BEYOND MAXIMUM ENTROPY: FRACTAL PIXON-BASED IMAGE RECONSTRUCTION

R. C. PUETTER AND R. K. PIÑA
Center for Astrophysics and Space Science,
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0111

ABSTRACT We have developed a new Bayesian image reconstruction method that has been shown to be superior to the best implementations of other competing methods, including Goodness-of-Fit methods such as Least-Squares fitting and Lucy-Richardson reconstruction, as well as Maximum Entropy (ME) methods such as those embodied in the MEMSYS algorithms. Our new method is based on the concept of the pixon, the fundamental, indivisible unit of picture information. Use of the pixon concept provides an improved image model, resulting in an image prior which is superior to that of standard ME. Our past work has shown how uniform information content pixons can be used to develop a "Super-ME" method in which entropy is maximized exactly. Recently, however, we have developed a superior pixon basis for the image, the Fractal Pixon Basis (FPB). Unlike the Uniform Pixon Basis (UPB) of our "Super-ME" method, the FPB basis is selected by employing fractal dimensional concepts to assess the inherent structure in the image. The Fractal Pixon Basis results in the best image reconstructions to date, superior to both UPB and the best ME reconstructions. In this paper, we review the theory of the UPB and FPB pixon and apply our methodology to the reconstruction of far-infrared imaging of the galaxy M51. The results of our reconstruction are compared to published reconstructions of the same data using the Lucy-Richardson algorithm, the Maximum Correlation Method developed at IPAC, and the MEMSYS ME algorithms. The results show that our reconstructed image has a spatial resolution a factor of two better than best previous methods (and a factor of 20 finer than the width of the point response function), and detects sources two orders of magnitude fainter than other methods.

BAYESIAN IMAGE RECONSTRUCTION

Bayesian image reconstruction estimates the best image by statistically modeling the imaging process. To do this, one factors the joint probability distribution of the triplet, D, I, and M, i.e., $p(D, I, M)$, (where D, I, and M are the data, unblurred image, and model respectively) using Bayes' Theorem to derive an expression for the most probable or M.A.P. (Maximum A Posteriori) image. In deriving the expression for the M.A.P. image, most Bayesian methods assume that all aspects of the model, M, linking image and the data are to be held

S. Terebey and J. Mazzarella (eds.), *Science with High-Spatial Resolution Far Infrared Data*, 61–67, 1994, Jet Propulsion Laboratory, Pasadena, CA.

constant during the image reconstruction process. More advanced methods, such as Weir's multi-channel method (1991, 1993) and the method presented here (also see Piña and Puetter 1993 and Puetter and Piña 1993), systematically vary certain aspects of the model to improve the quality of the image reconstruction. Varying the image and the model simultaneously, places I and M on a more equal footing and begins to blur the distinction between image and model. Indeed, our Bayesian formulation of the image reconstruction problem treats them as full equals; see Piña and Puetter (1993) and Puetter and Piña (1993). With this in mind, a Bayesian formulation of the image reconstruction problem which is symmetric with respect to the image and model can be obtained by factoring $p(D, I, M)$ in the following manner:

$$p(D, I, M) = p(D|I, M)p(I, M) = p(I, M|D)p(D) \tag{1}$$

which implies

$$p(I, M|D) = \frac{p(D|I, M)p(I, M)}{p(D)} = \frac{p(D|I, M)p(I|M)p(M)}{p(D)} \propto p(D|I, M)p(I|M) \tag{2}$$

where in the standard notation $p(X|Y)$ is the probability of X given that Y is known, and the proportionality in equation (2) is obtained since D is constant (hence $p(D)$ is constant) and since we have no prior basis on which to discriminate between models (i.e., $p(M) = const$). Equation (2) is an expression for the M.A.P. image/model pair, i.e., the image/model pair which maximizes $p(I, M|D)$. Our method assumes that deriving the M.A.P. image/model pair represents the optimal reconstruction.

The significance of the two terms in the proportionality in equation (2) is readily understood. The first 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 second term, $p(I|M)$, the "image prior", expresses the *a priori* probability of a particular realization of the image; note that equivalently we could use the image/model prior $p(I, M)$. In GOF image reconstruction, $p(I|M)$ is effectively set equal to unity, i.e., there is no prior bias concerning the image. A typical choice for $p(D|I, M)$ is to use $p(D|I, M) = \exp(-\chi^2/2)$, i.e., the standard chi-square distribution. This approach ensures a faithful rendition of the data, but typically results in images with spurious low signal-to-noise features. 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 a Lagrange multiplier that adjusts the relative importance of the GOF and image prior.

PIXONS AND A NEW IMAGE PRIOR

By varying the model simultaneously with the image during the reconstruction process, the solution to the reconstruction process is optimized over a significantly larger solution space. This solution space contains, for example, the solutions spaces of the standard GOF and ME methods. Consequently, the results of our methods theoretically should be no worse than those of competing

methods. While the advantage of varying the model seem clear, it is less obvious how to do this in a productive manner. Certain aspects of the model, for example, should not be varied. The model includes, for instance, a description of the physics of the imaging process, e.g., that the data is the convolution of the point-spread-function (PSF) and the blur-free underlying image. The model also includes the description of the noise processes, e.g., the noise might have a Gaussian or Poisson distribution. These aspects of the model should not be varied if ones wishes to remain faithful to the truth. However, the mathematical model or representation of the image can be varied. Furthermore, varying the mathematical model of the image can substantially improve the quality of the image reconstruction. One relatively successful approach to image modeling was suggested by Piña and Puetter 1993 and Puetter and Piña 1993. They pointed out that in the most general of terms, an image is a collection of distinguishable events which occur in distinct cells. Hence the value for the image prior can be determined from simple counting arguments. If there are N_i events in cell i, and a total of n cells, then the prior probability of that particular image is:

$$p(\{N_i\}, n, N) = \frac{N!}{n^N \prod_i N_i!} = p(I|M) \quad , \tag{3}$$

where $\{N_i\}$ is the set of all numbers of events in cells i, and N is the total number of events, i.e., $N = \sum_i N_i$. The cells used in equation (3) are quite general. In fact, we have not specified a size, shape, or position for the cells. The goal of selecting cells for a specific image is to maximize $p(I|M)$ subject to the constraint that an adequate GOF is maintained. Equation (3) clearly indicates the desirable properties of the model, i.e., the model should contain the fewest number of cells with each containing the largest number of events. We have termed these generalized cells *pixons*. This terminology is not simply whimsical. An image's pixons represent the smallest number of cells (of arbitrary shape, position, etc.) required to fit the data and represent the minimum degrees of freedom necessary to specify the image within the accuracy of the noise. If properly selected, this set is not reducible to a smaller set. In this sense, the pixons are the fundamental particles of information in the image. In fact, using a pixon basis is the fulfillment of *Occam's Razor* formalized in Bayesian terms. Intuitively, this is the goal of every modeling effort, i.e., to fit the data with the simplest model. Equation (2) is the embodiment of that goal with the $p(D|I, M)$ term insisting on a good fit and $p(I|M)$ acting for the cause of simple models, i.e., acting as an *Occam's Razor* term.

While the goal of pixon-based image reconstruction is now clear, there is still considerable uncertainty associated with how to actually select the pixon basis. Piña and Puetter (1993) suggested a particular formulation: The Uniform Pixon Basis (UPB). In the UPB, each of the pixons contain the same amount of information, i.e., N_i is identical for each pixon. Because of this, the UPB image representation provides a sort of "Super-ME" reconstruction. This is because in the UPB representation each pixon is identical, i.e., in this basis the image is exactly flat. Hence entropy is maximized exactly. All of the image's structural information is inherent in the pixon basis. Because entropy is maximized exactly, and because the number of cells in the UPB representation is typically much smaller that the number of data pixels used in standard ME reconstructions,

the formal value of the image prior and hence the value of $p(I, M|D)$ is vastly improved. Crudely speaking, using the UPB representation is the mathematical equivalent of imaging the sky with a magical CCD which can interactively change its number of pixels, their positions, sizes, and shapes in a manner in which each pixel collects the same number of photons.

A FRACTAL PIXON BASIS

While the UPB image representation is a major advance, the UPB basis is rather *ad hoc*. It can not be justified as an optimal pixon basis. A more satisfying pixon basis can be chosen using fractal concepts as we shall now describe–also see Puetter and Piña 1993.

There are numerous definitions for the fractal dimension of a geometric object. However 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) of a line in a plane surface is defined in terms of how the measured length of the line varies as the length of the ruler used to measure the length varies. These ideas are closely related to pixons and the image reconstruction problem since choice of a pixon basis is essentially asking the question: "How does the GOF change as the size of the pixons are varied?"

In forming their UPB image model, Piña and Puetter (1993) use a pseudo-image calculated on a pseudo-grid (usually equal to or finer than the data pixel grid). At each point of the pseudo-grid a local scale is determined. This is the local pixon size. The image is then set equal to the pseudo-image convolved with the pixon kernel (or shape) function of the local width. This image is then convolved with the PSF and compared to the data to determine the GOF. Using the above prescription, the pixons are not cells with sharp boundaries, but are "fuzzy". However, since the image values on the pseudo-grid are correlated, the number of degrees of freedom in the image are significantly less than the number of pseudo-pixels.

We are now ready to consider how to improve the UPB representation using fractal concepts. *Occam's Razor* tells us that we must use the fewest number of pixons each containing the largest information content. The GOF term tells us we must fit the data as well as possible. Using fuzzy pixons, we know that at each point in the image we are going to smooth a pseudo-image over a local scale. What aspect of the smoothing process will prevent an adequate GOF value? Clearly, if the underlying image is locally smooth on a given scale, smoothing the pseudo-image on scales less than or equal to this scale loses nothing. The GOF term is unaffected and the image prior improves. On the other hand, smoothing on scales larger than the local scale is detrimental. So how does one tell if the pseudo-image is being smoothed too much? If the true image is smoothed with pixons of very small width, the value of the smoothed image changes inappreciably. As the widths of the pixons are increased, deviations of the smoothed value from the unsmoothed value become appreciable when the pixon size exceeds the local smoothness scale of the image. This provides an over-smoothing signal. The only remaining issue is determining what size signal is significant. We know, however, by what tolerance we can allow the smoothed

image to depart from the unsmoothed image. The departure must not exceed the local noise. Hence the determination of the local pixon size is an inverse fractal dimension problem. What we are seeking is the largest local smoothing scale consistent with the tolerance set by the noise.

RESULTS: 60 MICRON IRAS IMAGING OF M51

As a test of the ability of FPB-based image reconstruction we have reconstructed an image from $60\mu m$ IRAS survey scans of the interacting galaxy pair M51. This data was selected for several reasons. First, M51 is a well studied object at optical, IR, and radio wavelengths. Hence "reality" for this galaxy is well known. Second, this particular data set was chosen as the basis of an image reconstruction contest at the 1990 MaxEnt Workshop in Laramie, Wyoming (see Bontekoe 1991). Consequently, there have been a number of serious attempts at performing image reconstruction on this data set by specialists in the field. Finally, the IRAS data for this object is particularly strenuous for image reconstruction methods. This is because first, all the interesting structure is on sub-pixel scales (IRAS employed relatively large, discrete detectors (1.5 ′ by 4.75′ 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′ 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 μm detectors (pixel angular response) is known only to roughly 10% accuracy, and finally, the data are not an image, but incompletely and irregularly sampled scans of individual detectors across the galaxy.

Our FPB reconstruction appears in Figure 1 along with Lucy-Richardson and Maximum Correlation Method (MCM) reconstructions (Rice 1993) and a MEMSYS 3 reconstruction (Bontekoe *et al.* 1991)–see Gull (1991) for a description of the MEMSYS algorithms. 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–see Bontekoe (1991) for a gray-scale picture of this reconstruction. However, Weir's solution is qualitatively similar to Bontekoe's solution (Weir 1993). 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 breakup 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 1, our FPB-based reconstruction is clearly superior to those 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 NE and SW arms into discrete sources. Nonetheless, the level of detail present in the FBP reconstruction is clearly absent, e.g., the weak source centered in

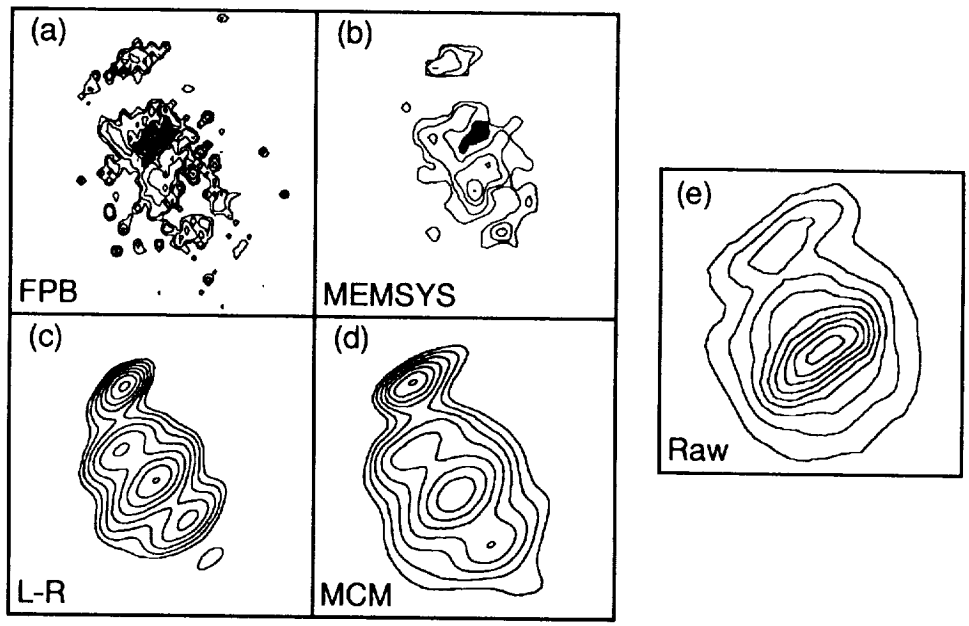

FIGURE 1 Image reconstructions of the Interacting galaxy M51. (a) FPB-based reconstruction; (b) MEMSYS 3 Reconstruction; (c) Lucy-Richardson Figure of panel (b) reproduced from Bontekoe (1991), by permission of the authors. Figures of panels (c) and (d) reproduced from Rice 1993 by permission of the author.

the emission hole (again, this feature corresponds to a known optical source), and the fainter sources around the periphery of the image (most of which are known radio or optical sources). One also notes that the resolution of the FPB reconstruction is roughly a factor of two greater than the MEMSYS 3 reconstruction. However, to be fair to the MEMSYS 3 reconstruction, the authors rebinned their reconstruction by a factor of two to eliminate any possible spurious features. However, we are quite confident that all of the features present in our reconstruction are real.

Aside from the fact that most of the sources can be identified with emission at other wavelengths, the residual errors in our reconstruction are much smaller than in the MEMSYS 3 reconstruction. As pointed out by Bontekoe *et al.* (1991) the peak flux in the MEMSYS 3 reconstruction is 2650 units. The residual errors are correlated with the signal and lie between 0 and 430 units. By contrast, the peak value in the FPB reconstruction is 3290 units, the residuals are uncorrelated with the signal, and the residuals lie between -9 and 17 units. (The contour levels for the MEMSYS 3 and FPB reconstructions are identical and are 150, 300, 600, 1200, and 2400 units.) Furthermore, the large deviation residuals in the FPB reconstruction are due to systematic errors involving incomplete scan coverage of M51. As mentioned above, these errors do not lie under the significant flux emitting portions of the M51 image. The residual errors associated with emitting regions in M51 are significantly smaller

and show a roughly Gaussian distribution. Consequently, the residuals in the FPB reconstruction represent a two order-of-magnitude improvement over the MEMSYS 3 residuals. This, of course, immediately explains the absence of the fine features in the MEMSYS 3 reconstruction. The weak sources present in the FPB reconstruction all lie in the 150 to 300 unit range, and hence are 30 or greater detections in the FPB reconstruction. By contrast, these same features would be less than one detection in the MEMSYS 3 reconstruction.

ACKNOWLEDGMENTS

The authors would like to thank Romke Bontekoe and Do Kester for useful conversations on image reconstruction, IRAS data in general, and the 60 μm IRAS data for M51 in particular. The authors would also like to thank Nick Weir for numerous fruitful conversations both past and present. A patent application on the pixon is currently pending by the authors and The Regents of the University of California. This work was supported by NASA and the NSF.

REFERENCES

Bontekoe, Tj., R. 1991, in *Maximum Entropy and Bayesian Methods*, eds. W. T. Grady, Jr. and L. H. Schick, (Dordrecht: Kluwer Academic Publishers), 319

Bontekoe, Tj., R., Kester, D. J. M., Price, S. D., de Jonge, A. R. W., and Wesselieus, P. R. 1991, *A&A*, **248**, 328

Gull, S. F., and Skilling, J. 1991, *MemSys5 Quantified Maximum Entropy Users Manual*

Piña, R. K., and Puetter, R. C. 1993, *PASP*, **105**, 630

Puetter, R. C., and Piña, R. K. 1993, *Proc. SPIE*, in press

Rice, W. 1993, *AJ*, **105**, 67

Weir, N. 1991, in *Proceedings of the ESO/ST-ECF Data Analysis Workshop, April 1991*, P. Grosbo and R. H. Warmels, Eds. (Garching: ESO), 115

Weir, N. 1993, *J. Opt. Soc. Am.*, in press

Weir, N. 1993, private communication