AIPS Memo 115

Auto-boxing for Clean in AIPS

Eric W. Greisen

June 29, 2009

Abstract

Cotton (EVLA Memo 116) has demonstrated the importance of constraining the Clean deconvolution to search for model components only within restricted regions ("Clean boxes") of the dirty image. An option to create such boxes in an unbiased and automatic fashion has been added to the AIPS task `IMAGR` and two new tasks have been written to find these boxes in images which have already undergone a preliminary Clean. This memo describes the implementation of auto-boxing in AIPS.

1 Introduction

Schwarz (1978[1]) has shown that the iterative deconvolution method Clean can converge completely (residual image essentially 0.0), but that the solution is not unique. It does this in an unconstrained Clean with Clean components located all over the image, which model not only real sources but also all noise bumps. The consequences of this — the so-called Clean bias — were discussed by Condon, et al.(1998[2]). W. D. Cotton in EVLA Memo 116 has examined the Clean bias in some detail. He shows that a lightly constrained Clean algorithm will reduce the apparent "noise" level to arbitrarily low levels, will reduce the apparent brightness on actual sources, and will create apparent sources in regions actually containing no detectable real emission. The primary method to reduce this Clean bias is to restrict the Cleaning to find model components only in regions of obvious signal. These regions have come to be known as "Clean boxes", although some implementations use a mask image rather than a list of rectangular and circular boxes. Without such constraints, the number of free parameters in the deconvolution can easily exceed the number of actually independent visibility samples. Traditionally, these boxes have either been developed interactively or have been implemented in ways that are not particularly restrictive. The former is labor-intensive and subject to bias, while the latter does not solve the problem.

Cotton describes briefly in EVLA Memo 116 an algorithm to find boxes in an unbiased, automatic fashion implemented in his Obit[3] software package. That algorithm searches the residual image at each major cycle for regions of brightness above the noise which are in areas allowed to be Cleaned but not currently in any of the Clean boxes. If the strongest such region meets certain criteria, a Clean box encompassing the region is added to the list of boxes. Since AIPS is more widely used than Obit, it was decided to implement some form of auto-boxing in AIPS.

[1]Schwarz, U. J. 1978, *Astronomy & Astrophysics*, **65**, 245

[2]Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q., F., Perley, R. A., Taylor, G. H., & Broderick, J. J. 1998, *Astronomical Journal*, **1115**, 1693.

[3]http:www.cv.nrao.edu/ bcotton/Obit.html

2 The auto-box algorithm

I have implemented in 31DEC09 *AIPS* three ways to obtain boxes automatically from the data. Two of these are new tasks which begin with an image, or set of facet images, of a region which has already been Cleaned. These images could be the result of an initial Clean with IMAGR using the default inscribed circles suggested by SETFC as the Clean boxes. Such inscribed circles prevent Clean from finding components along the edges and, most particularly, in the corners of images. Those areas may be badly affected by aliasing and by the magnification of numerical error when the image is corrected for the gridding convolution function. The simpler of the two tasks is SABOX which uses a fixed set of input parameters, finds all regions meeting simple island-selection criteria, and writes Clean-box descriptions of those regions to an output box file. The other task is an *AIPS* display based task named FILIT that allows a set of facet images to be viewed interactively and the current boxes if any to be edited and augmented. From its TV menu, the task allows selection of which facet to work on next and a full range of enhancement and data examination tools including a TVROAM-like display which allows images larger than the display screen to be examined. Boxes may be deleted or changed and increased by hand with the familiar DELBOX and REBOX functions. However, the auto-box algorithm may also be invoked, repeatedly if desired, to add boxes based on the parameters to be discussed below. These parameters may be changed while running FILIT and there is a limited capability to undo the new boxes should the auto-box function prove excessive. FILIT should be an efficient way to examine the results of all boxing approaches to make sure that the results are reasonable. A sample screen from FILIT is shown in Figure 1.

The third and most significant place for the new auto-boxing algorithm is in the *AIPS* imaging and deconvolution task IMAGR. At each major cycle, the current facet image, or all facet images depending on mode, are examined by the algorithm and new boxes created if appropriate. In the OVERLAP=2 mode, in which only one facet is Cleaned at a time, it is important to examine all facets periodically to insure that real sources in facets which have not been Cleaned recently are not overlooked.

The algorithm functions as follows. The main routine when invoking auto-boxing calls a top-level boxing subroutine (CLABOX in QCLEAN.FOR for IMAGR, IMABOX in FILIT.FOR). That routine gets the current Clean boxes, identifies the desired image(s), allocates dynamic memory sufficient to hold the largest facet image in memory, calls a lower-level routine (CLABXW in QCLEAN.FOR, ABOXIT in FILIT.FOR) to do the heavy-lifting, reports results, and squirrels away the Clean boxes in the Clean object for use elsewhere.

The heavy-lifting routine begins by filling the image array from disk. If the auto-boxing is to avoid the edges and corners, those edge and corner pixels are changed to a magic blank value and henceforth ignored. At this point, it is possible also to blank out all pixels currently included in Clean boxes. FILIT does this since it is not dealing with a residual image, but SABOX has no current Clean boxes and IMAGR considers that it is dealing with a residual image whose values inside Clean boxes will decrease between iterations. The routine then does a robust determination of the true image rms. It determines a mean and rms excluding values more than 3 times the current value of rms from the current mean. Starting with a very large rms, and repeating this only a few times, results in an excellent value for the true rms unless most of the image pixels have real source brightness.

Given the true rms, the heavy-lifting routine calls ISLAND[4], a routine originally designed for SAD by Walter Jaffe. This routine identifies all connected areas in the image which contain pixels greater than $n\sigma$ where σ is the rms. ISLAND returns the minimum and maximum x and y pixels for each island. The heavy-lifting routine then locates the maximum in each island and discards those that are

1. $< m\sigma$ where $m \geq n$,
2. $< fP$ where P is the current maximum residual, or
3. already located within a Clean box.

Note that these discards are based on the peak pixel value in the island and the location of that peak. The

[4]The *AIPS* file ISLAND.FOR contains subroutines ISLAND, ADDPK and MERGPK making nearly a stand-alone set.

new island dimensions are not compared to the size of the Clean box that causes the discarding. Pixels adjacent to a current Clean box will become strong enough with continued Cleaning (or repeated `AUTOBOX`es in `FILIT`) to produce another Clean box if they are sufficiently bright. See Figure 1 for an example.

Having pruned the list of islands, the heavy lifting routine then examines them beginning with the strongest and working down until at most N boxes are created. If an island is only one pixel wide in either x or y, it will be ignored unless its peak is quite high (2.5 times the larger of the two cutoffs above). If the x and y widths of the island are within 1 pixel of each other, then a circular Clean box is written with a diameter of the larger of the two widths centered on the middle pixel in the island. Otherwise, a rectangular Clean box is written using the island dimensions returned by `ISLAND`.

There are reasons behind all these parameters. Parameter n determines the size of the Clean box as a region containing emission. This is the area in which one would like to Clean. Parameter m insures that only regions with significant peak emission are included. Parameter f provides a cutoff to avoid including sidelobes of the brightest sources. Parameter N also helps prevent the including of sidelobes as sources to be Cleaned and prevents going too deep in any one iteration.

The actual parameters used by these tasks are conveyed by the *AIPS* array adverb `IM2PARM` as follows. `IM2PARM(1)` is N which also acts as a signal that auto-boxing is allowed ($N > 0$) or not. `IM2PARM(2)` is n with a default of 3.0. `IM2PARM(3)` is m with a default of `IM2PARM(2)` +2. `IM2PARM(4)` is f with a default of 0.1. `IM2PARM(5)` allows the user to extend each box outward by `IM2PARM(5)` pixels with allowed values from -1 to 6. `IM2PARM(6)` defines the number of rows and columns skipped at the edge of an image; the radius of the inscribed ellipse is `IMSIZE(1)/2 - IM2PARM(6)` in x and `IMSIZE(2)/2 - IM2PARM(6)` in y. The default is 5 and `FILIT`, but not `IMAGR`, allows a value of -1 to indicate that no ellipse is to be applied. `SABOX`, because of its different role, uses adverb `APARM` rather than `IM2PARM` and defines N as infinite, $m = n$, $f = 0$, and offers a choice of an inscribed ellipse or simply omitting edge rows and columns.

3 Conclusions

In general, this algorithm has been found to work well. It makes an honest estimate of the rms and places Clean boxes only on regions which are reasonably above the rms. I found that the boxes I had used in a wide-field L-band deep Clean included all of those found by the auto-boxing but also included some that, in hind sight, should not have been used. In practise, it appears that N should be larger than I had originally thought. Otherwise, it takes too many major cycles to encompass all real sources. `IMAGR` uses the largest pixel inside the inscribed ellipses of all facets for P and uses the largest value inside the ellipse of each facet separately in selecting which facets are imaged. When P approaches the noise cutoff ($m\sigma$), it was found necessary to zero these values, more to obtain reasonable selection of facets for Cleaning than for the auto-boxing.

A screen capture from `FILIT` is shown in Figure 1. It began with no boxes and generated the boxes shown in red after several selections of the `AUTOBOX` function. Note the menu of options plotted in cyan and the inscribed ellipse plotted in blue. The `REROAM` menu item is omitted since the particular image fits within the TV display. The auto-boxing has handled the very bright quasar, the extended woofly galaxy, and a wide variety of weaker L-band objects.

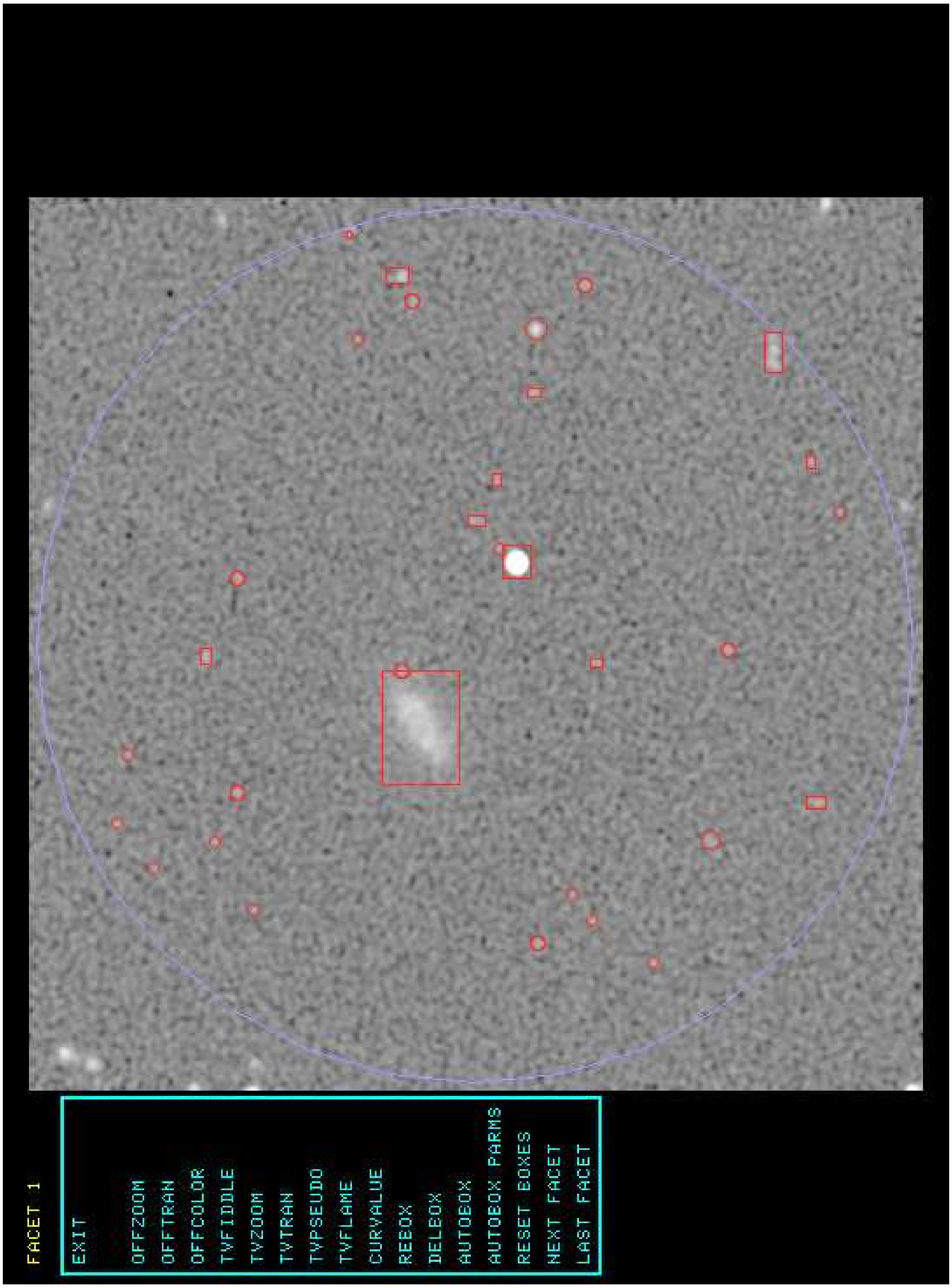

Figure 1: The central facet of the 1420-MHz continuum from the NGC 6503 region displayed and auto-boxed in `FILIT` along with the interactive menu.