Progress Report

26th April 2010

1 Introduction

This is the third progress report on the Austria In-kind Contribution in the context of the *Data Reduction Software Project*.

Currently, work is concentrating on the deliverables DR1 and DR2, which concern the astrometric and photometric correction (DR1) and the stacking and mosaicing of dithered images (DR2), and in particular sky subtraction for near-IR data and the determination of astrometric solutions.

The main events since the previous status report were:

- **Oliver Czoske** visited ESO headquarters from 1 to 12 February 2010.

- **Saptarshi Das** in collaboration with Czoske developed and coded an instrument-independent sky-subtraction recipe for near-infrared data. Details on the implementation can be found in Sect. 2 and in a separate report.

- **Saptarshi Das** and **Oliver Czoske** visited ESO headquarters from 15 to 24 March 2010. During the visit, Das presented the sky subtraction recipe and discussed code structure and organisation with the ESO partners. Czoske participated in the development of an ESO/Reflex workflow for the reduction of imaging data.

- **Saptarshi Das** visited ESO again 14 to 16 April 2010. The general structure of CPL recipes and the integration of CPL recipes into the ESO/Reflex workflow were the main topics dicussed during the visit.

- **Tomasz Hrycak** studied the problem of astrometry of imaging, in particular the algorithms used in SCAMP from the Terapix software suite (Sect. 3).

- **Darian Onchis** worked on the possibility of parallelizing the execution of cpl recipes (Sect. 5).

2 Sky subtraction from NIR imaging data

We have implemented a two-pass algorithm for the subtraction of the sky background from near-infrared imaging data and tested it on data from HAWK-I and ISAAC.

Briefly, the algorithm works as follows: Given a set of dithered exposures from an observing block, a separate sky correction image is built for each exposure by stacking exposures from a certain time window around the exposure. The main concern is the removal of any traces of objects in the sky-correction images. This is achieved by (i) normalizing the exposures by their median; (ii) taking the mean of these exposures while applying an outlier rejection algorithm; (iii) subtracting the sky-correction images from the exposures; (iv) combining the images based on the astrometric solution available at this point of the reduction process; (v) detecting objects in the combined image using `cpl_detect_objects` or SEXTRACTOR; (vi) masking objects in the original exposures by applying the inverse of the astrometric solution; (vii) taking the mean of the masked exposures as in (ii).

The algorithm has been implemented in two user-visible cpl recipes:

- `esosoft_compute_bkg`

- `esosoft_subtract_bkg`

and a number of invisible cpl functions implementing $\kappa\sigma$-clipping, the mean/median rejection algorithm, etc.

The code has been tested on data from HAWK-I (covering the two main observing strategies: sky estimated from the science exposures for observations of "empty" fields or sky estimated from dedicated sky exposures in a nodding strategy) and ISAAC.

A detailed description of the algorithm, code and tests that we have carried out will be given in a separate report.

3 Astrometry and distortion correction

The astrometry of imaging data will use SCAMP from the Terapix software suite.[1] We have studied the documentation of SCAMP and the algorithms used by the programme and found them to be generally adequate. From a mathematical point of view, it appears to be possible to improve the model used for the astrometric solution. SCAMP uses a polynomial model given as a sum of monomials. A representation by a sum of Chebyshev polynomials would improve the numerical stability of the algorithm. In practice, however, the model as used by SCAMP does not appear to give problems and there is no immediate need to act on this point.

SCAMP also determines the *relative* photometry between the exposures. In the infrared, *absolute* photometry can be determined if objects from the 2MASS catalogue are present in the observed fields. However, SCAMP does not appear to be able to do this. In the optical, absolute photometry is determined from observations of standard star fields, which is a much harder problem.

---

[1]`http://www.astromatic.net/software/scamp`

Overall, SCAMP is judged to be a good and reasonably well-tested solution to the problem of astrometry. With the availability of a preliminary ESO/Reflex workflow (Sect. 6), it is now possible to further test SCAMP on the test data sets from various VLT instruments as well as WFI, and to refine and determine default parameter settings.

As for the sky subtraction, we will discuss algorithms and tests in a separate report.

4 Development of instrument-independent recipes

At present, the esorex-executable recipes are instrument dependent. Going forward, the aim is to develop recipes which are independent of any particular instrument and are properly integrated within the ESO data flow system (dfs). Such recipes are required to have the capability of determining the instrument-specific values from the input images to be processed. Also, such recipes should save the data accordingly; for instance, depending on the supplied image the recipe must decide whether to save a primary header or not.

For such a requirement, we have designed a framework to develop recipes independent of any instruments. The general structure of each function that the recipe executes is as follows:

- Determine the detector specific numbers.

- Extract required framesets.

- Set appropriate dfs group to the frames.

- If required, save the primary header of the product file.

- Loop over the number of detectors.
    - Extract the frameset corresponding this detector.
    - Execute the recipe at the algorithm level.
    - Save the product (if required by extending the primary header).
    - clean up memory acquired by this detector.

- clean up memory, and return.

The algorithm at the recipe level is kept at the conceptual level, no data related I/O functionality is involved at this level. However, this strategy might change in order to have a better memory usage.

Examples of functions developed following this scheme are `esosoft_extract_frameset`, `esosoft_dfs_set_groups`, `esosoft_save_primary_header`.

5 Code parallelization

Some of the problems we are facing when running the calibration and data reduction recipes for astronomical instruments are related to the increasingly severe run time constraints. These

difficulties are related to the size of the FITS images, the number of images used and the performances of the implemented sequential algorithms. In order to efficiently solve part of these computational drawbacks, we are testing automatic parallelization strategies that will allow the user to run the recipe on a desktop PC or on a computer cluster without further modification of the code. The automatic parallelization mode will simply take the serial code and determine the best way to break it up among multiple processors. This approach will also cover some aspects that are missing in the astronomical code development up to now (see e.g. Noble, M. S.: "Getting More From Your Multicore: Exploiting OpenMP for Astronomy", ADASS XVII, 2008[2]).

The optimization (parallelization and profiling of the code) for the HAWK-I ESOsoft is done on the cluster MareNostrum at Barcelona. This cluster comprises 2560 JS21 computing nodes (blades) and 42 p615 servers. Every blade has two processors at 2.3 GHz running the Linux operating system with 8 GB of RAM and 36 GB local disk storage. All the servers provide a total of 280 TB of disk storage accessible from every blade through GPFS (Global Parallel File System).

We use IBM's XL C compiler to introduce a number of new enhancements to the HAWK-I ESOsoft recipes. We are also considering the use of the new GCC release that will have support also for automatic parallelization (using Graphite). Some initial tests showed good improvement using the OPEN MP as default parallel API. We were testing recipes using a relatively small number of processors (4–8) in order to avoid the waiting time in the queue with the following preliminary results:

Test1:
Run time without OMP...   180.0011 s
Run time with OMP...   96.2123 s

Test2:
Run time without OMP...   201.1201 s
. Run time with OMP...   129.2001 s

Test 3:
Run time without OMP...   427.4785 s
Run time with OMP...   245.9401 s

6 ESO/Reflex workflow

A preliminary version of the workflow for the reduction of imaging data using the ESO/Reflex environment was implemented by Armin Gabasch. This version uses existing CPL recipes for the sky subtraction and python scripts that act as wrappers around the Terapix programmes SEx-TRACTOR, SCAMP and SWARP that respectively create object catalogues, determine astrometric solutions and resample and combine the exposures into a combined output image.

This workflow has been installed on `serv2` in Vienna, and it has been verified that it is running. The workflow will serve as the backbone of the further development of software modules.

---

[2] `http://aspbooks.org/publications/394/730.pdf`

In particular, the sky subtraction recipes will shortly be replaced by the improved recipes developed for DR01, and the python scripts will be replaced by CPL recipes that will set up and run the Terapix programmes.

7 Revised schedule

Due to the delay in signing the contract between the Ministry and the University of Vienna, the hiring of the personnel also had to be delayed. The contract was signed beginning of July 2009 and the first installment (2009 budget) was transferred from the Ministry in October. The hiring process started after signing the contract. In the case of personnel at the Faculty of Mathematics, the 3 positions could be filled by October 2009. The astronomer position had to be announced twice and was filled by 15 January 2010.

These delays had an impact on the progress of the DR-01/02 projects. The time schedule, as defined in the Statement of Work is the following for both projects:

- Deliverable start: 01.04.2009
- Detailed specification: 30.04.2009
- Prototype review: 15.08.2009
- User documentation: 31.03.2010
- Acceptance release: 30.06.2010
- Final release: 30.09.2010

It is now aimed to achieve the status of the prototype review by 30.06.2010. The CPL-based instrument-independent background subtraction procedure is in a fairly advanced stage. A MATLAB-based prototype for the background subtraction of optical data is tested and is going to be converted into CPL code. The central effort now is the workflow-based prototype for astrometric and photometric corrections and final mosaicing based on the Terapix software modules SExtractor, scamp and SWarp.

The release of user documentation based upon the prototype review is planned for 30.08.2010. The acceptance release is aimed for 30.09.2010. Again, the central effort will be the conversion and adaption of the Terapix modules to CPL compliance. Final release of DR-01/02 is then scheduled for 31.12.2010. The time schedules of the other projects should not be affected.