NOAO DPP Document PL003

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

F. Valdes[1], R. Swaters[2], T. Huard[2]

National Optical Astronomy Observatory
Data Products Program

February 1, 2008

[1]NOAO Data Products Program, P.O. Box 26732, Tucson, AZ 85732
[2]Department of Astronomy, University of Maryland, College Park, MD 20742

Table of Contents

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.

Keywords: pipelines, algorithms, CCD, Mosaic

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.

This document describes the calibration. It does not describe the data parallel and distributed algorithmic implementation. The implementation is a complex subject in its own right because ... While this document does not try to describe the data flow orchestration of the pipeline there is some discussion of the datasets upon which an operation is performed. Datasets are groupings of data. The groupings change during processing and some datasets are subsets of other datasets.

In the description that are many parameters identified. In some cases these are indicated to be parameters or set by an external mechanism. In other cases the parameters are simple set in the implementation. When appropriate the values being used as of the writing of this document are specified. It should be obvious that these might be changed, either for specific datasets or as future recommendations change. While it is clear that parameters can be changes it should also be understood that fixed values in the code are also open to change.

There are literally thousands of parameters which define the behavior of the pipeline stages and support commands. Many of these are explicitly set in the modules, colloquially called ”hard wired”. Those that are not are defined in parameter files. The parameter files are collected in a parameter directory. If a parameter file is not included in the directory then the default parameters for the program are used. Different versions of this parameter directory are stored in the calibration library and the each pipeline selects a directory based on the instrument and observation time. Other categorizations supported by the calibration library, such as filter or mosaic image element, but a need for this level of customization has not been found necessary.

Most of the processing stages are implement as IRAF scripts calling key IRAF data reduction and image processing tasks. Additional detail is available in the technical documenation for those tasks. However, it is the goal of this document to provide a complete description of each algorithmic step.

We believe the NOAO Mosaic Science Pipeline is unique among science pipelines in its application of dark sky self-calibrations. This is an approach widely used by individual astronomers for reducing their own data. It is challenging for an automated pipeline because it requires care in selecting and rejecting exposures for making the dark sky stack. The methods and challenges are discussed in section 15. Most science pipelines either rely on twilight sky flats, which the Mosaic pipeline also uses in some cases, or on externally derived self-calibrations.

The pipeline methods have been informed and influenced by earlier work leading to tools and guides for Mosaic Camera observers to reduce their data. The references are [ADA], [User Guide], [NDWFS].

2 Pipeline Context and Datasets

The NOAO Mosaic Science Pipeline calibrates Mosaic Camera data taken by any observing program using the instrument. The processing occurs some time after the observations are completed. In regular operation this happens within a few days after the observations at a science pipeline data center; currently in Tucson.

This context of processing a variety of programs after completion of some or all of the nights leads to a particular pipeline operational model. This is discussed in greater depth by **?**). The main consequence of context and model is that the pipeline operates with many exposures taken over several days to hopefully include sufficient calibration material. And even in this case weather or observer protocols may result of a lack of calibration data. The context of the pipeline allows use of calibrations derived previously from another program through the use of a calibration library. The pipeline both stores any calibrations derived during processing and retrieves any calibrations previously stored.

3 Calibration Manager

The NOAO pipeline applications make use of a Calibration Manager (CM). This server application is described in more detail by **?**). In this document it is only necessary to understand that the CM maintains a calibration library where calibration data and processing rules are stored and retrieved. The calibration data is indexed by instrument, detector, filter, exposure time, and quality attributes. In addition, there is a range of applicable dates which are normally the time when the data used to create the calibration was obtained.

Calibration rules are also maintained by the CM. These are scripts which are indexed by applicable dates.

4 Remediation

The first thing the pipeline does after staging a dataset is perform verification and remediation steps. First there is an optional ”special” step. The operator may insert a command to be executed when the dataset name matches a particular pattern; for instance, data from CTIO. This is not common but one command that has been used on occasion checks for proposal identifications from split nights that were not remediated earlier and pauses the pipeline for manual operator remediation if needed.

The main verification and remediation step is performed by a pipeline developed command called dirverify. This uses a text database, obtained from the calibration library which is appropriate for the dataset, to verify the header keywords. Currently, there is no specialization and the database applies to all Mosaic Camera data. The database identifies keywords to be checked. The checks include whether the keyword must be present and a pattern the value must match. There may be a remediation strategy, either a specific method for the keyword or a more generic method, to be applied if a check fails or an error action to be taken. One error action is to exclude an exposure from further processing which means a keyword is required for the pipeline to process the exposure.

Examples of keyowrds that may be remediated are marginally malformed filter names which may still be identified with the standard filter name.

[dirverify]

5 Cross-talk

The 8 Mosaic camera CCDs are read out using 4 ARCON controllers. Each controller operates 2 CCDs through single or dual amplifiers. The CTIO camera uses dual amplifiers while KPNO camera uses a single amplifier per CCD due to non-working amplifier channels. There is electronic cross-talk between the channels in each controller, though no cross-talk between controllers. The signature of this cross-talk is a proportional response in other channels from signal in each channel. The proportionality coefficients are small so primarily very bright sources produce faint ghosts which are correlated by the read out order.

The correction applied by the pipeline consists of subtracting scaled copies of each contaminating amplifier image, in read out order, from each target image. A single scaling coefficient is determined between each pair of amplifiers from the same ARCON controller. These coefficients are determined outside of the pipeline using special software developed for this purpose and stored in a calibration database used by the pipeline because these cross-talk coefficients have been found to be fairly stable.

6 Read Out Bias

Each amplifier channel has an electronic bias which varies for each pixel value read out. The controllers sample the bias multiple times at the end of each line read out and record the bias values as overscan pixels. This results in the raw data including a set of overscan columns. For the Mosaic cameras there are 64 overscan columns for each amplifier image.

In processing the overscan columns are collapsed into a single column vector. The simple average of all columns define an *average overscan vector*. The average with the highest and lowest value excluded is called a *minmax overscan vector* while a single 5-sigma clipping iteration is call a *5-sigma overscan vector*. A polynomial function fit to the average overscan vector produces a smoothed calibration called a *polynomial overscan vector*.

The pipeline electronic bias removal follows the widely used CCD reduction model of subtracting the overscan vector from each column of the image followed by a two dimensional bias subtraction using a dark or bias calibration calibration image. This section describes the overscan vector subtraction and section 11 describes the 2D pattern subtraction. The overscan vector then calibrates rapid electronic bias changes and drifts between exposures, generally from amplifier temperature changes, while the 2D bias pattern, typically from voltage noise, calibrates a more static electronic pattern.

There are five overscan subtraction options: minmax, constant, polynomial, constant or minimax, and polynomial or minmax. The option used is defined by a rule in the calibration library. Currently, the rule selects the minmax method for all science and calibration data except for short exposures, those less than 120 seconds, where a polynomial fit is used unless there is evidence of rapid changes in the line bias, called *bias jumps*.

The ”minmax” option subtracts the minmax overscan vector from each image column. The ”constant” and ”polynomial” option subtracts the polynomial overscan vector where a 1st or 101st Legendre polynomial is used respectively. The constant option is mathematically the same as subtraction the average of all bias pixels from all image pixels.

The other two options apply a constant or polynomial subtraction, for minimizing sampling noise and low level line patterns, except when bias jumps, abrupt and erratic changes in the bias, are detected. When bias jumps are present the minmax subtraction is used where the bias for each line is independent.

Bias jumps are identified using the 5-sigma overscan vector. This vector convolved with an ”edge detection” filter of the form [-1,...,-1,1,...,1] where the transition is in the middle. The current filter has a half-width of 10 pixels. The convolved strip is checked for positive and negative peaks. If a peak is found, a straight line is fit around the peak with the same width as the filter. If the reduced χ^2 is larger than a fixed (empirically determined) number, indicating the region is not a steep gradient, it is identified as a bias jump. For each detected jump, the beginning and end of the points, the length, and the amplitude (both up and down) are recorded, as are the maximum lengths and amplitudes, both up and down. The probability that a detection is a jump is estimated, although usually the answer is 0 or 1.

Another problem (data quality) condition detected from the overscan is electronic drop-out. If the sixth lowest value in the 5-sigma overscan vector is negative or zero the image from the amplifier read out is flagged as bad. Often this condition consists of the entire read out having zero values. The exclusion of a small number of low values is to accept cases where the initial lines have problems but the majority of the data is usable. When an image is flagged as bad it still processed but excluded from some algorithms such as stacking or resampling.

After the overscan columns have been used for bias analysis and subtraction the image is trimmed to the useful science pixels. The bias columns and trim region are set by the data acquisition system. It would be a simple extension to obtain this information from the calibration library if a need arises to override the data acquisition system values.

When dual amplifier are read out from a CCD, resulting in two images with their own overscans, they are merged into a single CCD image after the overscan bias subtraction. From then on the pipeline operates on CCDs rather than amplifiers.

The core task which implements the overscan subtraction and trimming is ccdtool (a version of ccdproc) and the merging is performed by combine, both from the mscred package.

7 Pixel Masks

Detector pixel mask, exposure mask, object mask. Data products.

Pixel masks are an important component of the Mosaic pipeline algorithms as well as one of the data products. Pixel masks are conceptually integer raster images associated with the flux images whose values identify various properties of the flux pixels. They may also be used to identify regions like a stencil. The association of mask rasters and flux images is usually made through there pixel raster coordinates though more advanced usages may apply a mapping.

Generally one value, zero, is used to identify no property and this value is also typically the most common. This allows use of special formats where only values identified with some property are stored. The Mosaic pipeline uses IRAF pixel list storage format which provides only non-negative values and which is compact when the most common mask value is zero.

The pipeline uses three types of pixel masks. There are detector bad pixel masks which are applied to many exposures. These identify pixels which are not scientifically useful. They are fairly static and are created by instrument support personnel and entered into the calibration library for use by all exposures taken with those detectors. There are two non-zero mask values. A value of one is used for low level or large regions of bad pixels which are to be propagated to the final per exposure mask. A value of two are for bad columns or small defects which are also propagated to the final exposure mask but are also replaced in observational data by linear interpolation along lines or columns.

Exposure masks are those associated with a particular image data product. The mask values include not only the detector bad pixels but also additional pixels specific to the exposure. Currently these are saturated pixels (§**??**) and bleed trail pixels (§**??**). For resampled, mosaicked, and stacked data the masks identify areas with no data, such as gaps or corners, and interpolated pixels with significant contamination from bad pixels in the original CCD exposures. Currently, cosmic rays are not identified by the pipeline but when they are they would also be included in the final data product mask.

The last type of mask used in the pipeline are called object masks. These are the result of segmenting the image data into pixels with identifiable source flux. Also each unique source has a particular mask value and codes related to the segmentation process. The converse, that is pixels with no object mask value, obviously identifies the sky or background pixels. Object masks are important when combining exposures, particularly when creating deep sky self-calibrations (§15).

8 Saturation

Saturated or significantly non-linear pixels are defined as having values above a threshold provided by the calibration library. These are detected in the raw images before any other processing. Neighbors, defined by a cartesian distance between pixel centers along lines and columns, of saturated pixels may also considered as affected by saturation as noted in the subsections. Pixels identified in the detector bad pixel mask will not be tested for saturation.

8.1 Saturation in Bias, Dome Flat, and Twilight Flat Calibrations

Individual calibration images are checked only for the fraction of saturated pixels not counting neighbors. If the fraction of saturated pixels in a CCD (that is after merging of multiple amplifiers) is greater than a value set by a calibration rule (see §12.1) the CCD for that exposure is rejected for further use in calibrations. This is designed to eliminate dome and twilight flats which are over exposed, something that happens when observers are first experimenting with the exposure times.

The saturated pixels are also replaced by linear interpolation for cosmetic purposes and to avoid extreme behavior when applied. There is no propagation of these pixels into the science exposure pixel mask.

8.2 Saturation in Science Exposures

Saturated pixels and their neighbors within one pixel are identified. They are recorded in the exposure pixel mask with a value of 4. The pixels are also replaced by linear interpolation for cosmetic purposes.

9 Bleed Trails

Bleed trails, only in science exposures, are identifed as regions with some minimum number of consecutive pixels along a column having values above some threshold. The pixel values are those in the raw image before they modified by any other calibration. Neighboring pixels pixels along lines or columns, of the threshold selected bleed trail pixels are also identified as part of the bleed trail. The threshold used is 5000 ADU above the mean of the image and the minimum column extent above this threshold must be at least 15 pixels. Any pixels identified as bad in the detector pixel mask are excluded for consideration as bleed trail pixels.

The identified bleed trail pixels are recorded in the exposure mask with a mask value of 5. The pixels are also replaced by linear interpolation along the lines for cosmetic purposes.

10 Bad Pixel Interpolation

In earlier sections mention was made of replacing pixels in the flux images with interpolated values. This is done for two reasons: 1) to allow operations which do not check a bad pixel mask to avoid undesirable behavior from possibly extreme values and 2) for producing an esthetically appealing display. In all cases the modified pixels are identified in the exposure mask which indicates pixels that are not scientifically useful.

The interpolation is always linear interpolation between the first good pixels either along columns or lines. Bad regions that extend to the edges use constant extension from the nearest good pixel along the line or column. Depending on the application, the interpolation is either along the dimension with the smallest span of bad pixels or specifically along a particular dimension. An example of the latter application is the replacement of bleed trails which are always vertically oriented and so interpolation is always along lines.

11 Bias and Dark Count Pattern

After removal of the electronic bias pattern the standard CCD calibration model assumes a fairly stable two dimensional bias and dark count pattern. The pattern is removed by subtraction of a calibration image derived from calibration exposures with the shutter closed.

If the pattern is dependent on the exposure time, primarily due to the accumulation of signal from dark counts, then the calibration should be a ”dark” calibration with an exposure time matching the science exposure. The Mosaic Camera CCDs do not have significant dark counts so the shortest exposure time, a ”zero” or ”bias” calibration, is equivalent.

11.1 Bias Calibrations

The pipeline will create bias calibration images from all bias and dark exposures.

12 Dome Flat Fielding

Dome flats provide the first level of gain calibration.

The dome f

Mention twilight flats.

12.1 Dome and Twilight Flat Calibrations

After all the bias sequences have been processed and stored in the calibration library the next stage is the processing and creation of dome and twilight flat calibrations. The flat exposures are identified and grouped by filter. The exposures in each filter are then grouped as sequences. The sequences are identified heuristically by successive dome flat exposures of the same filter with a time between the end of one exposure and the start of next exposure less than ten minutes.

Each sequence is first corrected for overscan and bias, trimmed, and bad pixels are replace. Any value less than 1 ADU is set to 1 ADU. If there are multiple amplifiers they are merged into a single CCD image. As described in §**??** the fraction of saturated pixels is computed and used to reject CCDs with too many saturated pixels.

The next step is to average all the exposures in a sequence into a single master flat field.

Some data quality rules are applied from the calibration library. If an individual flat fails a rule a flag is set that the image is unreliable. An image fails if 1) mean pixel value is less than 2000 ADU for broadband filters or 1000 ADU for narrowband filters (as classified by the calibration library) or twilight flats, 2) if the mean is greater than 90% of the saturation level for the CCD (as defined by the calibration library), and 3) if the precentage of saturated pixels is more than 0.01flats or 0.5% for twilight flats.

j

is flagged as unusable.

the end of one exposure and the start of another that is less than

The dome flat exposures are grouped by filter

13 Data Units

The flux calibration includes adopting a data unit convention for the linearized pixel values. This carries through to all flux data products and calibrations. The Mosaic pipeline calibrates the pixel values to detected photons per second, assuming a conversion of one photon to one electron in the CCDs. What this means is that the observed digital data numbers (DN), after overscan, bias, and dome flat calibrations, are multiplied by a gain in electrons per DN and divided by the exposure time. The gain is determined external to the pipeline by instrument scientists and entered into the calibration database.

The metadata for the exposure includes the effective and observed gain and exposure time. In our choice of data unit the effective gain and exposure time are 1.

14 Astrometric

Astrometric calibration consists of determining the celestial coordinates of every pixel in the exposure. This is done using a non-linear function fit mapping the measured pixel positions of sources in an image to the celestial coordinates of those same sources obtained from a reference catalog. This mapping is generally referred to as world coordinate systems (WCS). Though separate WCS functions must be determine for each image in the mosaic instrument we will generally refer to the collection of functions as the WCS for the exposure. The most challenging algorithm in the astrometric calibration is matching detected sources in the exposure to reference sources in a catalog.

Initially the raw data has a WCS set by the data acquisition system. The pipeline overrides this if there is another applicable WCS in the calibration library. In either case the WCS is only approximate because of uncertainties in the telescope pointing and offsets to the instrument field, small mounting errors, and atmospheric distortion effects.

The astrometric calibration begins by creating two catalogs. The first is a catalog of sources extracted from a reference catalog for the field covered by the exposure. Because of the telescope pointing uncertainties a field larger that that of the exposure is extracted. In the Mosaic pipeline a circular field from the USNO-B catalog is used. The catalog contains the right ascension and declination of the sources along with g, r, and i magnitudes computed from the J, F, and N photographic material defining the USNO-B catalog.

The second catalog is a detection catalog containing the pixel centroids and digital instrumental magnitudes of sources in the exposure. The matching of sources from the two catalogs uses a coarse offset determination from the bright sources followed by a fine centroiding of all sources. The coarse offset determination

[Flags] [Algorithm info here]

A raster of potential pixel offsets from the nomimal position of a reference source is used to accumulate ”votes” for the offset. The raster size is set by a parameter defining the maximum postion error. The raster bin size is one pixel in the target coordinate system. The votes are cast by computing the pixel coordinates for each reference celestial coordinate in the target image. All target catalog sources within the raster then cast a vote for the offset.

Because this is an N*M operations and a high density of sources takes the power out of the method, the magnitudes of the reference and image sources are used to limit the set of sources used.

There are uncertainities in the source position from the WCS solution, proper motion, and seeing. The vote is distributed over an region by convolving the position of the catalog by a Gaussian. In effect, the delta function of the catalog position is turned into a Gaussian profile. The width of the profile is a parameter. For MOSAIC data the full-width-half-max of the convolution is 1 arc second. In implementation the convolution is applied to the vote array after all the delta function votes are cast.

This algorithm was initially developed to work from image data rather than catalog data. The convolution of the catalog source positions is basically equivalent to simulating an stellar image for this algorithm.

After all the votes are cast and the vote array is convolved it is searched for maxima. If there is no point which is at least 20 sigma above the average the algorithm reports a failure. A threshold is then set at a point half way between the maximum value and the average value. Then a weighted centroid is computed using the value above the threshold. The centroid relative to the center of the vote array gives a coarse offset.

[Wriggle for rotations]

Once the coarse offset is determined a simple closest radius match between the source catalog and the reference catalog is performed. The nearest source to a reference catalog that is within 2 arc seconds forms a match between a source and reference. These matches produce a ”matched catalog”.

The matched catalog is used to produce the photometric and astrometric calibrations.

The advantage of this algorithm is that is fast, it works in the presence of significant distortions.

This algorithm allows parallelization by generating the source catalogs in parallel, requires only the catalogs to be brought together, and the vote accumulation can be done over all the pieces of a mosaic.

14.1 Photometric Calibration

The photometric calibration produced by the Mosaic pipeline is intended as a rough guide to the depth of the exposure. It is based solely on the single exposure so there are no color terms. It also depends only on the sources in the exposure rather than reference to a standard star exposure. In the future the pipeline may identify and use standard star observations within the dataset.

Since the primary reference catalog is USNO-B and there is a wide variety of principle investigator programs using the full complement of Mosaic Camera filters the photometric system is not a close match. What the calibration does is to determine a photometric scale or zero point magnitude that matches, on average, the instrumental fluxes to the reference fluxes of the matched catalog sources in a reference flux photometric system.

The USNO-B catalog provides photograhic magnitudes. For much of the sky these are based on the J, F, and N emulsions. Where this is not available the O and E emulsions are available. Rather than provide the zero points in one of these bands the published calibrations to the SDSS g, r, and i systems are used.

The calibration library provides for a transformation from (g, r, i) to some other system based on the Mosaic filter. Currently only identity transforms are used with the particular SDSS system set in the library being defined by wavelength proximity. For example, blue broad or naorrow band filters use g, yellow and red to r, and far red to i. In the event that there is no N reference magnitude for far red filters the calibration is made to r where a transformation from O and E is defined.

A selection is made from the matched catalog to eliminate sources with questionable or low signal-to-noise instruental magnitudes. First all sources which the catalog identifies as being impacted by saturation, blead trails, or bad pixels (see §**??**) are excluded. Remembering that the instruental magnitudes in catalog are calibrated to the expected zero point for the filter using the calibration library, large magnitude differences (greater than 20) are rejected. Next the brightest 10% and faintest 20% of sources (ref & cat mags?) are excluded. Finally, the remaining matched sources are averaged to yeild the approximate photometric zero point.

Both the mean and sigma are computed and recorded for data quality reference. Sources in the output matched catalog which are more than 3 sigma from the mean are flagged.

The zero point and sigma are computed both independently for each CCD and over all CCDs. Note that this is not the average of each of the calibrations for each CCD but the values using all sources in all CCDs. This simply means that the weighting of the CCDs for the final exposure zero point is by the number of matched and accepted sources.

Because the reference pbotometry is typically not in the same photometric system as the

14.2 Astrometric Calibration

in the image using the initial WCS as a guide and produces an accurate WCS for the data.

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.

The detection catalog and catalog matching are performed using tasks from the ACE (Astronomical Cataloging Environment) package. The WCS function fitting and evaluations are from the core IRAF system tasks.

15 Dark Sky Self-Calibrations

Dark sky self-calibraion refers to a technique that uses a set of disregistered observations to eliminate astronomical sources and produce an approximation of a blank sky exposure. Under the assumption that a blank sky exposure should be uniform, any structure is due to instrumental artifacts which must be removed either as extraneous light or as non-uniform response. The calibrations derived from this are generally better than dome flat fields because the color (the spectrum across the filter band pass) and illumination of the dome flat field is a poorer match to the science observations than blank sky.

Because of astronomical sources in any given exposure only a subset of the pixels are exposed to blank sky. The principle behind dark sky self-calibration is that in the ensemble of exposures there are some exposures that sample blank sky for every pixel. When this is not the case for the most pixels the data cannot be self-calibrated and either a calibration from another time (such as from an earlier program) must be used or the data is calibrated using only dome calibrations.

A consideration in determining the calibrations is whether they degrade the data because of noise in the calibration is worse than the instrumental pattern being removed or because unresolved sources and large-scale extended light introduce artifacts. Ideally every pixel will have a large number of clean samples of blank sky. But when this is not the case, there is an intermediate level of calibration where large scale instrumental patterns may still be removed even if individual pixel responses may not. Therefore, another aspect of the dark sky self-calibration is deciding whether a full pixel response, a large scale pattern, or no calibration is best.

Observers who self-calibrate their data in this fashion examine the individual exposures and apply their judgement in selecting a set of exposures for the calibration. The challenge for the Mosiac pipeline is to heuristically determine the set of exposures to use or, conversely, not use and which level of calibration is supported by the data. This is difficult because the criteria are subjective and observers apply complex trade offs and the ability to recognize the content of the exposures. This is made more difficult because, as noted in the introduction, the observational programs and protocols are unconstrained so that some sets of data are not suited to this type of self-calibration.

Nevertheless we have attempted to develop a set of criteria and algorithms to automatically determine whether self-calibration may be performed, at what scale, and with what set of exposures. This section describes these methods.

There are currently three types of dark sky self-calibrations performed by the Mosaic pipeline. These are pupil ghost removal, fringe subtraction, and flat fielding. The first two are structures which must be subtracted and the last is a response behavior.

15.1 The Approach

The dark sky calibration portion of the pipeline begins with a dataset consisting of science exposures from one or more consecutive nights taken with the same filter. The exposures are winnowed through a series of tests. At the end of these tests a set of exposures for creating blank sky images is identified.

If there are insufficient exposures then self-calibration is not done and, instead, a library calibration is used. When no suitable library calibration is available the exposures are left calibrated only by dome calibrations.

When there are enough exposures the exposures are scaled and combined using rejection methods based on statistics and source segmentation masks to eliminate data which does not sample blank sky. The instrumental structure in the blank sky exposure is used to calibrate each exposure.

For the response calibration, the dark sky flat field, the signal-to-noise of the calibration is evaluated and when it is poor a filter is applied to suppress the pixel noise while retaining the large scale illumination response calibration.

15.2 General Considerations in the Dark Sky Self-Calibrations

This section discuss various general considerations in the dark sky calibration method and algorithmic and implemenation choices that apply to all the instrumental effects.

The problem of calibrating exposures which are affected by both additive (pupil ghost and fringe patterns) and multiplicative (pixel responses and illumination patterns) is difficult and not strictly decomposable. What is done is to use the dark sky calibration process iteratively. This means forming a dark sky exposure, using it to extract the strongest instrumental feature to be removed, calibrating all of the exposures, and then repeating the steps for the next calibration. For the Mosaic camera the calibrations are done in the following order. For those datasets affected by a pupil ghost (Mayall telescope exposures in certain filters) the pupil pattern is determined and subtracted from the exposures. For those datasets affected by fringing (red filters from both telescopes) the fringe pattern is extracted and subtracted from the exposures. Finally, all datasets create a dark sky flat field which is used to calibrate the pixel responses (if appropriate) and large scale illumination patterns across the field.

In the process of evaluating each exposure there are quantities that are measured independently, and in parallel, for each CCD of the mosaic; for example the maximum number of pixels segmented as part of a single object. One might consider accepting and rejecting a different set of exposures for each CCD. However it is possible for subtle continuity effects to occur between CCDs when the calibrations derived for each CCD are based on different exposures. Continuing with the large single object example, when a large source such as a galaxy or very bright star falls in one or two CCDs while the rest are relatively sparse and might be good candidates for dark sky self-calibration. The Mosaic pipeline currently takes the conservative approach of rejecting an entire exposure if only a subset of CCDs fails to meet some of the criteria described below. The criteria reported in the diagnostic then are the maximum or minimum over the set of CCDs.

Making a dark sky image consists of scaling each contributing exposure using the sky mean, computed earlier, and averaging the pixels with rejection of non-sky pixels. There are two aspects to the rejection. The first is use of the object segmenation masks for each exposure. The second is statistical rejection from the ensemble of pixel values. The rejection algorithm is a parameter with the default being a percentile clipping method. In brief, this method determines a measure of the distribution width from the pixels below the mean or median and clips pixels above the mean or median which significantly exceed this width. The implementation used for this operation is

IRAF's `imcombine`. This is a complex application with many options and features. For more information consult the documentation for this program.

As will be described later each of the calibration patterns are extracated from the dark sky image. These patterns are weak and so they are sigma clipped to identify outlier pixels. The outlier pixels are replaced the minimum or maximum value of the remaining pixels. This is done to avoid having extreme pixels affect the application of the calibration to the data. In particular, the pupil ghost and fringe pattern removal must scale the weak pattern to each exposure and the scaling calculation can be biases by a few extreme values in the pattern.

Each of the calibrations have a common pattern in their use of the calibration library. One feature of the pipeline is a configuration parameter that allows the generation of calibrations even if the analysis of the exposures determines that the dataset is inadequate for creating dark sky self-calibrations. These are generated from all the data [ARE ANY NOT USED?] but a quality parameter flags this as an "evaluation" calibration. These appear in the final review documentation and data products but are not stored in the calibration library nor applied to the data.

Another common pattern is that if a self-calibration cannot be derived from the dataset a calibration from the calibration library is used if available.

In principle, a different set of images may suitable for each calibration – pupil ghost, fringe, and flat field – and the pipeline is designed to allow this. However, currently the set of exposures for all dark sky calibration are the same.

Before describing the criteria a note about the mechanics. Ultimately whether a particular image is used for a particular type of calibration depends on image header keywords. There are two levels of specification; one that cannot be overridden ("Y!" or "N!") and one that can ("Y" or "N"). The former is intended for observers, operators, upstream software to prejudge the data. However, the usual case is that various distributed pipeline stages can veto the image.

15.3 Identifying Exposures for Dark Sky Calibrations

This section describes a variety of requirements and heuristics that lead to a subset of exposures from a dataset for creating a blank sky exposure. As noted previous, the approach is mostly one of winnowing exposures at each step in a series of steps. The steps are presented in the order in which the are evaluated and, unless otherwise indicated, excluded exposures are not considered in subsequent steps.

External acceptance: For special cases, particularly for reprocessing datasets for privileged investigators, the acceptance flag can be set to "Y" or "N" in the input data externally or in a special processing stage at the beginning of the pipeline. Those set to yes will be used in all subsequent test and not rejected by those tests while those set to no will never be used.

Dome calibrations: If the basic dome calibrations – overscan, bias, dark, and dome flat – fail for some reason the image is marked for exclusion. Currently the only data quality criteria of this type are a bad amplifier as identified by too many zero ADU overscan values and saturation as identified by the fraction of pixels above a saturation threshold.

Exposure time: Short exposures provide poor measures of blank sky because of low counts. A rule, which could be indexed by filter but is not currently, is applied to the exposure time. The current rule is a simple threshold of 120 seconds that is independent of any other factor such as filter.

Note that an exposure criteria is also implicit in the initial definition of a dataset. The pipeline divides short and long exposures into different datasets with the long exposures being processed first. The current division uses the same threshold exposure time of 120 seconds though the this can be easily changed. The consequence of using the same time is that all exposures in the long exposure dataset are accepted and all exposures in the short exposure dataset are rejected when creating dark sky calibrations. There for only long exposure datasets can produce calibrations. This is why the short exposures datasets are processed by the pipeline only after the long exposures so that any dark sky calibration exposures added to the calibration library for that night are available to be applied to the short exposures.

Fraction of sky pixels: Exposures with a high source density have fewer pixels observing blank sky. As part of the image segmentation step the fraction of pixels not associated with sources is determined. As mentioned earlier, this is computed independently for each CCD. The current rule is that the fraction of sky pixels must be greater than 85% in all CCDs.

Maximum object area: In addition to eliminating a large number pixels from sky, exposures with large sources are poor candidates for sky calibrations because they have large regions of low surface brightness wings that typically show through when making the master blank sky image. The current rule is that exposures with any single segmented source having more than 500,000 pixels in any CCD are excluded. Note that large sources often span more than one CCD. The Mosaic pipeline does not currently identify this situation so if the parts of a single source spanning multiple CCDs do not have more than 500,000 pixels in any CCD then the exposure will not be exclude based on this criterion.

Sky region: Certain regions of the sky have little if any blank sky. These are regions of nebulosity or very large galaxies. The Mosaic pipeline has the capability of consulting a database of positions on the sky and rejecting exposures that overlap these positions. There is currently no such database in the calibration library.

Magnitude zeropoint discrepancy: For filters that have an expected magnitude zeropoint for photometric conditions stored in the calibration library, the difference in magnitude zeropoint for the exposure relative to this photometric value provides an indication of clouds. The current rule is no brighter than 1 magnitude from the photometric value.

Positive mean sky rate: Exposures whose mean sky rate is less than 0.01 electrons/second are excluded. Partly this is because some sky is needed to make blank sky calibrations but this is required for scaling exposures by the reciprical of this rate.

Empirical magnitude zeropoint discrepancy: The previous check of the magnitude zeropoint was against a calibration reference value which may only be available for certain filters. There is also an empirical test for exposures with anomalously low zeropoints within a dataset. This consists of iteratively computing the mean zeropoint and standard deviation and excluding values more than twice the standard deviation. When the mean converges, meaning that no more values are reject, then exposures with zeropoints more than one magnitude brighter than the mean are excluded.

Sky brightness criteria: A set of exposures may include a minority inadvertantly taken through significant clouds or too far into twilight. These are identified and excluded by sigma clipping on the mean sky rates as measured over all CCDs. The sigma clipping is performed as with the magnitude zeropoint using an interative two standard deviation clipping to determine the typical sky rate. Exposures are then excluded which are more than three standard deviations from the mean.

Low resolution sky structure: Low resolution sky maps are produced by segmentation of the images. Note that this is done after dome calibrations but before the dark sky calibrations stages. The sky maps for the CCDs are tiled together to make a single small image of the field and normalized to unit mean sky rate. The sky maps are sensitive to large, low surface brightness light from large sources or clusters of bright source. In this context large sources also includes very bright stars.

Sky maps, and therefore the exposures, are used to exclude those with significantly different structure from the average. The average sky map is formed by fitting and subtracting a planar gradient across the whole field from each map and then averaging the results. This average is subtraated from each sky map and the root mean square (RMS) over all map pixels is computed. An iterative rejection of large RMS values is performed on the list of RMS values using a two standard deviation threshold above the mean RMS. The exposures corresponding to the sky maps whose RMS value are rejected are then excluded from consideration.

bf Rejection of low resolution, non-flat field, structures: A new average of the surviving, normalized and gradient subtracated, sky maps is formed. This is treated as a low resolution dark sky flat field which is divided into each of the sky maps. This flat fielding is done because some of the structure in the sky maps is the flat field and pupil pattern features we are trying to address by dark sky calibration. One feature that is easy to see in some data is the pupil ghost pattern which has not yet been removed. Though this feature must be subtracted for a correct scientific calibration in this step of identifying suitable images for a dark sky calibration it is removed by flat fielding.

What is left in the sky maps after this flat fielding are structures caused by undetected faint light from large sources in the exposures. While many exposures with large galaxies or extremely bright stars were removed earlier using the full images, there are some situations that are not caught. These include sources spanning more than one CCD or groups of moderately large sources, especially bright stars.

To identify the exposures with this large scale faint light, the ACE source detector is run on the sky maps with parameters appropriate for the small size and low resolution of the maps. A fairly large threshold (7 sigma) with no convolution is used to find only significant structures. Those maps detected source are rejected. However, if fewer than four exposure survive then the result of this rejection step are ignored.

Overlap statistics: Once a tentative set of exposures is identified, the number of exposures that sample blank sky at each pixel is determined using the object segmentation masks. A cumulative histogram is created of the number of pixels with fewer than one sky sample, fewer than two sky samples, and so forth up to the number of exposures. This cumulative histogram is expressed in terms of the percentage of exposures and the percentage of the total number of pixels. Figure **??** shows a histogram for the case of eight exposures along with the acceptance threshold. To be accepted the points have to all lie below the curve as in this example.

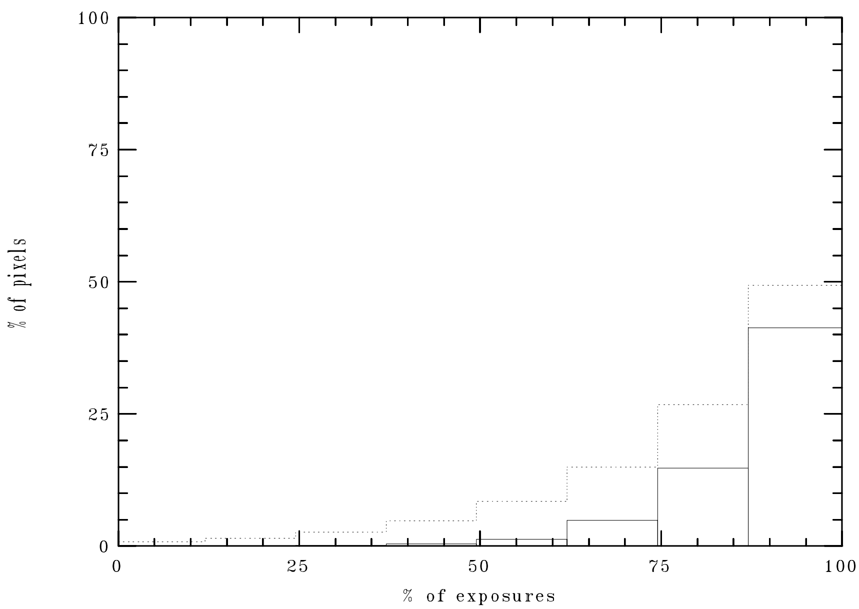

Figure 1: Example overlap histogram for eight Mosaic exposures. The solid line is from the sky pixels identified in the exposures and the dotted line is the upper limit for acceptance of the exposures for a dark sky stack. Each bin represents the percentage of the total number of pixels, where 100% is $8192^2 = 67108856$ pixels, which have fewer than X percent of exposures sampling sky, where X is ordinate at the right of the bin. In this example 41% of the pixels have fewer than 100% of sky pixels which means 59% of the pixels have no sources in any exposure.

Minimum number for deep sky stack: The last criteria is a rule based on the number of surviving exposures. There may be different rules for each class of dark sky calibration and the rules may depend on the filter. Currently there is a single, simple rule that there be a minimum of five (5) exposures. If there are any externally selected exposures then this rule is not applied.

15.4 Pupil Ghost and Fringe Pattern Subtraction

In this section we describe the removal of pupil ghost and fringe patterns from the science exposures. These are combined in one section because they require similar methods. The basic idea is that a pattern template is extracted from the dark sky image and the template is scaled in amplitude to each exposure and subtracted. The challenges are extracting the templates from the background and scaling these faint templates to the science exposures in the presence of astronomical sources and background.

The Mosaic cameras have optical surfaces which can reflect a small fraction of the light such that a large out-of-focus ghost pattern, called the pupil ghost pattern, reaches the detector. The amplitude of this ghost depends on the optics path and coatings of the camera and telescope and on the field illumination. The first dependency results in the pattern only being significant at Kitt Peak, only in some filters, and only affecting some CCDs.

As with many high performance, ”thinned”, astronomical CCDs, the Mosaic camera detectors exhibit fringing patterns at some wavelengths. This pattern is also filter dependent and occurs in data from both cameras and all CCDs.

A commonality of these effects is that they are detector and filter dependent. Whether these calibrations are required for a particular dataset is defined by a calibration parameter in the calibration library. If the calibration is not required the pipeline simply skips the operations.

When the calibration is required, the pipeline determines whether a suitable dark sky self-calibration image can be created from the dataset as described previously. If it can then the pattern template extraction steps described below are performed. But if a self-calibration is not possible the pattern removal may still be accomplished using templates previously stored in the calibration library from other datasets. The nearest in time with the highest quality is retrieved and the template scaling and subtraction methods described later in this section are carried out.

Extraction of a pattern template from a dark sky image consists of specifying the location of the pattern, removal of the background, and minimizing the effects of any residual, non-pattern features which may remain. The location of the pattern is provided by a mask. This is only necessary for the pupil ghost pattern which occurs near the center of the mosaic and affects only four of the CCDs.

If the pupil ghost is to be removed from the exposures for the dark sky calibration are combined using their associated object masks and scaled to a common mean sky. A template for the pupil pattern is extracted using a algorithm developed for this purpose and described in the IRAF documentation. Basically a data file defines the location of the pattern, a background is determined and subtracted based on inner and outer rings, Because the pattern is faint and, despite the object masks, there may be light from sources or cosmic rays it is important to exclude any large features. This is done by s

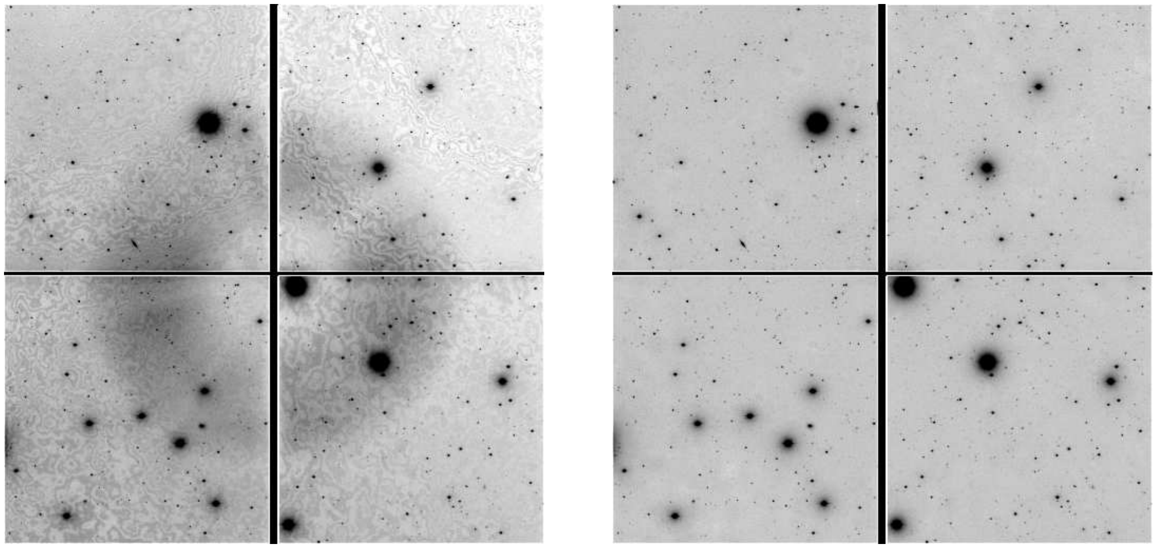

Figure 2: The two panels are negative contrast displays from the same central 4K x4K region (covering 4 or the 8 CCDs) of a KPNO Mosaic z'-band exposure. On the left is the result after dome calibration and on the right after dark sky self-calibration. The instrumental effects seen in the left panel are the pupil ghost pattern, fringing, and an illumination pattern. This is a challenging example because bright sources overlap the pupil ghost pattern. The bright lines on the edges of the CCDs are an anti-aliasing display artifact.

The pixels within the pupil pattern, as defined by a mask are, clipped (replaced by an upper or lower threshold value) where the thresholds are determined by an three iterations of sigma clipping method using a 3 sigma lower rejection and a 5 sigma upper rejection. three iterations. The clipping excludes pixels from the statistics in each iteration and then any pixels outside the final limits are set to

The outlier pixels are replace in the template by the threshold values when storing the template in the calibration library. This is because if this template is used by other data where the pupil pattern template cannot be determined there will be no mask for these pixels.

Once a pupil pattern is determined by self-calibration or, failing that, from the calibration library. Note that if there is no suitable library template then the pupil ghost removal is not performed and the pipeline data products will not be fully calibrated.

While the structure of the pupil pattern is relatively stable the amplitude may vary from exposure to exposure. This is because the amount of reflected light forming the ghost will depend on the amount of source, sky, and even light from outside the field of view. Therefore, the pupil pattern template must be scaled and subtracted from each exposure. The pipeline allows this to be done per CCD but normally a common scaling is determined from all the affected CCDs even though this results in a synchronization in the data parallel processing. This is compensated for by processing multiple datasets in parallel.

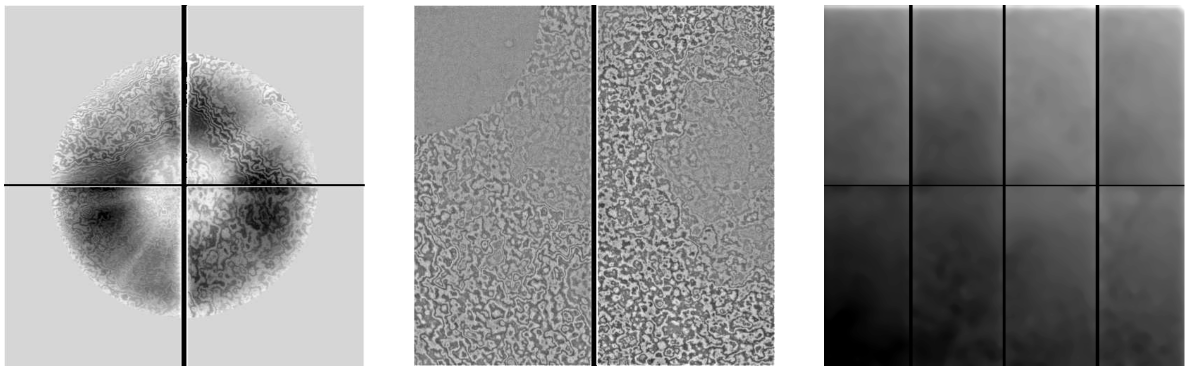

Figure 3: These panels show the three calibrations extracted, successively, from a dark sky image produced from KPNO Mosaic z'-band exposures taken on one night. The exposure shown in figure 2 is from this dataset and the calibrations shown here were used on that exposure. The left panel shows the pupil ghost template occupying the central region of the exposure. The middle panel shows the fringe template for two of the eight CCDs. The right panel shows the illumination flat field for the whole field. The left and middle panels cover one-fourth of the field (4K x 4K). Note the presence of fringes in the pupil template and the corresponding absense of fringes in the upper left corner of the middle panel where the pupil pattern was removed. This illustrates the difficulty of disentangling effects.

The fitting of the template to an exposure is hard because the pattern is faint and diffuse and the exposure may have sources anywhere in the pattern. The fitting is optimized by using the template mask to look only at pixels containing the pattern and using a weighting that is gives more weight to places where the template is stronger. The result of the fitting is a single number that is multipled into the pattern prior to subtracting from the exposure.

15.5 Dark Sky Flat

Dark sky flat fields provide the ultimate response calibration on all scales. At one extreme is the calibration of individual pixels to a uniform response to the sky background and at the other is correction for illumination patterns. At an intermediate scale there is normalizing the response of the amplifiers and CCDs to the sky.

The calibrations must avoid introducing more noise. This is hardest with the pixel response. The pipeline provides the possiblitly of applying only a large scale calibration when the data are not adequate for the individual pixel calibration.

Forced decision: There is a pipeline configuration parameter for requiring smoothing, requiring no smoothing, or make the decision based on data quality criteria.

Smoothing: If smoothing of the dark sky flat field is selected, either by a fixed requirement or by the result of the data quality analysis, another pipeline configuration parameter defines a median filtering window.

Once a dark sky flat field is identified (or not) it is applied to each exposure within the data set.

If a new dark sky flat field was produced and used it is saved in the calibration library, made available to the observers, and made public.

Smoothing verse no smoothing

16 Photometric Characterization

17 Astrometric Transformations

The images have thus far been flux and astrometrically calibrated with their original pixel sampling and an exposure has been kept as a set of separate images. The set of flux calibrated images, packaged as a multi-extension file, is one of the data products of the pipeline. However, the pixel sampling on the sky is complex and, for a number of reasons including producing a simple mosaicked image data product and combining overlapping exposures, it is desirable to transform the set of images to simple standard sampling.

The stacking of exposures could bundle the registration, resampling, and combining of data from a collection of single flux calibrated mosaic images into a single operation. However, for high performance the pipeline separates the process into multiple steps. Each image is first transformed independently allowing the operation maximally parallel and distributed. From the transformed pieces the two data products of a single mosaicked image per exposure and a stacked image from overlapping exposures can be created by shifting, matching flux scales, and combining without the complexities of image interpolation.

The key to allowing simple mosaicking and stacking after transformation is to chose a sampling grid on the sky that is common to all images which may be combined. Therefore, there are two algorithmic steps, defining the grids and resampling the pixels.

This step and all steps using the transformed image require successful flux and astrometric calibrations. When the calibration failed there will be no contribution to the resampled data products or contribution to stacks for the exposure. Some flux calibration failures may affect only some CCDs so the mosaicked image for the exposure will have sections missing. An astrometric calibration failure affects the entire exposure and so there will be no mosaicked data product.

17.1 Setting a Common Sampling

The tangent points of all exposures in a ”day” dataset are used to define groups of images which will share the same sampling grid. Note that the tangent points for all CCDs from a common exposure are the same so all one may think fo this step as applying to the ”ftr” dataset though the computations are done per CCD group.

At this time stacking of multiple exposures is limited to those which are in the same ”ftr” dataset. This mainly means dither sequences taken on the same night. When an ftr dataset spans multiple nights due to a small number of exposures in a night then dither sequences and stacking of overlapping exposures may also span multiple nights.

The first step is grouping the dataset by overlaps. This uses a nearest neighbor clustering algorithm. The coordinates are first sorted by declination and gap between consecutive declination values greater than 15 arc minutes delimit declination groups. Each coordinates in each declination group are then sorted by right ascension and gaps in angular separation greater than 15 arc seconds delimit final overlap groups. Special steps are taken to handle the wrap-around discontinuity in right ascension.

The second step is to assign each overlap group to the same tangent point. This is done using the coordinates of the first image in a group and finding the nearest point on a grid of the celestial sphere. The grid has points with approximately one degree separations in right ascension and declination. The precise details of the grid are not critical as long as it is well defined.

In addition to a tangent point the sampling requires an orientation and scale. The most common ”standard” orientation for astronomical images is declination increase with the line index and have right ascension decrease with column index with mimimal cross-terms. Visually, if the image is displayed in raster order with the origin at the lower-left of the display then north is up and east is to the left. The pixel interval is set to a nice number for the data. For the Mosaic Camera this is 0.25 arc seconds per pixel. The final piece of the sampling is the projection from a rectangular grid to the sky. For optical astronomy the most common projection is the FITS ”tan” projection [REF].

The idea behind the use of a grid of tangent points is that for data from different epochs resampled images or stacks which are nearby have a significant chance of being sampled with the same tangent point and, thus, can be stacked without additional interpolation. This is not perfect because at the border between grid points nearby images may be split between two grid points. This drives the grid points to be widely spaced. On the other hand the further a tangent point is from the center of a field the more the image departs from having columns aligned with north.

17.2 Resampling Images

The transformation of an image from one pixel sampling to another requires rebinning or interpolation of the data. There are various

The pipeline defines the flux of each pixel in the output grid by interpolating from the original image raster. The WCS is used to compute the point in the input image corresponding to the desired output pixel center. The nput pixels around that point are used to define an interpolation function which is then evaluated at that requested point. There are many interpolation functions that can be used including sinc, polynomial, and ”drizzling”. The pipeline has an operator defineable configuration parameter for changing this but the recommended inerpolation is a truncated sinc interpolation with a tabulated kernel (for computation efficiency) of 17 pixels.

With any interpolation and complex mappings, such as from optical distortions, there is the problem of interpolation at the edges of the image data. The pipeline uses a reflection boundary condition which effectively eliminates ringing at the edges. However, the setting of the output image size is based on a trimmed version of the input. The trim is eight pixels which is appropriate for the sync interpolation. The effect of this is that no output pixel will map to an input pixel closer than eight pixels from the edge, thus avoiding needing to apply boundary extension to the input data. While this results in a slightly smaller output image it eliminates potential artificial edge effects. For later stacking of dithered exposures the trimming is like having slightly larger gaps in the individual mosaic exposures.

Sinc interpolation is one of the more compute intensive functions but it has the property of minimizing correlations in the initially uncorrelated noise. A disadvanatage of this interpolation is the ringing from undersampled features, namely cosmic rays and detector defects. The algorithm does not have the option to exclude bad pixels from the interpolation function which is an important computational reason why pixels replacement is important beyond just cosmetic appeal.

The final transformed image is stored in an image raster that is just big enough to include all the pixels with science data. The point is that one does not need to pad the images to some large common ”canvas”. Defining the final image when mosaicking and stacking is left to later with the knowledge that only integer shifts along columns and lines is needed to align data which have the same resampling grid.

The exposure mask is also transformed. The mask values are not propagated and the assumption is that any pixel flagged in the exposure mask should be flagged in the transformed mask. The input mask is converted to a boolean mask. Because mask values are integers the non-zero value is set to 10,000 for a reason that will become apparent, The mask is resampled using the same transformation but using bi-linear interpolation. The use of linear interpolation is to minimize ringing effects and so that the interpolated value at an output pixel can be interpreted as the degree of contribution from a bad pixel. The output mask values will be integers between zero and 10,000. So any interpolated value whose floating point value is less than one will become zero which means, in a sense, that the contribution of a bad pixel to the output is less than 0.01%. After the reampling the non-zero mask values are set to one to make a final boolean mask.

The resampling transformation is performed by the IRAF task mscimage from the MSCRED package. Further details may be found in the documentation for this task.

18 Mosaicking

A mosaic image is a simple raster image where all pieces of an exposure are put together with a common world coordinate system. Pixels where there is no data because of gaps between the CCDs and because of distortions, rotations, alignments, and celestial projects are handled by setting the values to a defined blank value and recording them in a pixel mask.

Starting with the individual resampled CCDs which have been transformed to a common coordinate grid is a simple matter. First no adjustments are needed for the pixel values because the flux calibration is assumed to have brought all pixels to the same scale and effective exposure. Second since there is no overlap in the original CCD geometry there is no need to worry about combining overlapping pixels. So to make a mosaic simply requires determining the integer shifts for the origins of each resampled CCD image, determining an output size that just fits all the data, and populating the output pixels. At the same time a mask for the output mosaic is created and marked where there are no pixels. The mask also is filled with the bad pixels identified for the resampled image from the original flux calibration.

The task with does this is imcombine.

19 Overlap Stacking

The pipeline combines overlapping exposures when it identifies them. They must be within a signle ”ftr” dataset since that is what is available at one time within the pipeline data flow. Currently, the identification of overlapping exposures is limited to dither sequences. These are sets of exposures taken specifically by the observer, using a special observing command, for filling in the mosaic gaps and detector defects.

There are three possible approaches to selecting data to combine. One is to stack the mosaic exposures described in §18. Another is to stack all the individual resampled CCDs at one time. The last, and the one implemented by the pipeline, is to first stack all resampled images from a single CCD and then stack those into the final product.

he pipeline uses the last approach because it fits well with the way the pipeline parallelizes and distributes data in our high performance system. It is the case that the resampling steps (§17) are distributed by CCD in ”day” datasets. In that stage overlapping data, which includes dither sequences, are identified and the resampling done to a common tangent point. It is a simple matter to follow this with a step that takes the dither sets from each ”day” dataset, makes subset ”stk” dither datasets, and stacks them. If desired, the ”stk” datasets may be operated on in parallel and distributed.

The second step of combing the ”stk” datasets is done in the same manner as for the mosaics of individual exposures except now the stacked images for each CCD will overlap. Combining the CCD stacked images for a final field stack is equivalent to either stacking the mosaicked images for an exposure or stacking all the individual resampled CCD images provided exposure masks are maintained and used.

19.1 Missing vs Bad Data

A goal of dither stacks, besides the obvious one of deeper imaging, is to fill in detector defects and missing data, such as from the gaps in mosaic instruments. This unobserved or bad data is fixed in the focal plane and so dithering on the sky fills them in during stacking.

A related reason for dither stacks, sometimes not really dithers but split exposures, is to handle pixels compromised by cosmic rays or satellite trails.

Dither stacking depends on registering exposures on the sky. In many cases, particularly with wide-field cameras and mosaics, this requires resampling of each exposure to a common pixel grid. This inevitably introduces another type of missing data at the edges of the images due to the storing of data as rectangular image rasters. As with the gaps, this type of missing data is also tied to the focal plane and filled in by dithering except at the extremes of the dither field.

However, another kind of bad data follows the sky. This is usually related to bright sources causing saturation and, for CCDs, ”bleed trails”. Dither stacks generally cannot fill this in except through a strategy of variable exposure times, which is infrequently done though the pipeline handles this case.

The way these absent, bad

The Mosaic pipeline includes algorithms using static detector defect masks, a WCS describing the geometry of the mosaic (which implicitly define the gaps), resampling to an undistorted pixel grid, and identify and flag saturated and bleed trail pixels. For cosmetic and software reasons pixels suffering from detector defects and bright source effects are replaced by ”reasonable” values using interpolation. These substituted values are identified as non-science data in associated masks.

A fuzzy situation is when data is of marginal quality. An example of this is with IR detectors where non-linearity calibrations get gradually worse (meaning progressively less accurate). One wants to avoid setting a hard division between good and bad. In this case the data is not replaced by artificial data but is still flagged. Like saturation these pixels will be fixed to the sky near bright sources.

This discussion leads to a categorization of bad pixels in stacked images, defined as where there is no contributions from good (calibrated) pixels, into those where the sky was not observed and those where the observations are of poor quality or compromised. For simplicity we call the two types ”no data” and ”bad” pixels.

The pipeline produces bad pixel masks for the stacked images which discriminate these two types. It also allows the stacked images to also have cosmetically cleaner pixels, avoiding sharp artifacts as well as producing a stack of original pixels when they are of ”poor quality” but not needing to be replaced.

There are three possiblities for handling the bad stacked pixels, setting them to a ”blank” value, interpolating across the final stack as is done with the original exposures, or combining the input replaced pixels as if they were good data. The third approach has two advantages, it avoids another pipeline step and it randomizes the pixels somewhat to reduce the ”zipper” pattern of simple linear interpolation. It also results in possibly interesting pixel values for the input bad pixels which were of ”poor quality” but not needing to be replaced.

Figures X, Y, and Z show the distinction between treating the two categories the same, producing output blank values, and differently, producing blank values when there is no data and combined input pixel values for the bad data. The pipeline stacks data using the latter approach.

The main application used in the stacking pipeline is IRAF’s IMCOMBINE.

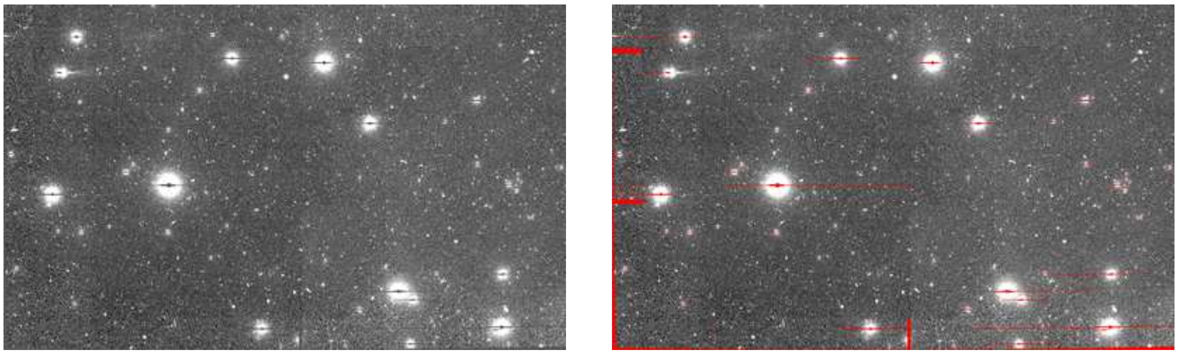

Figure 4: Large region of a Mosaic dither stack using masktype="goodvalue" showing how bad pixels associated with the bright stars result in no overlap data: a) output image has blank values of zero, b) output mask has values of 1 overlayed as red. The "ticks" at the left and bottom are incomplete filling of the gaps by the dither.

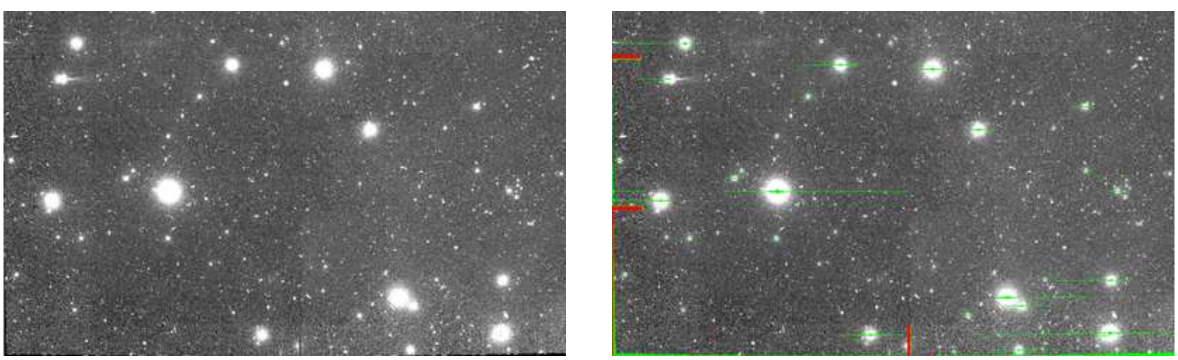

Figure 5: The same stack as the previous figure but using masktype="novalue": a) output image with pixels filled in by combined interpolated values, b) overlaid output mask with values of 1 (red) for no data and 2 (green) for interpolated values.

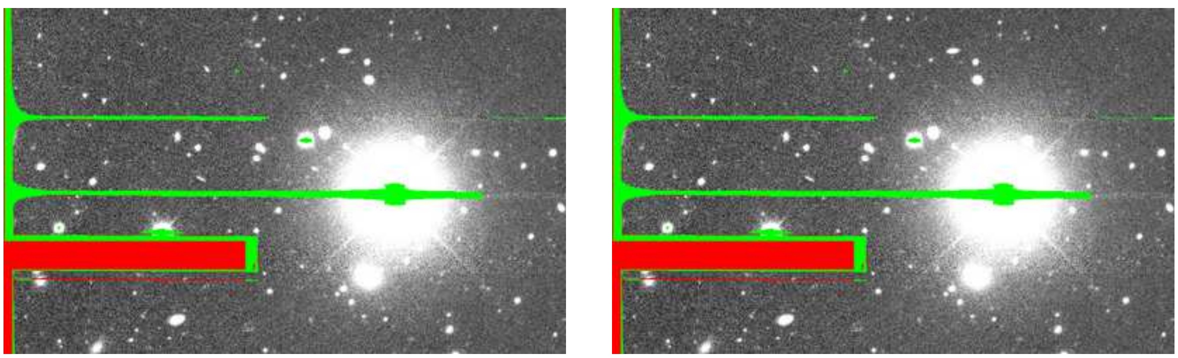

Figure 6: A blow-up of previous figure showing edge overlap effects and the remnant of the bleed trail when one input exposure did not have the bright star and, hence, no trail. The exposure mask identifies this region of low significance.

20 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.

21 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.

22 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 astrometric 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

[1] Cline, T. et al. 2007, to appear in ASP Conf. Ser. ADASS XVI, [P4.19]

[2] Monet, D. et al. 2003, AJ, 125, 984

[3] Smith, C. et al. 2007, to appear in ASP Conf. Ser. ADASS XVI, [O6b.1]

[4] Swaters, R. & Valdes, F. 2007, to appear in ASP Conf. Ser. ADASS XVI, [P4.20]
[Reference the ADA paper]