The NOAO High-Performance Pipeline System: The Mosaic Camera Pipeline Algorithms

Francisco G. Valdes

National Optical Astronomy Observatories, Data Products Program, P.O. Box 26732, Tucson, AZ 85726

Robert A. Swaters

Department of Astronomy, University of Maryland, College Park, MD 20742

Abstract. The NOAO Mosaic Camera Pipeline produces instrumentally calibrated data products and data quality measurements from all exposures taken with the NOAO Mosaic Imagers at the KPNO and CTIO telescopes. The types of calibrations and data products are described in a companion paper (Swaters 2007). This paper describes the algorithms and methods used for many of the calibrations applied to the data.

1.   Introduction

The NOAO High-Performance Pipeline System (NHPPS: Cline 2007) provides a framework for developing parallel and distributed pipeline applications using conventional host-callable data processing components. Our first pipeline based on this system is for instrumental calibration of data from the two NOAO Mosaic Imager cameras. These consist of 8-2048x4096 CCDs providing a 8192x8192 pixel format with minimal gaps and a 0.5 degree or 1 degree field of view at the NOAO 4m and 0.9m telescopes respectively. The instrumentally calibrated data products from these cameras are to be provided to the investigators and, after a proprietary period, to the community through the NOAO Science Archive and Portal (Smith 2007).

In order to make best use of the NHPPS, the data flow and processing algorithms must be structured to benefit from data parallelism. The NOAO Mosaic Camera Pipeline (henceforth the Mosaic Pipeline) structure and performance is summarized by Swaters (2007). In this paper we outline the various algorithms which make up the pipeline.

As noted previously, the Mosaic Pipeline applies instrumental calibrations to produce standard data products. The calibrations fall into the categories of photometric uniformity, astrometric evaluation, data quality characterization. The data products include images and pixel quality masks in both the observed pixel sampling and in a standard spatial orientation and sampling combining all the detectors of the Mosaic camera.

2.   Photometric Uniformity

CCD digital cameras suffer from variable pixel response. The key instrumental calibrations correct every pixel to a common linear and uniform flux system. Most of the algorithms are standard CCD ones with a few additional constraints due to the mosaic nature of the camera, such as balancing the gains between detectors and removing cross-talk signal produced by bright sources. For pixels that cannot be calibrated, CCD defects and saturated pixels, the pipeline replaces some with reasonable data and flags all pixels in a bad pixel mask.

The Mosaic Pipeline applies calibrations using overscan pixels provided by the controller, bias exposures, dome flat field or twilight sky flat exposures, and flat fields, generally referred to as dark-sky flat fields or super-flats, derived from the set of observations over one or more nights. The dark sky flat fields are needed to remove several instrumental photometric patterns; namely, fringing, out-of-focus light due to corrector reflections, and the color and illumination dependent response patterns not corrected by dome and twilight flat fields.

The biggest challenge in the photometric calibration is selecting exposures from one or more nights to form a deep sky stack with sources eliminated. It is a challenge because the pipeline must automatically process data from a wide range of observing programs. The approach is basically to identify and reject exposures which are not suitable and then determine if there are enough exposures with sufficient sky coverage to make an acceptable stack. The areas of rejection are exposures with instrumental problems, low sky levels, low transparency and fields which are crowded or have large sources. Then we look at sky maps for evidence of structure from unresolved sources and analyze the overlap statistics for sufficient sky coverage at each detector pixel. The final decision is based on a minimum number of remaining exposures and, if insufficient, make use of the fact that as an observatory pipeline, data from other nights and programs may be used as a fall back.

3.   Astrometric Calibration

Astrometric calibration means determining the celestial coordinates of every pixel in the exposure. This is done using non-linear functions fit to the measured pixel positions of stars from a reference catalog. The most challenging algorithm in this calibration is matching detected sources in the exposure to reference sources in a catalog.

The astrometric calibration is one area that cannot be totally parallelized. This is because, at least for our data, using sources from all the images produces a more robust match to the reference catalog. The initial world coordinate system (WCS) includes the stable physical relationship between the CCDs. The uncertainty in the astrometric calibration is then a zero point offset, a small rotation from remounting the camera between maintenance periods, and atmospheric distortions. The parallelization of the astrometric calibration consists of generating source catalogs for each CCD in parallel. This is typically distributed across nodes with the bulk image data. The catalogs are then combined and matched on one pipeline node against a reference catalog selected using the assumed telescope pointing. Because only the catalogs are used to determine the

astrometric calibrations the data transfers are small. The penalty is primarily the need to synchronize the processing for an exposure at this point. This is offset by allowing other parallel calibrations steps to proceed at their own pace after the catalogs are generated and sent to the global astrometric calibration node. Once the solutions for each CCD are determined they are sent back to the nodes where the calibrations are being done.

The matching algorithm is too complex to describe here. Basically since the camera orientation is fixed and the scales are known and stable a correlation offset in x and y is used as opposed to a more sophisticated triangle matching algorithm required for arbitrary relative orientations.

4.   Data Quality Characterization

The primary purpose of the data quality characterization in the Mosaic Pipeline is to provide accurate enough information about the depth and image quality of the exposure to allow instrument scientists to track the performance of the camera and to allow archival researchers to determine if the data are of use to them. There are additional data quality measurements and algorithms for internal characterization and trending which we cannot cover in this paper. Accurate enough means that the pipeline does not perform the complete photometric characterization that the investigators will perform using standard star observations across several filters. Instead it uses measurements of detected sources matched to reference sources to find a first order zero point relative to the photometric system or a standard transformation of that system.

The image quality is characterized by the average FWHM about the peak of the histogram of the FWHM over all sources. Note that the classification of sources as stellar is implicit in the histogram. The photometric depth is computed based on the sky statistics to estimate the instrumental magnitude of a five sigma detection through an aperture matched with the seeing. The photometric zero point, and its use in converting the photometric depth to magnitudes, is the catalog of matched image and reference sources. Currently this is a catalog of matched instrumental magnitudes and three color magnitudes from the USNO-B Catalog (Monet 2003). The photographic magnitudes are converted to Sloan magnitudes using the published calibration. The most appropriate Sloan magnitude for the Mosaic filter of the exposure is selected and a simple zero point which minimizes the residuals between the instrumental and reference magnitudes is computed.

5.   Data Products

There are two primary data products from the Mosaic Pipeline. One is the set of individual CCD images and the other is a resampling of the images into a single mosaic image of the exposure. The first format avoids introducing complex correlations between pixels while the second provides an easier to use format for stacking and inter-comparisons without the need to handle the instrument specific astrometric distortion mapping.

The principle algorithms involved here are determining the astrometric system of the resampled version and the resampling algorithm. The basic astro-

metric system is the standard "north up and east to the left" at a uniform pixel scale of 0.25 arcseconds/pixel. However, because of the inherent projection of an image to the celestial sphere we require a projection function. We chose the common "tangent plane" projection which introduces a "tangent point" parameter. Rather than simply using the center of each exposure as the tangent point we wish to allow potential stacking of data products without additional resampling. In particular, many of the datasets are observed in dither patterns with the intent of stacking the exposures for greater depth and to eliminate the gaps in the mosaic format. There is also the possible serendipitous overlap of exposures between different observers and programs.

The Mosaic Pipeline handles the case of observer intent by recognizing dithered observations taken using the standard NOAO Mosaic data acquisition command for dither sequences. The second case of unrelated exposures of the same field are handled by selecting tangent points from a grid on the sky which has roughly points every degree on the sky. Naturally this will fail when two pointings cross a grid boundary, but it will work for most cases, certainly more than not using a grid.

The actual resampling algorithm is a compute intensive sinc interpolation which minimizes spatial correlation patterns in the random noise. We can afford to use this more complex method even under high data rate requirements because of the parallel and distributed structure of the pipeline. The data parallel aspect of the resampling problem is that each CCD element can be resampled in parallel and then the pieces are added together, using only integer pixel origin offsets, into a single image when the final data product is created. The only requirement for this is that the same tangent point is selected for each CCD image of an exposure.

One other algorithm to mention is that the pipeline image data products are compressed by digitization to 16-bits. The images are in standard FITS format using 16-bit signed integers with keywords defining a linear mapping to physical values. The key aspect of this algorithm is using only unsaturated good pixel values to define the range of physical values to be digitized into the full range of 16-bit integers and then insuring that the 1 sigma uncertainties in the pixel values are sampled by at least a factor of 30. If a digitization satisfying this criterion cannot be obtained then the data product is created as a FITS 32-bit IEEE floating point image. For the Mosaic camera data the compression is successful more than 90% of the time.

References

Cline, T. et al. 2007, this volume, [P4.19]

Monet, D. et al. 2003, AJ, 125, 984

Smith, C. et al. 2007, in ASP Conf. Ser., Vol. 125, ADASS VI, ed. G. Hunt & H. E. Payne (San Francisco: ASP), [O6b.1]

Swaters, R. & Valdes, F. 2007, this volume, [P4.20]