Astronomical Data Analysis Software and Systems VII
ASP Conference Series, Vol. 145, 1998
R. Albrecht, R. N. Hook and H. A. Bushouse, eds.

The STScI NICMOS Calibration Pipeline

Howard A. Bushouse

Space Telescope Science Institute, Baltimore, MD 21218, Email: bushouse@stsci.edu

Elizabeth Stobie

University of Arizona, Steward Observatory, Tucson, AZ 85721

Abstract. The NICMOS data reduction and calibration pipeline employs file formats and software architecture techniques that are new and different from what has been used for previous HST instruments. This paper describes the FITS file format used for NICMOS data, which includes error estimate and data quality arrays for each science image, and describes the approach used to associate multiple observations of a single target. The software architecture, which employs ANSI C language algorithms and C bindings to IRAF libraries, is also described.

1. File Formats

The NICMOS archival and run-time file format is FITS, with extensive use of IMAGE and BINTABLE extensions. Each image from NICMOS is represented by five data arrays that are stored as five IMAGE extensions in a single FITS file. The five data arrays are: 1) the science (SCI) image from the instrument; 2) an error (ERR) image giving estimated uncertainties for the SCI image values; 3) a data quality (DQ) array containing bit-encoded flag values that represent known problem conditions with the SCI image values; 4) an array containing the number of samples (SAMP) used to derive the SCI values; and 5) an exposure time (TIME) array giving the total effective exposure time for each pixel in the SCI image.

This group of five data arrays is called an image set, or "imset" for short. NICMOS operation modes that produce multiple readouts of the detector during a single exposure have multiple imsets in a single FITS file; one imset per readout. Multiple observations of a single target are stored in separate FITS files, but the data are logically associated through the use of an "association table", which is a FITS file containing a single BINTABLE extension, listing the names and other relevant information for each observation in the association. One-dimensional spectra that are extracted from NICMOS grism images are also stored in a BINTABLE extension of a FITS file. There is never any image data in the primary header-data unit (HDU) of NICMOS FITS files; all data are stored in IMAGE or BINTABLE extensions. The primary HDUs only contain header keywords with information that is relevant to all extensions in the file.

2. Software Design

All of the NICMOS pipeline calibration software has been written in the ANSI C language. A set of data I/O routines, known as "hstio", has also been written which is used to easily perform the complex mapping between the disk file formats for imsets and their representations as NICMOS-specific C data structures within the pipeline code. The hstio library, in turn, uses a set of C bindings to IRAF image I/O libraries to perform the actual file access and data I/O. C bindings to other IRAF, STSDAS, and TABLES package libraries also provide access to existing mathematical and analysis algorithms, as well as FITS BINTABLE data I/O. The algorithms that apply the actual calibration steps to the science data are completely isolated from and independent of the data I/O routines. The calibration tasks are built in such a way that they can be run from either the host operating system level or as an IRAF native task from within the IRAF cl.

3. Image Associations

Multiple NICMOS observations of a single target field are often obtained for one of several reasons. Multiple exposures at a single sky position can be obtained in order to allow for rejection of cosmic ray hits when the images are later combined by the calibration pipeline. Small-angle dithering of a compact target within the instrument's field of view is useful for removing the effects of defective pixels and for averaging out residual flat fielding uncertainties. For observations at longer wavelengths, where the thermal background signal from the telescope and instrument begins to be significant, it is necessary to alternately "chop" the telescope field of view between the target and nearby blank sky positions in order to obtain measurements of the background signal. For very extended sources an image of the entire target can be created by "mosaicing" individual observations obtained at many sky positions.

These sets of observations are known as image "associations". The associations are a *logical* grouping only, in that the individual observations that make up an association are initially stored and processed separately, but are eventually combined by the pipeline software into a final product. The individual observations are first processed, one at a time, by the `calnica` task, which performs instrumental calibration, and are then processed as a group by the `calnicb` task, which combines the images, measures and removes background signal, and produces final mosaic images. Information about the association is stored in an association table, which is a FITS file with a single BINTABLE extension. The table extension lists the names of the observation datasets contained in the association and their role in the association (target or background image).

4. CALNICA Pipeline

The NICMOS pipeline calibration is divided into two tasks. The first task, `calnica`, applies instrumental calibration to individual exposures, including associated and non-associated observations. The task is completely data driven in that all information necessary to accomplish the calibration is read from image header keywords. These keywords contain instrumental configuration parameters, such as camera number, filter name, and exposure time, as well as "switches" that indicate which calibration steps are to be performed, and the names of reference files (e.g., flat fields, dark current images) that are needed for the calibration.

The major steps performed by `calnica` include the masking (setting DQ image values) of defective pixels in the images, the computation of uncertainties (setting ERR image values) for the image data, which includes both detector readnoise and Poisson noise in the detected signal, dark current subtraction, correction for non-linear detector response, flat fielding, and conversion from detected counts to count rates. Each calibration step that involves the application of reference data (e.g., dark subtraction, flat fielding) propagates estimated uncertainties in the reference data into the ERR images of the data being processed.

Another major step performed by `calnica` is the combination of images generated by the MultiAccum operational mode of the NICMOS detectors, in which multiple non-destructive detector readouts are performed during the course of a single exposure. The images generated by each readout are first individually calibrated and are then combined into a single, final output image. The image combination is accomplished by fitting a straight line to the (cumulative) exposure time vs. detected counts data pairs for each pixel, using a standard linear regression algorithm. Cosmic ray hits in individual readouts are detected and rejected during this processes by searching for and rejecting outliers from the fit. The fit and reject process is iterated for each pixel until no new samples are rejected. The slope of the fit and its estimated uncertainty are stored as the final countrate and error, respectively, for each pixel in the output image SCI and ERR arrays. In addition, the number of non-rejected data samples for each pixel and their total exposure time are stored in the SAMP and TIME arrays, respectively, of the final image.

5. CALNICB Pipeline

The second phase of the NICMOS pipeline is performed by the `calnicb` task, which processes associated observations, each of which must have already received instrumental calibration by `calnica`. This task is also completely data driven, with processing information coming from image header keywords and the association table. The three major steps performed by `calnicb` are: 1) combine multiple images that were obtained at individual sky positions (if any); 2) measure and subtract the thermal background signal from the images; and 3) combine all the images from a given observing pattern into a single mosaic image.

Steps 1 and 3, which both involve combining data from overlapping images, have many common features. First, proper registration of the input images is accomplished by estimating the image-to-image offsets from world coordinate system (WCS) information contained in the image headers, and then applying a cross-correlation technique to further refine the offsets. The cross-correlation uses a "minimum differences" technique which seeks to minimize the difference in pixel values between two images. Second, shifting the images to their proper alignments is done by simple bi-linear interpolation. Third, the combining of pixels in overlap regions uses the input ERR array values to compute a weighted mean, rejects pixels flagged as bad in the DQ arrays, rejects outliers using iterative sigma clipping, and stores the number of non-rejected pixels and their total exposure time in the SAMP and TIME arrays, respectively, of the combined image.

The process of background measurement varies depending on the exact makeup of the image association that is being processed. If observations of background sky locations were obtained by using an observing pattern that includes chopping, then only these images are used to compute the background signal. If background images were not obtained, then an attempt is made to estimate the background from the target images, with an appropriate warning to the user that the background estimate may be biased by signal from the target. A scalar background signal level is measured for each image by computing its median signal level, excluding bad pixels and iteratively rejecting outliers from the computation. The resulting background levels for each image are then averaged together, again using iterative sigma clipping to reject discrepant values for individual images. The resulting average value is then subtracted from all images in the association, both background (if present) and target images. Provisions have also been made for subtracting a background reference image which could include spatial variations in the background signal across the NICMOS field of view, but to date no such spatial variations have been seen.

At the end of `calnicb` processing an updated version of the association table is created which contains information computed during processing, including the offsets for each image (relative to its reference image), and the scalar background level computed for each image. An example of such a table for a "two-chop" pattern is shown in Table 1. This pattern contains one position on the target and two separate background positions. Two images have been obtained at each position. The two images at each position are first combined, the background is measured from the combined images at each background position, the average background is subtracted from all images, and then three final images are produced, one for the target and one for each background position.

Table 1. Output Association Table.

MEMNAME	MEMTYPE	BCKIMAGE	MEANBCK (DN/sec)	XOFFSET (pixels)	YOFFSET (pixels)
n3uw01a1r	EXP-TARG	no	INDEF	0.00	0.00
n3uw01a2r	EXP-TARG	no	INDEF	-0.15	0.00
n3uw01a3r	EXP-BCK1	yes	0.28	0.00	0.00
n3uw01a4r	EXP-BCK1	no	INDEF	0.00	0.20
n3uw01a5r	EXP-TARG	no	INDEF	0.32	0.00
n3uw01a6r	EXP-TARG	no	INDEF	0.00	0.00
n3uw01a7r	EXP-BCK2	yes	0.26	0.00	0.00
n3uw01a8r	EXP-BCK2	no	INDEF	0.00	0.12
n3uw01010	PROD-TARG	no	INDEF	INDEF	INDEF
n3uw01011	PROD-BCK1	no	INDEF	INDEF	INDEF
n3uw01012	PROD-BCK2	no	INDEF	INDEF	INDEF