Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

The HST Science Data Archive as a Discovery Tool: First Experiment

A. Micol[1], D. Durand[2], D. Schade[2], S. Gaudet[2], P. Stetson[2], B. Pirenne[3], P. Benvenuti[1], G. Malloci[4], G. Raviv[5]

Abstract. The Space Telescope European Coordinating Facility and the Canadian Astronomy Data Centre have started a project aimed at augmenting the way an astronomer can query the Hubble Space Telescope science archive.

Today the HST archive contains more than 70,000 WFPC2 exposures. We are preparing an automatic pipeline that will extract several parameters from those images. A list of all the sources seen by WFPC2, along with photometrical, morphological and astrometric measurements of both the detected astronomical sources and of the background will be extracted.

Astronomers will make use of such a database to query the archive in a more sophisticated and scientific manner, that is in the way they think, not in the way the archive is built.

This new approach will allow queries of the kind: search for high latitude fields having a certain limiting magnitude and where the density of extended sources is greater than a certain threshold.

Nowadays, with the advent of huge detectors over big fields of view, such an approach becomes a must. The catalogue of extracted objects will be immediately useful to prepare a wide range of scientific programs for the 8 - 10 meter class telescopes, e.g., multi-object spectroscopy of HST sources.

1. Rationale

Every time a scientist, looking for information on some astronomical field of interest, tries to access a science data archive, s/he has to deal with the intrinsic structure of the archive itself.

Modern astronomical archives are trying to hide all the technicalities to the astronomer, but yet, usually, only a catalogue of observations can be queried.

[1]Space Telescope European Coordinating Facility / ESA

[2]National Research Council of Canada / CADC

[3]Space Telescope European Coordinating Facility / ESO

[4]Universitá di Cagliari

[5]In sabbatical at National Research Council of Canada / CADC

Such a catalogue is a mere collection of pointings, with some associated information, like, e.g., coordinates, exposure time, filter, principal investigator, instrument mode.

Only rarely some more useful information is added to the catalogue; for example in the IUE case a classification field is available. Through such a field the scientist can query the archive for all objects of a certain type.

This approach is valid if the archive was to serve only the primary targets; it is certainly not prone to be a discovery tool.

As a consequence while an archive could contain very interesting and unexploited data, scientists are offered with very limited means of finding them.

2. WFPC2 Associations

The WFPC2 associations were introduced in the HST archive at the beginning of 1998. Since then, astronomers can retrieve from the ST-ECF and CADC archives WFPC2 on-the-fly re-calibrated (Crabtree et al. 1996) co-added images which have already been cleaned of cosmic rays.

The offsets used to co-add the images from an associations are computed from the *jitter* files (Micol, Pirenne & Bristow 1997, 1998). Unfortunately not all the HST observations have associated jitter files, mainly because of telemetry problem. Therefore many WFPC2 frames (55%!) cannot be grouped into an association.

In early 1999, the development of a new association pipeline was started. Standard cross-correlation techniques were used to measure the offsets among WFPC2 observations belonging to the same proposal and using the same filter.

Comparison of the cross-correlation offsets with the jitter offsets, along with manual inspection of the WFPC2 frames, showed an unexpected first result: the jitter pointing information is affected by a drift in its zero-point. The effect could be as high as 0.2 WF pixels during 3 HST orbits ($1orbit = 96minutes$). So we had to limit the allowable jitter offset for the current implementation.

The cross-correlation method works well for observations showing enough features to correlate with. Basically this happens for the redder filters. When the cross-correlation fails (i.e., there is no agreement among the offsets computed with different chips), the jitter offsets are used.

The computed offsets are then the input for the stacking algorithm. A co-added, cosmic ray free image is produced for each association. This high level product is therefore an image ready to be analyzed and reduced.

In order to check the cross-correlation results, we ran on some selected stellar fields the DAOPHOT suite. After tuning the fitting parameters, we were able to reproduce very well the cross-correlation results, giving us a third independent algorithm for the offsets. At the same time, stellar photometry is also performed and elementary information on sky background for each exposures part of an association is available.

There are, at the time of writing, about 20.000 WFPC2 associations ready to undergo the next step: data mining.

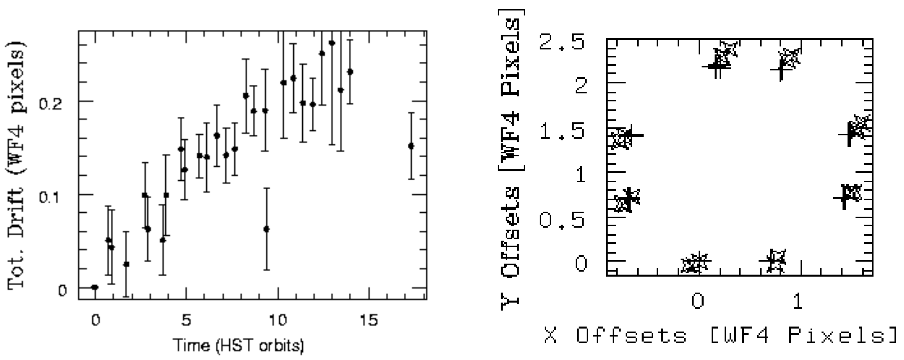

Figure 1. Left: Drift of the jitter zero point - Right: Comparison of offsets measured with cross-correlation (stars) and daomatch (pluses) techniques. The observing strategy (dither) is clearly reproduced.

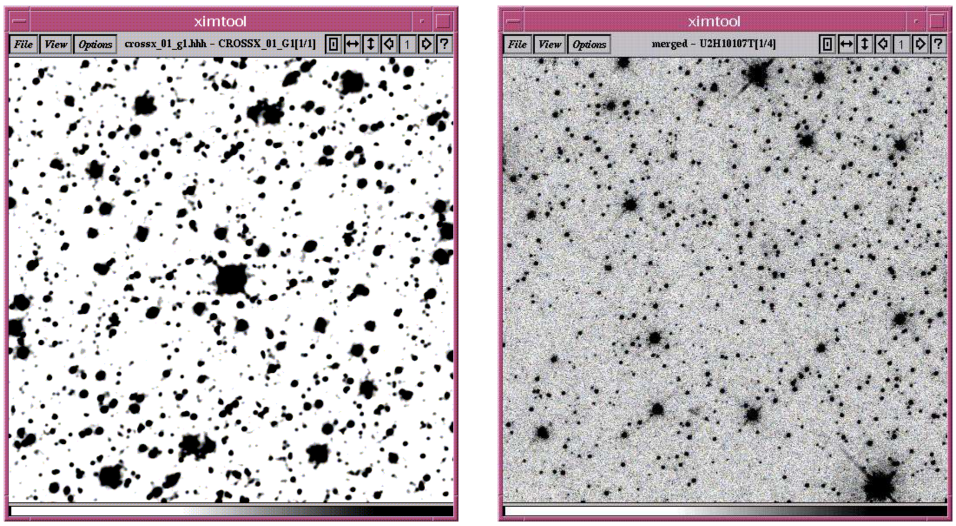

Figure 2. Left: The cross-correlation image - Right: A WFPC2 association product (co-added, cosmic-ray free).

3. Extraction of Scientific Parameters

To deal with the massive number of images, and of objects on those images, an automatic pipeline is being constructed. Different levels are foreseen in such a pipeline:

- The WFPC2 association products described above will be used;
- Object detection: Objects will be detected on those associations, along with preliminary rough measurements of both objects and background, using standard packages (e.g., SExtractor (Bertin & Arnouts 1996), DAOFIND, etc.). Statistical information will be extracted and will populate a database aimed at describing the contents of the WFPC2 archive. Method used: SExtractor;
- Object measurements: Various pipelines can be plugged in at this stage, depending on the image characteristics. E.g., *DAOPHOT* can be run onto stellar fields, *GALPHOT* onto galactic fields, etc.;
- Data Mining: The measurements will be stored in a database. Different database systems (SYBASE IQ, Objectivity, etc. are under evaluation. Dedicated user interfaces are being designed (Dowler et al. 2000, Schade et al. 2000);
- Quality Control: The quality of the extracted magnitudes has been preliminary assessed onto some Medium Deep Survey fields (Malloci 1999); the agreement of the SExtractor magnitudes with the MDS magnitudes is generally good, and show a maximum discrepancy of about 0.6 mag at the fainter end. More assessments against published measurements are required (more fields, different filters).

References

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Crabtree, D. R., Durand, D., Gaudet, S., Hill, N., & Pirenne, B. 1996, in ASP Conf. Ser., Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes (San Francisco: ASP), 505

Dowler, P., Schade, D., Zingle, R., Durand, D., & Gaudet, S. 2000, this volume, 211

Malloci, G., master thesis, in preparation

Micol, A., Pirenne, B., & Bristow, P. 1997 "Associations of WFPC2 Exposures" HST Calibration with a New Generation of Instruments, ed. S. Casertano, (Baltimore: STScI), 349

Micol, A., Pirenne, B., & Bristow, P. 1998, in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 45

Schade, D., Dowler, P., & Durand, D. 2000, this volume, 215