A brief introduction to image reconstruction is made and the basic concepts of the maximum entropy method are outlined. A statistical inference algorithm based on this method is presented. The algorithm is tested on simulated data and applied to real data. The latter is from a 1024 × 1024 Hubble Space Telescope image of the binary stellar system R Aquarii, which suffers from both spherical aberration and detector saturation. Under these constraints the maximum entropy method produces an image that agrees closely with observed results. The calculations were performed on the MasPar MP-1 single-instruction/multiple-data computer.

Introduction

Traditionally the inversion of the image formulation equation for the object is both an ill-conditioned and ill-posed problem. In most cases noise can complicate the task, leading to ill conditioning. It can enter both before and during the collection of data, and it presents a major obstacle to perfect restoration. Perhaps the most common problem confronting researchers is that of missing data or incomplete information, which is sometimes referred to as the ill-posed problem, that is, one that has no mathematically unique solution. This then becomes an estimation problem, so it is necessary to find the appropriate criteria with which to determine the best solution from all the possible ones that are consistent with the observations. In applications, ill-posed problems are the rule rather than the exception. In biology, geophysical exploration and optical systems, for example, a well-posed problem is virtually unknown. In image reconstruction this frequently leads to propagation errors, whose origins can be traced back to the spread function matrix. Elements in the matrix that are close to zero can cause erratic and often large oscillatory behavior in the solution for the unknown object scene. Stability can usually be established by incorporation of a priori object information into the restoration. It often requires a little more effort however, to convert such problems into those in which there is sufficient information for determination of a unique solution. An excellent source of the art of digital image processing is provided by Andrews and Hunt.

Since the image-restoration problem is basically an ill-posed and ill-conditioned one, statistical estimation methods have been employed to address these issues. Those based on the maximum entropy principle of Jaynes have produced impressive results in applications as diverse as radio astronomy, power spectra, and medical tomography. The maximum entropy principle has its roots in the research of Boltzmann and Gibbs in the latter part of the nineteenth century and Shannon in 1948. It has to do with drawing inferences from incomplete information. Fundamentally it states that any inferences made concerning the outcome of any natural process should be based on the probability distribution that has the maximum entropy permitted by the data taken during observation of the process. If the data can be defined as ensemble averages,

\[ d_k = \sum_{j=1}^{n} p_j A_{kj}, \quad 1 \leq k \leq m, \quad (1) \]

where \( A_{kj} \) defines the nature or physics underlying the measured quantities and \( p_j \) defines the distribution upon which the ensemble averages are imposed as constraints. As shown by Gibbs and Jaynes, with the partition function

\[ Z(\lambda_1, \ldots, \lambda_m) = \sum_{j=1}^{n} \exp\left( -\sum_{k=1}^{m} \lambda_k A_{kj} \right) \quad (2) \]
and the method of Lagrange multipliers, the maximum entropy distribution is

$$p_j = \frac{1}{Z(\lambda_1, \ldots, \lambda_m)} \exp\left(-\sum_{k=1}^{m} \lambda_k A_{kj}\right), \quad 1 \leq j \leq n. \quad (3)$$

The Lagrange multipliers $\lambda_k$ are obtained from

$$\frac{\partial \ln Z}{\partial \lambda_k} + d_k = 0, \quad 1 \leq k \leq m, \quad (4)$$

a set of $m$ simultaneous equations for $m$ unknowns. Any other distributions permitted by the constraints of Eq. (1) necessarily have entropy values less than those determined by Eq. (3). An extensive discussion on the rationale of the maximum entropy method (MEM) is provided by Jaynes.\textsuperscript{16}

The adaptation of this method to the problem of image reconstruction requires the identification of the probability distribution with an appropriate image parameter such as the fraction of the total luminance received at a pixel of a scene,

$$f_j = \frac{N_j}{N}, \quad (5)$$

where $N$ is the total luminance and $N_j$ is the number of elements of luminance at the $j$th pixel.

From this the distribution is

$$p_j = \frac{f_j}{\sum_{j=1}^{n} f_j} \quad (6)$$

In the maximum entropy algorithm, $p_j$ is then replaced by $f_j$. In addition, the recovered object considered as an energy distribution is positive at each point. This condition, reflected as a positivity constraint

$$f_j \geq 0, \quad (7)$$

is guaranteed by the exponential in the maximum entropy solution. The desire for inclusion of this positivity in image restoration has been the subject of much work.\textsuperscript{3,17-19} Together with the normalization and additivity constraints for probabilities, the information content of the image is defined as the entropy

$$S = -\sum f_j \ln f_j.$$ 

Image restoration then proceeds by determination of the sequence of positive numbers $\{f_j\}$ that has maximum entropy consistent with the imposed observational constraints. Furthermore, the inversion is stabilized through the use of the constraints.

In this paper we utilize the maximum entropy method in a specific algorithm on single-instruction/multiple-data-stream (SIMD) -type computer architecture to reconstruct a blurred Hubble Space Telescope (HST) image that also contains large portions of oversaturated data. After an introduction to the problem, we discuss the HST imagery, the MasPar architecture, the implementation of the algorithm, and, in the section, we discuss the results.

Statement of the Problem

The R-Aquarii data used in this research was obtained with the Faint Object Camera (FOC) operating in the [O III] -line emission at 5007 Å. The R-Aquarii symbiotic system, one of the first objects observed with the Hubble Space Telescope,\textsuperscript{20} has shown interesting resolution-limited features with all ground-based imaging techniques applied thus far. For example, the system is observable at x-ray, ultraviolet, optical, infrared, and radio wavelengths and contains a 1.5-M$_\odot$ long period variable of 387 days with a suspected 1.0-M$_\odot$ hot companion with an attendant accretion disk. Moreover, the accretion disk is believed to give rise to a symmetrical jet that extends ~2500 AU toward the northeast and southwest. The dimension assumes that a distance of 250 pc is correct but, in any case, the jet’s observable extent in the optical is ~ 10 arcsec in approximately opposing directions.\textsuperscript{21} The accretion disk and the hot companion it surrounds can only be inferred from ultraviolet observations of hot nebular lines emanating from the jet since no significant ultraviolet continuum emission is present.\textsuperscript{22} One explanation for the unseen companion and accretion disk is that the system is viewed nearly edge-on and the accretion disk thickness precludes direct detection of the source of hot ionizing radiation.

Figure 1(a) shows two problems that are associated with the FOC R-Aquarii data: (1) a large elliptical region indicative of the effects of at least two point sources of emission, each characterized by a point spread function (PSF) that is extended owing to the spherical aberration of the HST primary mirror, and (2) a region of saturation or decrement of recorded intensity with an increase in light level, caused by the presence of apparently two intense features emanating from the central region of R Aquarii. Saturating characteristics of the FOC detector are described by Paresce\textsuperscript{23} and Greenfield.\textsuperscript{24}

For problems such as described here, restoration can become a difficult task at best. Many of the existing maximum-entropy-based algorithms, although quite sophisticated, were designed for specific applications, and we found it undesirable to attempt to adapt them to this specific problem. Instead we used a straightforward approach: we first masked the region of oversaturation in the range, thus eliminating the unusable data. Then by utilizing information resident in the regions outside of the mask as constraints, that is, by including contributions of the PSF from the unsaturated pixel regions, we reconstructed the central region in an iterative fashion. The procedure is described in the section on implementation of the method.

Hubble Space Telescope Imagery

HST imagery of R Aquarii was made available through the Science Assessment and Early Release Observations program of the Space Telescope Science Institute. R Aquarii was observed from the FOC in the f/96 zoom mode of operation with the [O III] filter centered at 5010 Å with a 74-Å FWHM intensity and
Fig. 1. HST FOC image of R Aquarii in [O III] at 5007 Å: (a) raw data showing over exposure, (b) with the dark areal mask covering the saturated pixels, (c) the MEM result of restoration after 500 iterations, (d) the MEM restoration after 1000 iterations.

an exposure time of 597 s on August 23, 1990. The 1024 × 1024 pixel image of R Aquarii provided by this program (file name x0c9010t.cvt.clh) had a ground-based flat field and geometrical corrections applied; from Reseau marks spaced at 1.5 mm, the spatial extent of an individual square pixel was calculated to 0.02238 arcsec on a side. Additional details of the observations and data reduction are contained in Ref. 20. The 22 arcsec × 22 arcsec field of view, resulting from the FOC zoom mode, primarily encompasses the dominant northeast jet and the central objects (long period variable, hot companion, and accretion disk) as shown in Fig. 1. In order to apply the MEM, we used the PSF data provided (file name x0cj010bt.cvt.clh) in this restoration technique. The PSF observation was a 721-s exposure of the star BPM 16274 on August 28, 1990 with the same observational parameters as the R-Aquarii observation except that the FOC was employed in its normal mode rather than its zoom mode of operation.

Owing to the spherical aberration of the primary mirror, the diameter core of the PSF contains ~16% of the incident energy (as opposed to the design specification of 70% cited in Leckrone25), while the remainder spreads out over a region ~20 times wider than that of the diameter core. It is precisely because the PSF is so large that the saturated region of the R-Aquarii image can be restored since the light...
from the wings of the PSF falling outside of the saturated region permits statistical inference.

**MasPar Architecture**

MasPar manufactures a family of massively parallel computer systems capable of attaining peak processing speeds up to 26,000 million instructions/s (MIPS) (32-bit integer adds) and 1300 million floating-point instructions/s (MFLOPS) (32-bit floating-point adds and multiplies). The MP-1 family of computers obtains its performance by using an array of processing elements (PE’s). The PE array comprises from 1024 to 16,384 processors, which operate in a single-instruction/multiple-data (SIMD) fashion. There are three major components to the machine: the PE array, the array control unit (ACU), and the Unix subsystem.

Computational power is attained by use of a massively parallel array of PE’s. Architecturally each PE is a reduced instruction-set computing (RISC) processor with a 64-bit-wide accumulator, 48 32-bit registers, and 16 kbytes of data memory. Hardware support is provided for both integer and floating-point operations as well as direct and indirect addressing. All PE’s execute instructions broadcast by the ACU in lockstep on data stored in its local memory. Each PE can either enable or disable itself for part or all of a computation based on a logical expression for conditional execution. For sharing data with other PE’s, there are three communication mechanisms available: the Xnet, the router, and the global tree.

Xnet, or 8-way nearest-neighbor communications, provides a fast path for moving data between a PE and its eight neighbors in the N, E, W, S, NW, NE, SE, and SW directions. Since the MP-1 operates in a SIMD fashion, all PE’s send data in the same direction at the same time. Similar to a PE’s ability to enable or disable itself for computation, a PE can be disabled for any communication operation. Xnet provides an aggregate bandwidth of up to 18 Gbytes/s (16,000 PE’s).

In addition to Xnet, the MP-1 has an alternative multistage circuit-switched network for global or random communication patterns. This network, or route, provides a PE the ability to send or fetch data from any other PE in the array. The aggregate bandwidth of router communications is 1.3 Gbytes/s (16,000 PE’s). In addition to providing PE-to-PE communications, the router provides the link to a high-performance input-output (I/O) subsystem, which can move data to and from the I/O random-access memory (RAM) buffer into the array at speeds approaching peak.

A global OR tree exists for moving data from the PE array into the ACU. If more than one PE is actively sending data at any time, words from active PE’s are OR’ed together. This OR tree permits the MP-1 to perform fast global reductions, such as finding the maximum or minimum data value in the array.

The array control unit is a 32-bit Harvard-style RISC processor with separate data and instruction memories. The ACU performs two functions: execution control and scalar computation, and broadcasting instructions and/or data to the PE array. The ACU is the master and controls all processing in the MP-1 computer. Programs are written to control the ACU and hence the PE array. The Unix subsystem provides applications engineers with a programming and run-time environment. The Unix subsystem manages jobs targeted for the MP-1. MasPar provides a high-level C language called MPL for programming MP-1 systems. In addition, a graphically oriented symbolic debugger is provided. The algorithm described above was implemented with the MPL programming language.

**Implementation**

Following Eqs. (1)–(6), we obtain the maximum entropy distribution for the estimate of the object, given by

\[
f_j = \frac{\exp \left( -\sum_{k=1}^{m} \lambda_k \ast PSF_{kj} \right)}{Z(\lambda)} , \quad 1 \leq j \leq n, \tag{8}\]

where \( PSF_{kj} = A_{kj} \) and the Lagrange multipliers are found from Eq. (4). Since we are dealing with a million or more pixels, determining the \( \lambda_k \) that maximize the Shannon information measure becomes a formidable task. To simplify matters, we use an approach developed by Agmon et al. who introduce potential function \( F \), which is concave for any trial set of Lagrange multipliers. The values of the multipliers are determined as that set for which \( F \) is a minimum,

\[
F(\lambda) = \lambda_0 + \sum_{k=1}^{m} \lambda_k \ast d_k , \tag{9}\]

and \( \lambda_0 \) is obtained from the normalization condition on the maximum entropy probability distribution. This is given by

\[
\lambda_0 = \ln \left[ \exp \left( -\sum_{k=1}^{m} \lambda_k \ast PSF_{kj} \right) \right] = \ln Z(\lambda_k) . \tag{10}\]

To find the minimum of \( F(\lambda) \), we set to zero its partial with respect to \( \lambda_k \):

\[
G_k = \frac{\partial F}{\partial \lambda_k} = -\sum_{j=1}^{n} PSF_{kj} \ast \exp \left( -\lambda_0 - \sum_{k=1}^{m} \lambda_k \ast PSF_{kj} \right) + d_k = 0 . \tag{11}\]

This is merely the pointwise difference between the original image (blurred image) and the object image convolved with the spread function (reblurred image) when the object was formed from the set of \( \lambda_k \).

Agmon et al. use a modified Newton method to solve for \( \lambda_k, k = 1, \ldots, m \). A traditional Newton-Raphson convergence approach cannot be used effec-
tively since there are $2^n$ functions $G_k$ for a 1024-square image along with $2^n \partial G_k / \partial \lambda_k$ partial derivatives to manipulate and store. This is beyond the computational capabilities of our system. Since we are dealing with large data sets, we must use another approach. Instead we use the more storage-efficient method of successive approximation. Our recursion relation initially was defined as

$$\lambda_k(\text{New}) = \lambda_k(\text{Old}) - G_k.$$ (12)

This converges slowly. However, it was noticed that the $\lambda_k$ are of $O[\log(f_j)]$. Thus a new recurrence is formed. $\lambda_k$ is incremented by log of the ratio of the magnitude of the blurred image and the reblurred image.

$$\lambda_k(\text{New}) = \lambda_k(\text{Old}) - \ln \left( \frac{d_k}{d_k - G_h} \right).$$ (13)

Since $G_k$ and the log term are both positive, both negative, or both zero for any given $\lambda_k$, $k = 1, \ldots, m$, the latter must converge toward the same set of $\lambda_k$ as the former. This relation converges much more rapidly and is very stable.

Along with the impracticality of taking the inverse of a matrix of size $2^n$, it is also impractical to hold a matrix of that size in memory. Therefore a matrix-vector product such as $\lambda \cdot \text{PSF}$ is performed by the convolution of a row of the point spread function with unity. This limits the flexibility of the point spread function, since it must be the same for every point in the image, but it makes the computation tractable.

We also address the capability of ignoring invalid data with this algorithm, hereafter referred to as masking. This is done by our setting $\lambda_k = 0$ for each point containing bad data and never updating them. This in effect treats those particular $\lambda_k$ in the same way as it would those points totally outside the image.

Determining the scaling of the data is a problem in MEM. If the sum of the pixels of the PSF is normalized to one, then the reblurred image, which is the convolution of the two, sums to one. This is important because the original blurred image may not sum to one. It may appear that the solution to this problem is to normalize the sum of the values of the original blurred image to one. This is not possible at initialization if some of the original blurred image is masked. In effect, we implement dynamic scaling during the deblurring process by dividing the sum of the values of the original blurred imaged in the unmasked region at every iteration. This is then used to adjust the values of the deblurred image when the $\lambda_k$ are updated.

This algorithm then models the real scene (deblurred image) for determination of the most probable scene that conforms to both the original blurred image and the PSF. This method starts with all of the $\lambda_k = 0$, which corresponds to a scene that is a flat image (an image with the maximum possible entropy). The entropy of the scene decreases with each iteration as it converges on the scene with the maximum entropy that satisfies the constraints of the original blurred image and the PSF. During this process, noise in the data is represented in the difference between the original blurred image $B$ and the reblurred image $R$ formed during each iteration.

**Pseudocode of the Algorithm**

The algorithm is summarized by the following pseudocode, where $D$ is the deblurred image, $\lambda$ is the set of Lagrange multipliers, $R$ is the reblurred image, $B$ is the original blurred image, and $P$ is the point spread function. The loop can be terminated either after a given number of iterations, when the results no longer converge (owing to lack of floating-point precision) or when the entropy value changes by less than some amount, epsilon. The function $\text{sumof}(A)$ returns the sum of the values of $A$. The function $\text{Convolve}(A, B)$ convolves $A$ with $B$. The function $\text{exp}(A)$ returns the exponent of each value of $A$. If $A$ and $B$ are arrays of $n$ elements, then the function $B = \text{flip}(A)$ reorders the values of $A$ such that $B(i) = [[n - i + 1 \mod n] + 1]$ for $i = 1, \ldots, n$. All other operations are performed elementwise between arrays of the same size or between the elements of an array and a single value.

**Initialization:**

```
read(masked); read in mask
read(B); read in original blurred image
read(P);
\lambda = 0;
P = P/\text{sumof}(P);
\text{flip}(P);

; for x and y arrays of n elements
; such that y(i) = x(\{n - i + 1 \mod n\} + 1)
```

**Loop:**

```
D = \text{Convolve}(\text{flip}(P), \lambda);
D = \exp(D);
D = D/\text{sumof}(D);
entropy = \text{sumof}(D*\log(D));
R = \text{Convolve}(P, D);
If (not masked)
scale = \text{sumof}(B)/\text{sumof}(R); calculate scaling
error = \text{sumof}(\text{abs}(B/scale - R));
\lambda = \lambda + \log(B/scale) - \log(R);
output(D*scale);
```

end of loop
Results and Discussion

The MEM was applied to R-Aquarii original FOC data shown in Fig. 1(a), with the given PSF. Any contribution from Reseau marks and the large central region of saturated raw data were masked and ignored. This is shown in Fig. 1(b). Several central masks of various sizes were tried, all yielding similar global results. The central mask shown in Fig. 1(b) was determined so that the detector count rate did not exceed 0.12 (counts/pixel)/s in the zoom mode, thus ensuring that data used for the restoration would be largely within the linear portion of the intensity transfer function (see Fig. 9 of Greenfield). Figure 1(c) and 1(d) show the progress of the iterative restoration of the central saturated region (which fills in because of the contributions of the PSF from unsaturated pixel regions) as well as the extended filamentary [O III] structure of the dominant northeast jet and some hint of the weaker southwest jet after 500 and 1000 iterations. Using the Richardson-Lucy algorithm, Adorf restored part of the R-Aquarii image outside of the saturated region, and his result compares favorably with our restoration in the same limited region. In the later stages of iteration, the restoration of the central region stabilizes in morphological detail. Although changes in the entropy value by less than some arbitrarily small value can be used to stop the iterative process, there is no general criterion for stopping the continually converging restoration, which will eventually tax the computational precision. To put our computational requirements into perspective, Fig. 1(d) shows the result of 3 h of processing time, in which the MP-1 has a peak speed of $6 \times 10^8$, 32-bit floating-point operations/s.

To test our methodology on simulated data, we took several PSF's (scaled to several different multiplicative amplitudes) and deposited them across a field at specific close intervals and used the result as an unsaturated test image. We deconvolved the image to retrieve the amplitude and resolution, using of course the very irregular mask function used in the R-Aquarii image restoration. Then we doubled the iterations to show that it does not make any difference when you terminate the solution; you still achieve essentially the same results as long it is continually converging. Finally we deconvolved the deconvolved image with the original PSF to show that this result is the same as the simulated image we started out with. This is a particularly convincing methodology test since any extended object can be approximated by an infinitude of points.

The maximum-entropy-restored HST imagery compares favorably with the National Radio Astronomy Observatory Very Large Array 2-cm continuum emission imagery at the limiting resolution of 0.1 arcsec for the central saturated [O III] region. It is suggested that certain features in the restoration such as the southwest protruding ridge of material [Fig. 1(d)] are also consistent with speckle interferometry in the H$_\alpha$-line emission (6563 Å).

Figures 2(a) and 2(b) show the three-dimensional rendering of Figs. 1(a) and 1(d). The outer ring or envelope in Fig. 2(b) was taken to 300 machine iterations, while the center peaks were subject to the full 1000 iterations. In this problem, the MEM solution produces an improvement in dynamic range of 2 orders of magnitude.

In this paper we have presented the results of one particular maximum-entropy-based restoration method applied to blurred and oversaturated HST FOC data. It is precisely because of the distorted...

(a) (b)

Fig. 2. Three-dimensional rendition of (a) Fig. 1(a) and (b) Fig. 1(d).
PSF arising from the spherical aberration of the primary mirror that maximum entropy was able to produce this reconstruction. For data, however, that are blurred and noisy and also encompass regions of sharp contrast in brightness, a contrast model entropy function, that is, one in which no a priori knowledge is included, may exact a penalty during reconstruction. This may appear as a small oscillation, or ringing, in parts of the image. One approach to resolving this is to apply maximum entropy on a larger space of multiple sample distributions, that is, by inclusion of pixel-to-pixel correlations. This is discussed in detail by Gull. Because of the nature of the algorithm used in this paper, however [see Eq. (11)], and because of the fact that the corresponding recurrence relation [Eq. (12)] is quite stable, any such oscillations in our final converged image are effectively removed. This may be apparent by comparison of the solution for the outlying regions of Fig. 2(b), which was taken to 300 machine iterations, with that of the two central hot spots, which was carried to 1000 iterations.

The authors thank Ramarao Inguva of the Department of Physics and Astronomy, University of Wyoming, who was on a fellowship with Symbiotic Technologies, Inc., of Riverdale, Maryland, for reading the manuscript and for several suggestions on this paper. We also extend our gratitude to Robert Hanisch of the Space Telescope Science Institute in Baltimore, Maryland, for informative discussions that contributed to the paper.

References