

The WFPC2 Associations Science Products Pipeline at CADC and ST-ECF

David Schade, Alberto Micol, Daniel Durand, Benoit Pirenne, Luc Simard, Marcus Dolensky, Peter B. Stetson

Canadian Astronomy Data Centre, National Research Council of Canada and Space Telescope-European Coordinating Facility

Supported by the Canadian Space Agency and the European Space Agency

The WFPC2 Associations Science Products Pipeline

The Virtual Observatory will be thirsty for high-quality science products. The WFPC2 Associations Science Products Pipeline (WASPP) has been developed by the Canadian Astronomy Data Centre (CADC) and the Space Telescope-European Coordinating Facility (ST-ECF) with the goal of producing science-quality images and catalogs. Association information allows 74,787 individual images to be processed into 21,260 combined images, or “stacks”. Data quality is greatly improved by this procedure that removes cosmic rays and increases sensitivity. Source detection and photometry is done on the stacks to produce catalogs. Corrections are made to the WFPC2 astrometry by reference to the USNO2 catalogue. The result is ready-to-use science content. This project has consumed approximately 3 CPU-years of processing time and has yielded a massive dataset with 25 million source detections that represents a new class of archival science products for Hubble Space Telescope and ensures that WFPC2 data will play an important role in the launching of the Virtual Observatory.

Data centers need to initiate projects like this one. As part of their evolution they must move toward creating improved science content, thus building the foundation of the Virtual Observatory. The WASPP represents the most recent of a long line of innovations by CADC and ST-ECF, including the first WWW interface for astronomy archives, previews and associations for HST data, and the concept of “On-the-Fly Recalibration” for astronomical data (Crabtree, D. Durand, D., Gaudet, S., & Hill, N. 1996 *Astronomical Data Analysis Software and Systems V*, ASP Conference Series, Vol 101, p. 505). The Virtual Observatory is a compelling vision but, in this case, the medium is *not* the message; in the absence of excellent science content flowing through it, the VO will be a dry, scientifically infertile riverbed.

Motivation for associations and WFPC2 pipeline

The accumulated archive of nearly a decade of WFPC2 observations is vast and varied, a rich resource for astronomy. As of May 2002 there are 21,260

associations after calibration observations are removed (see article by Micol & Durand, ST-ECF Newsletter January 2002) with a total exposure time of 42.7 million seconds. Figure 1 shows the distribution of associations across only the 10 most popular filters. Roughly 85% of the associations have 4 or fewer members (58% have two members). The value delivered to users of the WFPC2 archive is enhanced by applying the WASPP procedure. It obviates the need for labor-intensive procedures to transform raw or calibrated data into a science-ready product. The normal sequence of steps needed to use archive data (query, download, determine “dither” pattern by hand, combine images) is a reasonable investment for a few datasets but it is impractical if a large number of datasets need to be sifted through and evaluated to find a few of interest.

We recognize that it is not possible to design a single pipeline that will be optimal for all science applications. Nevertheless, we believe that it is possible to save a lot of effort on the part of a lot of users and we believe there are many science applications that can use our pixel and catalog data directly.

How are the associations constructed?

Individual exposures are grouped into associations in a variety of ways (see references or http://archive.eso.org/archive/hst/wfpc2_asn). The association process identifies logical sets of images and determines their offsets from one another via one of three routes: cross-correlation, jitter information, or World Coordinate System (WCS) information. Dither patterns are known reliably for 93% of the potential “stacks”. The **association file** includes all ancillary information like gain and exposure time for each member of the association. The associations used here are based on a common proposal identification, filter, roll angle, and position on the sky. Offsets are limited to no more than 100 WFC pixels to avoid geometrical distortions problems. This scheme can be extended to associate data from different telescopes at different wavelengths or even to associate data from widely divergent sources that address a common science theme.

How does the production pipeline work?

The starting point for WASPP is the **association file** which contains all of the information necessary to do the image combination. A series of scripts accesses this file and executes the WASPP procedures in sequence.

Recalibration on-the-fly is the first step of the process. Members of the association and their calibration data are retrieved from the archive and placed in a work directory where standard recalibration is performed.

Image shifting, scaling, zeropoint: We adopted a shift-and-add approach for combining the individual members of the association. Images are shifted to a

common reference frame using fractional pixel shifts normally accurate to 0.015 arcseconds or better and then combined using a weighted average. The frames are scaled to the average exposure time and a zeropoint offset is added to correct for background differences between the images.

Artificial Skepticism (AS) (Stetson 1989, V Advanced School of Astrophysics [Univerisidade de Sao Paulo], p.1.) is a method of computing a robust average image using a continuous weighting scheme that is derived from the data themselves.

The adaptive weights are given by the equation:

$$w_i = \frac{1}{\sigma_i^2} \frac{1}{1 + \left(\frac{|r_i|}{\sigma_i}\right)^2}$$

where w_i is the weight of the i^{th} pixel, σ_i is the sigma of the i^{th} pixel in the stack as derived from the read noise and gain. The “natural” pixel weights $1/\sigma_i^2$ are therefore modified by a Lorentzian-like function that puts decreasing weight on increasingly suspicious values such as cosmic rays without fully clipping any value. The r_i term, which is the residual between the current average pixel value and the value of the i^{th} pixel, is recomputed at each iteration. Convergence is achieved quickly, and the weights only need to be iterated a few times (we adopt five iterations).

Weight Maps: After the first pass through the AS stacking is complete the resulting stack is used to back-predict the variance in each pixel yielding an improved, more robust estimate that is free of cosmic rays. The AS stacking is then repeated with this improved weight map. An output weight map is produced for the stack by propagating the AS weights for the final image.

Astrometric corrections: The WFPC2 World Coordinate System has good internal, or relative, precision but individual frames exhibit a dispersion of 1.6 arcseconds (systematic offset) relative to the USNO2 reference frame. This is not adequate to cross-identify sources. For each WFPC2 image we build a mosaic (using `stdas.hst_calib.wfpc2.wmosaic`) and retrieve all USNO2 catalogue stars from the Vizier clone at CADC that might be on the frame given the original WCS information. We search for correspondence between the catalogue and observed bright sources and calculate offsets in RA and DEC and errors in the offset where more than 1 star was found. The uncertainty of the USNO2 data is claimed to be 0.25 arcseconds (root-mean-square or r.m.s.) but that was at an epoch in the 1950s. Accumulated proper motions bring the current r.m.s into the range 0.3-0.4 arcseconds depending on the field (Stetson, private communication) Figure 2 shows the offsets and the errors in those offsets. This

reduces the systematic errors to less than 0.25 arcseconds for 50% of the WFPC2 pointings and less than 0.34 arcseconds for 85%.

Image Content classification

There are many types of observations in the WFPC2 archive collection. There are fields that are dominated by a single bright star and fields that are dominated by numerous faint stars or galaxies. There are pointings with 20,000 detected sources and there are blank fields. Some fields have extended objects that are larger than the field-of-view of WFPC2. It is very useful to determine an “image content” classifier to allow querying for specific types of pointings and so that automated processing pipelines can “tune” the choice of processing parameters to match the image content. We have implemented a simple scheme to do this.

Each WFPC2 chip is analyzed to determine the total flux per second and the ratios of the fluxes in stars, small galaxies, and extended objects to the total flux in the chip. If a single chip differs dramatically in these ratios from the other chips this might indicate that a bright target is centered on that chip (often the PC).

Source catalogs were produced using the widely-used SExtractor software (Bertin, E. & Arnouts, S. 1996 A&AS, 117,393). SExtractor was run on a variety of image types (for example, rich galaxy clusters, deep extragalactic pointings, globular clusters, and Cepheid fields in galaxies) in order to determine a good general set of configuration parameters that would yield robust detections and object deblending. The results were inspected to decide which detection and deblending parameter values worked best.

Our source catalogs for the 6 reddest filters contain detections for approximately 18.4 million and the full dataset containing all filters will contain roughly 25 million sources (not all distinct objects).

Products: In addition to the combined image we provide a weight map, an image showing the SExtractor apertures, the segmentation image (showing which pixel belongs to which source), and the background image. The catalogs themselves are ingested into a database and delivered through a query interface.

Data verification

A verification process needs to be designed and executed to demonstrate the quality of the products of the WFPC2 Association Science Products Pipeline. Science users need to be able to trust the data and to point to proof that their just is justified. Elements of the verifications process are:

- Photometric accuracy

- zero points have been set correctly
 - header information (filter, exposure time, gain) is correct
- Astrometric accuracy
 - zero point adjustments are determined and implemented correctly
- Basic calibration (flat-fielding etc.) is reliable
- Weight maps are correct
- Stacking has been implemented correctly
 - Scaling, shifting, zero offsets
- Sextractor results are accurate photometrically, astrometrically

This part of the work is in progress at the time of writing.

Release and future development

We are going to release the results of this project early in a prototype interface and this means that data quality may not always be optimal during the initial release. We will collect reactions from users and we will evaluate the system and the data quality ourselves and we will compile a list of improvements to the processing and delivery systems. This may include changes to the products themselves and to the user interface. Then we will do a Phase II release of this project. The early release in mid-2002 will include processed associations for the 6 most popular filters comprising 14,089 stacked WFPC2 images (66% of the total) and catalogs with 18.4 million sources.

References

{2002 January

ST-ECF Newsletter 30, p. 14

Micol, A. Durand, D.

"Enhanced WFPC2 Associations"

}

{2000adass...9..223M,

author = {{Micol}, A. and {Durand}, D. and {Schade}, D. and {Gaudet}, S. and
 {Stetson}, P. and {Pirenne}, B. and {Benvenuti}, P. and {Mallocci}, G. and
 {Raviv}, G.},

title = "{The HST Science Data Archive as a Discovery Tool: First
 Experiment}",

booktitle = {ASP Conf. Ser. 216: Astronomical Data Analysis Software and
 Systems IX},

year = 2000,

volume = 9,

pages = {223+},

```
url = {http://esoads.eso.org/cgi-bin/nph-  
bib_query?bibcode=2000adass...9..223M&db_key=AST},  
adsnote = {Provided by the NASA Astrophysics Data System}  
}
```

```
{1998adass...7..304D,  
author = {{Dolensky}, M. and {Micol}, A. and {Pirenne}, B. and {Rosa}, M.  
},  
title = "{Enhanced HST Pointing and Calibration Accuracy: Generating HST  
Jitter Files at ST-ECF}",  
booktitle = {ASP Conf. Ser. 145: Astronomical Data Analysis Software and  
Systems VII},  
year = 1998,  
volume = 7,  
pages = {304+},  
url = {http://esoads.eso.org/cgi-bin/nph-  
bib_query?bibcode=1998adass...7..304D&db_key=AST},  
adsnote = {Provided by the NASA Astrophysics Data System}  
}
```

```
{1998adass...7...45M,  
author = {{Micol}, A. and {Pirenne}, B. and {Bristow}, P.},  
title = "{Constructing and Reducing Sets of HST Observations Using Accurate  
Spacecraft Pointing Information}",  
booktitle = {ASP Conf. Ser. 145: Astronomical Data Analysis Software and  
Systems VII},  
year = 1998,  
volume = 7,  
pages = {45+},  
url = {http://esoads.eso.org/cgi-bin/nph-  
bib_query?bibcode=1998adass...7...45M&db_key=AST},  
adsnote = {Provided by the NASA Astrophysics Data System}  
}
```

```
{1997hstc.work..349M,  
author = {{Micol}, A. and {Bristow}, P. and {Pirenne}, B.},  
title = "{Associations of WFPC2 Exposures}",  
booktitle = {The 1997 HST Calibration Workshop with a new generation of  
instruments /edited by Stefano Casertano, Robert Jedrzejewski, Charles D.  
Keyes, and Mark Stevens. Baltimore, MD : Space Telescope Science Institute  
(1997) QB 500.268 C35 1997, p. 349.},  
year = 1997,  
month = jan,  
pages = {349+},  
url = {http://esoads.eso.org/cgi-bin/nph-  
bib_query?bibcode=1997hstc.work..349M&db_key=AST},
```

adsnote = {Provided by the NASA Astrophysics Data System}
}