MULTISCALE INTENSITY ESTIMATION FOR MARKED POISSON PROCESSES

Rebecca Willett

Department of Electrical and Computer Engineering Duke University, Durham, NC 27708

ABSTRACT

Inference on severely data-starved Poisson processes can be dramatically improved by using auxiliary information about measured discrete events in the form of "marks". Marks are widely available in many applications, and can take the form of photon energy, time delay information, packet size, or other forms of characterizations. Effectively using marks results in innovative signal processing methods and dramatic error reductions for Poisson intensity estimation. The efficacy of the proposed method is demonstrated in the context of photon-limited spatio-spectral intensity estimation.

Index Terms— Poisson processes, Wavelets transforms, Multidimensional signal processing

1. MARKED POISSON PROCESSES

Marked Poisson processes arise when indirect observations of a complex physical phenomenon are collected by counting discrete events and measuring an associated auxiliary feature vector for each event. Such information is available in fluorescence microscopy, hyperspectral imaging, solar spectroscopic imaging, and other applications. Multiphoton laser scanning microscopy (MPLSM), for example, is a cutting-edge tool for high-resolution imaging of living tissue and organisms used to study cellular dynamics [1]. The result of this powerful technique is a series of detected photons resulting from fluorescence events occurring at each raster laser scan location. For each observed photon, we also have access to the time at which it was observed, its characteristic emission spectrum, and its characteristic excited state lifetime. The challenge here is to use small numbers of random, discrete, marked events such as these to perform accurate inference on the underlying phenomenon.

The marks associated with each observed event, while relevant to application-specific objectives, do not *directly* lead to improved reconstruction capabilities in these data-starved scenarios. When the number of observed events is severely limited, our ability to perform very high-resolution or highdimensional reconstruction can rapidly become critically impaired. This is often counteracted by reducing the dimensionality of the problem in a linear, highly suboptimal manner via binning observations spatially and/or regardless of their marks. This approach runs counter to the entire multiscale philosophy of using the data to determine the optimal spatially-varying resolution of the estimate. However, as this paper demonstrates, it is possible to use marks in a principled, non-linear, and non-parametric framework to dramatically improve reconstruction performance. In particular, reconstruction is often most accurate when prior knowledge of the underlying phenomenon's smoothness or other characteristics are incorporated in the form of a regularization term or prior probability model. However, standard regularization techniques do not easily incorporate key information contained in the mark of each observed event, and hence it is typically discarded to the detriment of reconstruction performance.

In this paper we focus on one example of marked Poisson processes: photon-limited spatio-spectral analysis. In a variety of application domains, sensors collect photon-limited observations of the intensity of interest, which has threedimensions - two spatial, and one spectral. (In some contexts these observations may be indirect, such as tomographic projections. We will focus on the denoising problem in this paper, with the understanding that it can be incorporated into an Expectation-Maximization framework for solving inverse problems [2].) Binning the observed photons into small spatial and spectral bins results in a data cube which has zero observations in almost all bins. Performing reconstruction on this data by generalizing image-based methods to multiple dimensions typically results in regularization terms which treat all dimensions as equal. The structures prevalent in spatial dimensions, however, are rarely the same as structures prevalent in spectral dimensions. Thus, specialized regularization methods are required for accurate and effective analysis.

2. MULTISCALE SPATIO-SPECTRAL INTENSITY ESTIMATION

Assume that we measure Poisson observations of a spatiallyand spectrally-varying intensity over a relatively small time window, such that the number of observed photons is small. In particular, let f denote the $N \times N \times M$ true threedimensional intensity, where the first two dimensions correspond to the spatial location, and the third dimension corresponds to the marks, which in this case are photon wave-

The author was partially supported by DARPA, grant no. HR0011-06-1-0026.

lengths. Our observations are of the form

$$\boldsymbol{x} \sim \operatorname{Poisson}(\boldsymbol{f}),$$

where x has the same dimensions as f. Our goal is to infer f from the measurements x as accurately as possible, effectively utilizing the mark information contained in the third dimension. One approach towards analyzing this threedimensional data cube would be to independently reconstruct an image using the method described above for each spectral band *independently*. Such a method, however, ignores correlations between spectral bands, so that when spectral bands are narrow and photon counts are small in each band, the reconstruction error is unacceptably high, as we will see in Section 3.

The approach described in this paper will be referred to as a hereditary TI-Haar method because it is Translation-Invariant and roughly based upon denoising Haar wavelets with a hereditary constraint. Taking advantage of correlations in the data between both marks and spatial locations, the proposed method entails the following two stages:

- **Stage 1:** Perform hereditary TI-Haar Poisson intensity estimation in the spatial dimensions, with each leaf of the resulting unbalanced quad-tree decomposition corresponding to a spectra (or series of marks).
- Stage 2: Smooth each spectrum (or mark dimension) using 1D hereditary TI-Haar Poisson intensity estimation.

We will elaborate upon both these stages below.

2.1. Stage 1

In the first stage we compute an initial estimate by determining the ideal partition of the *spatial* domain of observations (assumed to be $[0, 1]^2$) and use maximum likelihood estimation to fit a *single, mean* spectrum to each square in the optimal partition. The space of possible partitions is a nested hierarchy defined through a recursive dyadic partition (RDP) of $[0, 1]^2$, and the optimal partition is selected by pruning a quad-tree representation of the observed data. (This quad-tree is referred to as a complete RDP.) This gives our estimators the capability of spatially varying the resolution to automatically increase the smoothing in very regular regions of the intensity and to preserve detailed structure in less regular regions.

In general, the RDP framework leads to a model selection problem that can be solved by a tree pruning process. Each of the terminal squares in the pruned spatial RDP could correspond to a region of intensity which is spatially homogeneous or smoothly varying (regardless of the regularity or irregularity between the spectral bands or marks). Such a partition can be obtained by merging neighboring squares of (*i.e.* pruning) a complete RDP to form a data-adaptive RDP \mathcal{P} and fitting spectra to the data on the terminal squares of \mathcal{P} . Specifically, given the partition \mathcal{P} , $f(\mathcal{P})$ can be calculated by finding the average spectrum fit to the observations over each cell in \mathcal{P} . Thus a Stage 1 intensity estimate, \hat{f} , is completely described by \mathcal{P} .

This provides for a very simple framework for penalized likelihood estimation, wherein the penalization is based on the complexity of the underlying partition [3]. The goal here is to find the partition which minimizes the penalized likelihood function:

$$\widehat{\mathcal{P}} \equiv \arg\min_{\mathcal{P}} \left[-\log p(\boldsymbol{x} \mid \boldsymbol{f}(\widehat{\mathcal{P}})) + \operatorname{pen}(\widehat{\mathcal{P}}) \right]$$

$$\widehat{\boldsymbol{f}} \equiv \boldsymbol{f}(\widehat{\mathcal{P}})$$
(1)

where

$$p(\boldsymbol{x}|\boldsymbol{f}) = \prod_{i,j=0}^{N-1} \prod_{k=0}^{M-1} \frac{e^{-f_{i,j,k}} f_{i,j,k}^{x_{i,j,k}}}{x_{i,j,k}!}$$

denotes the likelihood of observing x given the image estimate $f(\hat{\mathcal{P}})$ and where pen $(\hat{\mathcal{P}})$ is the penalty associated with the estimate $f(\hat{\mathcal{P}})$. We penalize the estimates according to a codelength required to uniquely describe each model with a prefix code; the penalties are discussed in detail in [3, 4]. The resulting estimator \hat{f} is referred to as the penalized likelihood estimator (PLE).

This approach is very similar to the image estimation method described in [3, 5], with the key distinction that *Stage I as described above forces the spatial RDP to be the same at every spectral band*. This constraint makes it impossible for the method to perform spatial smoothing at some spectral bands but not others. In other words, when a tree branch is pruned in the proposed framework, it means spatial cells are merged in every spectral band simultaneously at the corresponding spatial location. This approach is effective because an outlier observation in one spatio-spectral voxel may not be recognized as such when spectral bands are considered independently, but may be correctly pruned when the entirety of the corresponding spectrum is very similar to spatially nearby spectra.

2.2. Stage 2

While Stage 1 effectively used mark information to constrain spatial smoothing, it did not perform any smoothing between different marks. If smoothness in the mark dimension is not an appropriate *a priori* assumption, Stage 1 alone can produce accurate results. However, in many applications smoothness or piecewise smoothness in the mark dimension is very reasonable, and Stage 2 exploits that smoothness to improve the accuracy of the result of Stage 1.

In particular, Stage 2 consists of pruning an RDP representation of the spectrum at each spatial location *independently*. This procedure, detailed in [4], is highly analogous to the procedure described in Stage 1, again using a minimum description length/coding theoretic approach to regularization [3]. For each spatial location, we build and prune a complete RDP of the spectrum. In this case, the RDP takes the form of a binary tree representation, and the model fit to each terminal cell in the RDP is simply the empirical mean of the spectrum across the corresponding spectral bands.

2.3. Translation Invariance

The accuracy of these estimates can be augmented by a process called cycle-spinning, or averaging over shifts, resulting in translation-invariant (TI) estimates [6]. Cycle-spinning, as originally proposed, requires $O(N^2M \log NM)$ operations, but was derived in the context of undecimated wavelet coefficient thresholding in the presence of Gaussian noise, and is difficult to implement efficiently in the case of Poisson noise. The above multiscale tree-pruning methods can be modified to produce the same effect by averaging over shifts, but the increase in quality comes at a high computational cost; naïve algorithms require $O(N^4M^2)$ operations. Novel computational methods, however, as described in [5], can be used to yield TI-Haar tree pruning estimates in $O(N^2M \log NM)$ time.

3. EXPERIMENTAL RESULTS

As described above, while the first stage of the proposed method exploits spatial homogeneities, the second stage is designed to take advantage of correlations between neighboring spectral bands. This significantly reduces over-smoothing across spatial boundaries and edges while improving the reconstruction of each spectrum. To see the effectiveness of the proposed method, consider a portion of the Shepp-Logan phantom; we can assign a spectrum to each unique intensity level in the phantom to get a phantom spatio-spectral intensity function, as displayed in Figure 1. In this figure, each different spectral region is assigned a label in the image to the left, and the corresponding spectra are plotted in the image to the right. Each spectrum is a mixture of five Gaussians with random amplitudes, means, and variances.



Fig. 1. Spatio-spectral intensity. (a) Labeled unique spectral regions in phantom image. (b) Labeled corresponding spectra.

The spatially-varying intensities for two representative spectral bands are displayed in Figure 2(a) and (e). Note the faint contrast between some of the features. Photon-limited

observations in the same two spectral bands are displayed in Figure 2(b) and (f); from this data, several small and/or faint features are not easily discernible by the eye. The data cube is $128 \times 128 \times 128$, and the total number of photons observed is 1, 445, 524, resulting in an average of 0.6893 photons per voxel. In this simulation study, the penalty term in (1) was multiplied by a scalar weight to improve empirical performance. The weight was not selected to minimize any particular error metric, but rather to yield generally accurate reconstructions.

If we were to reconstruct this data cube by performing hereditary TI-Haar *image* estimation (as described in [5]) on each spectral band, we would achieve the result displayed in Figure 2(c) and (g); the mean squared error associated with this imaging-based data cube estimate, denoted \hat{f}_{I} , is $\|\boldsymbol{f} - \hat{\boldsymbol{f}}_I\|^2 / \|\boldsymbol{f}\|^2 = 0.0868$. Spatio-spectral analysis offers dramatic advantages over processing individual spectral bands independently, as shown in Figure 2(d) and (h). The mean squared error associated with this marks-based data cube estimate, denoted \widehat{f}_M is $\|f - \widehat{f}_M\|^2 / \|f\|^2 = 0.0258$, significantly lower than the MSE for \hat{f}_{I} . This simulation result shows several spatial features, including faint boundaries and small-scale inhomogeneities, which are present in the true and spatio-spectrally estimated intensities (Figure 2(d) and (h)) but which are difficult or impossible to visualize in the noisy observations or independently processed spectral bands (Figure 2(c) and (g)). Finally, Figure 3 shows five representative spectra and their estimates, further demonstrating the strength of the proposed method. This experiment was repeated one hundred times. The average MSE of \widehat{f}_I was 0.0863 and the average MSE of \widehat{f}_M was 0.0261.

4. DISCUSSION

This paper demonstrates that the use of marks is highly useful in data-starved Poisson process analysis. The results are particularly relevant for photon-limited imaging systems in which we have access to supplementary information about each observed photon, such as its observation time, wavelength, emission spectrum, or excited state lifetime.

The proposed method consists of two key stages. During the first stage, we initially build a quad-tree decomposition of the data cube in the spatial dimension, so that each leaf node of the tree corresponds to a series of marks (e.g. a spectrum). We then prune back this tree using a penalized likelihood criterion. In contrast to processing the "image" at each mark (e.g. spectral band) independently, this method imposes the constraint that spatial smoothing be the same across all marks. This can then be augmented with a second stage, which consists of denoising the intensity across marks for each spatial location independently.

In the context of image estimation, multiscale methods based on translation-invariant (TI) Haar wavelets are near



Fig. 2. Spatio-spectral analysis results. (a) Spatial variation of intensity in 32^{nd} spectral band. (b) Poisson observations. (c) Hereditary Haar image reconstruction result, \hat{f}_I ; cube MSE = 0.0868. (d) Result after proposed reconstruction, \hat{f}_M ; cube MSE = 0.0258. (e) Spatial variation of intensity in 88th spectral band. (f) Poisson observations. (g) Hereditary Haar image reconstruction result, \hat{f}_I . (h) Result after proposed reconstruction, \hat{f}_M .



Fig. 3. Spatio-spectral analysis results. (a) True spectra at five different spatial locations. (b) Observed spectra at same locations. (c) Estimated spectra at same locations after performing intensity estimation on each spectral band independently, resulting in \hat{f}_I . (d) Estimated spectra at same locations after performing proposed spatio-spectral intensity estimation, resulting in \hat{f}_M .

minimax optimal reconstruction techniques [4]. This suggests that the proposed method potentially has similar optimality characteristics for some classes of marked Poisson processes. Theoretical analysis of the proposed method is an important component of our ongoing work.

5. REFERENCES

- W. Denk, J.H. Strickler, and W.W. Webb, "Two-photon laser scanning fluorescence microscopy," *Science*, vol. 248, no. 4951, pp. 73–76, 1990.
- [2] R. Nowak and E. Kolaczyk, "A multiscale statistical framework for Poisson inverse problems," *IEEE Trans. Info. Theory*, vol. 46, pp. 1811–1825, 2000.
- [3] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Stat.*, vol. 32, pp. 500–527, 2004.
- [4] R. Willett and R. Nowak, "Multiscale poisson intensity and density estimation," Submitted to IEEE Trans. Info. Th., 2005.
- [5] R. Willett and R. Nowak, "Fast multiresolution photon-limited image reconstruction," in *Proc. IEEE Int. Sym. Biomedical Imaging — ISBI '04*, 15-18 April, Arlington, VA, USA, 2004.
- [6] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.