each parent-child group. This factorization fits into the same framework. However, the levels where the third dimension is split can be either start from the coarser side or from finer side, leading formally two different methods as explained in Table 1.

Table 1: Possible partitions for a $64 \times 64 \times 16$ image.

Scale	Finer Smoothing	Coarser Smoothing
1	$2 \times 2 \times 1$	$2 \times 2 \times 2$
2	$4 \times 4 \times 1$	$4 \times 4 \times 4$
3	$8 \times 8 \times 2$	$8 \times 8 \times 8$
4	$16 \times 16 \times 4$	$16 \times 16 \times 16$
5	$32 \times 32 \times 8$	$32 \times 32 \times 16$
6	$64 \times 64 \times 16$	$64 \times 64 \times 16$

2.1. Prior Distributions

Since it is important to find structures in an image which can be thought of as constant intensity of the source of the radiation at several contiguous voxels, we specially design a prior distribution which can generate ties among the intensity parameters $\lambda_{(l,j,k)}$ at such voxels. This is done again by using the multi-scale representation and tying some of the values of the relative intensity parameters $\rho_{s+1,(l',j',k')}$, $l' = 2l - 1, 2l, j' =$

$2j - 1, 2j, k' = 2k - 1, 2k$, within each-parent-child group, and independently across different parent-child groups of all possible levels of the multi-scale representation, thus randomly creating a “configuration” of tied relative intensity parameters. The approach also has the local build-up property — voxels that are nearby are more likely to share a common value of their intensity parameters. Such a locality dependent tying mechanism has been proposed recently in an image segmentation problem by using a complex notion of distance dependent Chinese restaurant process [Ghosh et al. (2011)], but here the local feature arise automatically from the multi-scale decomposition and separately tying intensity parameters in each parent-child group. After a configuration of ties has been selected in a parent-child group, we distribute their prior masses respecting the configuration using a Dirichlet distribution, which is conjugate for the multinomial likelihood. Note that when some values are tied together, the prior must sit in a lower dimensional space to respect the configuration, leading to a mixture of Dirichlet distributions over varying dimension. Nevertheless, the resulting prior, being a mixture of conjugate priors, will also be conjugate, leading to an explicitly computable posterior distribution, completely avoiding slower Markov chain Monte-Carlo algorithms. To induce a distribution on the configurations in each parent-child group, we use a Chinese restaurant process described below.

The Chinese restaurant process (CRP), as described in Pitman (1995), is a model for feature sharing, where the N th observation is exactly equal to one of the K previously observed values or is a completely new value according to probabilities $N_t/(M + N)$, $t = 1, \dots, K$, and $M/(M + N)$ respectively, where N_1, \dots, N_K are the

multiplicities of occurrences of the previously observed values and M is a smoothing parameter controlling the extent of ties. The significant aspects of the CRP are that bigger blocks attract more values to them forming bigger clusters, and smaller values of M encourage more formation of clusters in general. The first property is an attractive feature for our prior assignment, since in images, we would like to form significantly big structures. Actually, the CRP is generally used in the context of a potentially unlimited sequence of observations, but in our context, we consider a finite version of it consisting of m observations forming ties with each other according to the probabilities given above, to be called an m -person CRP below. In a parent-child group consisting of 8 children, we shall create ties a priori using a 8-person CRP while a 4-person CRP will be used for a reduced size parent-child group consisting of 4 children. To illustrate the idea more clearly, consider a parent-child group of size 8, whose elements 1–8 are arranged as shown in Figure 1a. We shall use the parenthesis notation to describe configurations, namely, by $(1234)(5678)$, we mean that 1–4 share a common value while 5–8 share a different common value. Then the probability of the configuration $(1234)(5678)$, under the CRP is given by

$$\frac{M}{M} \times \frac{1}{M+1} \times \frac{2}{M+2} \times \frac{3}{M+3} \times \frac{M}{M+4} \times \frac{1}{M+5} \times \frac{2}{M+6} \times \frac{3}{M+7}.$$

In a similar way, the tying probabilities for a parent-child group consisting of 4 children (see Figure 1b) can be described.

While actually assigning prior probabilities on configurations, we slightly modify the CRP probabilities to address the following appealing principle for images: In a given

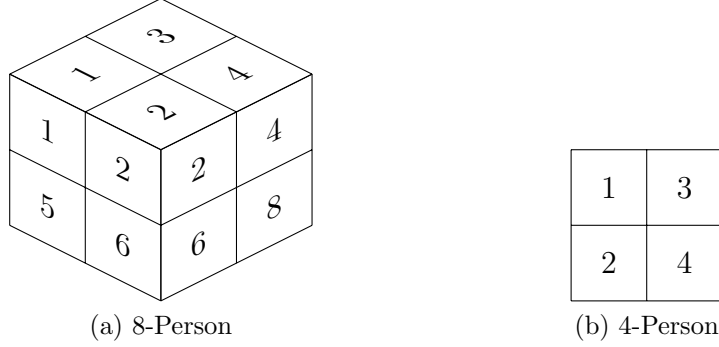


Figure 1: Parent-child groups

parent-child group, clusters can only be formed by children that are all adjacent to one another. Thus for a 4-member parent-child group, 1 cannot be tied with 4 without also being tied with either 2, or 3 or both. Similarly in an 8 member parent-child group, a tie among only 1, 4 and 8 only is prohibited. This reduces the number of configurations to 12 from 15 in a 4-member parent-child group, and from 4,140 to 958 in an 8-member parent-child group. Although this is still a large number of configurations in the latter case, the restriction nevertheless reduces the number by nearly a factor of 5, has more intuitive appeal and allows a more stable estimation of smoothing parameters. After removing the non-complying configurations, the remaining possible configuration probabilities are re-scaled to add up to 1.

Once a configuration \mathcal{C} has been chosen by the prior in a (l, j, k) parent-child group in level s , conditional on \mathcal{C} , a Dirichlet distribution is placed on the cluster sums of the level s relative intensity values $(q_{s,(l,j,k)}^{(1)}, \dots, q_{s,(l,j,k)}^{(g)})$ with dimension equal to the number of groups g and parameters equal to the vector of cardinalities \aleph_{C_m} of these

clusters C_m , $m = 1, \dots, g$:

$$P(q_{s,(l,j,k)}^{(1)}, \dots, q_{s,(l,j,k)}^{(g)} | \mathcal{C} \in \mathcal{C}) \sim \text{Dirichlet}(g; \aleph_{C_1}, \dots, \aleph_{C_g}). \quad (5)$$

This places equal weight on each child *a priori* in a parent-child group regardless of the configuration, thus demonstrating objectiveness of the resulting prior on the intensities. Note that if two relative intensity values $\rho^{(1)}$ and $\rho^{(2)}$ share a common value and sum to $q^{(1)}$, then $\rho^{(1)} = \rho^{(2)} = q^{(1)}/2$.

We do not impose a prior distribution on the total intensity $\lambda_{0,(1,1,1)}$ but set it to $X_{0,(1,1,1)}$. This preserves the total flux of the observed image in its denoised form.

2.2. Posterior Distributions

With the specification of the prior distributions, the posterior mean of the configuration given the observed data can be obtained. For each configuration, a generic expression is available for any parent-child group. Let \mathbf{Y} represent the vector of photon counts of children in a parent-child group and $\sum \mathbf{Y}$ stand for the sum of their components. Then the posterior distribution of the configuration is given by

$$P(\mathcal{C}|M, \mathbf{Y}) \propto P(\mathcal{C}|M)P(\mathbf{Y}|\mathcal{C}, \sum \mathbf{Y}), \quad (6)$$

where $P(\mathcal{C}|M)$ is given by the modified CRP. Let g be the number of clusters formed by \mathcal{C} , \mathbf{q} the vector of cluster-sums of the level-wise relative intensity parameters and $\boldsymbol{\rho}$ the level-wise relative intensity parameters obtained from \mathbf{q} . Further let $\sum \mathbf{Y}!$ denote

the sum of the factorial of each component of \mathbf{Y} and let $\sum(\log(\mathbf{Y}!))$ denote the sum of log-factorial of each component of \mathbf{Y} . Then we may write

$$P(\mathbf{Y}|\mathcal{C}, \sum \mathbf{Y}) \propto \int_{\Delta_g} P(\mathbf{Y}|\mathcal{C}, \sum \mathbf{Y}, \boldsymbol{\rho})P(\mathbf{q}|\mathcal{C})d\mathbf{q}, \quad (7)$$

where Δ_g stands for the unit simplex in \mathbb{R}^g . This integration is performed to get a generic closed form solution given by

$$\exp \left(\log \Gamma(d) - \sum_{i=1}^g (\log \Gamma(\aleph_i)) + \log(\sum \mathbf{Y}!) - \sum \log(\mathbf{Y}!) \right. \\ \left. - \log \Gamma(\sum \mathbf{Y} + d) - \sum_{i=1}^g (S_i \log \aleph_i) + \sum_{i=1}^g \log(\Gamma(S_i + \aleph_i)) \right),$$

where S_i is the sum of counts of members of the i th cluster, \aleph_i is the cardinality of the cluster and d is either 8 or 4, depending on whether or not the third dimension is split in that parent-child group. Multiplying the above expression by the modified CRP probabilities and renormalizing, posterior probabilities of each possible configuration \mathcal{C} are obtained.

The posterior distribution of $\boldsymbol{\rho}$ is a mixture of the posterior distributions obtained from those on \mathbf{q} given \mathcal{C} . The posterior distribution of \mathbf{q} , conditional on \mathcal{C} , is the conjugate Dirichlet distribution

$$P(\mathbf{q}|\mathbf{Y}, \mathcal{C}) = \text{Dirichlet}(g; S_1 + \aleph_1, \dots, S_g + \aleph_g).$$

Combining with the expression for $P(\mathcal{C}|\mathbf{X})$, this gives an analytically computable

expression for $E(\boldsymbol{\rho}|\mathbf{X})$. Now applying (4), $E(\lambda_{(l,j,k)}|\mathbf{X})$ will be obtained. Note that $E(\lambda_{0,(1,1,1)}|\mathbf{X}) = X_{0,(1,1,1)}$ since we did not assign a prior distribution for $\lambda_{0,(1,1,1)}$, the overall intensity. This will provide point estimates of the underlying intensities $\lambda_{(l,j,k)}$ giving denoising of the observed image $X_{(l,j,k)}$ at each voxel (l, j, k) .

Due to the nature of the recursive partitions, there will be staircase-like artifacts created if the denoising method is performed only once. A standard solution is to “cycle-spin” the image to remove these artifacts. This requires multiple iterations of a computing algorithm where the original image is circularly shifted, denoised, and then inversely shifted to obtain an estimate. An average over these circular shifts will remove the staircase-like artifacts. Details of this process are outlined in Coifman and Donoho (1995). In the present context, because of the added variation in the third dimension, cycle spinning must include averaging over circular shifts along that third dimension as well as the other two dimensions.

2.3. Smoothing Parameters

The only smoothing parameter necessary in this algorithm is a parameter M for each scale controlling the extent of ties in a CRP. Intuitively, within each level, the same amount of smoothing is reasonable, while the extent of smoothing should increase at the finer levels. Thus, we shall use a common value of M in each level, but allow the value to vary over different levels. We determine the value of M for each level by maximizing the marginal probability of obtaining the given data as a function of M . Due to the multi-scale structure, each maximization can be carried out separately.

Note that the marginal probability of obtaining the given sample values in a generic parent-child group is given by

$$\sum_{\mathcal{C} \in \mathcal{C}} P(\mathcal{C}|M)P(\mathbf{Y}|\mathcal{C}, \sum \mathbf{Y}). \quad (8)$$

The expressions for the CRP probabilities of different configurations are explicit function of M while $P(\mathbf{Y}|\mathcal{C}, \sum \mathbf{Y})$ was obtained in the previous section. The marginal probability of each parent-child group can be calculated separately and then multiplied by the other probabilities of each parent-child group in the same level to obtain the overall marginal probability for that level. Indexing different parent-child groups in the s th level by z , the logarithm of the marginal probability of obtaining the given sample at the s th level is therefore

$$\sum_{z=1}^{d^s-1} \log \left(\sum_{\mathcal{C}_z \in \mathcal{C}} P(\mathcal{C}_z|M)P(\mathbf{Y}|\mathcal{C}_z, \sum \mathbf{Y}) \right),$$

which would be maximized to obtain the common value of the smoothing parameter M at the s th level using the Newton-Raphson method.

Intuitively, the smoothing parameter M should become smaller as the scale becomes finer. To comply with that, we impose a monotonicity constraint when optimizing with respect to M . This is a very effective tool for reducing overfitting and it also leads to more numerically stable maximization.

The proposed denoising method has some appealing asymptotic properties. If the exposure time increases indefinitely so that the total photon count approaches infinity,

the posterior distribution of the relative intensity parameters concentrate around their true values. Moreover, the chance that the posterior will falsely determine a structure in the image decays to zero exponentially fast and the chance of missing a genuine structure also decays to zero polynomially. This properties can be shown easily by following the arguments given in the proof of Theorem 1 of White and Ghosal (2011) verbatim.

3. Simulations

In order to perform a simulation study, we use as the true image a three-dimensional Shepp-Logan phantom image and simulate photon count data using Poisson sampling. In two dimensions, the Shepp-Logan phantom is a commonly used simulated image to test image processing methods. It was adapted to a three-dimensional approach in Schabel (2006). This image creates ellipsoids in three dimensional space with values that are used as the underlying intensities.

Since it is difficult to view a three-dimensional image, slices of the third dimension can be examined to show the denoising performed by the algorithm. In the following figures, we show slices of the third dimension at the halfway and the first quarter stage, respectively. To compare the accuracy of the denoising method, we use two metrics — the mean absolute difference (MAD) defined by

$$\text{MAD} = 2^{-2S-S^*} \sum_{l=1}^{2^S} \sum_{j=1}^{2^S} \sum_{k=1}^{2^{S^*}} |\hat{\lambda}_{(l,j,k)} - \bar{\lambda}_{(l,j,k)}|$$

and the root mean squared error (RMSE) defined by

$$\text{RMSE} = \sqrt{2^{-2S-S^*} \sum_{l=1}^{2^S} \sum_{j=1}^{2^S} \sum_{k=1}^{2^{S^*}} |\hat{\lambda}_{(l,j,k)} - \bar{\lambda}_{(l,j,k)}|^2}$$

on the relative intensity values.

We compare the distance of the observed noisy image and the denoised image from the underlying true image. A third comparison is made with the image obtained by setting the values of the relative intensity parameter to zero for any voxel having less than 5% of the maximum voxel intensity estimate. The rationale behind this hard thresholding is that often a constant mild background noise is present in images, so the additional thresholding may get rid of the noise. A 20-shift cycle-spinning procedure was used and the whole experiment was replicated 5 times. Table 2 provides the error associated with each of these images. The MAD of the smoothed image is reduced to be about 28% of the MAD of the observed image when comparing each to the known true image. Similarly, the RMSE is reduced to be about 35% of that of the observed image. After removing intensity estimates less than the 5% threshold, the MAD is now reduced to be about 25% while the RMSE is only barely reduced further. Although the thresholding reduces the error, a more formal approach to that problem could be necessary. When the background is removed in this fashion the Monte-Carlo standard errors increase. Figure 2 show a simulated phantom image that is $128 \times 128 \times 128$ with a maximum intensity of 25.

Table 2: Simulation results for $128 \times 128 \times 128$ Shepp-Logan phantom with 20 cycle-spins.

Results after 5 simulations, Maximum intensity = 25

Method	MAD($\times 10^{-7}$) (SE)	RMSE($\times 10^{-6}$) (SE)
Observed Image	1.28 (0.001)	3.44 (0.002)
Bayesian CRP	0.36 (0.001)	1.20 (0.006)
Bayesian CRP Threshold	0.32 (0.002)	1.19 (0.013)

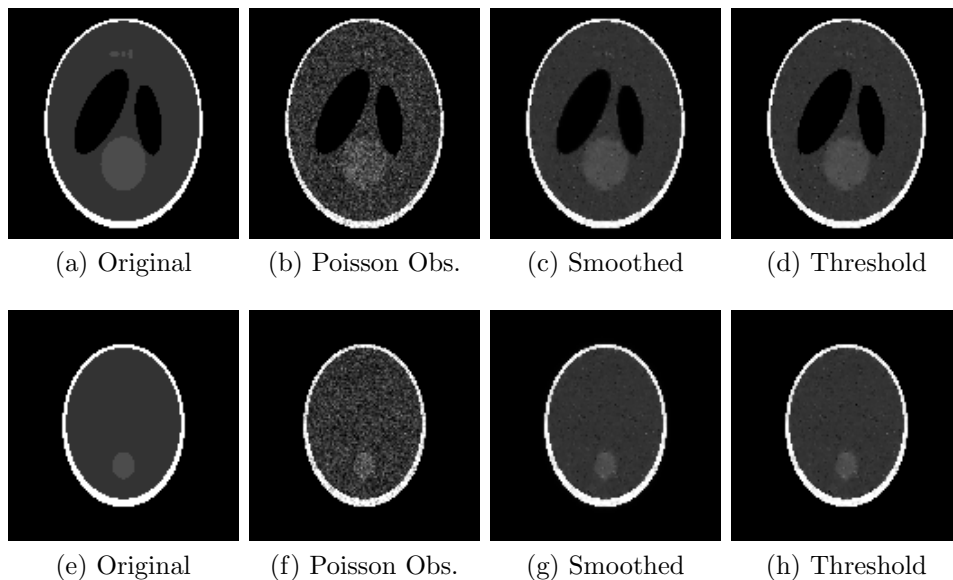


Figure 2: Simulation images of Shepp-Logan 3D phantom with two different three-dimensional slices.

4. Astronomical Example

When capturing high energy photons using a detector in space, each photon’s energy level is also recorded by the device. These values are binned in channels to give counts like the two spatial dimensions. For instance, an X-ray observatory may use 1024×1024 channels to capture spatial information of the photon sources and further 1024 channels for the spatial dimension.

The image shown in Figure 3 is the G1.9+0.3 supernova remnant. On the left, the

observed noisy image collected by the Chandra X-ray observatory is shown. The count data is based on binning in $256 \times 256 \times 128$ channels. To view the three-dimensional image, we bin all energy spectra into three distinct bands that are given different colors: red, green and blue in analogy with the energy spectrum of visible light. The colors in the energy spectra were assigned by looking at the histogram over the energy spectrum which appeared to be trimodal.

The image in Figure 3 was denoised using the proposed method. We notice that the denoised images are much more detailed with clearer boundaries than the originally observed image. There are many different ways these images can be examined. Slices of the images from a spatial perspective can be examined for spectral characteristics. Slices of the spectral dimension can be taken to view the physical shape of the image at those given spectral levels. Below, we display three spectral slices of the image corresponding to the three colors.

Since the spectral dimension is half of each physical dimension, the image was denoised using both approaches described in Table 1. The denoised image using the coarser smoothing method is further separated in the three different colors and the corresponding images are shown in Figure 4. These images carry valuable information about the spatial distribution of energy in G1.9+0.3 and hence important clues about the formation of the supernova.

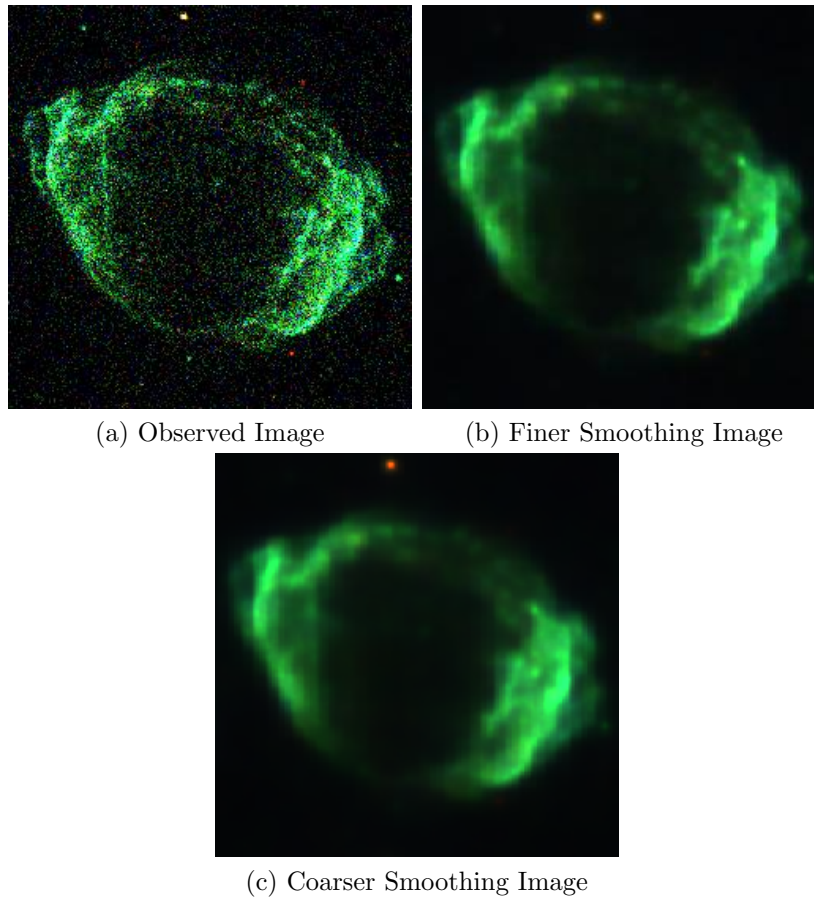


Figure 3: G1.9+0.3 with three energy bands in colors.

References

- Coifman, R. R., Donoho, D. L., 1995. Translation-invariant de-noising. In: Antoniadis, A., Oppenheim, G. (Eds.), *Wavelets and Statistics*. pp. 125–150.
- Crouse, M., Nowak, R., Baraniuk, R., 1998. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing* 46.
- Ghosh, S., Ungureanu, A. B., Sudderth, E. B., Blei, D. M., 2011. Spatial distance dependent chinese restaurant process for image segmentation, preprint.

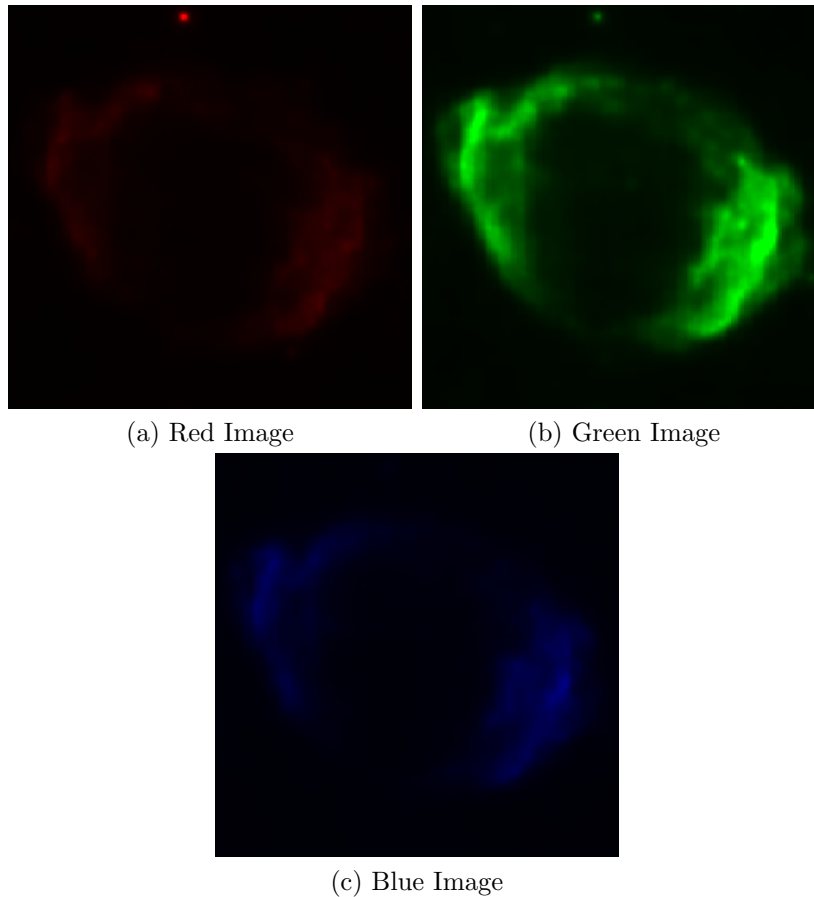


Figure 4: G1.9+0.3 coarser smoothed image split into three separate colors.

Kolaczyk, E. D., 1999. Bayesian multiscale models for Poisson processes. *Journal of the American Statistical Association* 94, 920–933.

Kolaczyk, E. D., Nowak, R. D., 2004. Multiscale likelihood analysis and complexity penalized estimation. *Annals of Statistics* 32, 500–527.

Krishnamurthy, K., Raginsky, M., Willett, R., 2010. Multiscale photon-limited hyperspectral image reconstruction. *SIAM Journal of Imaging Sciences* 3.

Pitman, J., 1995. Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* 102, 145–158.

Schabel, M., 2006. 3D Shepp-Logan phantom. Mathworks.

Starck, J., Murtagh, F., 2006. Astronomical Image and Data Analysis, 2nd Edition. Astronomy and Astrophysics Library. Springer.

White, J. T., Ghosal, S., 2011. Bayesian smoothing of photon limited images with applications in astronomy. *Journal of the Royal Statistical Society, Series B* 73, 579–599.

Willett, R. M., Nowak, R. D., 2003. Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *IEEE Transactions On Medical Imaging* 22, 332–350.