Astronomical Data Analysis Software and Systems XXIII
ASP Conference Series, Vol. 485
N. Manset and P. Forshay, eds.

The DECam Community Pipeline

Francisco Valdes,[1] Robert Gruendl,[2] and the DES Project[3]

[1]*National Optical Astronomy Observatories, P.O. Box 26732, Tucson, AZ 85719, USA*

[2]*University of Illinois at Urbana-Champaign, Dept. of Astronomy, MC-221, 1002 W. Green Street, Urbana, IL 61801*

[3]*Fermilab, P.O. Box 500, Batavia, IL 60510*

Abstract. The Dark Energy Survey Project and NOAO have produced a world class, large format camera for the CTIO Blanco Telescope. This camera, called the Dark Energy Camera or DECam, with 62 CCDs and a ~20 second readout is capable of producing a very large amount of data. NOAO and the DES Data Management (DESDM) team have worked together to develop a Community Pipeline (CP) to produce calibrated data products for community users (the DES has its own dedicated pipeline with the same or similar components). First light and commissioning occurred in the fall of 2012 and the camera and pipeline have had nearly a year of operation and evolution. The CP has recently reach a production level with good data quality products. This paper provides an overview of this DECam Community Pipeline.

Introduction

The DECam Community Pipeline is a high performance, dedicated cluster, software system for producing basic calibrated data products for both principle investigator and archival researchers. It was developed over many years and by many people in the DES Project and at NOAO. The collaboration between the DES and NOAO is both synergistic and a condition of undertaking the 525 night survey on the NOAO CTIO Blanco telescope. This contribution provides a brief description of the CP. It attempts to touch on as many aspects of the CP as possible within the limit of these Proceedings by omitting most details.

Architecture and Software

The pipeline has the common astronomical pipeline architecture of an orchestration component and plug-in calibration modules. The orchestration component is a little unusual in that it makes use of two orchestration systems (not counting DAGMAN). One is the NOAO High Performance Pipeline System (NHPPS) and the other is the DESDM-CP pipeline system (DESCP). In the CP the DESCP is run by the NHPPS as a type of module.

The NHPPS is used as the higher level orchestrator for several reasons. One is to provide the same environment for NOAO developers, operators, and infrastructure as the other NOAO instrument pipelines (Mosaic and NEWFIRM). Another reason is that it more flexible and efficient in its use of NOAO computer resources.

The DESCP orchestration system was developed by DESDM using NCSA elements. It makes use of the ELF application container with the Ogrescript language, and the HPCondor/DAGMAN job scheduling and execution system. The application structure in this orchestration is a series of *blocks* which are run sequentially. The blocks include various housekeeping stages around the HPCondor/DAGMAN stage that provides the data parallel job execution on the NOAO DECam compute cluster. A block typically is performing one type of calibration across all exposures in an input dataset. (Parenthetically, a redesigned DESCP orchestration component is under development.)

A core aspect of the DESCP is use of a database to track the location of all files and to store exposure and job metadata. This aspect means that data to be processed in a block must be identified through the database making the processing *database driven* in concept. At NOAO a Postgres database running on a dedicated database machine is used.

Another core aspect of the DESCP is a dependence on a shared file system. At NOAO this is a GPFS implementation using two dedicated machines to serve the 15 node/180 core compute cluster.

The plug-in calibration components are drawn from a large suite of DESDM developed programs (primarily in C and C++), the Astromatic programs by E. Bertin (primarily in C), and IRAF (CL and SPP). Perl, IRAF, and SH/TCSH are used in various places for wrapper and glue tasks.

Calibrations

The CP applies a variety of calibrations to transform the raw exposures from the data acquisition system to archived data products. A list of the principle calibrations is given in table 1. Many of the calibrations are fairly standard for CCD cameras. In this section we touch on some of the more challenging ones for DECam. As noted in the introduction, the description is necessarily quite brief. There are also some even more subtle characteristics (e.g., *tree rings*, correctible columns) which are being studied.

Table 1. Principle calibrations performed by the DECam CP

• overscan	• photometric zero points
• cross-talk	• remapping
• non-linearity	• stacking with outlier exposure rejection
• bias	• masking / cosmetic interpolation
• dome flat	- non-linear pixels; i.e. bad
• star flat (pupil/illum)	- saturation
• fringe	- bleed effects; trails, serial edge
• background subtraction	- cosmic rays
• astrometric WCS	

Cross-talk occurs between the two amplifiers of a CCD as well as some interactions between CCDs. The cross-talk has a number of challenges including a non-linear behavior. The cross-talk coefficients have been determined, added to the calibration library, and applied at an early stage of the calibrations. The DECam CCDs also exhibit variable amounts of non-linearity at low and high light levels. The behavior of each CCD has been quantified, recorded in a calibration library, and a correction is also applied at an early stage.

An out-of-focus, additive light pattern from optical reflections needs to be corrected in dome flat and sky exposures. This is a challenge because of the need to discriminate CCD gain variations from this additive light. Dealing with this pattern is particularly critical in dome flats to avoid introducing photometric errors. A method has been developed based on photometric mapping, i.e., using star flats, which provides gain calibration using dome flat fields and the star flat information. Implicit in this method is an illumination gain calibration for differences between dome flat and on sky responses. This method does not remove the reflected light from the gain calibrated science exposures. A simple attempt is made to subtract the pattern as part of the background subtraction but residual light is evident in the final data. However, since the effect on the gains has been accounted for, photometry with local sky subtraction will be correct.

Bright stars produce saturation and *bleed trails* which are detected, masked, and interpolated away. This is not unusual for CCDs. What is unusual for DECam is bleeding out from the serial register edge for extremely bright sources. The detection algorithm had to be designed to detect this type of edge effect.

Astrometric calibration and photometric characterization are performed by matching sources to external catalogs. The world coordinate system (WCS) mapping is currently based on the 2MASS catalog and the photometric zero point on the USNO-B1 catalog. As with other mosaic cameras, the astrometric solution must be a careful balance between local distortions and global continuity.

Sky subtraction, including a pupil pattern model, is applied prior to remapping and stacking/coadding. This calibration is challenging in the community pipeline context because of the variety of fields observed by different programs. For fields with only stars and small galaxies, better matching across CCDs, and hence better stacks, is achieved by fitting a low order background surface to each CCD. However in fields with large galaxies (arc minutes or more) or large nebulosity, individual CCD fitting produces gross effects. In that case only a global sky gradient can be subtracted. Determining which to use in an automatic pipeline is beyond the ability of the CP. Instead data products using both methods are applied leaving the best choice to the investigator. It is not reasonable to archive every individual exposure in two versions but two versions of coadded stacks is possible.

Another common challenge in stacking is which exposures to coadd. The CP produces stacks in which outlier exposures in seeing, zero point, and sky brightness are excluded. These coadds are provided for possible use by PI programs as well as for archival investigators. However, PIs may want to include or exclude exposures differently from the automated CP. To accommodate this, individual resampled exposures, with just a global gradient removed, are provided. Investigators may then apply their own treatment of the background and choice of exposures when coadding exposures.

The thick DECam CCDs are strongly subject to cosmic ray events. There are two levels of cosmic ray masking. The first is a single exposure algorithm based on sharpness. When there are multiple exposures the stacking processing includes detection of sources which differ significantly from the median of the exposures.

Data Products

The data products of the CP are categorized by processing type and product type. Processing types are different classes or stages of calibration while product types are various associated data from a particular class. The processing types are master calibrations, instrumentally calibrated exposures, distortion corrected exposures, and field stacks/coadds. The product types are flux images, data quality maps, weight maps, and exposure/coverage maps.

Master calibrations which are produced by the CP (as opposed to externally provided calibrations) are bias and dome flat stacks for a night. These are multi-extensions FITS (MEF) files where each extension is data from a single CCD. The product types are the flux data and weight maps. There are currently no data quality maps.

Instrumentally calibrated exposures are MEF files consisting of CCD extensions where the pixel data have not been interpolated. Note that the CP has chosen to eliminate one CCD which is of problematic utility. The CCD images are trimmed of bad edge pixels and the header metadata attempt to capture the scientifically useful information from the telescope and processing. The associated data products, each in its own MEF file with matching extensions, are the basic flux images, data quality maps, and weight maps.

Resampled/remapped exposures are MEF files where each extension is an image of a single CCD that has been resampled/remapped/interpolated/warped (take your pick of terms) to a standard orientation and pixel scale at a standard tangent point. The point of this is to make different exposures congruent in pixel sampling. While this data product doesn't provide new information relative to the instrumentally calibrated version, it is provided to allow investigators to apply their own sky and/or gain adjustments and make coadds with a different subset of exposures without needing to remap. If the tangent points are the same then images only need to be registered by an integer offset before combining. The associated data products are the flux images, data quality maps, and weight maps where all are MEF files.

Stacked/coadded fields are MEF files where each extension is a piece of the field. The full overlapping stacked field is divided into abutting *tiles* whose size is less than 16K on a size. The investigator can work with individual tiles or recreate a single large image by pasting the tiles together. The associated data products are the flux files, data quality maps, weight maps, and exposure or coverage maps.

There are two versions of the stacked flux images provided to investigators; one with only a planar background removed and one with a low order spline fit to a grid of mode values in each CCD.

The data products described above are archived in the NOAO Science Archive (NSA) using FITS compression. Investigators may then either use software that directly understands the compressed format or use standard tools for uncompressing to more classic FITS files. The image and weight data products use *lossy* tile compression which provides roughly a six-fold compression that is important for storage and data transport. This type of astronomically specialized compression has been carefully studied and has become a de facto standard. Without compression the NSA would become prohibitively expensive for NOAO and the ability to provide individual remapped versions and two versions of stacks to help observers easily optimize stacks would not be possible.