Astronomical Data Analysis Software and Systems XVI
ASP Conference Series, Vol. 376, 2007
R. A. Shaw, F. Hill and D. J. Bell, eds.

An All-Sky 2MASS Mosaic Constructed on the TeraGrid

A. Laity, G. B. Berriman, J. C. Good

Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA, USA

D. S. Katz, J. C. Jacob

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

L. Brieger, R. Moore

San Diego Supercomputer Center, University of California, San Diego, CA, USA

R. Williams

Center for Advanced Computing Research, California Institute of Technology, Pasadena, CA, USA

E. Deelman, G. Singh, M.-H. Su

Information Sciences Institute, University of Southern California, Los Angeles, CA, USA

Abstract. The *Montage* mosaic engine supplies on-request image mosaic services for the NVO astronomical community. A companion paper describes scientific applications of Montage. This paper describes one application in detail: the generation at SDSC of a mosaic of the 2MASS All-sky Image Atlas on the NSF TeraGrid. The goals of the project are: to provide a value-added 2MASS product that combines overlapping images to improve sensitivity; to demonstrate applicability of computing at-scale to astronomical missions and surveys, especially projects such as LSST; and to demonstrate the utility of the NVO Hyperatlas format. The numerical processing of an 8 TB, 32-bit survey to produce a 64-bit, 20 TB output atlas presented multiple scalability and operational challenges. An MPI Python module, MYMPI, was used to manage the alternately sequential and parallel steps of the *Montage* process. This allowed us to parallelize all steps of the mosaic process: that of many, sequential steps executing simultaneously for independent mosaics and that of a single MPI parallel job executing on many CPUs for a single mosaic. The Storage Resource Broker (SRB) was used to archive the output results in the Hyperatlas. The 2MASS mosaics are now being assessed for scientific quality. Around 130,000 CPU-hours were used to complete the mosaics. The output consists of 1734 plates spanning 6° for each of 3 bands. Each of the 5202 mosaics is roughly 4 GB in size, and each has been tiled into a 12×12 array of 26 MB files for ease of handling. The total size is about 20 TB in 750,000 tiles.

1. Project Goals

The goals of the project were to provide a value-added Two-Micron All-Sky Survey[1] (2MASS) product that combines overlapping images to improve sensitivity; to demonstrate applicability of computing at-scale to astronomical missions and surveys, especially projects such as the Large Synoptic Survey Telescope; and to demonstrate the utility of the NVO Hyperatlas[2] format.

The scientific and technical applications of wide-area mosaics involve visualization of the content and observational coverage of large regions. Examples of science questions, and the role of the mosaic in addressing them, are:

Q: *I am planning an observing proposal of a cluster of galaxies. What observations have been made in this area?*

- Superimpose footprints of observations on a mosaic of the cluster

Q: *I am studying a large molecular cloud. Are there candidate sources not recorded in individual source catalogs?*

- Superimpose locations of sources from existing catalogs on mosaics from surveys at different wavelengths
- Examine multi-wavelengths mosaics for new candidate sources seen in at least two wavelengths

Q: *Have I found a high proper motion object or an artifact in an image? Is it a T-dwarf candidate or is it in a region of higher than average extinction?*

- Examine mosaics from 2MASS, SDSS and DPOSS, all at different epochs
- Examine mosaics at different wavelengths to study distribution of color across images.

Another application of a mosaic is in quality assurance. Wide area coverage simplifies identification of artifacts present over spatial scales of several degrees, which may be caused by diffraction spikes, ghost images, and filter and dichroic glints. Figure 1 shows such artifacts image defects in a 3-color cutout, 5° on a side, from the portion of the all-sky mosaic centered on Wolf 359.

2. Calculation of the Mosaic

The mosaic was calculated with the *Montage* image mosaic toolkit on the Teragrid. Montage is highly parallelizable, processes FITS images with any WCS projection, re-projects images while preserving flux and astrometry, and rectifies variable sky emission to a common level.

The mosaic was computed from the 4,121,440 files (12.4 TB; 32-bit) of the 2MASS atlas image data at J, H, and Ks; each 2 MB in size (with 32-bit pixels).

---

[1] http://www.ipac.caltech.edu/2mass

[2] http://www.us-vo.org/pubs/files/hyperatlas.pdf

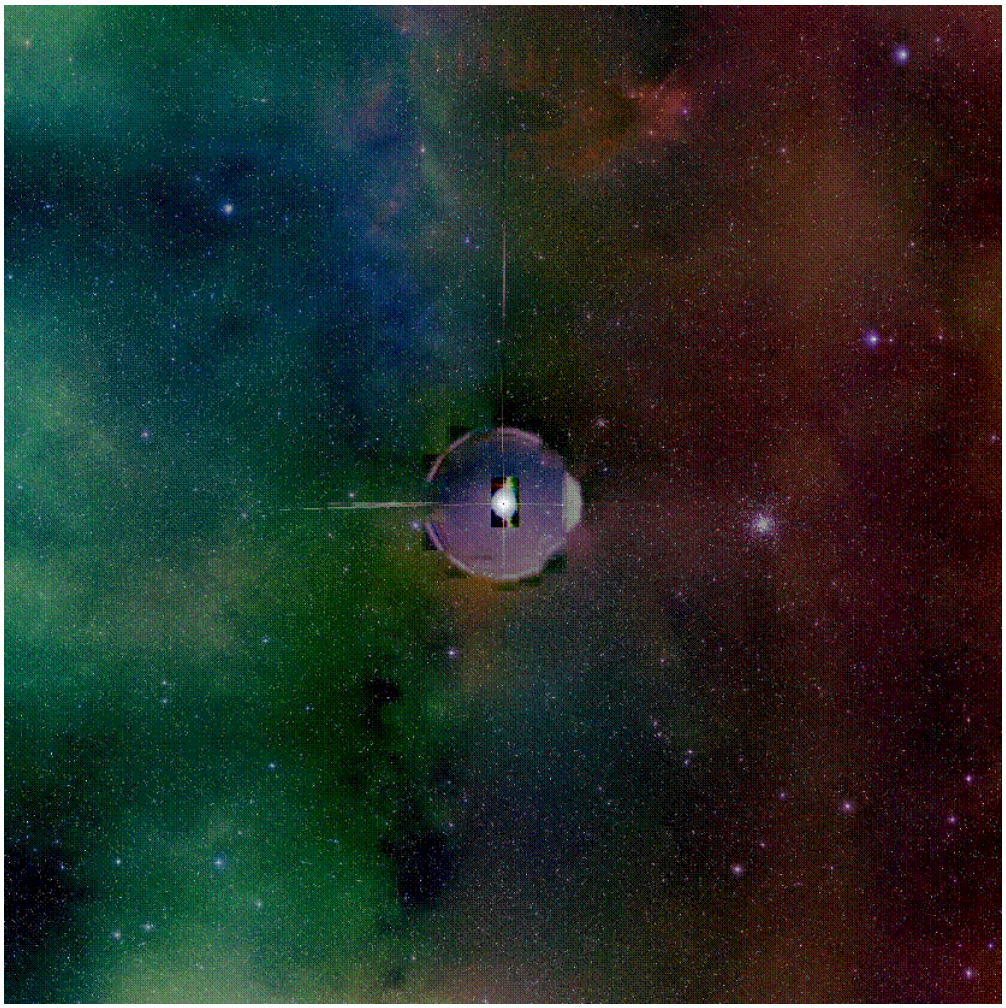

Figure 1. Three-color cut-out from the 2MASS all sky mosaic, centered on Wolf 359. The image is 5° on a side.

The all-sky mosaic was performed on the Distributed Terascale Facility[3] node at San Diego Supercomputer Center (SDSC)[4], with parallelization through the Message Passing Interface (MPI). The images distributed to multiple nodes of a cluster for parallel steps. The input files that fell on mosaic boundaries were opened, read, and used multiple times in the processing of adjacent mosaics, so that a total of 14 TB in 6,275,494 files were actually opened and read. The processing time to complete the mosaics was more than 100,000 CPU-hours on a Teragrid Linux cluster (1.5 Ghz Intel Itanium-2 processors); this included time for a shake-down of the hardware.

[3]http://www.teragrid.org

[4]http://www.sdsc.edu

Further details of processing and data management are as follows: First, the entire 2MASS image archive was replicated onto Teragrid disks. An archive of the images was then created on the IBM HPSS High Performance Storage System (HPSS) after aggregating the images into 147,000 containers to minimize the impact on the HPSS name-space. Check-sums were used to validate the integrity of the images. Continued validation was essential to recognize and correct problems with data transport, RAID controllers, and media failure. The computation was decomposed into fifty 20-hour runs on 128 processors. The mosaic process generated 64-bit image files, resulting in 20 TB of output data The output images were truncated back to 32-bit image files, registered into the Storage Resource Broker, and replicated into the IBM HPSS archive.

3. Processing Summary

The total size of the output mosaic over all three bands is 10 TB (64-bit). It is organized in 1734 6° plates per band, 5205 plates total, 2 GB in size, each tiled into a 12×12 array of 13 MB files for ease of handling. There are 750,000 files in total. The output is organized in a Hyperatlas structure, implemented using the SDSC Storage Resource Broker[5] data grid. The Hyperatlas organizes images in an Atlas consisting of multiple pages that are standard WCS sphere-to-plane projection mappings defined by a pointing center, projection type rotation, and scale. The all-sky mosaic uses the TM-5-TAN-20 Atlas of 1734 pages covering the celestial sphere, with every point of the sphere at most $3\overset{\circ}{.}5$ ($5^\circ/\sqrt{2}$) from a pointing center, with a scale of $1''$ pixel^{-1}, and a tangent-plane WCS projection.

Acknowledgments. This work made use of Montage, funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computational Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology. The 2MASS project is a collaboration between The University of Massachusetts and the Infrared Processing and Analysis Center (JPL/ Caltech). Funding is provided primarily by NASA and the NSF. The 2MASS image data products generated by *Montage* are not endorsed by the 2MASS project. Part of the work was done under funding support from the National Virtual Observatory, through the National Science Foundation.

[5]http://www.sdsc.edu/srb/