Astronomical Data Analysis Software and Systems VI
ASP Conference Series, Vol. 125, 1997
Gareth Hunt and H. E. Payne, eds.

The Data Handling System for the NOAO Mosaic

Doug Tody

IRAF Group, NOAO,[1] PO Box 26732, Tucson, AZ 85726

Abstract. This paper presents the data handling system being built for the NOAO Mosaic by the IRAF group at NOAO. This system consists of a Data Capture Agent which assembles and saves the disk images, a Real Time Display and Mosaic viewer used to view the very large (134 MB) Mosaic images during and after readout, and IRAF software for CCD processing, quick look, and general data interaction. The system architecture is based on general message bus and distributed shared object facilities. The Real Time Display and Mosaic viewer is a general, user extensible image viewer based on the existing Ximtool and SAOtng display software from NOAO and SAO. Companion papers describe the Mosaic data format and data reductions. Like IRAF itself, all of this software is portable and controller independent hence suitable for use by other observatories, particularly if they already use IRAF in the observing environment.

1. Introduction

The NOAO Mosaic is a large-field detector consisting of eight 2K×4K CCDs arranged in a 4×2 mosaic, for a total detector size of 8K×8K. At 16 bits per pixel, raw images are 134 MB in size. The instrument will be used on the 4-meter and 0.9-meter telescopes on Kitt Peak in the northern hemisphere and on the 4-meter telescope on Cerro Tololo in Chile. The field of view of the Mosaic ranges from $36'$ to 1° depending on the telescope, with a scale of $0.26''$/pixel at the 4-meter. Because the field is so large, optical field correctors are required. Similar instruments are being built at a number of other observatories, e.g., CFHT/U.Hawaii, Keck/Lick, and McDonald. The software described in this paper is being developed in collaboration with these and other groups.

While there is little fundamentally new about Mosaic data handling, the overall system accentuates old problems to a degree rarely before seen in ground based telescopes. The use of multiple CCDs causes problems with misaligned grids and gaps between the CCDs, requiring interpolation, image combination, and dithering to rectify the data. The CCDs can have different bias and flat characteristics requiring calibration before they can be viewed together on the display. The large field and the use of optical correctors mean that field distortions are significant, and combined with the misalignment of the CCDs this complicates coordinate determination and astrometry. The use of multiple CCDs

[1] National Optical Astronomy Observatories, operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation.

requires that data be read out simultaneously from all CCDs, hence the raw data is interleaved as it arrives from the controller and must be "unscrambled" before being written to disk or displayed. Finally, the images are very large. A powerful computer system and efficient software is required to be able to handle such large images. Even viewing the data is difficult since the image is a composite of a number of smaller images, and at 8K×8K or 64 megapixels, the area of the full Mosaic is about 50 times that of the typical workstation screen.

2. The Mosaic Data Handling System

The Mosaic data handling system takes the raw data as they are generated by the CCD controller during frame readout and does all subsequent processing of the data, including capture to disk, real time image display, quick look and data quality assessment, pipeline data reductions, taping, queueing of data to the data archive, and, if desired, re-reduction of the data at the observer's home institution using IRAF.

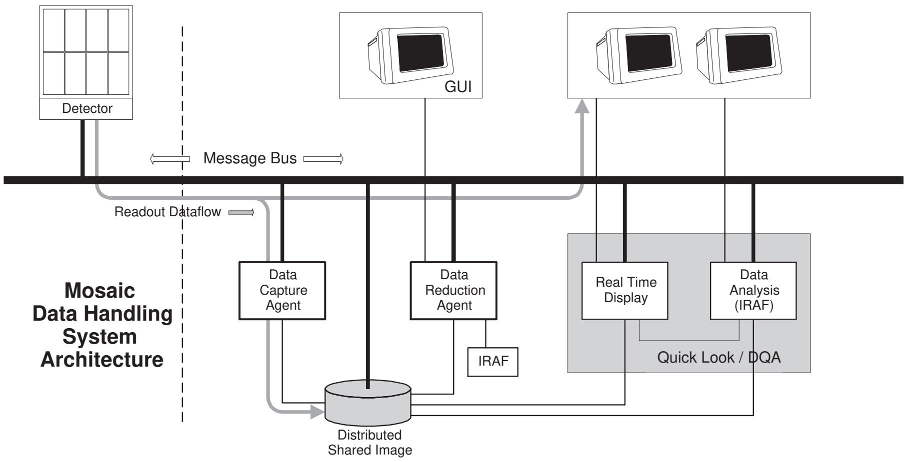

Figure 1. Data Handling System Architecture.

Figure 1 illustrates the software architecture of the Mosaic Data Handling System (DHS). As the Mosaic is read out, pixel and header data packets are written to the *message bus* which connects all elements of the data system. The *Data Capture Agent* (DCA) captures these data packets and builds an observation file on disk. At the same time data is sent to the *Real Time Display* (RTD) which displays the Mosaic during frame readout. Quick look is provided by the RTD and by IRAF, which can interact with the RTD during and after frame readout. The *Data Reduction Agent* (DRA) directs the post-processing of each observation file, applying standard calibrations and writing the data to tape and to the data archive.

2.1. Messaging and Shared Data Access

The heart of the Mosaic data handling system is the message bus, which connects all data system components. The message bus provides flexible and efficient facilities for components to communicate with each other. The message bus (which is a software facility) supports both distributed and parallel computing, connecting multiple host computers or multiple processors on the same host.

The message bus provides two methods for components to communicate with each other. *Producer/consumer events* allow components to listen for (consume) asynchronous event messages produced and broadcast by other components. *Requests* allow synchronous or asynchronous remote procedure calls (method invocations) to be directed to services or data objects elsewhere on the message bus. Discovery techniques can be used to determine what services are available and to query their methods. Host computers and components can dynamically connect or disconnect from the bus. The bus can automatically start services upon request; or services and other components can be started by external means, connecting to the message bus during startup.

An important class of message bus component is the *Distributed Shared Object* (DSO). DSOs allow data objects to be concurrently accessed by multiple clients. The DSO provides methods for accessing and manipulating the data object and locking facilities to ensure data consistency. DSOs are distributed, meaning that clients can be on any host or processor connected to the message bus. In the case of the Mosaic DHS, the principal DSO is the *distributed shared image* which is used for data capture, to drive the real time display, and for quick look interaction from within IRAF. The distributed shared image uses shared memory for efficient concurrent access to the pixel data, and messaging to inform clients of changes to the image.

The Mosaic DHS uses a custom message bus API which is layered upon some lower level messaging system. At present we are using the Parallel Virtual Machine (PVM) facility; in the future we might use other facilities such as CORBA. The use of a custom API provides isolation from the underlying messaging facility and aids development of a standard framework and set of services for integrating a set of applications.

2.2. Data Capture

When the Mosaic is read out the controller writes a stream of message packets onto the message bus. These take the form of requests to the *Data Capture Agent* (DCA), which operates as a service on the message bus. The DCA sits in a loop handling requests from the message bus. Incoming header data is buffered internally within the DCA. Incoming pixel data blocks are unscrambled and written directly to the output image using the distributed shared image facility. When the readout is finished a table driven keyword translation module (implemented as a configurable Tcl script) transforms the input device dependent detector keywords as necessary to conform to the data format required by the DHS. Ultimately a new observation file is written to disk and passed off to the DRA for post-processing. The observation file is a multi-extension FITS file containing one IMAGE extension for each amplifier of the Mosaic. The DCA can handle multiple simultaneous readouts from different clients; the clients can be on any host computer connected to the message bus.

2.3. Real Time Display and Quick Look

The primary function of the *Real Time Display* (RTD) is to display the Mosaic in real time during readout; pixels appear in the display as soon as readout begins. As noted above the DCA does not just write to a disk image, it writes to a *distributed shared image*, a type of DSO. At the same time that the DCA is writing to the DSO, the RTD is reading from it and displaying the incoming data. The DCA receives an incoming write-pixel request on the message bus, obtains locks on a set of regions in the output image (e.g. 16 regions for 8 CCDs with 2 amps each), copies the input data to the output regions, and then frees the regions. This causes the DSO to send messages to all clients, such as the RTD, which want to be informed of changes to the image. The RTD then performs any on-the-fly calibration or other processing and updates the displayed image.

To the user the RTD is an image browser displaying to one or more workstation screens, two in the case of the NOAO Mosaic. One screen shows the full mosaic dezoomed at 50-to-1. The second screen shows a zoomed up region of the mosaic. Multiple zoom windows can be active on multiple screens.

The RTD is not just a real time display, it is a fully functional image viewer with extensive builtin functions for quick look image analysis. Additional functionality (possibly quite extensive) can be added via a dynamic "plug-in" facility, which allows users or projects to easily customize the display or tailor it for existing data systems. Finally, extensive image analysis is available via IRAF or any other external image analysis system which interfaces with the RTD and DSO. IRAF sees the DSO as if it were a conventional disk image, allowing any IRAF task to be used. This allows tight integration of IRAF quick look or analysis tasks with the RTD. It is even possible to use an IRAF task to operate upon the incoming image during readout, before readout has completed.

2.4. Data Processing

The Mosaic DHS includes a full pipeline data reduction capability, plus facilities for taping, archiving, viewing and managing the data set. All data reduction is performed by IRAF tasks under the direction of the *Data Reduction Agent* (DRA). The DRA is driven by a device dependent script hence is user configurable and easily adapted for new instrument configurations. Companion papers by Frank Valdes (Valdes 1997a, 1997b) describe the Mosaic data format and data reductions.

3. Summary

The primary function of the Mosaic data handling system is to process data from the NOAO Mosaic. The significance of the project is much greater however. The DHS itself is applicable to any type of data and when completed will be used for general data acquisition within NOAO and at other observatories. The message bus, DSO, and plug-in image display (RTD) technology used by the Mosaic is being developed as a more general facility for use in the IRAF system and by other projects. This work is supported in part by grants from the NASA Astrophysics Data Program and from the NASA Applied Information Systems Research program. See the Mosaic project Web page (http://iraf.noao.edu/projects/mosaic) for additional information on these efforts.

References

Valdes, F. 1997a, this volume, 455

Valdes, F. 1997b, this volume, 459