PIXON-BASED MULTIRESOLUTION IMAGE RECONSTRUCTION AND QUANTIFICATION OF IMAGE INFORMATION CONTENT

R.C. PUETTER

Center for Astrophysics and Space Sciences
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0111[†]

Abstract. This paper describes the theory of pixon-based image reconstruction. After a brief introduction of the basic concepts of the pixon, the paper concentrates primarily on our current implementation of the techniques along with the approximations and shortcuts that seem to provide practical software algorithms. We then present an example of application of the method to astrophysical data.

Key words: pixons, image reconstruction, Bayesian estimation, multiresolution algorithmic complexity, maximum entropy, maximum likelihood

1. Image Reconstruction and the Inverse Problem

This paper presents a practical tutorial on how to implement pixon-based image reconstruction. Pixons (or informatons—their extension to generalized data sets), represent an optimized, information theory-based coordinate system for performing 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 methods and Maximum Entropy methods. Like many of these methods, pixon-based reconstruction can be viewed within the Bayesian estimation framework. Hence before going into the particulars of pixon-based image reconstruction, we shall describe Bayesian methods briefly and the interpretation of pixon-based concepts within this formalism. Since we intend this paper to provide a practical tutorial for the implemetation of pixons, we shall gloss over much of the theoretical discussion presented in other papers on pixons. For readers more interested in these topics, we refer you to the papers [1–4], and especially my review article [5].

[†]Email: rpuetter@ucsd.edu

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:

$$\begin{aligned} p(D, I, M) &= p(D|I, M)p(I, M) && (1) \\ &= p(D|I, M)p(I|M)p(M) && (2) \\ &= p(I|D, M)p(D, M) && (3) \\ &= p(I|D, M)p(D|M)p(M) && (4) \\ &= p(MI|D, I)p(D, I) && (5) \\ &= p(MI|D, I)p(D|I)p(I) \quad , && (6) \end{aligned}$$

where $p(x|y)$ is the conditional probability of x given y, i.e. the probability of x given that y has a particular value. Equations (1) through (6) can be rearranged to give:

$$\begin{aligned} p(I|D, M) &= \frac{p(D|I, M)p(I|M)p(M)}{p(D|M)p(M)} && (7) \\ &= \frac{p(D|I, M)p(I|M)}{p(D|M)} && (8) \\ &\propto p(D|I, M)p(I|M) \quad , && (9) \end{aligned}$$

or

$$\begin{aligned} p(I, M|D) &= \frac{p(D|I, M)p(I, M)}{p(D)} && (10) \\ &\propto p(D|I, M)p(I|M) \quad . && (11) \end{aligned}$$

The formulae given in equations (7)-(9) are the basic starting point for Bayesian image reconstruction. These equations essentially assume that the appropriate model, M, is known and will not be varied during the reconstruction process. This assumption is made more explicit in equation (9) where the term $p(D|M)$ is dropped (it is assumed to be constant—M and D are not varied during the reconstruction). Our preferred formulation, however, is given in equations (10)-(11). Here, we do not assume that all of the parameters associated with the model are "nuisance" parameters. Indeed, in pixon-based reconstruction the model parameters associated with the local smoothness scale (see below) are extremely important. Hence we have chosen to keep some of the "model" parameters on an equal footing with the image, and to seek the most likely estimate of their values simultaneously with the image.

3. Choice of Image Representation and Pixon-Based Methods

Notable successes of previous methods of Bayesian image reconstruction, e.g. Maximum Likelihood (ML) and Maximum Entropy (ME) methods are greatly reduced noise propagation errors relative to linear methods (e.g. Fourier convolution inversion) and consequently greatly enhanced resolution and spurious source rejection. While these methods represent a significant advance, they still display significant problems. ML methods typically over-fit the data and ME methods systematically under-estimate source brightnesses. Both methods still suffer from spurious source production although at lower levels than linear methods.

One of the fundamental sources of the problems with these methods is their selection of the coordinate system in which to represent the image. For most modern electronic imaging systems, the seemingly natural choice for the image coordinate system is the same rectangular pixel coordinate grid that is used for the data. However this is actually a very arbitrary coordinate system. This is easily seen to be so if the data is very noisy or the underlying image is intrinsically very smooth. In either case the collected data typically is inadequate to constrain as many independent numbers as are contained in the pixel coordinate grid. Such a situation will give rise to numerical artifacts since the inversion algorithm cannot distinguish between solutions contained in the huge solution space. Furthermore, since the typical member of the solution space has many spurious sources, the selection of such a solution is virtually guaranteed. The only way to prevent these problems is to constraint the solution in some manner.

Both Bayesian methods and concepts from information science suggest parts of the solution to this problem. Modern Bayesian interpretations of Occam's razor indicate that all that is required to solve the spurious source problem is to use the simplest hypothesis to explain the data consistent with its noise properties. In other words we need to optimize the prior for the problem while keeping the Goodness-of-Fit criterion, $p(D|I, M)$, optimized. The information science concepts of Algorithmic Information Content (AIC) further tells us that the terseness of a description (and hence its simplicity) is a strong function of the complexity (descriptiveness) of the language in which it is expressed [6–9]. Thus it seems in order to achieve the Bayesian ideal of Occam's razor we need to select a descriptive language in which the image can be described in the most concise manner. Since the language for describing the image is normally associated with the model, i.e. part of the nuisance parameters, we see that in order to achieve this goal we must be willing to vary the "image description language", which is part of M, as part of the problem.

We are now left with the problem of how to select a highly descriptive language for generic images. Again, work in other disciplines that deal routinely with images suggest at least one practical solution. In a variety of endeavors (e.g. image classification, morphology, and segmentation; computational methods; image compression; image restoration; video encoding and high definition TV; wavelet theory and applications; medical imaging; neural networks, and motion detection—see [5] and references therein) multiresolution image languages are receiving considerable attention. 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—the grain size is smaller than the finest scale required to be captured. It is quite clear, for example, that one approach to concisely describing the information in an image is to use 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. Indeed, each of these degrees-of-freedom might 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. In the section below, we shall outline in more detail our specific implementation of a multiresolution, pixon language for image description.

4. A Multiresolution Pixon Language for Image Description

In the section above, we argued that a multiresolution image description language would be suitable (i.e. concise) for describing generic images. In fact, what we discovered is that in order to optimize the Bayesian prior we need the simplest image/model combination possible. While detailed prior knowledge of the properties of the image would allow even greater optimization of the prior, our knowledge is usually incomplete. That is why we are talking about *generic* images and *generic* languages to describe images. Use of detailed prior knowledge will always allow us to construct better image models, but if we are clever about our construction of models for generic images, we should not do too badly for most types of images we shall encounter.

Again, since our goal is to optimize the image/model prior, $p(I, M)$, we see that this can be accomplished by model simplification, i.e. using the fewest degrees-of-freedom necessary to fit the data within the accuracy demanded by the noise [remember the product of the terms $p(D|I, M)$ and $p(I, M)$, i.e. $p(I, M|D)$ is the quantity to be optimized]. Thus any scheme which controls the local density of DOFs used to describe the local image information is suitable. Thus 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

$$I(\vec{x}) = (K \otimes I_{pseudo})(\vec{x}) \tag{12}$$

$$(K \otimes I_{pseudo})(\vec{x}) = \int dV_{\vec{y}}\ K(\vec{x}, \vec{y}, \delta(\vec{x}))I_{pseudo}(\vec{y}) \tag{13}$$

$$K\left(\vec{x},\vec{y},\delta(\vec{x})\right) = K\left(\frac{\|\vec{x}-\vec{y}\|}{\delta(\vec{x})}\right) \quad \text{;for radially symmetric pixons} \quad ,(14)$$

where equations (12) through (14) 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 (13) is carried out over volume in pseudo-image space. We have also indicated in equation (14) 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 ratioed 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.

$$K\left(\vec{x},\vec{y},\delta(\vec{x})\right) = \begin{cases} \left(1-\frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) / \int dV_{\vec{y}} \left(1-\frac{\|\vec{x}-\vec{y}\|^2}{\delta(\vec{x})^2}\right) & ;\ \|\vec{x}-\vec{y}\| \le \delta(\vec{x}) \\ 0 & ;\ \|\vec{x}-\vec{y}\| > \delta(\vec{x}) \end{cases} \qquad (15)$$

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 below. 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.

5. 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 $p(I,M|D)$. As is clear from equation (11), we need to maximize the product of $p(D|I,M)$ and $p(I,M)$. The appropriatel choice for $p(D|I,M)$ in the case of pixelized data with independent Gaussian noise is $p(D|I,M) = (2\pi\sigma^2)^{1/n}\exp(-\chi^2/2)$ 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). A choice for $p(I,M)$ is less obvious. The choice of Maximum Entropy enthusiasts is something like

$$p(I|M) = \frac{N!}{n^N \prod_{i=1}^n N_i!} = e^S \qquad (16)$$

$$p(M) = constant \quad , \qquad (17)$$

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 with the only exception being that the pixel cells are not the most appropriate choice for an image's basis. A more appropriate choice would be the multiresolution pixon basis discussed above. While the ME prior of equations (16)-(17) states that the most likely *a priori* image is flat in the pixel coordinate system, evaluating equation (16) using fuzzy pixon cells rather than pixels, would state that once the structure of the image is known (i.e. all the local smoothness scales are determined) we would guess *a priori* that each of the cells would have equal brightness.

In essence then, 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 (16)-(17) remain valid. It is just that we now recognize that the pixons are the appropriate coordinate system and that we will probably use vastly fewer pixons (i.e. DOFs) to describe the image than there are pixels in the data. To make the formulae explicit, we would define the pseudo-image on a psuedo-grid which is as least as fine as the pixel grid (we use this finer grid for the image too) and then use the following substitutions:

$$\begin{aligned}
p(I|M) &= \frac{N!}{n^N \prod_{i=1}^{n} N_i!} && (18)\\
N &= \sum_{i=1}^{n_{pixons}} N_i = \sum_{j=1}^{n_{pixels}} N_j^p && (19)\\
N_i &= \int dV_{\vec{y}}\ k(\vec{x},\vec{y},\delta(\vec{x})) I_{pseudo}(\vec{y}) && (20)\\
p(\vec{x}) &= 1 \ / \int dV_{\vec{y}}\ k(\vec{x},\vec{y},\delta(\vec{x})) && (21)\\
k(\vec{x},\vec{y}=\vec{x},\delta(\vec{x})) &= 1 && (22)\\
p(M) &= constant && (23)
\end{aligned}$$

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 $\vec{y}=\vec{x}$.

To obtain the M.A.P. image/model pair, one could now proceed directly by minimizing the product of $(2\pi\sigma^2)^{1/n_{pixels}} \exp(-\chi^2/2)$ and equation (18) 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 have done in practice. Instead, we have divided 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. 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 the the pixon prior of equation (18) increases rapidly as the number of pixons are decreased and the remaining

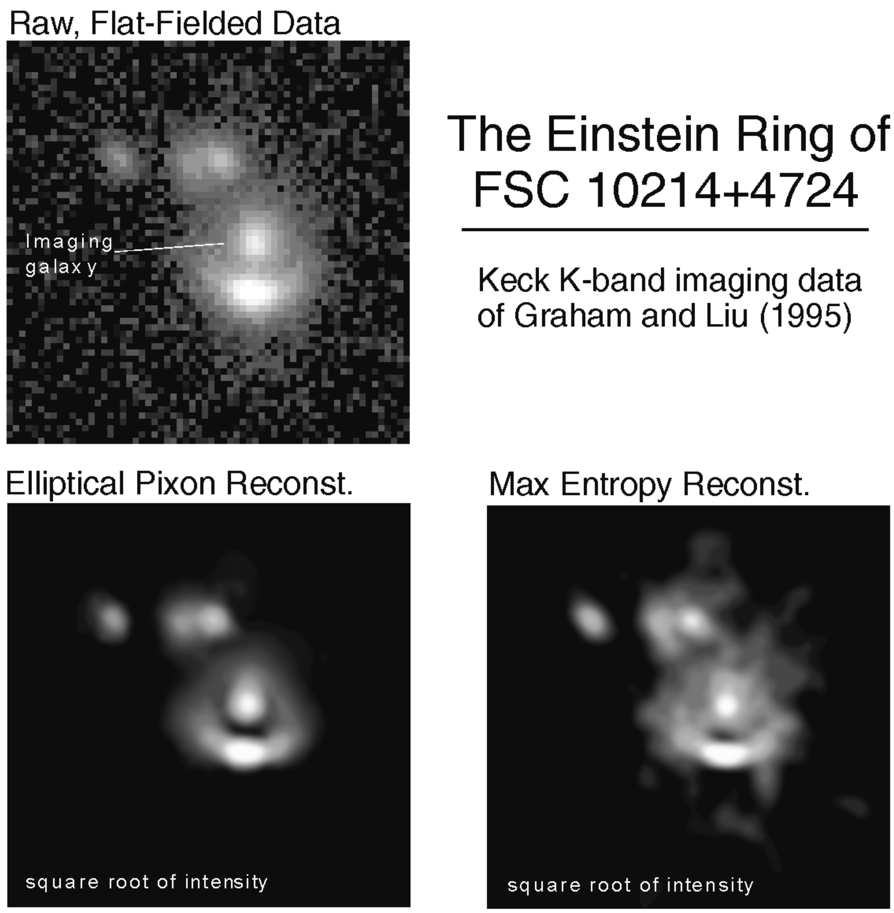

Figure 1. Comparison of image reconstruction techniques for the Einstein ring of FSC 10214+4724. The elliptical pixon reconstruction used 196 different elliptical pixon kernel functions. Relative to radially symmetric reconstructions, the elliptical pixon reconstruction used roughly 50% fewer DOFs in the image model.

pixons are packed as full of signal as possible. So what we do is, 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 pixons scales. We then use our pixon-calculator [the code that performs step (2) by selecting the largest scales that still provide an adequate fit to the data] to determine new scales. The pixon-calculator reports 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 use the smallest scales requested. Instead, we let the scales get somewhat smaller over the portion of the image which was determined to require smaller scales 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. The process has proven to be very robust.

6. A Sample Restoration: The Einstein Ring in FSC 10214+4724

Figure 1 presents an elliptical pixon-based reconstruction of the Keck Telescope K-band imaging data of [10]. Also shown is an ME reconstruction of the same data from [10]. As can be seen, the pixon-based reconstruction reveals considerably more of the Einstein ring. This improved sensitivity is possible because of the ability of pixon-based methods to develop a critical image model. Also note the rejection by the pixon-based reconstruction of the many extended fine structure features in the ME reconstruction. The apparently finer resolution in the ME reconstruction is unwarranted since the pixon-based solution has an equally good statistical fit.

7. Acknowledgements

The author would like to acknowledge the many valuable contributions from his collaborators and colleagues, epecially R. Piña and N. Weir. The author also thanks J. Graham for making his data and ME reconstruction available. This research was supported by the NSF, NASA, the California Association for Research in Astronomy, and Cal Space.

References

1. 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
2. Puetter, R.C., and Piña, R.K. (1993a) The Pixon and Bayesian Image Reconstruction, *Proc. S.P.I.E.*, **1946**, pp. 405–416
3. Puetter, R.C., and Piña, R.K. (1993b) Pixon-Based Image Reconstruction, *Proc. MaxEnt '93*, in press
4. Puetter, R.C. (1994) Pixons and Bayesian Image Reconstruction, *Proc. S.P.I.E.*, **2302**, pp. 112-131
5. Puetter, R.C. (1995) Pixon-Based Multiresolution Image Reconstruction and the Quantification of Picture Information Content, **Int. J. Image Sys. & Tech.**, in press
6. Solomonoff, R. (1964) *Inf. Control*, **7**, p. 1
7. Kolmogorov, A.N. (1965) *Inf. Transmission*, **1**, p. 3
8. Chaitin, G.J. (1966) *J. Ass. Comput. Mach.*, **13**, p. 547
9. Zurek, W.H. (1989) Thermodynamic Cost of Computation, Algorithmic Complexity and the Information Metric, *Nature*, **341**, p. 119
10. 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