Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

WFPC2 Associations Pipeline: Publishing HST archives within VO

D. Durand[1], Alberto Micol[2], David Schade[1], Luc Simard[1], Michael Corbin[3], Richard Hook[2,3], Anton Koekemoer[3], Megan Donahue[4]

1 *Canadian Astronomy Data Centre, Herzberg Institute of Astrophysics, National Research Council Canada, Victoria, Canada*

2 *Space Telescope - European Coordination Facility, ESA, Garching bei Muenchen, Germany*

3 *pace Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, U.S.A.*

4 *Dept. of Physics & Astronomy, Michigan State University, East Lansing, MI 48824-1116. , U.S.A.*

Abstract. After releasing the first version of the new HST WFPC2 associations in November 2002, the CADC, ST-ECF and STScI are jointly releasing Version 2. This presentation will discuss the software pipeline steps that were put in place to construct these associations, and the quality tests we will perform on them. Our goals are the following: Firstly, to make available higher quality products for scientists using the HST archives. Secondly, to translate all the necessary parameters used for the processing into a data model for inclusion in the Virtual Observatory. We will then present the different products which are accessible to the potential user.

1. History and definitions of associations

The WFPC2 associations concept was started back in February 1998 by an ST-ECF based team (Micol et al., 1998). Their goals at the time was trying to combine images originating from very similar pointings in order to minimize the number of cosmic ray hits on the final image. Since that time, multiple enhancements to the basic algorithm were brought up by CADC and ST-ECF, working jointly on this problem. The first joined releases occurred in November 2001 and then in November 2002 (Micol et al., 2000). At present, the team is working toward a final release which, with the help of another team based at STScI, should generate very high quality products ready for ingestion into the Virtual Observatory.

The current (released) associations are defined following these simple rules: *(1)* an association includes observations grouped by filter and observing program; *(2)* these observations occupied the same position on the sky (within 100 Wide Field Camera pixels, or 10 arc seconds); *(3)* finally, the observations were

obtained with the spacecraft at the same roll angle within 5 arc seconds (in order to avoid rotation of images when stacked).
The actual release is available at CADC[5], ST-ECF[6] and STScI[7].

2. Goals

In order to produce reliable high quality products for the final release of the WFPC2 associations, we put together the following requirements:

1. The associations will facilitate archive browsing.
2. They should be ready for science usage, i.e. the should be well calibrated photometrically and astrometrically.
3. They should be well described, with their processing history present in their headers.
4. They should be as deep as possible, given the fact that all the pipelines are running unattended.
5. Proper error propagation should be implemented
6. Weight maps should be available.
7. Powerful processing facilities required (21,000 stacks $\times$ 30 minutes each).
8. Enhanced metadata for data mining should be available for each data products (see below).

3. The WFPC2 association pipeline

3.1. Association discovery

The very first step of our WFPC2 pipeline is to run a discovery agent to identify and classify the associations. This somewhat complex process looks into the HST observing log and, according to the specified association rules, group together individual observations. This process runs every week automatically. Sometimes some associations got destroyed while more are created, depending on the availability of new or otherwise re-archived observations. The "center" of an association is defined as the deepest exposure of the group. Of course this process is running only on public datasets.

3.2. Shift Finding

Next, we need to find the relative shift between every observations part of a given association. This process is quite CPU intensive and consists in:

1. Masking the strongest cosmic rays using a *Sextractor* based algorithm
2. Running a 2-D cross-correlation and find the maximum and its error of the resulting array
3. Comparing the different shifts finding method (WCS, Jitter info and 2-D cross-correlation) and store the best ones

---

[5] `http://cadcwww.hia.nrc.ca/wfpc2/`

[6] `http://archive.eso.org/archive/hst/wfpc2_asn/3sites/`

[7] `http://archive.stsci.edu/hst/wfpc2/`

3.3. Association Products

Once the shift values are available, we then run the following steps to create the final products:

1. On each observation, we run OTFR, the calibration pipeline using latest software and calibration files (including warm-pixel correction)
2. Then we shift the images using standard *IRAF* tasks
3. Once all images have been registered to the master image, we stacked them using Stetson's artificial skepticism algorithm (Stetson, 1989)
4. For the astrometric correction we search for matching objects between the image and the USNOB2 catalogue. From the matched list, we compute a linear shift in RA and DEC and apply it to the product.
5. Then we augment each product image with relevant information about the processing history like the version of each modules and packages used and other important processing parameters, like the number of stars identified for the astrometric correction.
6. Finally, the different products are ingested within the CADC storage system which is accessible through a special procedure from ST-ECF and STScI). At this point in time, the association is accessible through all partner web site.

4. Quality assessment

Quality assessment of the WFPC2 associations is an important but difficult part of the project. For these images to be useful for research, they must be verified to be photometrically and astrometrically accurate. For the first release of the WFPC2 associations, we limited our tests to only a few hundred images. The results of these tests are presented on the joint CADC / ST-ECF / MAST web interface. We will perform these and additional tests for the second release:

4.1. Photometry test

1. Comparison between instrumental magnitude between final stack and individual exposures.
2. Comparaison with already published material.

4.2. Astrometry test

Automated astrometry tests are not possible. When the astrometric correction is performed within the pipeline, a measure of the correction accuracy as well as the number and the list of known stars is available within the header of each product. Since the internal precision of the USNO2 stars is not sufficient to provide the perfect solution for a given WFPC stack, trying to estimate each solution more accurately is impossible. From our first release, we estimated that the astrometric correction error was essentially equivalent to the internal error of the USNO catalogue after examining a subset of images with known USNO2 stars. We will perform the same test on the second version.

4.3. MultiDrizzle

Another test which we will perform on the second version of the products will involve comparing them with versions produced using the new STSDAS task multidrizzle (Koekemoer et al., this conference). In principle, the images produced by multidrizzle and the associations pipeline should be equivalent, and thus difference would indicate flaws in either the multidrizzle or associations pipeline software. A set of multidrizzled images has already been produced by MAST, and is awaiting comparison with the corresponding images produced in the second release.

5. Virtual Observatory publication

For the WFPC2 associations extra steps are actually running at CADC to allow us to publish the products within the Canadian Virtual Observatory (CVO). Similar steps are being taken within our collaborating institutions. In addition of publishing the product itself (space, time and wavelength parameters) in the CVO, we are running other steps to augment the information on those products: *(1)* an image characterisation software measures some basic statistical parameters like total flux and ratio of extended objects over point sources; *(2)*the position and the flux of each source are measured on the image.

6. Conclusions

The second release of the WFPC2 Associations will be made available to the public soon by CADC, ST-ECF and STScI. Richer metadata will be available for deeper and better calibrated products for WFPC2. Better uniformity will ensure better usage and inclusion within a Virtual Observatory in spite of the pointing mode nature of these products. The working group is applying a similar technique for creating ACS associations (Hook et al., 2004)

References

Hook, R., et al. 2004, this volume, 62

Micol, A., Pirenne, B., Bristow, P. 1998, in ASP Conf. Ser., Vol. 145, ADASS VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 349

Micol, A., et al. 2000, in ASP Conf. Ser., Vol. 216, ADASS IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 223

Stetson, P. 1989, "V Advanced School of Astrophysics", Universidade de Sao Paulo, p. 1