Software Modules for data Reduction
Projects DR-01 & DR-02

Progress report to be discussed Feb. 1/2, 2010

1. Introduction

The purpose of this project is to provide a set of pipeline recipes based on the ESO Common Pipeline Library (CPL) for the astrometrically and photometrically correct stacking and mosaicing of imaging data. Currently a lot of observers split observations of a given target among different observation blocks. A set of CPL recipes will provide a seamless interface to the existing pipelines and to the execution of front-end tools esorex, Gasgano and the Reflex workflow engine. This includes in detail:

- evaluate existing software modules and packages and develop a prototype data reduction workflow to achieve stacking, mosaicing of (jittered) images taken from different observation blocks;

- design a set of instrument independent CPL functions for the stacking, mosaicing, and variance estimates of dithered/jittered imaging data taken in different observation blocks;

- validate the methods using VLT archive data for different imaging instruments;

- integrate the recipes in the ESO Reflex scientific workflow environment;

- provide user documentation and validation reports.

2. Detailed requirements

The project has the following detailed requirements:

- Develop a prototype processing chain using existing software packages like the ESO/MVM package or the TERAPIX tools (SCAMP and SWarp). The prototype will be reviewed on the existing data sets before starting the coding in CPL.
- The input data for this task are pipeline reduced data that are basically free of instrumental signature at the single-chip level. The input data are either single-chip FITS data or multi-extension FITS data.
- Allow the inclusion of fringe-correction maps.
- Source detection should make use of the TERAPIX SExtractor task.
- Implement the recipes in CPL, provide user documentation and Reflex workflows.
- The CPL routines shall deliver the photometry and astrometry solutions with error estimates and quality indicators.

3. Software related issues

All the relevant evaluations, testing and development are carried out on a dedicated 64bit server machine running Scientific Linux version 5.3. The following CPL-5.0.1 based pipelines are installed and used for tests:

- fors-4.6.3;
- hawki-1.6.3;
- isaac-5.10.2;
- vimos-2.2.3;
- wfi-0.8.3.

The astronomical software collection scisoft-7.3.1 is used for generic data handling, image processing and image visualization.

Three software packages were evaluated for the purpose to derive a astrometrically and photometrically corrected stacked/mosaiced image from an input list of dithered/jittered images.

3.1 TERAPIX / Astromatic pipeline software

The installed software modules consist of

- SExtractor: creates object catalogues from images (version 2.5.0);
- SCAMP: computes astrometric and photometric solution (version 1.4.6);
- SWarp: re-sampling and coadding of images (version 2.17.1).

SCAMP operates on catalogues created by SExtractor requiring that the values for the WCS, pixel scale and position angle are already quite accurate contained in the FITS header of the image. An astrometric reference catalogue (default USNO-B) for the corresponding sky area is downloaded from the CDS/Vizier database. In a first step pixel scale and position angle are calculated by cross-correlation of 2D histograms of source pair coordinates between source and reference lists. The correction for the shift with respect to the true position is done by means of a cross-correlation in the space of the projected coordinates. At the end all sources are cross-matched and each set of overlapping exposures is projected o a local, common stereographic projection. The astrometric solution is computed by chi-square minimization of the differences in position between overlapping detections from pairs of fields of which one can be either a reference field or another uncalibrated field. The photometric solution is calculated by means of a chi-square minimization of the differences in magnitude between overlapping detections.

An upgrade of SCAMP to release 1.6.2 is considered which supports the UCAC3 astrometric catalogue and allows selecting astrometric references within a restricted magnitude range.

The module SWarp resamples the image pixels to a perfect projection system and co-adds them an basis of their weight maps. For the image resampling the Lanczos3 interpolation kernel with 6×6 pixels in size is used.

3.2. ESO Multi-scale vision model (ESO/MVM)

The ESO implementation of the MVM software package (esomvm-1.3.5), which is an end-to-end, high-throughput image reduction, was installed with exception of the MVM configuration editor. The Python/Gtk-based user interface yields an input error of a missing libxml2mod module which could not be recovered.

The ESO/MVM package is set up for an end-to-end data reduction of ISAAC data. The pipeline processes raw data, i.e. the science data and their associated calibration images and ancillary files (e.g. for absolute photometric calibration). All images should origin from the same observation block or a group of observation blocks. The science images are grouped in reduction blocks according to their calibration data and observing dates. Reduction modes are defined on basis of the instrument, its observing modes (e.g. filters, detector characteristics) and scientific content of the images (e.g. point or extended sources, deep imaging, crowding, etc.), which are used constrain the relevant parameters for the data reduction pipeline. The pipeline reduction consists of the following steps:

- removal of instrumental effects (bias/dark and flat field correction);
- background pixel-based subtraction (including fringe correction);
- background large-scale subtraction;
- relative astrometry based upon a reference grid;
- absolute astrometry based upon a reference catalog.

Background subtraction is done in 2 steps: a first estimate of the local background based on the median of the intensity of all images is used to create a mask for detected objects. The mask is then applied to the science images for a second estimate of the background, discarding all pixel

intensities within the masked regions. The large-scale background removal is accomplished by subtraction of smooth background whose smoothing scale is larger than the largest object in the image. For each image a background is computed using a pyramidal median transform according to the adopted smoothing scale.

The relative astrometry module produces images which are relatively astrometrically calibrated to a grid defined by a reference image (distortion correction). The reference image is selected from the images of the reduction block according to the number of sources and best S/N. This is done by means of the last scale of the wavelet transform. The wavelet scale depends on the instrument and type of observation and ranges typically between 5 to 7. The images are also searched for artifacts: cosmic rays are removed using a sigma clipping technique, satellite tracks are identified using the Hough transform and the affected pixels are masked.

The absolute astrometry module uses as input all distortion corrected images and aligns them to a reference grid defined by an astrometric catalogue. A reference image is created on basis of the WCS information of the input images, the reference grid of the relative astrometry and by converting the astrometry catalogue entries into pixel coordinates. Each source is represented by a 2D Gaussian function converting the respective magnitudes into (arbitrary) fluxes. The input images are then aligned with the reference images in the same way as described in the relative astrometry module.

The final mosaicing of the images (also from different reduction blocks) includes a photometric correction, where either a relative zero point is calculated between the images or an externally calculated absolute zero point is used. The procedure to perform the mosaicing is basically identical to the SWarp module of the TERAPIX software.

3.3. Astrometry engine developed by astrometry.net

The astrometry engine (version 0.2) was installed. Using the USNO-B catalog as reference, the program makes use of the inverted index of features for the astrometric calibration of images. The features have to be invariant to scale, rotation and translation. The selected features consist of nearby quadruples of stars. Based upon the features detected in the image, the inverted index is used to search which possible views from the catalog have also the same feature values. The features of the inverted index act as hash codes for locations on the sky. No further information on the input image is required.

4. Test data

Several data test data sets consisting of raw and/or pipeline calibrated data (i.e. bias/dark and flat field corrected) were either supplied by ESO or extracted from the ESO data archive. It is intended to cover a broad range of instruments (HAWK-I, FORS, ISAAC, VIMOS, WFI) and different types of astronomical objects. The data sets contain examples of faint extended sources, bright extended sources and fields of various degrees of crowding.

- Chandra Deep Field South (CDF-S), extracted from the GOODS data base:
  - ISAAC JHK´ band images;
  - VIMOS URI band images;
  - WFI R band images.

- galaxy cluster Abell 1689:
  - HAWK-I J band images;
  - FORS I band images.

- field around quasar QSO-J220702.4-314501:
  - HAWK-I K´ band images.

- field around the nearby bright elliptical galaxy NGC 5090:

- VIMOS R band images.

- field around the nearby bright spiral galaxy NGC 7793:
    - HAWK-I K´ band images

- field centered on globular cluster Messier 4:
    - HAWK-I J band images.

5. Evaluation of the software packages and data reduction pipelines

The test data were used to evaluate performance and accuracy of the various software packages described in section 3 in view of science-ready data. For the purpose of the test, the input data were selected to belong to the same observation block. The input data consists of calibrated (bias/dark and flat field) corrected science images. In the case that raw images were extracted from the ESO data archive, the appropriate modules from the respective data reduction pipeline was used to calibrate the data. Weight maps associated to the calibrated science frames were generally not available. Assuming that all typical instrumental signatures have been corrected, the following reduction steps were performed:

- background subtraction;
- source extraction;
- distortion correction;
- astrometric correction;
- photometric correction;
- resampling and coadding.

The software packages IRAF and MIDAS were used for the evaluation of intermediate stage data products. Generally, the visualization of the images was done with DS9 and Skycat.

The TERAPIX and the astrometry engine of astrometry.net software packages were tested with all relevant instruments. The ESO/MVM package could only be tested in the case of ISAAC data.

5.1. Background subtraction

The task of background subtraction is presently implemented differently in the various data reduction pipelines:

- **FORS:** no specific background implementation is implemented;

- **HAWK-I:** two-step background subtraction implemented in the recipes hawki_step_compute_bkg and hawki_step_subtract_bkg following the procedure of background subtraction described in the ESO/MVM package (see section 3.2.). The sources to be masked for background subtraction are computed in the recipe hawki_step_detect_obj;

- **ISAAC:** background filtering as part of the recipe isaac_img_jitter;

- **VIMOS:** no specific background subtraction is implemented;

- **WFI:** integral part of the recipe wfi_science following closely the procedure described in the ESO/MVM package (see section 3.2.).

5.2. Source extraction

Source extraction is implemented in a number of pipelines as part of the science data calibration such as vmimobsjitter and wfi_science based upon the SExtractor algorithm. For the remaining instruments the SExtractor was used directly on the background subtracted images on a script basis to create a source list.

5.3. Distortion correction

Distortion correction is only implemented in the HAWK-I pipeline using the recipe hawki_step_apply_dist to apply a reference grid to science images. The reference grid is generated by auto-calibration from suitable reference images computed with recipe hawki_cal_distortion. The procedure is follows the relative astrometry module described in the ESO/MVM package. Distortion correction is also included for the WFI in the recipe wfi_science by means of pattern matching. Distortion correction for images of other instruments was performed as part of SCAMP using sources in the overlapping image areas of the science data.

5.4. Astrometric and photometric correction

Presently the astrometric and photometric corrections are not implemented in any data reduction pipeline. SCAMP was used based upon the SExtractor generated source lists on the science image lists. In the case of ISAAC data, the ESO/MVM package was used.

5.5. Stacking and mosaicing

Stacking is implemented in various instrument data reduction pipelines such as hawki_sci_jitter, isaac_img_jitter, vmimobsjitter and wfi_science. Resampling is implemented only in the ESO/MVM package where the same functionality as in SWarp is used.

5.6. Results

Based on upon the above described constraints only a limited comparison of the software packages was possible.

A proper pixel-to-pixel and large-scale background subtraction is essential both for the identification sources in the images and of the quality of the final coadded science image. The two-step background subtraction as described in the ESO/MVM package and implemented in the HAWK-I pipeline proved to produce reliable results. An essential part is the source detection to create the object mask for the final background subtraction (e.g., hawki_step_detect_object or using SExtractor). The source detection level and source growing radius depend on the properties of the sources in the image and need to be individually set. The current implementation of the background subtraction uses pixel wise median calculation from a series of images with a minimum-maximum rejection for the running mean of the jittered images. Improvements could be achieved in a number of cases (e.g. images of crowded/overlapping sources) allowing to set the sky level to a certain percentile of the intensity distribution (default value is the 50th percentile which corresponds to the median value). Complementary to the minimum-maximum rejection of the jittered images, a kappa-sigma clipping rejection has been added.

The astrometric (including distortion) and photometric correction modules in SCAMP and ESO/MVM could only be evaluated on basis of the final coadded images. Internal astrometric calibrations based on suitable stars in overlapping fields yield only conditionally good results. External reference data (suitable distortion maps, astrometric reference stars) are necessary to achieve the desired astrometric accuracy. SCAMP and ESO/MVM yield similar results which are usually at the level of about 1´´ over the image area. The limiting factor for the accuracy is the quality of the external reference data, in particular the astrometric reference star catalogue. In all packages the USNO-B catalogue is used as default input. The present version of the catalogue has 1,042,618,261 entries down to V=21 mag with a positional accuracy of 0.2´´. The data originate from photographic plates of the Palomar and ESO sky surveys, which have a significantly lower spatial resolution than the science data to be corrected. This could result in mismatches, e.g. in the case of double stars resolved in the science images, but not resolved in the reference list or wrong distinction between point sources and extended sources. There are the following possibilities for improvement:

- ***Improving the quality of the astrometric reference stars:*** the UCAC3 astrometric catalogue covers stars in the 8 to 16 mag range with positional errors in the range of 0.015´´ to 0.020´´ containing however only 100,766,420 entries;
- ***Improving the SExtractor-generated source list:*** processing the source list to be better compatible to the astrometric reference list. The clipping procedure to exclude mismatches could be possibly improved;
- ***Algorithmic improvements:*** the astrometric solutions are generally obtained by least-square equations (normal equations). The numerical solution of the normal equations should involve regularization, e.g. by using the Wiener filter. Regularization prevents an amplification of noise. Suitable regularization parameters, depending on the noise level, need to be defined for the input science data. Error maps could provide such an input.

It is obvious that wide-field imagers, as WFI and – to some extent – VIMOS are more easily calibrated than imagers with field of views smaller than 10´×10´ because of a higher probability of finding a sufficient number of evenly distributed reference stars.

The photometric corrections are also obtained by least-square equations and have consequently the same shortcoming as the astrometric solution. Also here the implementation of a regularization would prevent noise amplification could improve the results.

The resampling for the final mosaicing is done by applying the Lanczos3 interpolation kernel with a size of 6×6 pixels. Tests with other interpolation kernels are planned.

The astrometry engine of astrometry.net cannot be directly compared to the astrometry modules of the TERAPIX and ESO/MVM packages. It searches for an astrometric solution using the USNO-B catalogue as default input without requiring any prior WCS, pixel scale and position angle. The program is focused on a fast location of the science image on the sky. Astrometry solutions turned out to be generally less accurate compared to respective TERAPIX and ESO/MVM modules. Especially when image sizes are smaller than 7´×7´ a significant drop in performance (i.e. number of solved fields) is noted.

6. Workflow design for a science data reduction pipeline

The scope of the workflow is to describe a science data reduction pipeline to obtain "science-ready" images from a general list of (instrument calibrated) science images. The proposed workflow was designed mainly on basis of the HAWKI-I data reduction pipeline. This pipeline is, compared to others, modular allowing adapting already existing recipes for the science data reduction pipeline. The development should be sufficiently general to allow also the calibration of science data of the other instruments listed in section 3. A prerequisite of the pipeline is that most of the recipes should run in an automated way without interaction of the user. User interaction is required at the following levels, apart:

- ***Data organizer for the input images:*** as the input list of science images may consist of observations from different periods and/or different observing conditions, the images need to be grouped to reduction blocks of common properties (e.g. observing dates and observing conditions) for the background subtraction module. A sanity check on the completeness and consistency of the relevant FITS keywords should also be performed. The user should have also the possibility for a first order evaluation of the science data to check their quality and consistency within a reduction block (e.g. estimate of the background, homogeneity, etc.). Relevant parameters for the background subtraction will be set.
- ***Evaluation of intermediate stage science data after background subtraction:*** the results of the background subtraction need to be evaluated, possibly repeating the background subtraction with more suitable input parameters. Relevant parameters for the distortion, astrometric and photometric corrections will be set.
- ***Evaluation of the final mosaic image.***

Image visualization is performed using standard tools such as DS9 and Skycat. The display of 1D graphics will rely on the tools of the REFLEX workflow engine. Suitable CPL functions to produce relevant data files need to be written.

The figure gives a schematic overview of the individual (instrument-independent) recipes of the science data reduction pipeline.
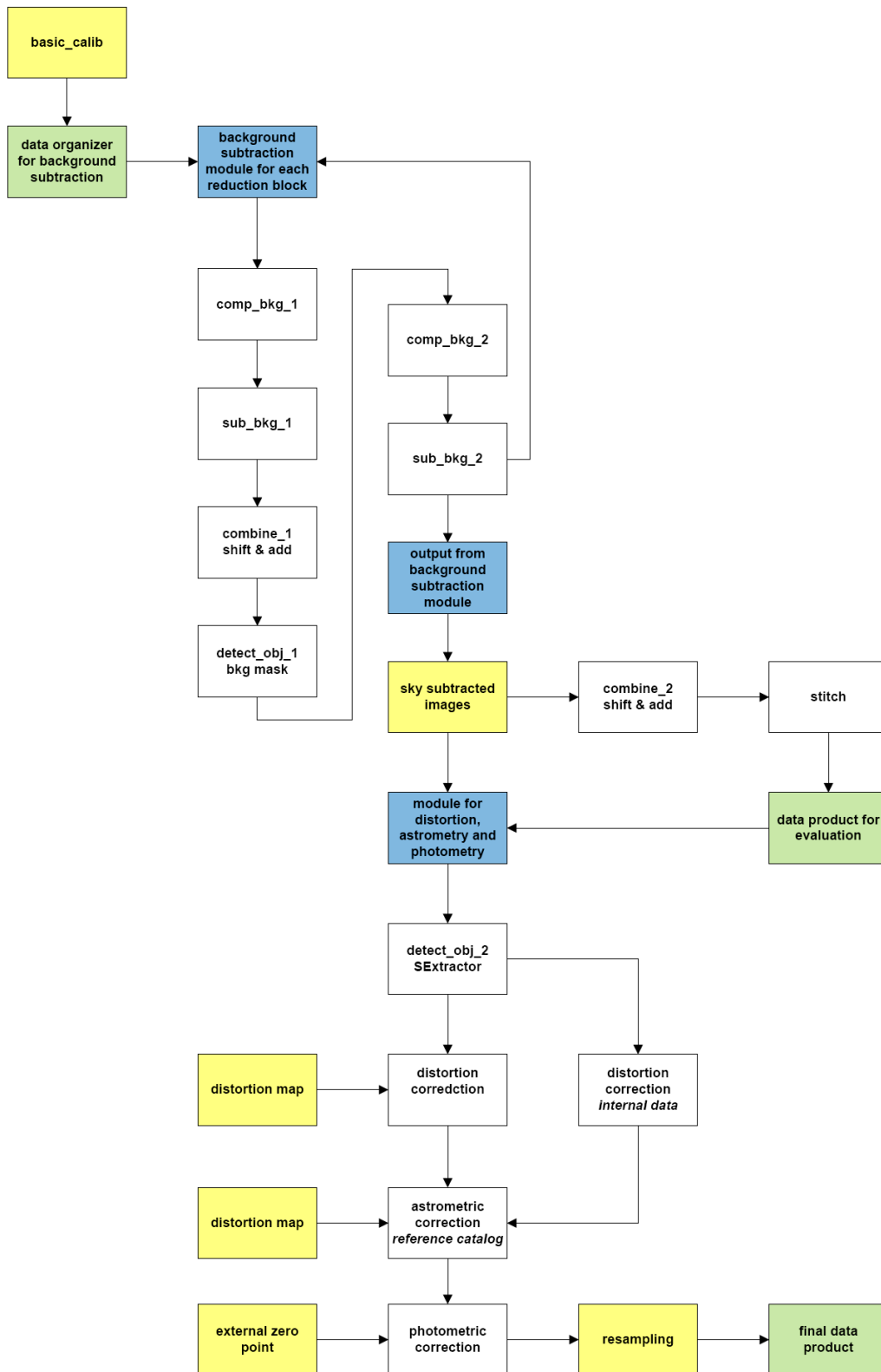

Fig. 1: *Schematic workflow of the science data reduction pipeline*

7. Description of the modules

The description of the modules for the science data reduction pipeline is based upon the present version of the HAWK-I data reduction pipeline, which contain with the exception of the fringe corrections all steps necessary to achieve "science-ready" data. The pipeline consists of the following modules:

- background subtraction;
- correction for distortion, absolute astrometry and photometry;
- resampling and coadding of the images to a final mosaic image.

Recipes for the background subtraction are implemented in the present version of the HAWK-I pipeline. For the remaining modules scripts on the basis of the TERAPIX software is used for the prototype workflow.

Computing performance is an issue for basically all modules. In order to take advantage of multi-core machines task parallelism using OpenMP will be implemented for the recipes.

7.1. Specification of input data

The input data are science frames which have been bias/dark and flat field corrected with the respective error maps. The science data may cover different observing periods and consequently may originate from different observation blocks.

7.2. Data organizer

As the input list of science images and associated error maps may consist of observations from different periods and/or different observing conditions, the images need to be grouped to reduction blocks of common properties (e.g. observing dates and observing conditions) for the background subtraction module.

A sanity check on the completeness and consistency of the relevant FITS keywords should also be performed. The user should have also the possibility for a first order evaluation of the science data to check their quality and consistency within a reduction block (e.g. estimate of the background, homogeneity, etc.).

Relevant parameters for the background subtraction will be set such as the method for background level estimate, the detection threshold and the growing radius for source detection and rejection method for running mean.

The data organizer will be implemented as part of the REFLEX work flow.

7.3. Background subtraction

As outlined in Fig. 1, the sky subtraction module uses essentially the recipes of the HAWK-I pipeline with adaptations described in section 5.6. Each reduction block created by the data organizer is processed individually by the sky subtraction module. The output is a list of background (and fringe corrected) science images with associated error maps.

The option of fringe correction in the case of optical images, by providing a fringe map needs to be implemented. Propagation of errors needs to be implemented.

7.4. Analysis of intermediate stage results

The output of the background subtraction module is analyzed as described in section 6 checking quality of the results. At the same time parameters relevant for the distortion, astrometric and photometric corrections are configured at this step.

7.5. Source extraction for reference data

Source extraction is performed using the SExtractor algorithm constraining the source list by the half-light radius of source and the star/galaxy separation parameter in order to create a list of bona-fide stars suitable for the distortion, astrometric and photometric correction.

7.6. Distortion correction

There are two options for the correction of the distortion:

- a distortion map is provided;
- the calibration is performed internally using a suitable reference image (e.g. high S/N image having sufficient overlap from the same reduction block as described in the ESO/MVM software package (see section 3.2.).

Error propagation needs to be implemented. It will be investigated to include the distortion correction into the background subtraction module in order to provide at the intermediate level science data, where all instrumental signatures are corrected, however with the disadvantage that the resampling needs to be done twice.

7.7. Astrometric correction

The astrometric correction will be applied on basis of the algorithms used in SCAMP using an external astrometric reference catalogue. It is proposed to use the CDSCLIENT tool to access the CDS/Vizier database. It will be investigated how to take advantage of higher resolution catalogues such as UCAC3 compared to the default USNO-B catalogue. Improvements outlined in section 5.6. will be implemented. Propagation of errors will be implemented.

7.8. Photometric correction

Flux correction will be derived on basis of matching source pairs in overlapping image areas as implemented in SCAMP. As alternative the use of a reference image created from the astrometry catalogue with fluxes converted from catalogue magnitudes will be investigated.

Correction of differences of angular resolution caused by seeing effects needs to be implemented. The correction will derived on basis of the SExtractor generated list, using a mean seeing value for each images. Alternatives to that will be investigated.

Improvements outlines in section 5.6. will be implemented. Propagation of errors will be implemented. The option to provide an external photometric zero point will be included.

7.9. Resampling and coadding

Resampling and coadding of the individual science images will be done on basis of the SWarp program yielding a final mosaic image with an associated error map. Alternative interpolation kernels to the Lanczos3 kernel will be investigated.