Performance Enhancement of the autoWindow technique

W. D. Cotton, May 26, 2009

Abstract—This memo describes performance enhancements to the autoWindow technique in Obit for automatically specifying CLEAN boxes. Allowing multiple boxes to be added to a facet CLEAN window per major CLEAN cycle ran in 66% of the time required for the single box per cycle method when applied to a large test case.

Index Terms—interferometry, performance

I. INTRODUCTION

THe next generation of instruments (EVLA, eMerlin, MeerKAT) will have such sensitivity that the full primary beam of the antennas will frequently need be imaged and deconvolved to remove the side-lobes from the sources widely scattered across the field of view. At high resolution, the number of degrees of freedom in the images derived will not be a trivial fraction of the degrees of freedom in the data. In this regime, deconvolution techniques such as CLEAN can turn "noise" into "sources" and in general will remove flux density from real sources and create spurious sources; see [1] for a discussion of this effect.

The generic solution to this problem is to restrict the regions over which CLEAN is allowed to work; traditionally this was done manually in an interactive deconvolution. An automated way of selecting the windows in which CLEAN is allowed to operate is described in [2] and [3]. The main drawback to this technique is that it forces the CLEAN to go more slowly, requiring a major cycle for each box added to the CLEAN window. The following describes an enhancement to the autoWindow technique in Obit [3] [1] allowing the addition of multiple boxes to the CLEAN window per major cycle. This technique is applied to a 100 GByte simulated dataset using the procedure described in [4].

II. MULTI-BOX AUTOWINDOW

The original implementation of autoWindow allowed only a single new box to be added to a facet CLEAN window per major cycle. This has the disadvantage of requiring more CLEAN cycles with fewer components per major cycle; this can substantially increase the run time for a deconvolution.

The finding of CLEAN components is done using the BGC (Barry G. Clark) [5] minor cycle technique in which a subset of the most significant residuals and a truncated dirty beam is used. Without autoWindow, this process stops when the
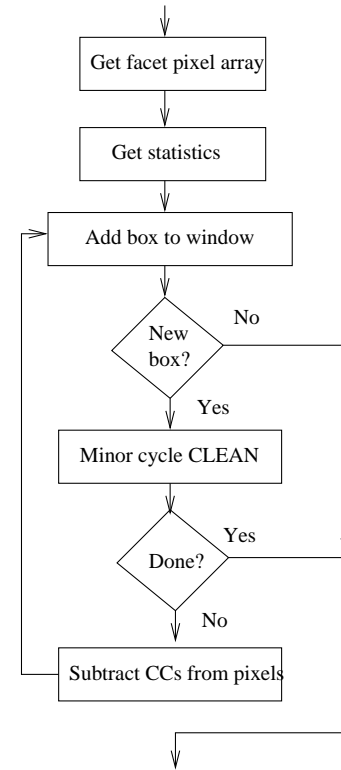
Fig. 1. Flow chart of the loop over adding boxes to the facet window. New box test no applied the first pass.

CLEAN reaches the level where the highest ignored side-lobe of the brightest initial residual is comparable to the most significant ignored residual. The autoWindow process modifies this by stopping the CLEAN when it gets to the brightest residual outside of the CLEAN box. Thus, the inner CLEAN stops before it reached the limitations inherent in the technique.

For the visibility–based or "Cotton-Schwab" CLEAN [6] [7] this can add substantially to the computational cost as the the number of major cycles increases. In this technique, the outer or "major" cycle involves subtracting the CLEAN model from the visibility data then re-imaging the facets to form the residual images. For large datasets, this major cycle can be expensive. In this case, it is desirable to have an image plane loop to do as much of the deconvolution as possible before having to reform the residual images from the visibility data.

As a test of this, an intermediate image plane "Window loop" was implemented in the visibility–based CLEAN in Obit. This implementation uses faceted imaging [8] to deal

with the non-coplanarity problem. This loop is shown schematically in Figure 1 and consists of components described in the following:

- **Get facet pixel array**
  The array of pixel values is read from the residual facet image; in Obit these are kept in an ObitFArray. The resulting array of pixels is used in the following operations.
- **Get statistics**
  The histograms of pixel values and beam side-lobe levels needed for the inner Clark Clean are derived.
- **Add box to window**
  The autoWindow technique for adding a single box to the facet window [2] is applied.
- **New box?**
  A test is made if a new box was added to the window. On the initial pass through the loop, processing is continued regardless of the outcome of this test. On subsequent passes, the loop is exited if no new box was added.
- **Minor cycle CLEAN**
  This is the Clark minor CLEAN in which the inner region of the dirty beam ("beam patch") is used to deconvolve the most significant residuals.
- **Done?**
  The loop is terminated on any of the following conditions: 1) the minor cycle stops for any reason other than that it reached the autoWindow limit (most significant residual NOT included in the CLEAN window), 2) no new box was added, or 3) 10 passes through this loop.
- **Subtract CCs from pixels**
  The beam patch is used to subtract the CLEAN components from the previous minor cycle from the residual image grid. This uses a shift and subtract technique for the component flux density times the beam patch from the residual image; hence is only as accurate as the beam patch is as an approximation of the dirty beam.

Such looping allows a more elaborate set of boxes to be added each major cycle. This procedure results in more-or-less the same depth of CLEAN as would be obtained if the same window was known for a single minor cycle CLEAN. The result is fewer major cycles with more CLEAN components being determined in each.

III. SIMULATED DATA

The simulated data as described in [4] was generated using the VLA "B" configuration and consists of 19 ten minute scans over 9.5 hours of a "target" source at $60°$ declination. Data samples were 2 second integrations and contained 1024 spectral channels divided among 32 "IFs" and spanning from 1.4 to 1.9 GHz. This generated a total of 2,323,269 visibility records of which 2,003,508 were on the "Target". In full precision (3 floats per visibility) this (including "calibrator" sources) is 106.3 GByte of data and in "compressed" (2 shorts per visibility) 35.5 GByte of data.

The sky model used for the target field was the CLEAN model derived from a moderately deep VLA survey (22 μJy RMS) pointing using a similar setup. This sky model consists of 116 facets Gaussian noise of 0.5 Jy was added to each visibility measurement; no primary beam gain or spectral index corrections were applied. The imaging in these tests used 147 facets and produced an image of 3599×3599 cells.

IV. TEST PROCESSING

Testing was performed on the Obit Development machine, mortibus, in Charlottesville. This machine has dual quad core Xenon processors for a total of 8 cores, a clock speed of 3 GHz, 8 GByte memory and a fast disk RAID system. Mortibus uses the Dell SAS/SATA RAID 5, PERC 6/i Integrated controller (made by LSI) with the Seagate 146GB15K RPM SAS 3Gbps 2.5-in HotPlug Hard Drive.

The general testing procedure was the same as described in [4]. Timing tests were made of several cases:

1) Image only - no CLEAN,
2) Image and CLEAN without autoWindow,
3) Image and CLEAN with single box per major cycle autoWindow
4) Image and CLEAN with multiple boxes per major cycle autoWindow.

The UNIX utility time was used to determine the CPU usage.

A. Timing Results

The timing results for the various executions are given in Table I. Eight cores were available so the maximum possible ratio is 8.

V. DISCUSSION

The comparison of the single box autoWindow CLEAN with the unconstrained CLEAN in Table I indicates that, for this test, the autoWindow technique imposed 81% extra runtime, or 32.1 hours vs. 18.8 hours. This is a very substantial increase in run time for larger data sets.

The timing result for the multi-box autoWindow run shown in Table I indicates a very substantial performance gain in this test over the single box per major cycle technique. The run time for this test for the multi-box test was 66% of the runtime for the single box test and represents a 20% increase over the unconstrained clean.

TABLE I
TIMINGS OF OBIT/IMAGER

Process	Real Time min.	CPU time min.	CPU/Real	Major cycles	Comments
Imager	436	2650	6.1	0	Image only
Imager	1066	6067	5.7	40	Image + CLEAN w/o autoWindow
Imager	1926	8368	4.3	110	single box autoWindow
Imager	1285	6539	5.1	85	multi-box autoWindow

REFERENCES

[1] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick, ""The NRAO VLA sky survey"," *AJ*, vol. 115, pp. 1693–1716, 1998.

[2] W. D. Cotton, "Automatic CLEAN windowing," *EVLA Memo series*, vol. 116, pp. 1–17, 2006.

[3] W. D. Cotton, "Obit: A Development Environment for Astronomical Algorithms," *PASP*, vol. 120, pp. 439–448, 2008.

[4] W. D. Cotton, "Testing Obit with a 100 GByte simulated dataset," *Obit Development Memo Series*, vol. 8, pp. 1–8, 2009.

[5] B. G. Clark, "An efficient implementation of the algorithm 'CLEAN'," *A&A*, vol. 89, pp. 377–+, Sep. 1980.

[6] F. R. Schwab, "Optimal Gridding of Visibility Data in Radio Interferometry," in *Indirect Imaging. Measurement and Processing for Indirect Imaging*, J. A. Roberts, Ed., 1984, pp. 333–+.

[7] W. D. Cotton, "Special problems in imaging," in *Synthesis Imaging in Radio Astronomy*, ser. ASP Conference Series, R. A. Perley, F. R. Schwab, and A. H. Bridle, Eds., no. 6. Astronomical Society of the Pacific, 1989, pp. 243–246.

[8] R. A. Perley, "Imaging with Non-Coplanar Arrays," in *Synthesis Imaging in Radio Astronomy II*, ser. Astronomical Society of the Pacific Conference Series, G. B. Taylor, C. L. Carilli, and R. A. Perley, Eds., vol. 180, 1999, pp. 383–+.