BANZAI Documentation

Release 0.23.0

Curtis McCully, Monica Turner, David Collom, Matt Daily, Las Cum

May 20, 2019

CONTENTS

This pipeline is designed to take raw data taken from Las Cumbres Observatory, and produce science quality data products. BANZAI is capable of reducing single or multi-extension fits files. For historical data, BANZAI can also reduce the data cubes that were produced by the Sinistro cameras.

STAGES

BANZAI is comprised of different stages, each corresponding to a single reduction step, which are chained together to process images. The individual stages are described below, in the order that they are executed.

1.1 Overscan Subtraction

We currently use the header the keyword `BIASSEC` to identify the overscan region. If this keyword is set to "Unknown", then we simply skip subtracting the overscan. We estimate a single overscan value for the whole image (rather than row by row). This is saved in the header under the keyword `BIASLVL`. If there is no overscan section, this value is derived from the average value from the bias frames.

1.2 Crosstalk

Currently, only 1-meter imagesr are read out using multiple amplifiers. These frames do have significant crosstalk between amplifiers, which is removed using linear coefficients that relate each quadrant to every other quadrant. These values are read from the `CRSTLK{i}{j}` keywords in the header. These coefficients are hard coded for historical data.

1.3 Gain

All pixels in the frame are multiplied by the gain, using the `GAIN` header keyword. Thus, the science frames output by BANZAI are all in units of electrons.

1.4 Mosaic

If the file it multi-extension BANZAI produces a single mosaiced frame. This relies on the `DETSEC` header keywords to mosaic the extensions together.

In some historical Sinistro data, there is a missing row in the center. This is because the top amplifiers (3 and 4) have more light sensitive pixels than the bottom amplifiers (1 and 2). While all of the parallel overscan rows are discarded, one edge row is discarded for the amplifiers that have an extra light sensitive row.

1.5 Trim

After being mosaiced, the data is trimmed to a useable region for science. This is set based on the `TRIMSEC` header keyword.

1.6 Bias Subtraction

Full frame bias images are subtracted from each of the darks, flat field images, and science frames. The master bias frame that was taken closest in time to the current data will be used. This will add a few counts of noise to each pixel, but it solves two problems. First, if there is systematic structure in the read out, this will be removed. Second, this will remove the bias values for images that do not have an overscan region.

If no bias frame exists for this instrument, the data will not be reduced and an exception will be raised.

1.7 Dark Subtraction

Full-frame master dark frames, scaled to the exposure time of the frame, are subtracted from all flat-field and science images. The most recent master dark frame is used. Often this is taken on the same day. If no dark frame exists for this instrument, the data will not be reduced and an exception will be raised.

1.8 Flat Field Correction

Master flat field images (normalized to unity using the inner quarter of the image) are divided out of every science frame. The most recent master flat-field image for the given telescope, filter, and binning is used. If no flat field exists, the data will not be reduced and an exception will be raised.

1.9 Source Detection

Source detection uses the "Source Extraction in Python" (SEP; https://github.com/kbarbary/sep). This is similar to Source Extractor, but is written purely in Python and Cython. This allows more customization.

We estimate the background by taking a 3x3 median filter of the image and the doing a 32x32 block average of the image.

We use the default match filter for source detection that is provided by SEP.

We do aperture photometry using an elliptical aperture that is set by 2.5 times the Kron radius. This produces approximately the same results as `FLUX_AUTO` from SExtractor.

We set the source detection limit at 3 times the global rms of the image. `MINAREA` is set to 5, the default. This should minimize false detections, but may miss the faintest sources.

1.10 Astrometry

The WCS is found by using Astrometry.net (Lang et al. 2012, ascl:1208.001, http://astrometry.net). We use the catalog from the source detection (the previous step) as input.

We have built our astrometry.net index files from the recent GAIA DR2 catalog. This is an exceptionally clean catalog that reduces false positives and improves the likelihood of finding a solve for a given frame.

We adopt a code tolerance of 0.003 (a factor of 3 smaller than the default), but increase the centroid uncertainty to be 20 pixels. The large centroid uncertainty allows the algorithm to find quads even if the initial guess is quite poor and even if there is significant distortion. However, decreasing the code tolerance forces the algorithm to only use high quality quads, making the solution more robust. We also go deeper into the catalogs (200 quads deep) to increase the chances of a successful astrometry solution.

Currently no non-linear distortion is included in the WCS (the current WCS solution only has a center, a pixel scale, and a rotation). At worst (in the image corners), the offset between coordinates with non-linear distortion terms included and those without are ~5 arcseconds.

MASTER CALIBRATION FRAMES

BANZAI also contains routines to create the master bias, dark and flat frames required for the reduction of science frames. Before we describe how these are created, we introduce an important statistical metric used throughout the BANZAI pipeline.

Sigma clipping is a standard technique to reject bad pixels. However, outliers artificially increase the standard deviation (std) of the points (which makes outliers appear to be fewer sigma away from the peak) making it difficult to produce robust results. We have adopted a slightly different method. We use the median absolute deviation (mad) to estimate the scatter of the distribution (traditionally characterized by the standard deviation). The mad is related to the std by

$\sigma \approx 1.4826 \times \text{mad}$

We have termed this the "robust standard deviation" (rstd). Using the robust standard deviation, we mask pixels reliably and take a mean of the remaining pixels as usual.

2.1 Master Bias Creation

For all instruments, we take many full-frame bias exposures every afternoon and morning. The afternoon and morning sets of bias frames are typically reduced together for quality control and to increase statistics.

When creating a master bias frame for the night, we first calculate the sigma clipped mean of each image. In this case, outliers that are 3.5 rstd from the median are rejected before calculating the mean. As the read noise is approximately Gaussian (to surprisingly high precision), the median is a robust estimation of the center of the pixel brightness distribution.

We then stack the individual bias frames. On a pixel by pixel basis, we reject 3 rstd outliers, and then take the mean. This should remove and structure produced by the readout process. The noise in each pixel should scale as sqrt(number of bias images). We take ~64 frames per night reducing the noise per pixel to read noise (RN) / 8. Thus, only a few counts of noise are being added to the frames in quadrature. This is much less than the ~10 electron read noise, meaning that this does not increase the noise in the science frames in any significant way.

2.2 Master Dark Creation

For all instruments, we take full-frame dark exposures every afternoon and morning. Like the bias frames, the afternoon and morning dark frames are combined together to increase statistics. Typically, a total of 10x900s images are taken.

When creating a master dark frame, each individual frame is scaled by the exposure time (read from the header). The sigma clipped mean of the scaled frames is then calculated on a pixel by pixel basis. We reject any 3 rstd outliers, similar to the master bias creation.

Our cameras have dark currents of 0.1-0.2 electrons / s per pixel. For 10x900s this corresponds to 1 - 3 electrons of additional noise per pixel added in quadrature (given the same length science frame, and not including the Poisson noise from the dark current itself). Again, this is much smaller than the read noise so it will not affect the noise properties of the final science frames.

2.3 Master Flat Field Creation

Twilight flats are taken every day. However, flat-field images for every filter are not taken daily, because twilight is not long enough to take all of them in a single night. Instead the choice of filter is rotated, based on the necessary exposure time to get a high signal to noise image and the popularity of the filter for science programs. Typically, a master flat field is produced about once every 3 days for any given filter. When a flat-field image is taken for a given filter is taken in the evening twilight, it is also taken in morning twilight for quality control. Typically, 5 flat field frames are taken in the evening and 5 taken in the morning per filter. The frames are dithered so that we can remove stars in the combined master flat field.

Each individual flat-field image is normalized to unity before combining them. The normalization is calculated finding the robust sigma clipped mean (3.5 rstd outliers are rejected) of the central region of the image. For the central region, we choose the central 25% of the field (the region has dimensions that are half of the full image).

The flat-field frames are then stacked using a sigma clipped mean, similar to the master bias and dark frames. We again choose to reject 3 rstd outliers.

QUALITY CONTROL

3.1 Header Sanity

The header sanity test first checks if any of the following principal FITS header keywords are either missing or set to `'N/A'`: `RA`, `DEC`, `CAT-RA`, `CAT-DEC`, `OFST-RA`, `OFST-DEC`, `TPT-RA`, `TPT-DEC`, `PM-RA`, `PM-DEC`, `CRVAL1`, `CRVAL2`, `CRPIX1`, `CRPIX2`, and `EXPTIME`.

This routine then verifies that the RA value (`CRVAL1`) is between 0 and 360 and that the declination value (`CRVAL2`) is between -90 and 90.

Finally, the header checker ensures that exposure time value (`EXPTIME`) is greater than 0. Note that this final test is not performed on bias frames, which can sometimes have negative exposure time values.

3.2 Thousands Test

There is a known issue with the Sinistro cameras where a large fraction of pixels report values of exactly 1000. This test measures the fraction of 1000-valued pixels in each Sininstro frame, and if this fraction is above 20%, the frame is rejected.

3.3 Saturation Test

A pixel is considered saturated if its values is greater than the `SATURATE` header kewyword. This test measures the fraction of saturated pixels in each Sininstro frame, and if this fraction is above 5%, the frame is rejected.

3.4 Pattern Noise Detector

Occasionally, if a camera is failing, it may exhibit highly structured electrical pattern noise. Although this is not a common occurrence, it is still desirable to detect the issue as soon as possible.

This algorithm computes a power array by taking the fourier transform of the full image, then taking the median of the absolute values along the vertical axis. Next, the SNR is computed as:

SNR = [power - median(power)] / MAD(power)

The method than searches for groups of 3 or more adjacent pixels that have an SNR above 15. If more than 1% of all pixels are in these groups, then the frame is considered to have pattern noise.

3.5 Pointing Test

This test computes the offset between the requested RA and declination from the header (given by either `OFST-RA` and `OFST-DEC`, or `CAT-RA` and `CAT-DEC`) with the actual RA and declination of the observation (`CRVAL1` and `CRVAL2`). The test is considered failed if the offset is above 300", and a warning is provided if it is above 30".

3.6 Master Calibration Comparison

When a calibration frame is processed by BANZAI, it can be compared to the temporally nearest previous master to check for significant variations, which can serve as an alert for e.g. major issues with the camera. Since this check also discards frames found to deviate significantly, it prevents the creation of bad master frames that can cause problems as they are propagated through the pipeline and used for the reduction of science data.

The algorithm works as follows. After some preprocessing that depends on the calibration type, the SNR at each pixel is computed as:

SNR = (individual_frame - master_frame) / noise

where the noise also depends on the type of calibration. The individual frame fails the comparison test if more than 5% of pixels have an SNR greater than 6.

The individual frame preprocessing steps and noise parameters for the different calibration types are listed below:

- bias:

    - preprocessing: bias level subtraction

    - noise = RN (read noise, from header keyword `RDNOISE`)

- dark:

    - preprocessing: bias subtraction, normalization by exposure time

    - noise = sqrt(RN^2 + PN^2) / exptime, where PN is the poisson noise, computed using the square root of the image counts prior to normalization

- skyflat:

    - preprocessing: bias and dark subtraction, normalization by the sigma clipped mean of image

    - noise = sqrt(RN^2 + PN^2) / normalization