Pixon-Based Image Restoration/Reconstruction

By R. C. PUETTER[1]

[1]Center for Astrophysics and Space Sciences,
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093-0111, USA
rpuetter@ucsd.edu

This lecture introduces pixon-based image restoration/reconstruction methods. The relationship between image Algorithmic Information Content and the Bayesian incarnation of *Occam's Razor* are discussed as well as the relationship of multiresolution pixon languages and image fractal dimension. Also discussed is the relationship of pixons to the role played by the Heisenberg uncertainty principle in statistical physics and how pixon-based image reconstruction provides a natural extension to the Akaike information criterion for Maximum Likelihood estimation. The lecture then discusses practical implementation of pixon-based reconstruction, and how these techniques can be applied to a variety of problems, including problems outside of astronomical imaging.

1. Introduction

As discussed in previous lectures of this series, while Goodness-of-Fit (GOF), or Maximum-Likelihood, and Maximum Entropy (ME) methods have improved dramatically on the performance of linear inversion methods for image deconvolution, they still suffer from signal correlated residuals and spurious source production. One of the fundamental sources of the problems with these methods is their selection of the coordinate system in which to represent the image. This lecture will introduce pixon-based methods which use an information theortic coordinate system which quantifies the image's Algorithmic Information Content (AIC), and uses this to optimally constrain the reconstruction process (see Piña and Puetter 1993, Puetter and Piña 1993a, Puetter and Piña 1993b, Puetter 1994, and especially the review article Puetter 1996a). We shall also see that pixons (or *informatons*–their natural extension to general data sets) are in effect quanta of information, and comprise the smallest collection of degrees-of-freedom (DOF) required to specify the data within the accuracy allowed by the noise.

2. Practical Methods for Implementing Pixons

How one might practically implement pixon-based methods is discussed in a number of previous papers. The material in this section follows closely the discussion in Puetter 1996b.

2.1. *A Multiresolution Pixon Language for Image Description*

In Lecture 2 we argued that a multiresolution image description language would be suitable (i.e. concise) for describing generic images. The success of multiresolution languages is not surprising. In fact, these ideas are familiar to most scientists and are fundamental to simple, well known concepts, e.g. the use of fine grain photographic film to capture fine detail. So our multiresolution pixon language aill use the idea that generic images can be concisely described by using fewer degrees-of-freedom (DOFs) per unit area in portions of the image which are smooth and a greater density of degrees-of-freedom where there is greater detail. Each of these degrees-of-freedom might then be likened to a single photographic grain or a generalized pixel. The value of this pixel represents the average brightness in a given region. This, in fact, is the origin of the name "pixon". Each pixon is a single DOF used to describe the image in a particular region. The "pix" part of the name recognizes its pixel heritage, while the "on" suffix recognizes its more fundamental nature (the pixon is fundamental to the image, not to the instrument that took the picture). However, since any scheme which controls the local density of DOFs used to describe the local image information is suitable for constraining the image reconstruction and optimizing the Bayesian prior, rather than using image signal contained in cells with hard boundaries, we have chosen to use a local correlation scale formulation to control the DOF density. We call these "fuzzy pixons". To formalize the definition, at each point in the image, $\vec{x}$, we write the image, I, as

$$I(\vec{x}) = (K \otimes I_{pseudo})(\vec{x}) \tag{2.1}$$

$$(K \otimes I_{pseudo})(\vec{x}) = \int dV_{\vec{y}}\; K(\vec{x}, \vec{y}, \delta(\vec{x})) I_{pseudo}(\vec{y}) \tag{2.2}$$

$$K\left(\vec{x}, \vec{y}, \delta(\vec{x})\right) = K\left(\frac{\|\vec{x}-\vec{y}\|}{\delta(\vec{x})}\right) \quad ; \text{ for radially symmetric pixons} \quad , \tag{2.3}$$

where equations (2.1) through (2.3) show that the image is a local convolution of a "pseudo-image" with a blurring function with a given local scale, $\delta(\vec{x})$. Note that the local scale varies with position $\vec{x}$ in the image and the integration in equation (2.2) is carried out over volume in pseudo-image space. We have also indicated in equation (2.3) that one suitable functional form for the pixon kernel function, K, might be a radially symmetric function that depends only on the distance between the kernel center and the image position relative to the local scale. We have found this functional form to work quite well for centrally peaked kernel functions with a finite foot-print. We normally use truncated paraboloids, i.e.

$$K\left(\vec{x}, \vec{y}, \delta(\vec{x})\right) = \begin{cases} \left(1 - \frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) / \int dV_{\vec{y}} \left(1 - \frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) & ; \; \|\vec{x}-\vec{y}\| \le \delta(\vec{x}) \\ 0 & ; \; \|\vec{x}-\vec{y}\| > \delta(\vec{x}) \end{cases} \quad . \tag{2.4}$$

Of course AIC concepts suggest that a richer language, e.g. elliptical pixons, would yield a more concise image description, which Occam's razor then says should have a more optimized prior. In fact, we have used elliptical pixons recently to perform some image restorations—see Lecture 4. Nonetheless, for generic images it seems clear that we have reached a point of diminishing returns, and it is unlikely that pixon kernels more complicated than ellipses are warranted for generic images.

2.2. *Solving for the M.A.P. Image/Model Pair*

Now that we have described a suitable language in which to solve the problem, we shall move on to the details of finding the M.A.P. (Maximum *A Posteriori*) image/model pair, i.e. the image/model pair that maximizes $p(I, M|D)$. To maximize $p(I, M|D)$, we need to maximize the product of $p(D|I, M)$ and $p(I, M)$. The common choice for $p(D|I, M)$ in the case of pixelized data with independent Gaussian noise is $p(D|I, M) = (2\pi\sigma^2)^{1/2n} \exp(-\chi^2/2)$ where χ^2 is the standard chi-squared value, [i.e. $\chi^2 = \sum_1^n (x_i - \langle x_i \rangle)^2/\sigma^2$], σ is the standard deviation of the noise, and n is the number of independent measure parameters (here the number of pixels). As already mentioned in Lecture 1, this choice of the Goodness-of-Fit criterion assumes that (1) within the chosen space of image/model pairs, a statistically acceptable fit to the data is obtainable, and (2) that near this solution, image/model space uniformly fills data space so that the resulting probability distribution of data realizations is dominated by the probability density of the Gaussian distributed measurement noise, i.e.

$$p(\text{Blurred Image+Noise}|I,M) = p(\text{Blurred Image}|\text{Noise},I,M)p(\text{Noise}|I,M) \quad (2.5)$$
$$\cong const \times p(\text{Noise}|I,M) \quad . \quad (2.6)$$

The "constan"t $p(\text{Blurred Image}|\text{Noise},I,M)$ is then generally dropped from the expression. This is generally not a bad an assumption, but it is by no means guaranteed, especially if the chosen space of image/models is selected to be restricted in some way.

While the functional form for the GOF criterion given above is almost always adopted, the various choices for $p(I,M)$ are far from standard. The choice of Maximum Entropy enthusiasts is something like

$$p(I|M) = \frac{N!}{n^N \prod_{i=1}^{n} N_i!} = e^S \quad (2.7)$$
$$p(M) = constant \quad , \quad (2.8)$$

where n is the number of pixels, N is the total number of counts in the image, N_i is the number of counts in pixel i, and S is the entropy. This choice is made on the basis of counting arguments. The arguments are basically sound, but operate within a fixed language, e.g. the pixel basis for the image. As discussed previously, a more appropriate choice for the image basis would be multiresolution pixons. This more critically models the data and provides a superior prior.

For our pixon prior we could use the same prior as used by ME workers with the pixons substituted for the pixels. The *a priori* probability arguments for the prior of equations (2.7)-(2.8) remain valid. It is just that we now recognize that the pixons are a more appropriate coordinate system and that we will use vastly fewer pixons (i.e. DOFs) to describe the image than there are pixels in the data. To make the formulae explicit, we define the pseudo-image on a psuedo-grid (we loosely refer to this as the "pixel grid" when talking about the pseudo-image) which is as least as fine as the data pixel grid (we use this finer "pixel" grid for the image too) and then use the following substitutions:

$$p(I|M) = \frac{N!}{n^N \prod\limits_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i!} \quad (2.9)$$

$$N = \sum_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i = \sum_{\substack{j=1, \\ j \,\epsilon\, Image}}^{n_{pixels}} N_j \quad (2.10)$$

$$N_i = \int dV_{\vec{y}} \; k(\vec{x},\vec{y},\delta(\vec{x})) I_{pseudo}(\vec{y}) \quad (2.11)$$

$$\prod_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i! = \prod_{\substack{j=1, \\ j \,\epsilon\, Image}}^{n_{pixels}} (N_j!)^{p_j} \quad (2.12)$$

$$p(\vec{x}) = 1 \Big/ \int dV_{\vec{y}} \; k(\vec{x},\vec{y},\delta(\vec{x})) \quad (2.13)$$

$$p(M) = constant \quad (2.14)$$

where N_i is the number of counts in pixon i, N_j is the number of counts in pixel j, p is the pixon density, i.e. the number of pixons per pseudo-pixel (in image space), and $k(\vec{x}, \vec{y}, \delta(\vec{x}))$ is the pixon shape kernel normalized to unity at $\vec{y} = \vec{x}$. [Note that while the formula of equation (2.9) can be easily justified, it is not the only justifiable expression. Other expressions might be more appropriate in certain situations.]

To obtain the M.A.P. image/model pair, one can now proceed directly by minimizing the product of $(2\pi\sigma^2)^{1/2n_{pixels}} \exp(-\chi^2/2)$ and equation (2.9) with respect to the local scales, $\{\delta(\vec{x_j})\}$—the pixon map, and the pseudo-image values, $\{I_{pseudo,j}\}$. However, this is not what we do in practice. Instead, we divide the problem into a sequential chain of two repeated steps: (1) optimization of the pseudo-image with the local scales held fixed, (2) optimization of the local scales with the pseudo-image held fixed. The sequence is then repeated until convergence–see Figure 1. To formally carry out this proceedure, in step (1) we should find the M.A.P. pseudo-image, i.e. the pseudo-image that maximizes $p(I|D, M)$—note that we are using the notation here that the local scales belong to M, while the pseudo-image values are associated with I. In step (2) we would then find the M.A.P. model, i.e. the scales that maximize $p(M|D, I)$.

While the above proceedure is quite simple, we have made still further simplifications. In neither step do we evaluate the prior. We simply evaluate the GOF term $p(D|I, M)$. So in step (1) we essentially find the Maximum Likelihood pseudo-image with a given pixon map. In step (2) we must take into account some knowledge of the pixon prior, but we simply use the fact that the pixon prior of equation (2.9) or any sensible prior increases rapidly as the number of pixons are decreased. So at each pseudo-grid point, j, we attempt to increase the local scale until it is no longer possible to fit the data within the noise. In other words at each pseudo-grid point we progressively smooth the current pseudo-image until the GOF criterion is "violated". We then select the largest scale at each point which was acceptable and use these values in the next iteration of step (1).

There is one more practical matter to consider. As with any interative method, there can be convergence problems. With the approach outlined above, we have noticed that if small scales are allowed early in the solution, then these scales become "frozen-in", even if they would have later proved inappropriately small. To solve this problem, we start out the pseudo-image calculation with initially large pixon scales. We then use our pixon-calculator [the code that performs step (2)] to determine new scales. The pixon-calculator, of course, will report that over some of the image the initial scales are fine, but over other parts of the image smaller scales are required. At this point, however, we do not allow the smallest scales requested. Instead, we let the scales get somewhat smaller over the portion of the image for which smaller scales were requested and proceed to step (1) of the next iteration. We repeat this process, letting the smallest scales allowed get smaller and smaller until the method converges. This procedure has proven to be very robust.

3. The Relationship of Pixon Concepts to Other Fields

Pixon-based concepts have a direct relationship to other fields of thought. Most of the subject matter in this section has already appeared in print (Puetter 1996a), but is repeated here for completeness.

3.1. *The Pixon and Fractal Dimensional Concepts*

Previous studies have introduced several different pixon schemes. The first scheme proposed (Piña and Puetter 1993) was called the Uniform Pixon Basis (UPB) since the AIC of each pixon was equal. The second scheme (Puetter and Piña 1993a) was called the

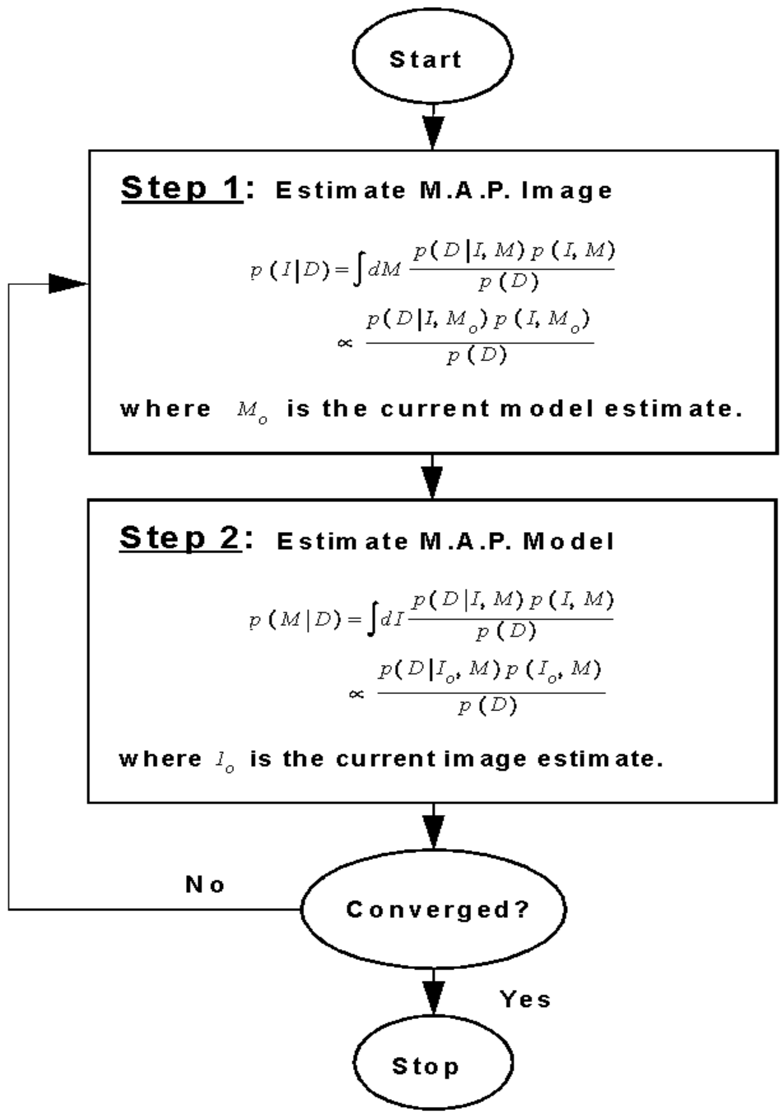

FIGURE 1. Iterative scheme for obtaining the M.A.P. image/model pair. This figure shows the pixon algorithm as currently implemented. The procedure is divided into two steps: (1) GOF optimization step in which the image is optimized while holding the pixon map (local smoothness scales) fixed, (2) Pixon optimization step in which a new pixon map is calculated while hold the current image estimate fixed. The method alternatively iterates between these two steps until convergence is obtained.

Fractal Pixon Basis (FPB). In fact, the pixon approach described in section 2 is essentially that used when calculating an FPB basis. In the FPB approach the pixon kernel function is radially symmetric. Because the FPB approach analyzes the AIC of the image in terms of a single spatial scale, it is conceptually similar to the definition of fractal dimension. While there are numerous definitions for the fractal dimension of a geometric object, all definitions have one thing in common. Each calculates a quantity that changes as a scale (or measurement precision) changes. For example, the compass dimension (also referred to as the divider or ruler dimension) is defined in terms of how the measured length of the line varies as one changes the length of the ruler used to make the measurement. The commonly used box-counting dimension is defined in terms of how many cells contain pieces of a curve as the sizes of the cells are changed. As can be seen from the above discussion, these ideas are closely related to the pixon representation since in order to calculate the local pixon scale, we ask: How does $p(I, M|D)$ change as the local size of the pixons are varied? In other words, $p(I, M|D)$ acts as our measure of length and the pixon scale acts as our ruler. (This is only true for radially symmetric pixons in which the only variable quantity is scale. For elliptical pixons, for example, we would also be interested in how $p(I, M|D)$ changes as the ellipticity and major axis angle changes.) So in this sense radially symmetric pixon kernel methods share much in common with fractal concepts. They each analyze and quantify geometric structure by how things change when varying a single spatial scale.

3.2. *Pixons and Data Compression*

Since pixons quantify the AIC of an image (or in the general case a data set), they are related directly to data compression. In fact, once a language has been chosen, the pixon representation is the most concise description of the image. Hence pixons provide optimal data compression. Indeed, it is precisely the fact that pixons are the natural and most concise coordinate system for the image, that gives pixon-based methods their computational power to provide an optimal solution to the inverse problem. So in effect, the image reconstruction and image compression problem are intimately related. Both can be optimized by using Bayesian techniques to maximize the fidelity to the data, while simplifying the model with which to represent it.

When one discusses image (or data) compression schemes, it is also of interest to understand whether or not the process is lossless or lossy. As should be clear, any data compression scheme based on pixon bases can be as lossless as desired. To adjust the fidelity to the data one need only adjust the GOF criterion. Such an adjustment will allow a uniform degradation or increase in the information content over the entire image. Alternatively, if it was desired to preserve certain sections of the data with higher fidelity, one need only to express this fact in the GOF criterion. Hence pixon-based methods provide an optimal method for data compression. As before, however, selection of a language suitable for the compression is still a key issue. Nonetheless, pixon-based compression with fuzzy, radially symmetric (or perhaps elliptical) pixons should produce excellent results for generic images.

3.3. *Pixons and Wavelets*

A current popular method for performing both image compression and image reconstruction is to use wavelets (see, for example, Press *et al.* 1994). Since the time a number of years ago when wavelets were first introduced, the theory of wavelet transforms has blossomed. This is with good reason. Wavelets provide many practical advantages over straight Fourier methods. Nonetheless, from the Bayesian image reconstruction theory developed above, it should be clear that the performance of wavelet data compression and image reconstruction will be inferior to pixon-based methods. There are several ways to see this. First, since standard wavelet bases are orthogonal, the basis functions have positive and negative excursions. Thus in order to construct part of a positive definite image, a number of wavelet components are required. This violates the principle of Occams Razor. Such a representation cannot hope to be minimal. Hence its Bayesian prior will be inferior to the pixon prior and it will be less likely from a probabilistic point of view. If this is not satisfying enough, the additional degrees of freedom represented by the many wavelet components needed to specify a local bump in the image may be inadequately constrained by the data. As with the pixel basis of more standard methods, this will give rise to spurious sources and signal correlated residuals. Hence pixons in any language which has positive definite pixon shape functions will provide a more optimal (i.e. more concise) description of the image.

3.4. *The Pixon and the Akaike Information Criterion*

In the 1970s, Akaike (Akaike 1973) introduced a penalty function for model complexity for maximum likelihood fitting. This formulation has come to be known as the Akaike Information Criterion (see Akaike 1973; Akaike 1977; Hall 1990; Veres 1990). The Akaike Criterion (AC) takes the log-likelihood of the fit and subtracts a term proportional to the number of parameters used in the fit. In this sense, the AC acts in the same manner as the pixon prior, i.e. acts as an *Occams Razor* term and works for the cause of simple models. One problem with the AC approach is that it is rather ad hoc. Each new parameter that is added to the model invokes an identical penalty independent of the innate merit of the parameter. In effect, the AC criterion uses a uniform prior for each new variable. This, however, often has serious flaws. For example, if it is known that one is fitting a data set that can be described by polynomial dependence on the variable, then introducing a new polynomial power in to the set of basis functions should be viewed with a different weight than adding an exponential function to the basis set. Furthermore, the AC method gives no suggestion as to the appropriate model variables that should be used. A pixon-based Bayesian approach does both, i.e. (1) an appropriate selection of the prior [e.g. that of equations (2.9)-(2.14)] invokes different penalties for each parameter of the model, and (2) pixon-based methods suggest exactly which DOFs are required to most succinctly model the data. In this sense, pixon-based methods are a direct generalization of the Akaike Information Criterion.

3.5. *Relationship to Statistical Mechanics/Heisenberg Uncertainty Principle*

The pixon has a very close connection to the concept of "coarse graining" in physics. Pixons, in fact, directly describe the "natural graininess" of the information in a data set due to the statistical uncertainties in the measurement. This also has a direct relationship to the role of the Heisenberg uncertainty principle in statistical mechanics. As is well known, the description of a system in statistical mechanics introduces a phase space to describe the state of a system. This is akin to the "language", basis, or coordinate system used to perform image reconstructions. Phase space in statistical mechanics is what is used to specify information about the system in the AIC sense, i.e. to completely specify the state of the system. Furthermore, all of the methods of statistical mechanics are in essence Bayesian estimation applied to physical systems. The partition functions used to make statistical predictions about the system are uniform priors (each volume of phase space is given equal probability), while the GOF probability distributions are normally taken to be delta functions since all of the macroscopic parameters of the system are usually given and assumed to be known exactly. This means that the state of the system is localized to a hyper-surface in phase space, e.g. that of constant energy, temperature, or particle number. Schematically, this might be written

$$p(Prop|Sys) \propto p(Sys|Prop)p(Prop) \quad , \tag{3.15}$$

where we have used the short hand notation that Sys stands for the state of the system and $Prop$ stands for the set of system properties in which one is interested. If, then, $Prop = \{T, E, N, . . .\}$, i.e. a set of macroscopic variables that might include temperature, T, total energy, E, total particle number, N, etc., then specification of the system temperature, $T_\circ$, would be equivalent to specifying the GOF term as

$$p(Sys|Prop) = \delta(T - T_\circ)p(E, N, \ . \ . \ .) \quad . \tag{3.16}$$

In this language all of the familiar concepts of temperature, Boltzman factors, etc., arise from the prior through definitions of the change in the volume of phase space with respect to extensive variables (e.g. $\frac{1}{kT} = \frac{\partial}{\partial E} \ln W$ where W is the volume of phase space and E is the energy exchanged between system and reservoir). Thus we would write

$$p(Prop|Sys) \propto \delta(T - T_\circ)p(E, N, \ . \ . \ .)\frac{1}{Z}Z(T, E, N, \ . \ . \ .) \ , \tag{3.17}$$

$$\propto \delta(T - T_\circ)p(E, N, \ . \ . \ .)\frac{\exp(-\Delta\sigma(T, E, N, \ . \ . \ .)}{\sum_{States, \ i} \exp(-\Delta\sigma(T_i, E_i, N_i, \ . \ . \ .))} \ , \tag{3.18}$$

$$\propto \delta(T - T_\circ)p(E, N, \ . \ . \ .)\frac{\exp\left(-\left(\frac{\Delta E}{kT} + \ . \ . \ .\right)\right)}{\sum_{States, \ i} \exp\left(-\left(\frac{\Delta E_i}{kT} + \ . \ . \ .\right)\right)} \ , \tag{3.19}$$

which contains the normal expression for the partition function.

What, then, is the role of uncertainty in statistical mechanics? The best known example of this is the role of the Heisenberg uncertainty principle. This principle declares that states within a particular hyper-cube [e.g. $(\Delta p \Delta x)^{3N} \sim (h/2\pi)^{3N}$ in the case of N free particles] are indistinguishable. This puts a natural graininess (or degeneracy) on phase space and directly affects calculation of the number of available states. Being unable to distinguish between states, however, can arise from other causes. The one of interest for the image reconstruction problem is uncertainty due to noise. The source of uncertainty, however, is inconsequential to the Bayesian estimation problem. Hence it is seen that the $(\Delta p \Delta x)^{3N} \sim (h/2\pi)^{3N}$ chunks appropriate to quantum phase space are nothing more than pixons induced by the uncertainty associated with the fundamental laws of physics.

Both the Heisenberg Uncertainty Principle and the uncertainty produced by measurement error in the image reconstruction problem cause the scientist to reevaluate the appropriateness of standard coordinate systems. This uncertainty in what might be an appropriate coordinate system can be seen from the above discussion to be a GOF induced problem. We might, for example, consider the effect on the statistical mechanics partition function if there were uncertainty in the system temperature. The direct outward affect of this would be to replace the delta functions in equations (3.16)-(3.19) with a broader distribution. However, we would then experience the same sort of effect that we see with pixons in the image reconstruction problem. Our old coordinate system for describing the properties of the system, i.e. that including the system temperature, becomes less appropriate. We can no longer use $T_\circ$ directly in our calculation of $p(Sys|Prop)$ or the Boltzman factor. The temperature $T_\circ$ has become an inappropriate variable. This is not because there is no appropriate value of the temperature. Temperature has become inappropriate because of uncertainty in its value. Just as in our use of pixons to smooth together adjacent pixel values (since having separately adjustable values with such a dense matrix was unjustifiable), a similar thing must be done with the statistical mechanics coordinate system (e.g. phase space). Phase space must be divided up into larger chunks, and this time the size of the chunks will have something to do with the uncertainty in the temperature. The specific size of the chunks will be determined by whether or not the states within the chunks (perhaps states with different temperatures) can be distinguished from each other.

4. Application of Pixon Based Methods

Pixon-based reconstruction has now been used by the astronomical community to perform a variety of image reconstruction problems. These include HST spectroscopy (Diplas et al. 1993), coded mask X-ray satellite imaging (Metcalf et al. 1996), far-IR airborne imaging (Koresco et al. 1995), IRAS far-IR survey data (Puetter and Piña 1994, Puetter 1996a), and mid-IR ground-based imaging (Smith et al. 1994). We shall concentrate on such applications in the next 2 lectures. Here, however, we would like to point out that pixon-based methods have application to a wide range of problems outside of astronomy. These include radar imaging, seismology, communication, medical imaging, laboratory spectroscopy, satellite earth-resource surveys, audio and video recording and play-back, and data compression, just to name a few. Most of these applications are obvious extensions of astronomical image reconstruction techniques to areas outside of astronomy, e.g. medical imaging. The data compression and communications aspects, however, are also quite obvious since pixon-based reconstruction begins by selection of a minimal representation, or coordinate system, in which to perform the reconstruction. This is effectively performing an optimal compression of the information contained in the data and so is directly applicable to image compression and compression of signals for efficient communication.

In the present section we would like to give 2 examples of pixon-based estimation that are not directly related to astronomical image reconstruction. In fact, in both of these examples, no reconstruction is done. Pixon-based methods have been used solely to perform an optimal spatially adaptive filtering of the data. The first example is presented in Figure 1. This example is again drawn from the field of astronomy. However this time we are interested in the temporal behavior of an X-ray source. Figure 1 displays a sample of timing data from the Vela satellites, along with the pixon-based estimate of the true underlying signal. Again, as already mentioned, in this case no reconstruction has been performed, i.e. we have not attempted to remove the temporal response of the Vela satellite X-ray detector. We have simply used pixons to filter the data and report only the statistically significant structure present in the data.

The second example is presented in Figure 2. This is a mammogram of a standard medical phantom. In this case a fiber of thickness 400 microns has been placed in a piece of material with X-ray absorption properties similar to the human breast. Again, in this example pixon-based methods were used solely to filter the data. No attempt was made to sharpen the image by removing the X-ray beam spread. As can be seen from this example, the pixon filtered image provides vastly superior feature detection contrast and would allow an X-ray technician to scan mammograms much more rapidly and with greater sensitivity to finding features such as cancerous tumors. It is obvious, for example, that even though the fiber present in this image is barely at the detection threshold (as evidenced by the fact that it breaks up into a chain of features due to noise), the pixon filtered image easily detects the presence of whatever statistically significant structure is present in the image.

5. Conclusions

The present lecture has shown how one might practically implement a multiresolution pixon-based image reconstruction algorithm. The relationship of pixons to other fields of study has been discussed, and it was shown that pixon-based (or pixon-like) concepts are already being used in a variety of fields. This lecture has also shown how pixon-based concepts can be used to extend the usefulness of these ideas. We have also argued that

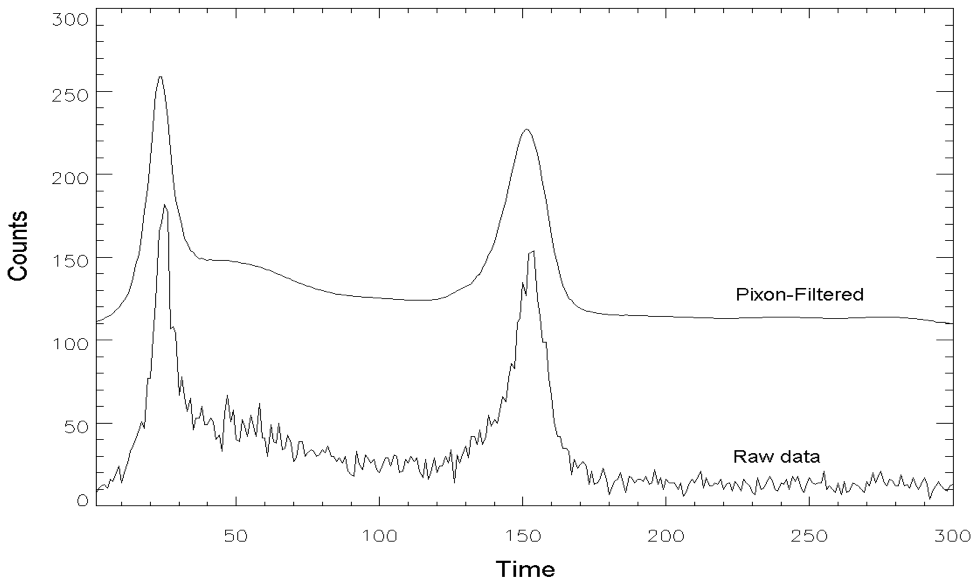

FIGURE 2. Pixon-based spatially adaptive filtering of X-ray timing data from the Vela satellites. Lower curve: Vela satellite data. Upper curve: pixon-filtered data with a constant added to provide a vertical shift. The noise in the data is given by the counting statistics in the signal. The pixon-filtered image displays only those structures calculated to be statistically significant relative to this noise.

pixon-based (or more generally informaton-based) methods can be applied to a variety of fields outside of astronomical imaging. In the following lectures we shall present a number of examples of pixon-based methods applied to astronomical studies. Lecture 4 will concentrate on applications to the image restoration problem, and Lecture 5 will concentrate on reconstruction of images from complexly encoded data.

REFERENCES

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., AND FLANNERY, B. P. 1994, *Numerical Recipes in C: The Art of Scientific Computing, Second Edition.* Cambridge University Press.

AKAIKE, H. 1973, Information Theory and an Extension of the Maximum Likelihood Principle, *Proc. Second International Symp. on Inf. Sci.*, eds. B. N. Petrov and F. Csáki, pp. 267-281. Akadémia Kiadó.

AKAIKE, H. 1973, On Entropy Maximization Principle, *Proc. Symp. on Appl. of Statistics*, ed. P. R. Krishnaiah, p. 267. North-Holland.

DIPLAS, A., BEAVER, E. A., BLANCO, P. R., PIÑA, R. K., AND PUETTER, R. C. 1993, Application of Pixon Based Restoration to HST Spectra and Comparison to the Richardson-Lucy and Jansson Algorithms, *The Restoration of HST Images and Spectra–II*, ed.s R. J. Hanisch and R. L. White, pp. 272-276. Space Telescope Science Institute.

HALL, P. 1990, Akaike's Information Criterion and Kullback-Leibler Loss for Histogram Estimation, *Prob. Th. and Rel. Fields* **85**, pp. 449-467.

KORESKO, C. D., HARVEY, P. M., CURRAN, D., AND PUETTER, R. C. 1995, Pixon Deconvolution of Far-Infrared Images from the UT Multichannel Photometer, Proc. Airborne Astronomy Symp. on the Galactic Ecosystem: From Gas to Stars to Dust, *A.S.P. Conference Series*, **73**, pp. 275-278.

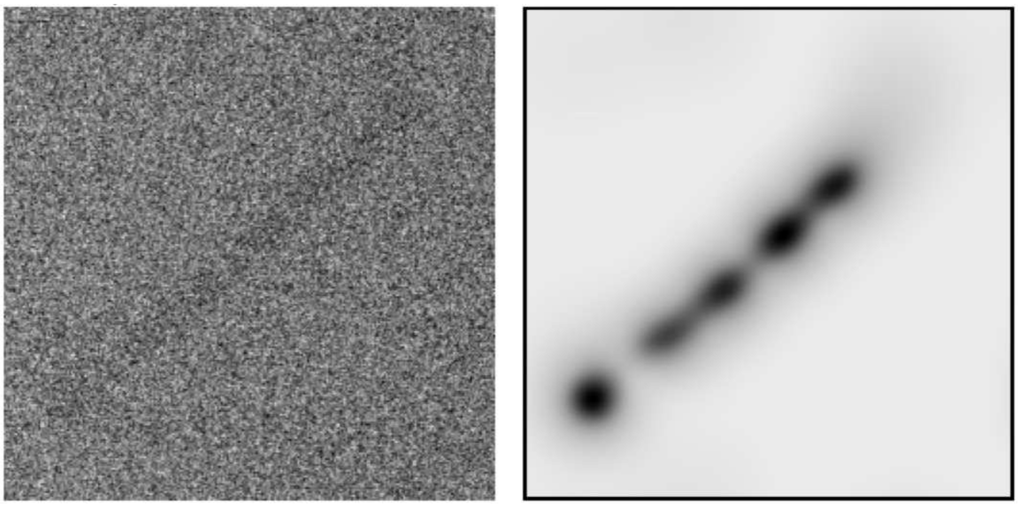

FIGURE 3. Pixon-based spatially adaptive filtering of a standard mammogram phantom. Left panel: raw phantom image. Right panel: pixon-filtered image. The phantom in this case is a 400 micron thick fiber placed in a piece of material with X-ray absorption properties similar to the human breast. The pixon-filtered image reports only those structures calculated to be statistically significant. Note that in this case the fiber is near the detection threshold–the fiber breaks up into a string of features because in some places the signal is statistically insignificant because of the large noise level. Noetheless, the pixon-filtered image easily detects the remaining statistically significant structure.

METCALF, T. R., HUDSON, H. S., KOSUGI, T., PUETTER, R. C., AND PIÑA, R. K. 1996, Pixon-Based Multiresolution Image Reconstruction for Yohkoh's Hard X-Ray Telescope, *Ap. J.* in press.

PIÑA, R.K., AND PUETTER, R.C. 1993, Bayesian Image Reconstruction: The Pixon and Optimal Image Modeling *P.A.S.P.* **105**, pp. 630-637.

PUETTER, R.C., AND PIÑA, R.K. 1993a, The Pixon and Bayesian Image Reconstruction, *Proc. S.P.I.E.* **1946**, pp. 405-416.

PUETTER, R.C., AND PIÑA, R.K. 1993c, Pixon-Based Image Reconstruction, *Proc. MaxEnt '93*, in press.

PUETTER, R.C., AND PIÑA, R.K. 1994, Beyond Maximum Entropy: Pixon-Based Image Reconstruction, *Science with High Spatial Resolution Far-IR Data*, ed/s S. Terebey and J. Mazzarella, JPL Pub. 94-5, pp. 61-68.

PUETTER, R.C. 1994, Pixons and Bayesian Image Reconstruction, *Proc. S.P.I.E.* **2302**, pp. 112-131.

PUETTER, R.C. 1995, Pixon-Based Multiresolution Image Reconstruction and the Quantification of Picture Information Content, *Int. J. Image Sys. & Tech.*, in press.

PUETTER, R. C. 1996, Pixon-Based Multiresolution Image Reconstruction and Quantification of Image Information Content, *Proc. MaxEnt '95*, in press.

SMITH, C. H., AITKEN, D. K., MOORE, T. J. T., ROCHE, P. F., PUETTER, R. C., AND PIÑA, R. K. 1994, Mid-infrared Studies of η Carinae–I: Sub-Arcsecond Imaging at 12.5 and 17μm, *M.N.R.A.S.*, **273**, pp. 354-358.

VERES, S. M. 1990, Relations Between Information Criteria for Model-Structure Selection. Part3. Strong Consistency of the Predictive Leaast Squares Criterion, *Int. J. Control* **52**, p. 737.