Memo 4

Source Finding and Measurement for VAST and EMU

Andrew Hopkins, Paul Hancock, et al.

September 7 2011

EMU/VAST Design Working Group 2 Memo

Source finding and measurement

Andrew Hopkins, Paul Hancock, Minh Huynh, Tara Murphy,
Bryan Gaensler, Russell Jurek, Ray Norris, Matthew Whiting

7 September 2011

EMU Design WG-2 and VAST WG-2

The efforts of EMU DWG-2 have been broadly split into (1) those focussing on optimising point-source identification and measurement, and (2) those focussing on developing or testing techniques for robust measurement of extended, complex, or low-surface-brightness sources. Reflecting this the working group has now been split into two complementary working groups (see:
`http://askap.pbworks.com/w/page/14049357/Working-Groups`
for the working group list). The progress is being tracked on the EMU wiki here:
`http://askap.pbworks.com/w/page/24421701/Source%20Finding%20Progress`

The WG-2 efforts for both VAST and EMU have to date had a particular focus on optimising compact source identification and measurement, and this has been pursued for both teams in a common project led by Paul Hancock at the University of Sydney. This has been complemented by an investigation led by Minh Huynh at the University of Western Australia.

This Memo

This memo presents recommendations for the ASKAP computing group that have arisen from the compact (point-source) source finding efforts, summarising the outcomes from two refereed journal papers recently submitted (or about to be submitted) for publication, led by Minh Huynh and by Paul Hancock. All discussion here refers to source-finding in 2D continuum images.

Recommendations

The recommendations arising from the investigations by Huynh et al., (2011) and Hancock et al., (2011) are summarised here.

- Background estimation (Huynh et al., 2011): Background estimation needs manual tweaking in existing tools to achieve optimisation. Automation of the process for identifying a suitable "local" scale over which to estimate the background mean and rms is desirable, but not currently implemented in any existing tool.
  Recommendation 1: The ASKAPsoft approach to "local" scale background mean and rms be tested, to assess whether a fixed "local" scale is sufficient for ASKAP imaging, or whether a fully automated algorithm for robust identification of the "local" scale for the background mean and rms in each ASKAP image (potentially varying across an image) needs to be developed.

- Thresholding (Huynh et al., 2011): All source-finders demonstrate increasing completeness and decreasing reliability with the reduction of the threshold to which sources are identified. Some tools perform better than others, though, in terms of the reliability, at a fixed level of completeness. Huynh et al., (2011) finds that the FDR-selected threshold approach of SFIND gives a higher reliability at a given level of completeness compared to SExtractor or Selavy. In practice for the ASKAP surveys, rather than having a single fixed threshold and omiting sources with peak flux densities at lower levels, what is most desirable is to return all sources down to some very low threshold level, but to also have a well-characterised measurement of completeness and reliability

as a function of S/N.
Recommendation 2: The ASKAPsoft source-detection be carried out to a low threshold level, recording the S/N of each source (from the measured flux-density and measured "local" rms value inferred or interpolated at the location of each source). The completeness and reliability of the ASKAPsoft source-finder should be robustly characterised as a function of S/N using simulated images.

- Deblending (Hancock et al., 2011): All source-finders tested have difficulty in correctly (1) identifying, and (2) characterising, sources that appear close in an image (within a common "island" of pixels above the threshold). Paul Hancock has implemented a tool, called "Tesla," that uses a Laplacian-filtering step to identify the number of peaks within islands, that substantially improves the way source-finders identify multiple closely-associated sources.
  Recommendation 3: The ASKAPsoft pipeline developers coordinate with Paul to explore how to implement this approach within the pipeline.

- Characterisation (Hancock et al., 2011): The position and flux-density information returned for sources has been shown by Hancock et al., (2011) to be robust in general, although the Selavy measurements are notably sub-optimal compared to others. Specifically, the positional accuracy is worse by about 50% compared to IMSAD, SFIND, and SExtractor, and the flux density estimates can be worse by factors of two. This result holds even for relatively bright sources, up to $\sim 100\sigma$. This seems to be a consequence of the fitting performed by Selavy preferring to fit multiple (rather than single) Gaussian components, even when a single component is sufficient.
  Recommendation 4: The ASKAPsoft pipeline developers implement an alternative Gaussian fitting routine from that used in Selavy/Duchamp (or modify the existing one) in order to obtain performance at least at the level displayed by the fitting routine from IMSAD (SFIND uses the same routine).

Terms of Reference for Design Study Review
Here we explicitly address the terms of reference for the upcoming review.

- *Report on source finding tools and the usefulness of the ASKAP simulations*.
  The ASKAP simulations have been valuable in highlighting the issues detailed above regarding background and local noise level estimation. We have tested a number of source finding tools and conclude that there are substantial limitations in the current implementation of Selavy/DUCHAMP that will need to be improved for compact continuum source identification and measurement.

- *Please include details of the requested updates to DUCHAMP and the results of testing the updates for your purposes*.
  A selection of requested updates are detailed above. There are further tools and techniques that are continuing to be explored, and we anticipate that there may be additional improvements that could also be incorporated in due course.

- *Please demonstrate that DUCHAMP can be used satisfactorily in your end-to-end pipeline and how are you progressing toward source characterisation*.
  Following the implementation of the above recommendations, we expect that the ASKAPsoft implementation of DUCHAMP will perform at a suitable baseline level, providing reasonable source identification and characterisation for compact continuum sources. We recognise the need for ongoing testing and additional exploration of other tools and techniques, however, and anticipate

additional interactions with the computing group to further optimise the source identification and characterisation performance.

Summary

It is clear that Selavy (DUCHAMP) as it is currently implemented is unlikely to be optimal for source identification and measurement for compact continuum sources. We are proposing a selection of improvements that need to be incorporated within the ASKAPsoft version of DUCHAMP, detailed above, that address existing limitations. In addition, we are also continuing to explore additional tools and techniques, and there may be further gains in optimisation that can be obtained.

Copies of the references used in this Memo can be obtained from their respective first authors on request prior to publication, in order to aid in implementing these recommendations.

References

Huynh, M., et al., 2011, PASA (submitted)
Hancock, P., et al., 2011, (in prep.)