Information, Language, and Pixon-Based Image 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 paper reviews pixon-based image reconstruction which uses information-theory based constraints to perform reconstructions with highly optimized Bayesian priors. The paper introduces the theory of Bayesian estimation and discusses the relationship between Bayesian estimations and and information theory, especially the important role played by the concept of Algorithmic Information Content in selecting image models. Finally the paper presents a variety of examples of pixon-based reconstruction, showing the significant performance advantages provided by these methods.

1. Image Reconstruction and the Inverse Problem

This paper reviews pixon-based image reconstruction. Pixons (or *informatons*—their extension to generalized data sets), represent an optimized, information theory-based coordinate system for parameter estimation. Pixon-based methods are now becoming recognized as one of the most successful methods for practical image reconstruction, with performance exceeding those of other commonly used techniques such as Maximum Likelihood and Maximum Entropy methods. Like many of these techniques, pixon-based reconstruction can be viewed within the Bayesian estimation framework. However, an important departure of pixon-based methods from other Bayesian schemes is the use of information theory concepts, specifically Algorithmic Information Content, to optimize the Bayesian prior and constrain the reconstruction. Much of this material has been presented before, e.g. Piña and Puetter 1993, Puetter and Piña 1993a, Puetter and Piña 1993b, Puetter 1994, 1995, 1996a-e. The current paper summarizes these discussions and presents a short, practical tutorial on implementation of these methods and presents a number of applications of the method to practical problems.

2. Bayesian Methods

For the general inverse problem, the method of Bayesian estimation seeks to infer the most probable set of "parameters-of-interest" from a set of measured or otherwise given parameters. In the case of image restoration, this normally means estimating a higher resolution (or undistorted) image from a blurry (or distorted) image. In order to use probabilitistic methods, one typically examines the joint probability distribution of the data, D, the underlying, unblurred image which is to be estimated, I, and the model, M, which describes the detailed relationship of the data to the image. This joint probability distribution, $p(D, I, M)$, can be factored using conditional probabilites in the following manner:

$$p(D, I, M) = p(D|I, M)p(I, M) \quad , \tag{2.1}$$

$$
\begin{aligned}
&= p(D|I,M)p(I|M)p(M) \quad , &(2.2)\\
&= p(I|D,M)p(D,M) \quad , &(2.3)\\
&= p(I|D,M)p(D|M)p(M) \quad , &(2.4)\\
&= p(M|D,I)p(D,I) \quad , &(2.5)\\
&= p(M|D,I)p(D|I)p(I) \quad , &(2.6)
\end{aligned}
$$

where $p(X|Y)$ is the probability of X given that the value of Y is known. By equating various terms, these equations give rise to the formulae

$$p(I|D,M) = \frac{p(D|I,M)p(I,M)}{p(D|M)} \propto p(D|I,M)p(I|M) \quad , \tag{2.7}$$

$$p(I,M|D) = \frac{p(D|I,M)p(I|M)p(M)}{p(D)} \propto p(D|I,M)p(I|M)p(M) \quad . \tag{2.8}$$

The formula of equation (2.7) is the typical starting place for Bayesian methods. This equation essentially assumes that the appropriate model, M, is known and will not be varied during the reconstruction. This assumption is made explicit in the proportionality of equation in (2.7) where the term $p(D|M)$ is dropped (it is assumed to be constant—D and M are not varied during the reconstruction). Our preferred formulation, however, is given by the formula of equation (2.8). Here we do not assume that all of the parameters associated with the model belong to the so called "nuisance parameters", i.e. parameters which might require estimation, but which are of no interest to the scientist. In pixon-based reconstruction, however, the model parameters associated with the local smoothness scale of the image (see later lectures) are extremely important.

The significance of the terms on the far right hand side of equations (2.7) and (2.8) is readily understood. The term $p(D|I,M)$ is a goodness-of-fit (GOF) quantity, measuring the likelihood of the data given a particular image and model. The terms, $p(I|M)$ and $p(M)$, are "priors", and incorporate our prior knowledge (or expectations) about the measurement situation or the nature of suitable selections for I and M. The term "prior" is used since each of these terms make no reference to the data, D, and hence can be decided on *a priori*, i.e. before the act of making the measurement.

In GOF (or maximum likelihood) image reconstruction, $p(I|M)$ and $p(M)$ are assumed to be constant, i.e. there is no prior bias concerning the image or parts of the model that might be varied. The standard choice for $p(D|I,M)$ is to use $p(D|I,M) = (2\pi\sigma^2)^{1/2n_{pixels}} \exp\left(-\chi^2/2\right)$, i.e. the joint probability distribution for n_{pixels} independent pixels with normally distributed (Gaussian) noise. This 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.

$$
\begin{aligned}
p(\text{Blurred Image+Noise}|I,M) &= p(\text{Blurred Image}|\text{Noise},I,M)p(\text{Noise}|I,M) &(2.9)\\
&\cong const \times p(\text{Noise}|I,M) \quad . &(2.10)
\end{aligned}
$$

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

In Maximum Entropy (ME) image reconstruction, the image prior is based upon "phase space volume" or counting arguments and the prior is expressed as $p(I|M) = \exp(\alpha S)$, where S is the entropy of the image and α is an adjustable constant that is used to weight the relative importance of the GOF and image prior. Again, usually the model is considered to be known and held fixed, i.e. $p(M) = constant$, and the usual assumption is to choose$p(D|I,M) = (2\pi\sigma^2)^{1/2n_{pixels}} \exp\left(-\chi^2/2\right)$. While many specific formulations for α and S appear in the literature (Kikuchi and Soffer 1976; Bryan and Skilling 1980; Narayan and Nityananda 1986; Skilling 1989), all ME methods capitalize on the virtues of incorporating prior knowledge of the likelihood of the image into the image restoration algorithm.

3. Bayesian Priors, Efficient Models, and Occam's Razor

While most scientists have a well developed intuition regarding the merits of achieving a good fit to the data, i.e. optimizing the GOF criteria, a well developed intution concerning the Bayesian priors and the effects of model selection on the likelihood of a particular image/model pair. Investigation into the general effects of model selection, however, turn up the interesting fact that simple, highly predictive models are highly favored. In fact, such considerations provide a very practical and real basis for the principle of *Occam's Razor*, i.e. the predilection of scientists to choose the simpler of two models that provide equally good fits to the data. One of the best discussions of the Bayesian embodiment of Occam's Razor is that presented by David MacKay (MacKay 1994). Our discussion shall parallel MacKay's discussion.

In order to judge the relative merits of two hypotheses, H_1 and H_2, for explaning the data set D, Bayesian's would form the relative probability ratio:

$$\frac{p(H_1|D)}{p(H_2|D)} = \frac{p(H_1)}{p(H_2)}\frac{p(D|H_1)}{p(D|H_2)} \quad . \tag{3.11}$$

The first ratio involves hypothesis priors, i.e. hypothesis probability densities with no functional dependence on the data. These can be used to express our prior conceptions of the relative likelihood of hypothesis H_1 relative to H_2. As pointed out by MacKay, this corresponds to the motivations behind Occam's razor. However, it is unnessary to make any prior assumptions about the relative probabilities $p(H_1)$ and $p(H_2)$. Equation (3.11) will still give rise to an Occam's Razor like effect. This is because the simpler hypotheses, say H_1, make the most precise predictions. Complex hypotheses embrace a much larger data space and hence spread their predictive probability $p(D|H_2)$ more thinly over this larger data space. Thus, in the case in which both hypotheses H_1 and H_2 are equally compatible with the data, the simpler hypothesis H_1 will be favored by equation (3.11), without having to express any subjective dislike for one hypothesis relative to the other. Of course even if one assigns an equal prior probability to certain aspects of the hypothesis, e.g. the model in image reconstruction, one can usually make sensible arguments for appropriate priors for the other aspects of the hypothesis, e.g. the image prior. For example, in the image reconstruction case, the hypothesis for a given data set is the pair (I, M). The prior is thus $p(I,M) = p(I|M)p(M)$. So even if one expresses no prior bias for a given model, e.g. $p(M) = constant$, one might easily assign a sensible prior probability distribution for $p(I|M)$, e.g. a ME prior. Such an assignment also acts in the cause of simple models, since any sample state drawn from a simple model typically has a higher probability simply because it represents a larger fraction of all possible states. Thus at every turn probability theory favors Occam's Razor, i.e. the selection of the simplest possible image/model (hypothesis) to explain the data. Thus in order to perform the very best image reconstruction, we need to simplify the image/model as much as possible. This will both optimize the Bayesian prior and provide the "highest possible peak" in the GOF term.

4. Successes and Failures of Classical Methods

The very fact that image restoration/reconstruction has been performed for many years is testimony to its basic success as a method. For PSF blurring problems, its primary advantage is its ability to provide enhanced resolution. The first attempts at PSF deconvolution, employed Fourier methods. However such method only provide suitable performance if noise associated with the measurement is absent or inconsequential. Nonlinear methods provide a significant improvement over Fourier deconvolution methods in moderate to low SNR situations by providing greater control over noise propagation. This is done by insisting (through the GOF criterion) that the modeled data (e.g. the data predicted by the inferred image and model) match the measured data within the uncertainty specified by the noise. This is quite different from Fourier methods were no such demand is made. Of course, insisting that the predicted data "fits" the measured data within the noise limits does not guarantee that the inferred image is without significantly larger noise. Indeed, as discussed above, each image reconstruction problem can have a significant "null space'.

Common problems with GOF (or Maximum Likelihood) and ME methods are overfitting of the data, the production of signal correlated residuals, and spurious sources. These problems are related to using an inappropriate number of degrees-of-freedom (DOFs) to fit the data and consequently are largely overcome by pixon-based methods—see later lectures in this series. Briefly, however, signal correlated residuals are produced by a "diluted" GOF criterion. Suppose we have imaged a galaxy in a large CCD frame. Assume that the CCD has dimensions 1024×1024 pixels. Further suppose that the galaxy only fills the central 100×100 pixels. If the astronomer were to use the entire data frame in his reconstruction and chose to represent the image as a rectangular grid of numbers with the same spatial frequency as the data (as is commonly done), then the reconstruction would use roughly 10^6 DOFs. If a standard GOF criterion were used, then one would stop the iterative procedure when there were less than 10^3 residuals larger than 3σ (0.1% of the residuals are 3σ or larger in a population of Gaussian distributed variables). Since iterative procedures that minimize GOF functions normally spend most of their time fitting the bright sources, they will be working on adjusting the bright source levels when the stopping criterion is met, resulting in the vast majority of the large residuals lying under the bright sources.

The origin of spurious sources in standard image reconstruction techniques is even easier to understand. Using the example above, it should be clear to most readers that the 10^6 DOFs used in the reconstruction are probably far more than are needed to describe the information present in the data. In fact, the data probably is inadequate to constrain the values of this many DOFs. Hence the vast majority of these DOFs are free to produce whatever bumps and wiggles they like, with the only requirement being that after they are smoothed by the PSF they average to zero within the noise. Hence their amplitude can be very large as long as they have a spatial scale small relative to the PSF and are arranged in such a manner that hills and valley tend to cancel each other out. Since this large number of unconstrained DOFs represents a huge phase space and the typical member has numerous bumps and wiggles, spurious sources are guaranteed.

One of the main motivations for developing Maximum Entropy methods was to help overcome the spurious source problem. After all, the ME prior acts to make unsmooth images undesireable. However, in their zeal to flatten spurious sources, ME methods can over flatten the true sources. This affects the ability of ME methods to perform statistically unbiased photometry. In fact, while ME methods are significantly more immune to spurious sources (they more effectively reject situations with large adjacent hills and valleys than GOF methods), they still tend to systematically underestimate source brightnesses (Sibisi 1990; Cohen 1991).

Most of the problems facing standard GOF and ME methods are over come by pixon-based methods. Pixon-based methods are a generalization of ME methods and perform the Bayesian estimation problem in a more natural language. The pixon language is information-based. Consequently it concisely and accurately describes the key variables in the problem and avoids the use of too many DOFs in the reconstruction. Because the number of DOFs is tuned to the number required to accurately describe the image, problems with stopping the reconstruction because a "diluted" GOF criterion is used are avoided. In addition, spurious source production are avoided since the pixon-basis automatically rejects spurious sources as being unneeded in the reconstruction. The generalization of the pixon to arbitrary data sets is the *informaton*, i.e. a quanta of information, and has a wide range of application to a variety of problems. To see how concepts from information theory can be used as a powerful constraint for the image reconstruction problem, we turn next to a discussion of information, itself.

5. Types of Information

The concepts of information and entropy have been discussed in numerous works. While the thermodynamic concepts of entropy have a very specific meaning and apply to the state of an ensemble of particles, the concepts of information have a much broader scope. Nonetheless, the laws of statistical physics can be derived directly from information theory concepts (Jaynes 1957; Jaynes 1963; Hobson 1971), and information and entropy (or negentropy) are often talked about as the same quantities. The most common definition of information, now known as Shannon information, was introduced in 1948 (Shannon 1948; Shannon and Weaver 1949) and specifies the average information contained in a string of symbols transmitted across a communication line. In this definition, the information per symbol is given by

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \quad , \tag{5.12}$$

where n is the total number of possible symbols, and p_i is the probability of occurance of the i^{th} symbol. There are many reasons why the Shannon information is a suitable definition for information. One of the most convincing arguments is given by the *Noiseless Source Coding Theorem* (see, for example, Rabbani and Jones 1991), which states that for a ergodic source with an alphabet of n characters, that as the size of the transmitted string approaches infinite length, it is possible to construct a unique random length binary code for the string that has an average length in bits per encoded symbol equal to, but no less than, the Shannon entropy.

While the *Noiseless Source Coding Theorem* in and of itself speaks strongly for Shannon entropy as a sensible information measure, there are still other motivations. For example it can be proven that the Shannon information is the unique and self-consistent measure which satifies conditions which would reasonably be expected of a measure of information (Khinchin 1957; Feinstein 1958; Hobson 1971). Using Khinchin's formulation of the properties uniquely defining an information measure, it is found that only the function H satisfies the conditions that (1) H takes it's largest value when all the p_i are equal, (2) H doesn't change it's value when additional impossible events are added—i.e. events for which $p_i = 0$, and (3) the information in two not necessarily independent events is given by the information of the first event plus the expectation value of the additional information provided by the second type of event after the first event has occurred.

Readers who have had classical training in physics will be interested that the formula of equation (5.12) is essentially equivalent to the physical definition of entropy, $\sigma = k \ln W$, (modulo a constant to convert logarithm types) where k is Boltzman's constant and W is the total number of available states. In fact, equation (5.12) is the Boltzman (or Gibbs) definition of entropy. Kittel (Kittel 1969—see page 118) gives a particularly fluid derivation of the Boltzman definition of entropy starting from the Boltzman factor.

While we have now provided motivation for the merits of Shannon information, of what use is this concept for solving the inverse problem in general or image reconstruction in particular. To be sure, ME methods make use of the entropy or Shannon information definition. However, Shannon information doesn't get to the heart of the problem. This definition of information doesn't allow us to describe the *information content* of an individual image or data set. As has been pointed out by several authors (Wicken 1987; Chaiten 1982), Shannon information is an ensemble notion. It is a measure of ignorance concerning possible realizations of events which have a given *a priori* probability distribution, and measures the statistical rarity or "surprise value" of a particular realization (Cherry 1978). It does not deal with the information content expressed by a given realization. For this we need a different concept, i.e. that of algorithmic information content (AIC), algorithmic randomness, or algorithmic complexity (Solomonoff 1964; Kolmogorov 1965; Chaiten 1966). As commonly used in information or computer science, the AIC of a string of characters is defined to be the size of the minimum computer program required for a universal computer to produce the specified string as its output. (A computer U is universal if for any other computer M there is a prefix program μ such that the program μp makes U perform the same tasks as performed by computer M running the program p, i.e. the program μ makes computer U simulate computer M.) As suggested by the name, AIC describes the information content of the specific item under scrutiny. In fact it does this is a very practical way, i.e. algorithmically. It provides a prescription for how the data can be reproduced. It should also be clear that the AIC represents the optimal compression of the data. There is no shorter description that will allow a complete reconstruction of the original information. In fact, it is this optimal compression of the information that will allow us to optimally extract an image in the image reconstruction problem in a practical manner.

6. The Role of Language in AIC and Bayesian Estimation

The term "randomness" comes into the definition of AIC since random strings of characters have maximum complexity or information content. This is because if the pattern of the string is non-random, there is some way of describing the string which is typically shorter than listing the string itself. For example, in strings of 1s and 0s, a string with 10^9 copies of the number 1 would take a lot of paper to write down, but can be described with a single sentence. Similarly this same string with a 0 placed in the millionth place can also easily be described without wasting paper. However, a totally random string of 10^9 digits cannot be so easily described, and if no rule is known (and this is what we mean by random), only the actual string will impart all of the information. It is well known that quantification of the AIC is a function of the "richness" of the language used to describe the character string or data set. In the example used above, we demonstrated how the English language could be used to briefly describe large strings of digits. Other languages, e.g. the binary digits 0 and 1, decimal digits, or even hexidecimal digits, do not provide such a terse description.

The above discussion makes it quite obvious that rich, i.e complex, languages allow a terse discription of a data set—the AIC is low for rich languages. This seems like an obvious advantage. It says that the data can be compressed to a higher degree and that the information can be more precisely located and identified. Similar concepts occur in Bayesian estimation, e.g. the concept of Occam's Razor. Clearly, in the Bayesian estimation problem, it is highly beneficial to develop concise, simple hypotheses (images or image/model pairs) to explain the data. Such simple hypotheses give rise to optimized priors [$p(I|M)$ or $p(I|M)p(M)$]. Consequently, they produce optimal reconstructions and a more likely M.A.P. image or image/model. Thus it is seen that minimizing the AIC of the image/model, optimally compressing the information in the data, and performing highly optimized Bayesian image reconstruction are directly related. Hence the goal for improved image reconstruction is clear. One must improve the language in which the reconstruction is performed, or equivalently the language in which the image is described. The language must be rich so that the description can be concise, the AIC low, and the Bayesian prior optimized. The fact that the description will be concise suggests that the language for describing the image must be "natural" in some sense, and we turn now to a discussion of possible languages for image description.

7. Languages for Images

There are many "languages" for images in current use. The most common of these are usually not even thought of as languages since they seem so natural. For example, in electronic or digital imaging we essentially expect to see every image in the form of a rectangular pixel array. This is so natural that we don't even consider the consequences of this language for the image. However there are consequences. The pixels are almost always rectangular and so is the picture format. However this may not be optimal for image reconstruction. Certainly in the case of astronomical imaging, many objects are not square, e.g. stars. Also images of star clusters or galaxies, etc., seldom uniformly fill a rectagular picture to its corners. Hence these clearly are not an optimal format or pixel shape for astronomical situations. There are usually too many pixels in the image and many pixels are used to describe single objects such as stars.

To motivate a discussion of what constitutes an optimal language for image description, et us first ask the question: "What is the nature of picture information?" This, of course, can have many answers, and the answer changes, depending on the object at which the photographer points his camera. We are interested, here, in "generic" images. So let us first ask the general question: "What do we really expect when we take a picture?" Thinking about this in the broadest sense, all we can really say is that we expect the picture to contain some limited amount of information. We expect, in particular, that the manufacturer of the film or the designer of the camera will have made the grain or pixel size fine enough to do service to the picture, i.e. not to leave out any important detail. We expect interesting features, perhaps people, with open spaces between them. In other words, we expect that at each point in the image there is a finest spatial scale of interest and that there is no information content below this scale. Indeed, this is how photographic grain sizes are chosen and why data is not sampled finer than the Nyquist frequency when pixel elements are at a premium.

How does one capture this prior expectation in mathematic form and incorporate it into a set of image basis functions? Again, the key comes from thinking about photographic grain sizes or Nyquist sampling. We would do just as well at recording the picture information with large photographic grains in portions of the image with coarse structure. We need only have fine grains when we need to record fine spatial structure. This means that the picture information can be dealt with by using variable sized cells, with the cell sizes set so as to capture the spatial information present. In other words, we are proposing that a multiresolution language is an appropriate choice, i.e. will be concise and hence fulfil the goals of Occam's Razor.

8. A Practical Multiresolution Pixon Language for Image Description

Above, 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

$$I(\vec{x}) = (K \otimes I_{pseudo})(\vec{x}) \tag{8.13}$$

$$(K \otimes I_{pseudo})(\vec{x}) = \int dV_{\vec{y}}\ K(\vec{x}, \vec{y}, \delta(\vec{x})) I_{pseudo}(\vec{y}) \tag{8.14}$$

$$K\left(\vec{x}, \vec{y}, \delta(\vec{x})\right) = K\left(\frac{\|\vec{x}-\vec{y}\|}{\delta(\vec{x})}\right) \quad ; \text{ for radially symmetric pixons} \quad , \tag{8.15}$$

where equations (8.13) through (8.15) 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 (8.14) is carried out over volume in pseudo-image space. We have also indicated in equation (8.15) 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.

$$K\left(\vec{x}, \vec{y}, \delta(\vec{x})\right) = \begin{cases} \left(1 - \frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) / \int dV_{\vec{y}} \left(1 - \frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) & ; \ \|\vec{x}-\vec{y}\| \le \delta(\vec{x}) \\ 0 & ; \ \|\vec{x}-\vec{y}\| > \delta(\vec{x}) \end{cases} \quad . \tag{8.16}$$

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.

9. 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 $p(I, M)$. As mentioned above, 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). While this 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

$$p(I|M) = \frac{N!}{n^N \prod_{i=1}^{n} N_i!} = e^S \tag{9.17}$$

$$p(M) = constant \quad , \tag{9.18}$$

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 (9.17)-(9.18) 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:

$$p(I|M) = \frac{N!}{n^N \prod\limits_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i!} \tag{9.19}$$

$$N = \sum_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i = \sum_{\substack{j=1, \\ j \,\epsilon\, Image}}^{n_{pixels}} N_j \tag{9.20}$$

$$N_i = \int dV_{\vec{y}} \; k(\vec{x}, \vec{y}, \delta(\vec{x})) I_{pseudo}(\vec{y}) \tag{9.21}$$

$$\prod_{\substack{i=1, \\ i \,\epsilon\, Image}}^{n_{pixons}} N_i! = \prod_{\substack{j=1, \\ j \,\epsilon\, Image}}^{n_{pixels}} (N_j!)^{p_j} \tag{9.22}$$

$$p(\vec{x}) = 1 \; / \int dV_{\vec{y}} \; k(\vec{x}, \vec{y}, \delta(\vec{x})) \tag{9.23}$$

$$p(M) = constant \tag{9.24}$$

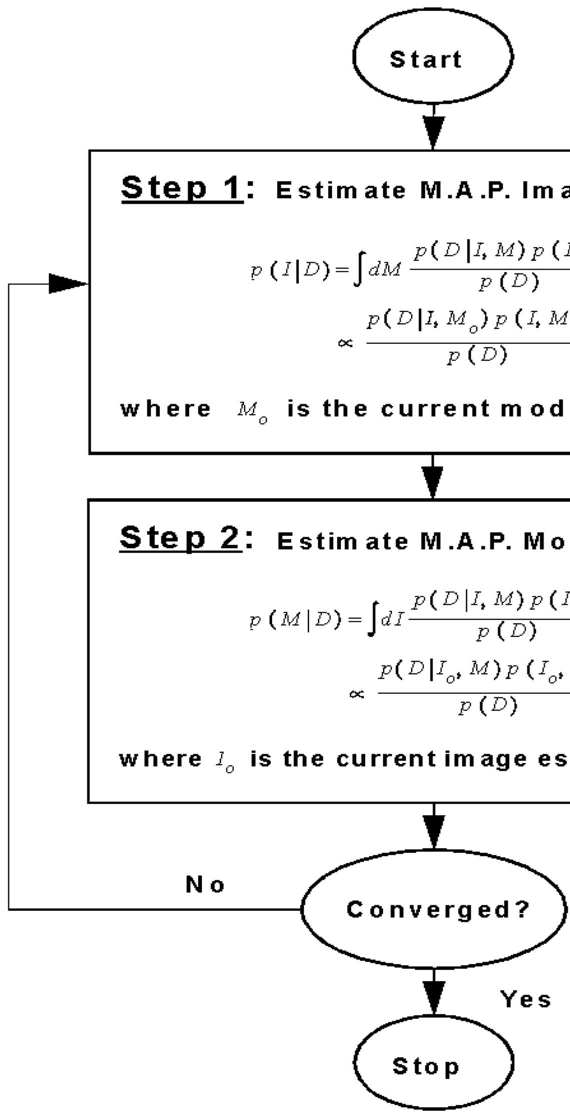

FIGURE 1. Iterative scheme for obtaining the M.A.P. image/model pair. This figure shows the pixon algorithm as currently implemented. The procedure is divided into two steps: (1) GOF optimization step in which the image is optimized while holding the pixon map (local smoothness scales) fixed, (2) Pixon 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.

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 (9.19) 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 (9.19) 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 (9.19) 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.

10. 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). In this section we present a number of applications of pixon-based methods, including adaptive filtering, image restoration, and image reconstruction. The first 2 examples present noise filtering applications, i.e. (1) temporally adaptive noise filtering of X-ray timing data from the Vela satellites, and (2) spatially adaptive filtering of a mammogram X-ray phantom. Next we present 2 examples of image restoration: (1) restoration of a mock data set, i.e. a completely made up data set in which the underlying image, PSF, and noise frames are manufactured and hence known precisely, and (2) an elliptical pixon-based reconstoration of Keck $2\mu m$ imaging data for the gravitational lens of FSC 10214+4724. Mock data sets are useful in comparisons of image restoration/reconstruction methods since there can be no argument about the goal of the reconstruction. The true, underlying image is known

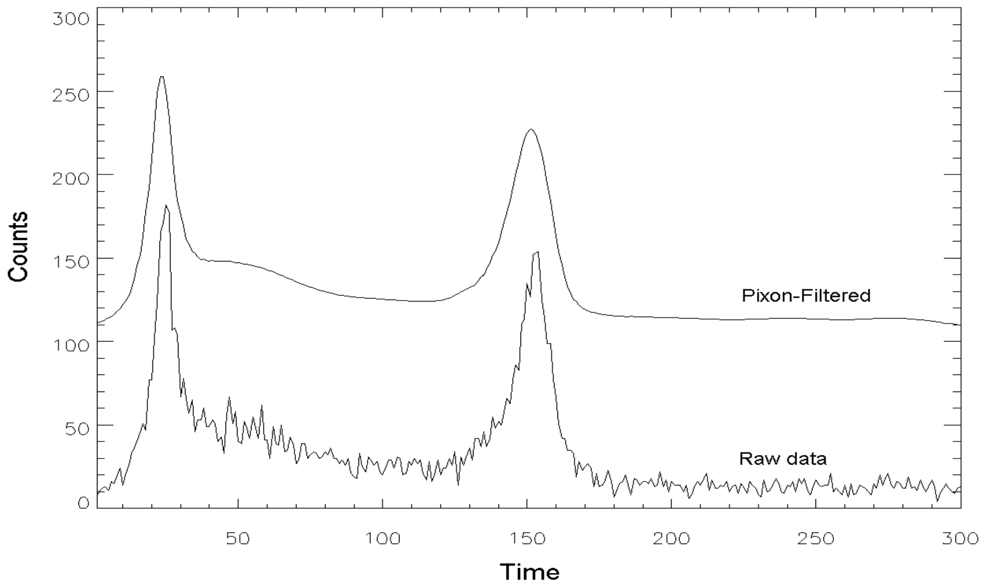

FIGURE 2. Pixon-based spatially adaptive filtering of X-ray timing data from the Vela satellites. Lower curve: Vela satellite data. Upper curve: pixon-filtered data with a constant added to provide a vertical shift. The noise in the data is given by the counting statistics in the signal. The pixon-filtered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

perfectly *a priori*. Furthermore, the noise and all parameters of how the input data was made are completely specified. The second example illustrates the use of non-radially symmetry pixon bases. We shall go into some detail in the discussion of this example principally because most pixon reconstructions presented in previous work have used radially symmetric basis functions and because the science allowed by the reconstruction is quite interesting. Finally, we present 3 examples of image restoration. The first example is of IRAS $60\mu m$ survey scans of the interacting galaxy M51. The second example is coded Fourier X-ray mask imaging of solar flare data from the hard X-ray telescope aboard the Yohkoh satellite, and the final example is scanned γ-ray collimator imaging data from the OSSE Virgo survey from the Compton Gamma Ray Observatory satellite. The examples of pixon-based applications used in this section have appeared before in Puetter and Piña 1994, Puetter 1995 and Puetter 1996c, Puetter 1996d, Puetter 1996e.

10.1. *Pixon-Based Adaptive Noise Filtering*

The present section presents 2 examples of pixon-based applications that are not directly related to astronomical image restoration/reconstruction. In fact, in both of these examples, no reconstruction is done. Pixon-based methods have been used solely to provide an optimal spatially adaptive noise filter. The first example is presented in Figure 2, which displays a sample of timing data from the Vela satellites, along with the pixon-based estimate of the true underlying signal. The lower curve shows the actual collected data as a function of time (arbitrary units). The upper curve represented the pixon-based best estimate of the true underlying signal. The curve is plotted with an added constant for vertical displacement. The noise in the data is due to photon counting statistics in the signal. As can be seen from the Figure, unlike box-car filtered or Fourier-domain filtered

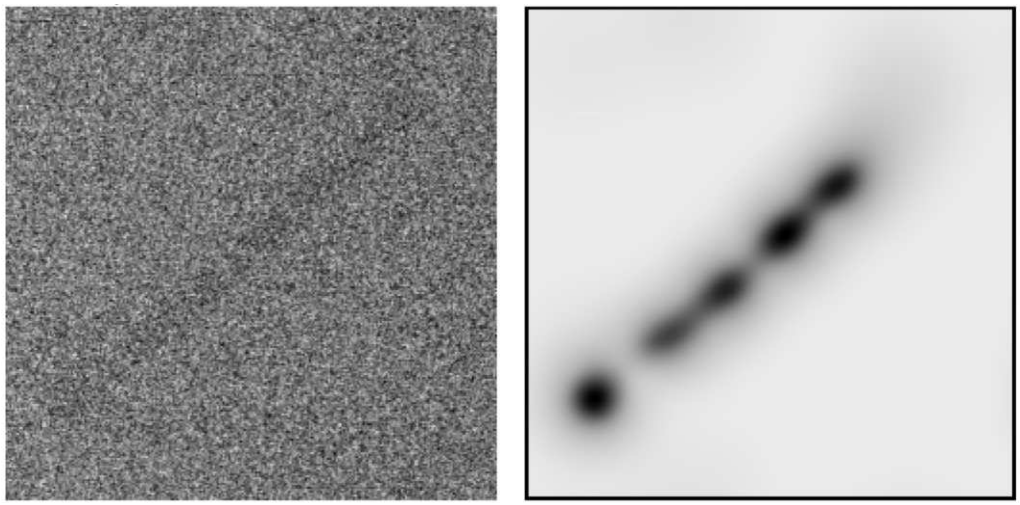

FIGURE 3. Pixon-based spatially adaptive filtering of a standard mammogram phantom. Left panel: raw phantom image. Right panel: pixon-filtered image. The phantom in this case is a 400 micron thick fiber placed in a piece of material with X-ray absorption properties similar to the human breast. The pixon-filtered 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. Nonetheless, the pixon-filtered image easily detects the remaining statistically significant structure.

data, the pixon filtered signal retains all of the high frequency information present in the strong pulses. This is because the filtering is temporally adaptive and preserves any statistically justifiable structure by using a broad filter on the low-lying noisy signal and a narrower filter on the stronger sources with just the appropriate widths to perserve the justifiable structure.

The second example is presented in Figure 3. 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.

10.2. *Pixon-Based Image Restoration*

We present in this section 2 examples of pixon-based image restoration. In the first example, we compare pixon-based and MEMSYS 5 restorations of a mock-data set. The ME algorithms chosen for comparison are those embodied in MEMSYS 5, the most current release of the MEMSYS algorithms. The MEMSYS code represent a powerful set of ME algorithms developed by Gull and Skilling (Gull and Skilling 1991). The MEMSYS algorithms probably represent the best commercial software package available for image reconstruction. The MEMSYS reconstruction was performed by Nick Weir of Caltech,

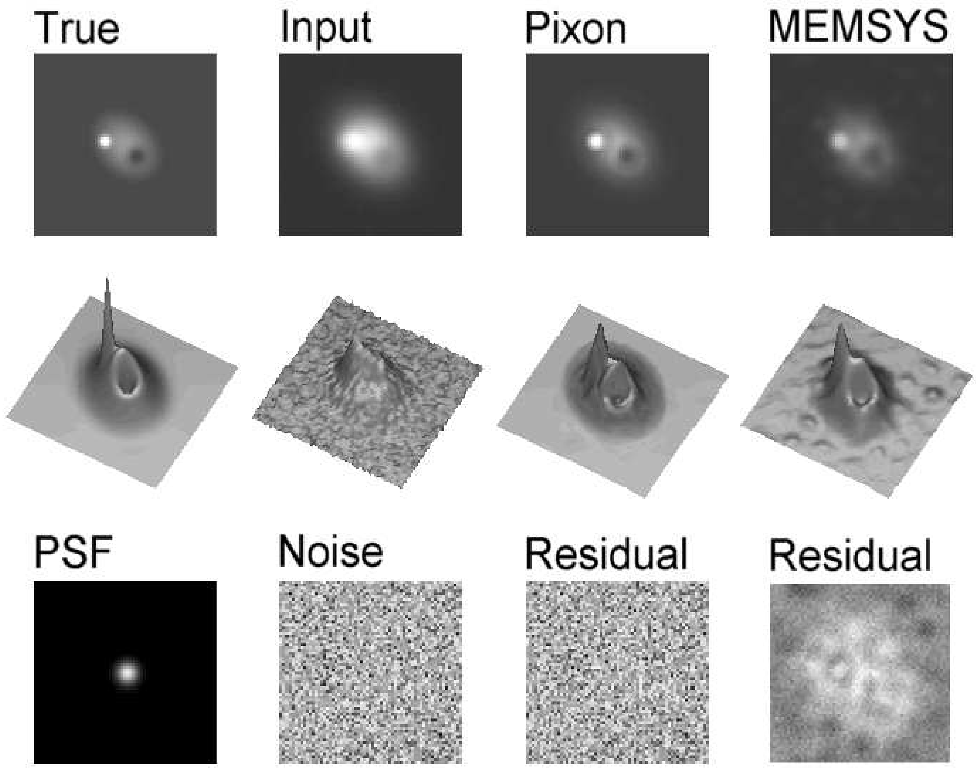

Figure 4. Comparison of pixon-based image restoration with Maximum Entropy restoration for a mock data set. Shown is the PSF (Gaussian with FWHM=6 pixels), the true, underlying image, and the Gaussian noise. The restored image and residuals for radially symmetric, truncated parabolic fuzzy pixons are shown. Also shown are the results and residuals for the multi-channel MEMSYS 5 method. Reproduced from Puetter 1994.

a recognized MEMSYS expert, and were supplemented with his multi-channel correlation method which has been shown to enhance the quality of MEMSYS reconstructions (Weir 1991; Weir 1994). The true, noise-free, unblurred image presented in the top row is constructed from a broad, low-level elliptical Gaussian (i.e. a 2-dimensional Gaussian with different FWHMs in perpendicular directions), and 2 additional narrow, radially symmetric Gaussians. One of these narrow Gaussians is added as a peak on top of the low-level Gaussian. The other is subtracted to make a hole. To produce the input image, the true image was convolved with a Gaussian PSF of FWHM=6 pixels, then combined with a Gaussian noise realization. The resulting input image is displayed in the top row. The signal-to-noise ratio on the narrow Gaussian spike is roughly 30. The signal-to-noise on the peak of the low level Gaussian is about 20. The signal-to-noise at the bottom of the Gaussian "hole" is 12.

As can be seen, the FPB reconstruction is superior to the multi-channel MEMSYS result. The FPB reconstruction is free of the low-level spurious sources evident in the MEMSYS 5 reconstruction. These false sources are due to the presence of unconstrained degrees of freedom in the MEMSYS 5 reconstruction and are superimposed over the entire image, not just in the low signal to noise portions of the image. Furthermore, the FPB reconstruction's residuals show no spatially correlated structure, while the MEMSYS 5 reconstruction systematically under-estimates the signal, resulting in biased photometry.

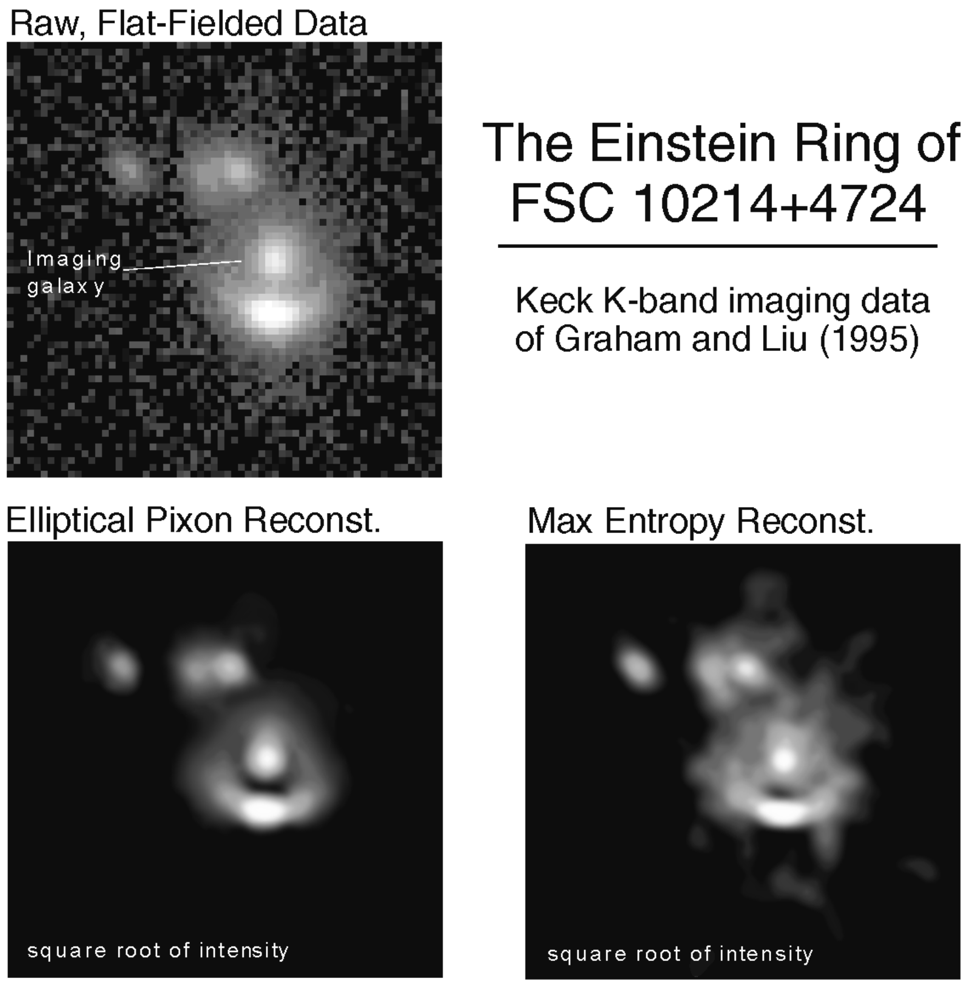

FIGURE 5. Comparison of image reconstruction techniques for the Einstein ring of FSC 10214+4724. The elliptical pixon reconstruction used 196 different elliptical pixon kernel functions. Relative to radially symmetric reconstructions, the elliptical pixon reconstruction used roughly 50% fewer DOFs in the image model. Reproduced from Puetter 1996.

In the second example of image restoration, we have applied pixon-based methods Keck K-band ($2.3\mu m$) imaging data of the gravitational lens system FSC 10214+7424 (data presented in Graham and Liu 1995). This object was initially thought to be the most luminous object in the universe with a luminosity approaching 10^{15} solar luminosities. However, it is now recognized as a gravitational lens with a magnificantion of roughly 100 (Broadhurst and Lehar 1995; Close et al. 1995; Graham and Liu 1995). Hence the luminosity is about 100 times fainter than initially estimated.

In order to make sure that our image description language would not be biased against faint Einstein rings, and was able to describe inherently thin structures, we decided it would be wise to use a set of elliptical pixon shape function as our language basis. Our elliptical pixon restoration of the Keck K-band image is shown in Figure 5. Also displayed in this figure is the Maximum Entropy reconstruction of Graham and Liu (Graham and Liu 1995). As can be seen from this figure, our reconstruction reveals the entire Einstein ring. This low-level ring was not detectable by the Maximum Entropy methods employed by Graham and Liu because of the large number and strength of the spurious sources produced by that method. These sources are obvious in the Maximum Entropy result displayed in Figure 5.

The science that our better reconstruction allows for this data is quite interesting. First, Graham and Liu concluded on the basis of their inability to see an entire Einstein ring, that the size of the emitting source (which has a redshift of 2.3) is 0.25 arcsec or less. They point out that given the lens geometry, if a complete ring was seen, then the size of the emitting region would have to be at least 0.5 arcsec. Hence our first conclusion is that the emitting region of FSC 10214+7424 has significant emission out to a size of 0.5 arcsec, roughly twice that deduced by Graham and Liu. Our second result is related to the nature of the gravitational lens. Recent HST images of the Einstein ring at far-UV rest-frame wavelengths ($0.25\mu m$) were obtained by Eisenhardt et al. (Eisenhardt et al. 1996). From the narrow nature of the arc, these authors conclude that the size of the UV source is of order 0.005 arcsec (40 pc), and that the best fit to the shape of the arc requires a contribution to the gravitational lensing potential from the next brightest elliptical galaxy near the main lensing galaxy. Our restored image directly confirms the gravitational contribution of this object to the lensing potential. The K-band Einstein ring is clearly warped in the direction of this source, directly confirming the speculations of Eisenhadt et al. Further discussion of these results will appear in a paper which is in preparation.

10.3. *Pixon-Based Image Reconstruction*

In this section we present 3 examples of pixon-based image restoration. The first example in given In Figure 6, were we present a reconstructed image from $60\mu m$ IRAS survey scans of the interacting galaxy pair M51. This particular data set is especially useful for comparing the performance of various different image reconstruction techniques since it was chosen as the basis of an image reconstruction contest at the 1990 MaxEnt Workshop (Bontekoe 1991). Furthermore, the IRAS data for this object is particularly strenuous for image reconstruction methods. This is because all the interesting structure is on "sub-pixel scales" (IRAS employed relatively large, discrete detectors—1.5 $arcmin$ by 4.75 $arcmin$ at $60\mu m$) and the position of M51 in the sky caused all scan directions to be nearly parallel. This means that reconstructions in the cross-scan direction (i.e. the 4.75 $arcmin$ direction along the detector length) should be significantly more difficult than in the scan direction. In addition, the point source response of the 15 IRAS $60\mu m$ detectors (pixel angular response) is known only to roughly 10% accuracy, and finally, the data is irregularly sampled.

In Figure 7, our FPB reconstruction is compared to the results from other methods, including the Lucy-Richardson, the Maximum Correlation Method (MCM)—see Rice 1993, and MEMSYS 3 reconstructions (Bontekoe et al. 1991). The winning entry to the MaxEnt 90 image reconstruction contest was produced by Nick Weir of Caltech and is not presented here since quantitative information concerning this solution has not been published—however, see Bontekoe 1991 for a gray-scale image of this reconstruction. Nonetheless, Weir's solution is qualitatively similar to Bontekoe's solution. Both were made with MEMSYS 3. Weir's solution, however, used a single correlation length channel in the reconstruction. This constrained the minimum correlation length of features in the reconstruction, preventing break-up of the image on smaller size scales. This is probably what resulted in the "winning-edge" for Weir's reconstruction in the MaxEnt 90 contest (Weir 1993).

As can be seen from Figure 7, our FPB-based reconstruction is superior to those

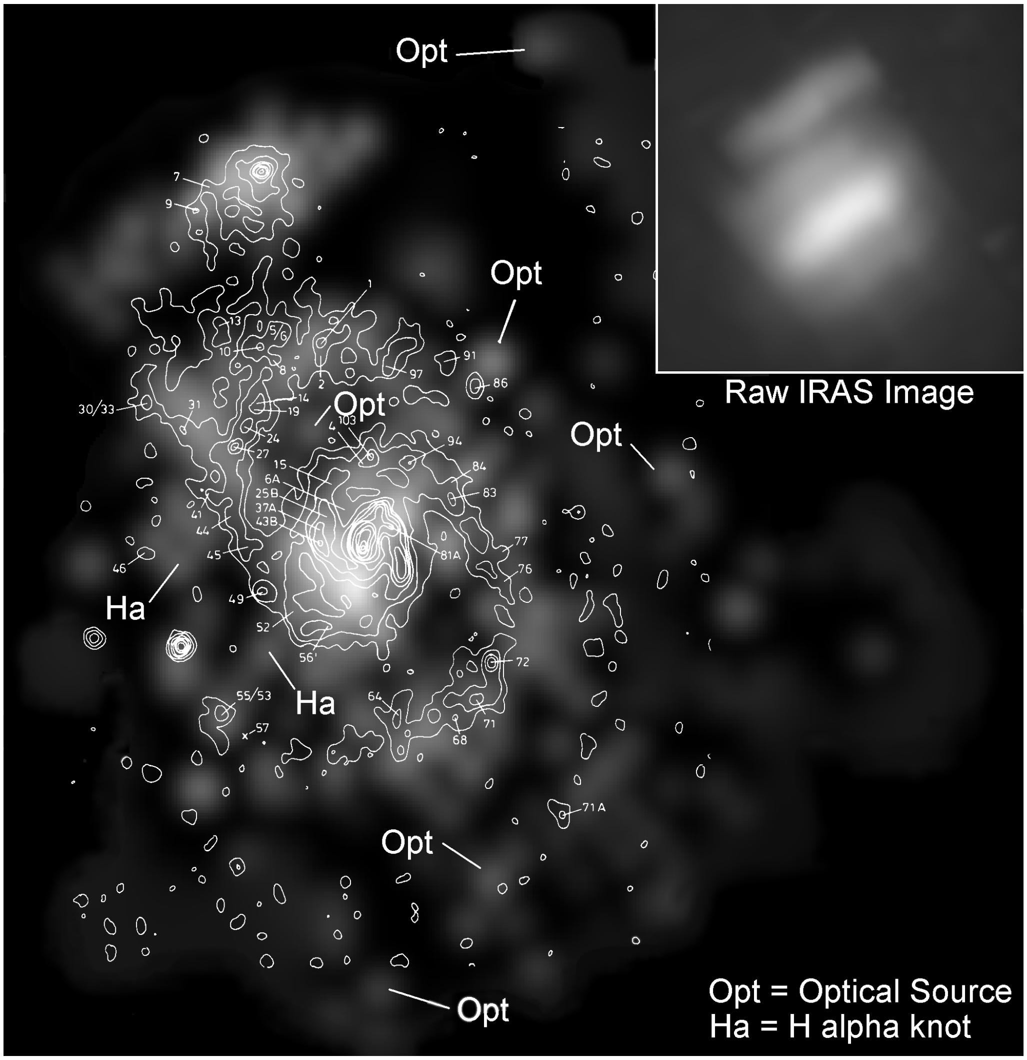

FIGURE 6. Pixon-Based reconstruction of $60\mu m$ survey scans of M51. Grey-scale image: The pixon-based reconstruction. The faintest sources visible have a formal SNR of roughly 30. Contours: The 5 GHz radio flux contours of van de Hulst 1988. Noted are several of the stronger features in the reconstruction that can be indentified with optical sources (labeled "Opt") and $H\alpha$ knots (labeled "$H\alpha$"). Figure reproduced from Puetter and Piña 1993b.

produced by other methods. The Lucy-Richardson and MCM reconstructions fail to significantly reduce image spread in the cross-scan direction, i.e. the rectangular signature of the 1.5 by 4.75 *arcmin* detectors is still clearly evident, and fail to reconstruct even gross features such as the "hole" (black region) in the emission north of the nucleus—this hole is clearly evident in optical images of M51. The MEMSYS 3 reconstruction by Bontekoe is significantly better. This image clearly recovers the emission "hole" and resolves the north-east and south-west arms of the galaxy into discrete sources. Nonetheless, the level of detail present in the FPB reconstruction is clearly absent, e.g. the weak source centered in the emission hole (again, this feature corresponds to a known optical source).

To assess the significance of the faint sources present in our FPB reconstruction, our reconstruction in Figure 6 is overlaid with the 5 GHz radio contours of van der Hulst

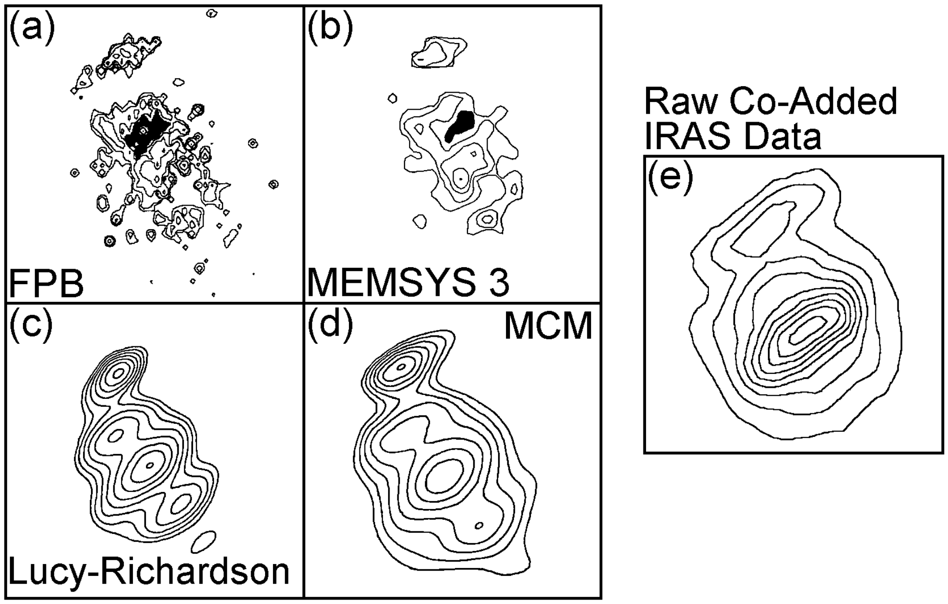

FIGURE 7. Comparison of the performance of various image reconstruction methods for the M51 IRAS survey data. Shown are the results for our pixon-based method (fuzzy, radially symmetric pixons), the Lucy-Richardson method, the "in-house" IPAC Maximum Correlation Method (MCM), and the Maximum Entropy Method (MEMSYS 3). Panel (b) is reproduced from Bontekoe 1991b by permission of the author, and panels (c) and (d) are reproduced from Rice 1993 by permission of the author. This figure first appeared in Puetter and Piña 1993a.

et al. (van de Hulst et al. 1988). The radio contours are expected to have significant, although imperfect, correlation with the far infrared emission seen by IRAS. Hence a comparison of the two maps should provide an excellent test of the reality of structures found in our reconstruction. Also identified in Figure 6 are several prominent optical sources and (hydrogen Balmer line emission) knots. As can be seen, the reconstruction indicates excellent correlation with the radio. The central region of the main galaxy and its two brightest arms align remarkably well, and the alignment of the radio emission from the north-east companion and the IRAS emission is excellent. Furthermore, for the most part, whenever there is a source in the reconstruction which is not identifiable with a radio source, it can be identified with either optical or knots. An excellent example is the optical source in the "hole" of emission to the north-east of the nucleus of the primary galaxy or the bright optical source to the north-west of the nucleus (both labeled "Opt" in Figure 6). Because of the excellent correlation with the radio, optical, and images, we are quite confident that all of the features present in our reconstruction are real. (The faintest sources visible in Figure 6 have a formal SNR of 30.)

We next present a sample image reconstruction of coded mask X-ray data from the Hard X-ray Telescope (HXT) on board the Yohkoh spacecraft (Kosugi et al. 1991). Figure 8 shows hard X-ray (23-33 keV) images of a solar flare which occurred on 20 August 1992. Since there is no effective method of manufacturing optics with which to focus hard X-ray light, the HXT instrument takes X-ray images by taking pictures through a series of coded masks. The series of coded images is then inverted to yield the underlying source structure. Figure 8 shows a time series of 3 images. Each row of panels shows three different reconstructions. All inversions were performed by Tom Metcalf (U. Hawaii). As can be seen, direct linear inversion produces an enormous amount of spurious structure. In addition to hiding low contrast features, the presence of such spurious structure is particularly worrisome since the flux conserving nature of the algorithm requires that flux placed in spurious sources must come from the true sources thereby grossly affecting photometry. In this regard, the ME and FPB reconstructions can be seen to be a great improvement over the direct linear inversion. Relative to the pixon-based reconstruction, however, the ME inversion still produces a wealth of spurious emission, resulting in poor photometry and often over resolves real features. (We know the resolution of the ME image is too high since the quality of the pixon fit is just as good—in fact slightly better for these images—and uses a lower resolution. Hence the ME resolution is unjustified.)

As a final example of pixon-based image reconstruction, we present a comparison of two image reconstruction techniques using 50 to 150 keV data from the Oriented Scintillation Spectrometer Experiment (OSSE) aboard the Compton Gamma-Ray Observatory (GRO). The OSSE instrument consists of four shielded detectors with a field of view of $3.8^\circ \times 11.4^\circ$ (FWHM). Each detector is mounted on an independent single-axis pointing system which allows for sub-stepping the detector field of view. Figure 9 presents a comparison of a pixon-based reconstruction of the OSSE survey data with the Non-Negative Least-Squares (NNLS) method developed at UCR. Both the NNLS and pixon-based reconstruction presented in Figure 9 were performed by D. Dixon of UC, Riverside (Dixon et al. 1996). Details of the UC, Riverside implementation of the pixon-based algorithm are given below. The dark area of the figure represents all of the points for which there is significant exposure time during the scanned observation. Each of the pixels in the reconstructed images in Figure 9 has an angular size of $2^\circ \times 2^\circ$.

As can be seen from Figure 9, the NNLS reconstruction produces an image which has the appearance of a random noise field. From this image, it is unclear whether or not there are any real detections of sources. On the other hand, the pixon-based reconstruction clearly finds the two bright sources expected to be seen in this data, i.e. 3C 273 and NGC 4388 (the active galaxy M87 is also in the pixel occupied by NGC 4388, but is not believed to contribute significant 50-150 keV emission), and it can be seen that the pixon-based method was very successful in suppressing spurious sources in the reconstruction.

The OSSE data, and γ-ray data in general, present an especially difficult challenge for image reconstruction methods. This is because the signal-to-noise of the collected data is very low. At γ-ray energies there are only very few photons to count. Consequently the standard pixon-based methods developed at UCSD were found to be inadequate—the conjugate gradient method could not find the global minimum. It was for this reason that D. Dixon (U.C., Riverside) decided to use a simulated annealing approach to finding the optimal pixon reconstruction. As previously discussed, simulated annealing is well known for its robust ability to find shallow global minima in the presence of numerous local minima. However, this ability has significant computational costs. To speed the method, Dixon adopted a mean field approach and used only two pixon scales, one appropriate for point sources and one appropriate for the diffuse background. Given the expected nature of the image, this is quite a suitable assumption and gives rise to the excellent reconstruction presented in Figure 9. It is clear from the success of this method, that pixon-based approaches such as that of Dixon et al. (Dixon et al. 1996) have dramatically increased the scientific capability of the OSSE instrument. Current plans at UC, Riverside are to apply these techniques to COMPTEL data as well, and to explore simulated annealing pixon-based approaches with multiple pixon scales.

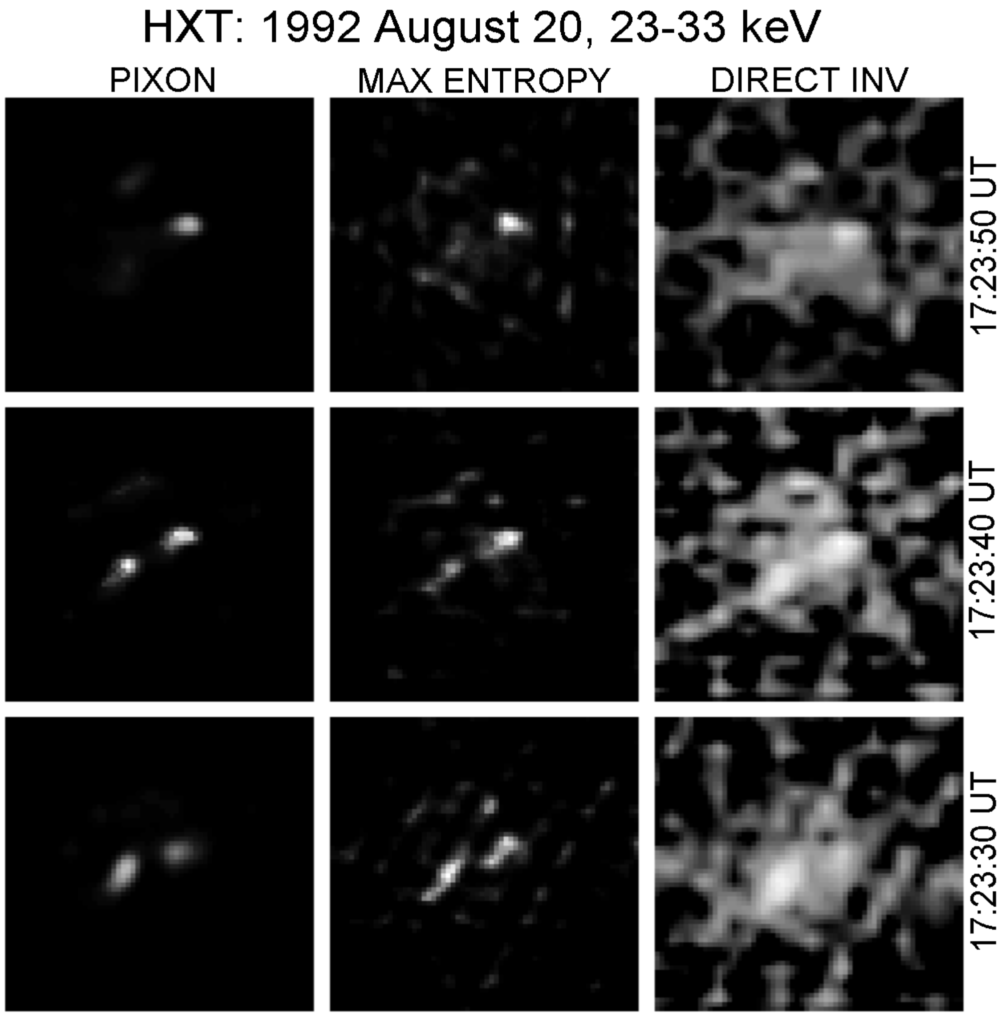

FIGURE 8. Hard X-ray (23-33 keV) images of the 20 August 1992 solar flare. The observations are from the hard X-ray Telescope (HXT) on board the Yohkoh satellite. Each column shows a time series of three image reconstructions. The background patterns which show up in the direct inversion and the ME inversion are artifacts resulting from the sparse UV-plane coverage in the HXT instrument. The pixon-based reconstruction essentially eliminates all of these artifacts. The reconstruction were performed by Metcalf of the University of Hawaii. Figure reproduced from Puetter 1995.

Acknowledgements

The author would like to acknowledge the many valuable contributions from his collaborators and colleagues, epecially Robert Piña, the co-developer of the pixon concept. The author also thanks N. Weir for insightful conversations and for gratiously performing the MEMSYS 5 reconstruction presented in Figure 4. The author thanks Romke Bontekoe and Do Kester for providing the M51 IRAS test data set and valuable conversations, Steven Gull and John Skilling for conversations regarding the role of the pixon in Bayesian image reconstruction, and James Graham for making his $2\mu m$ Keck data and ME reconstruction available. Finally, the author would also like to thank Tom Metcalf and Dave Dixon for extending pixon-based techniques to X-ray and γ-ray data and for

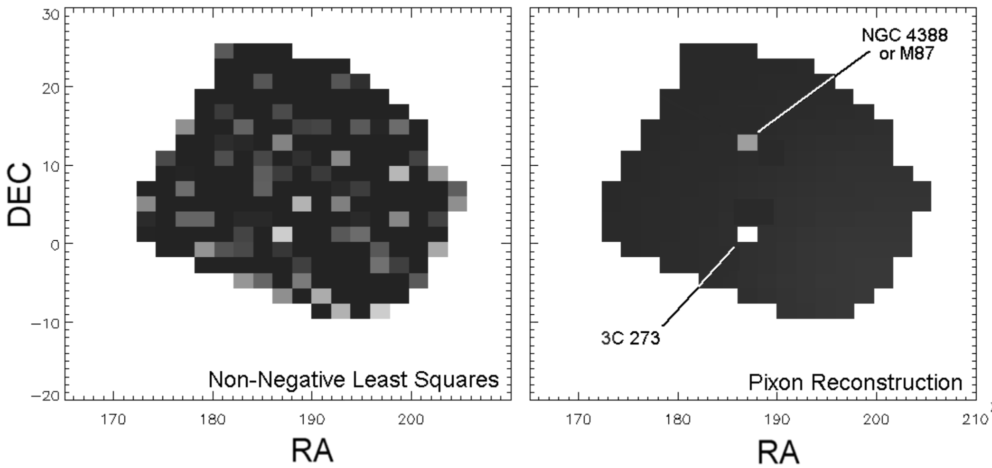

FIGURE 9. Non-Negative Least Squares (NNLS) and pixon-based image reconstructions from the 50-150 keV data of the OSSE Virgo Survey. The reconstructions were performed by Dixon of the University of California, Riverside. While the NNLS method gives a result that looks like random noise, Dixon's simulated annealing, pixon-based method clearly reveals the two strong sources expected in these data, i.e. NGC 4388 and 3C 273. It also detects a mild gradient in the γ-ray background. Figure reproduced from Puetter 1995.

contributing their insights to pixon-based methods. This research was supported by the NSF, NASA, the California Association for Research in Astronomy, and Cal Space.

REFERENCES

Bontekoe, T. R. (1991), The Image Reconstruction Contest, in *Maximum Entropy and Bayesian Methods*, eds. W. T. Grady, Jr., and L. H. Schick, Kluwer Academic.

Bontekoe, T. R., Kester, D. J. M., Price, S. D., de Jonge, A. R. W., and Wesselieus, P. R. (1991), Image Reconstruction from the IRAS Survey, *Astron. and Astrophys.*, **248**, pp. 328-336, ed.s W. T. Grady, Jr., and L. H. Schick, Kluwer Academic.

Broadhurst, T.; Lehar, J. (1995), A Gravitational Lens Solution for the IRAS Galaxy FSC 10214+4724, *Ap. J. (Letters)*, **450**, pp. L41-4.

Bryan, R. K., and Skilling, J. (1980), *M.N.R.A.S.*, **191**, p. 69.

Chaitin, G. J. (1966), *J. Ass. Comput. Mach.* **13**, p. 547.

Chaiten, G. J. (1982), Algorithmic Information Science, *Encyclopedia of Statistical Science*, **1**, pp. 38-41. Wiley.

Cherry, C. (1978), *On Human Communication,* 3rd edition, MIT Press.

Close, L.M.; Hall, P.B.; Liu, C.T.; Hege, E.K. (1995), Spectroscopic and Morphological Evidence that IRAS FSC 10214+4724 is a Gravitational Lens, *Ap. J. (Letters)*, **452**, pp. L9-12.

Cohen, J. G. (1991), in *Maximum Entropy and Bayesian Methods*, ed.s W. T. Grady, Jr. and L. H. Schick, Kluwer Academic Publishers.

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.

Dixon, D. D., Tumer, O. T., Kurfess, J. D., Purcell, W. R., Wheaton, W. A., Piña, R. K., and Puetter, R. C. (1996), Pixon-Based Image Reconstruction from OSSE Data, *Ap. J.*, submitted.

Eisenhadrt, P. R., Armus, L., Hogg, D. W., Soifer, B. T., Neugebauer, G., and Werner, M. W. (1996), HST Observations of the Luminous IRAS Source FSC 10214+7424: A Gravitaionally Lensed Infrared Quasar, *Ap. J.*, in press.

Feinstein, A. (1958), *Foundations of Information Theory*, McGraw-Hill.

Graham, J. R., and Liu, M. C. (1995), High Resolution Infrared Imaging of FSC 10214+4724: Evidence for Gravitational Lensing, *Ap. J. (Letters)*, **449**, p. L29

Gull, S. F., and Skilling, J. (1991), *MemSys5 Quantified Maximum Entropy User's Manual.*

Hobson, A. (1971), *Concepts in Statistical Mechanics*, Gordon and Breach.

Jaynes, E. T. (1957), Information Theory and Statistical Mechanics, *Phys. Rev.*, **106**, pp. 171-190.

Jaynes, E. T. (1963), Information Theory and Statistical Mechanics, in *Statistical Physics*, ed. K. W. Ford, pp. 181-218. W. A. Benjamin, Inc.

Khinchin, A. I. (1957), *Mathematical Foundations of Information Theory*, Dover.

Kikuchi, R., and Soffer, B. H. (1976), in *Image Analysis and Evaluation*, Society of Photographic Scientists and Engineers, Toronto, Canada, July 1976, p. 95.

Kittel, C. (1969), *Thermal Physics*, Wiley.

Kolmogorov, A. N. (1965), *Inf. Transmission*, **1**, p. 3.

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.

Kosugi, T., Makishima, K., Murakami, T., Sakao, T., and others (1991), The Hard X-Ray Telescope (HXT) for the Solar-A Mission, *Solar Phys.*, **136**, pp. 17-36.

MacKay, D. J. C. (1994), in *Models of Neural Networks III*, ed.s E. Domany, J. L. van Hemmen, and K. Schutten, Chapter 6, Springer-Verlag.

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.

Narayan, R. and Nityananda, R. (1986), Maximum Entropy Image Restoration in Astronomy, *Ann. Rev. Astron. & Astrophys.*, **24**, p. 127.

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. (1993b), Pixon-Based Image Reconstruction, *Proc. MaxEnt '93*, in press

Puetter, R. C., and Piña, R. K. (1994), Beyond Maximum Entropy, in *Science with High Spatial Resolution Far-IR Data*, ed.s S. Tereby 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.*, **6**, pp. 314-331.

Puetter, R. C. (1996a), The Image Restoration/Reconstruction Problem, in *Instrumentation for Large Telescopes*, in press.

Puetter, R. C. (1996b), Language and InformationContent, in *Instrumentation for Large Telescopes*, in press.

Puetter, R. C. (1996c), Pixon-Based Image Restoration/Reconstruction, in *Instrumentation for Large Telescopes*, in press.

Puetter, R. C. (1996d), The Restoration of Astronomical Images, in *Instrumentation for Large Telescopes*, in press.

Puetter, R. C. (1996e), Reconstruction of Complexly Encoded Images, in *Instrumentation for Large Telescopes*, in press.

Rabbani, M., and Jones, P. W. (1991), *Digital Image Compression Techniques*, SPIE Optical Engineering Press.

Rice, W. (1993), An Atlas of High-Resolution IRAS Maps of Nearby Galaxies, *A. J.*, **105**, pp. 67-96.

Shannon, C. E. (1948), A Mathematical Theory of Communication, *Bell Systems Tecnical Journal*, **27**, 379.

Shannon, C. E., and Weaver, W. (1949), *The Mathematical Theory of Communication*, University of Illinois Press.

Sibisis, S. (1990), iin *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, Kluwer Academic Publishers.

Skilling, J. (1989), Classic Maximum Entropy, in *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, p. 45. Kluwer Academic Publishers.

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 $17\mu m$, *M.N.R.A.S.*, **273**, pp. 354-358.

Solomonoff, R. (1964), *Inf. Control*, **7**, p. 1

van de Hulst, J. M., Kennicutt, R. C., Crane, P. C., and Rots, A. H. (1988), Radio Properties and Extinction of the H II Regions in M51, *Astron. and Astrophys.*, **195**, pp. 38-52.

Weir, N. (1991), Applications of Maximum Entropy Techniques to HST Data, in Proc. of the ESO/ST-ECF Data Analysis Workshop, April 1991, ed.s P. Grosbo and R. H. Warmels, p. 115. ESO

Weir, N. (1993), private communication.

Weir, N. (1994), A Maximum Entropy-Based Model for Reconstructing Distributions with Correlations at Multiple Scales, *J. Opt. Soc. Am.*, in press.

Wicken, J. (1987), Entropy and Information: Suggestions for a Common Language, *Philos. Sci.*, **54**, pp. 176-193.

Zurek, W. H. (1989), Thermodynamic Cost of Computation, Algorithmic Complexity and the Information Metric, *Nature*, **341**, p. 119