J. Astron. Space Sci. 20(4), 403–416 (2003)

SYSTEM ANALYSIS OF PIPELINE SOFTWARE - A CASE STUDY OF THE IMAGING SURVEY AT ESO

Young-Soo Kim
Korea Astronomy Observatory, Whaam-dong, Yuseong-gu, Daejeon 305-348, Korea
E-mail: ykim@kao.re.kr

(Received October 29, 2003; Accepted November 13, 2003)

ABSTRACT

There are common features, in both imaging surveys and image processing, between astronomical observations and remote sensing. Handling large amounts of data, in an easy and fast way, has become a common issue. Implementing pipeline software can be a solution to the problem, one which allows the processing of various kinds of data automatically. As a case study, the development of pipeline software for the EIS (European Southern Observatory Imaging Survey) is introduced. The EIS team has been conducting a sky survey to provide candidate targets to the ESO VLTs (Very Large Telescopes) observations. The survey data have been processed in a sequence of five major data corrections and reductions, i.e. preprocessing, flat fielding, photometric and astrometric corrections, source extraction, and coaddition. The processed data are eventually distributed to the users. In order to provide automatic processing of the vast volume of observed data, pipeline software has been developed. Because of the complexity of objects and different characteristic of each process, it was necessary to analyze the whole works of the EIS survey program. The overall tasks of the EIS are identified, and the scheme of the EIS pipeline software is defined. The system structure and the processes are presented, and in-depth flow charts are analyzed. During the analyses, it was revealed that handling the data flow and managing the database are important for the data processing. These analyses may also be applied to many other fields which require image processing.

Keywords: database, pipeline, software, survey, imaging, ESO, EIS

1. INTRODUCTION

Imaging on both space telescopes and ground telescopes has become more popular and important for surveying both celestial objects and ground targets. There are common features between imaging surveys in the fields of astronomy and remote sensing. The techniques of image processing, such as flat-fielding, radiometric (photometric) correction, geometric (astrometric) correction, and superposition, are commonly used in these fields, though their terminologies are different from each other.

As the image sizes become larger and larger, a vast amount of image data is accumulated from observations. Managing and processing large amounts of data is a big issue in both fields. One solution for handling large amounts of data is to apply pipeline software which can automatically process it. A case study of pipeline software was conducted for the ESO Imaging Survey (EIS), by defining the whole structure, and by analyzing the data processing procedure.

The EIS is an ESO initiative to provide astronomical objects for the ESO Very Large Telescope (VLT), by surveying interesting sky areas (Benoist et al. 1999, da Costa et al. 1999, Nonino et

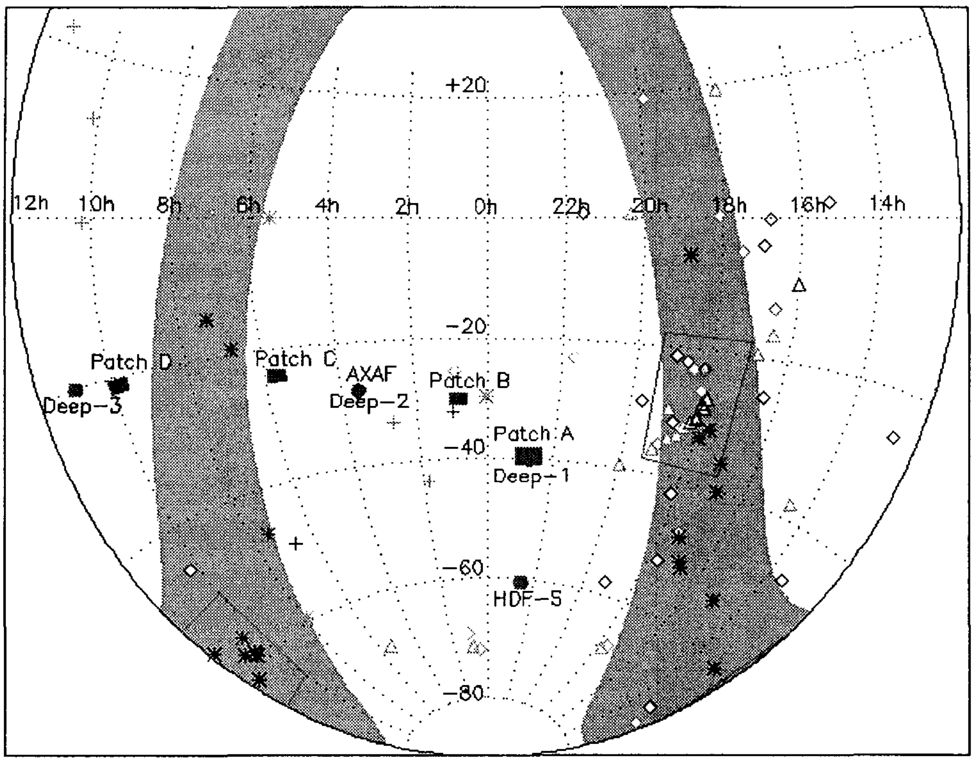


Figure 1. Area map for EIS observations. Rectangles represent the Patch areas, circles are the Deep fields, and the other symbols are targets for the Pre-Flames.

al. 1999, Olsen et al. 1999a,b, Pradoni et al. 1999, Renzini & da Costa 1999, Scodeggio et al. 1999, Zaggia et al. 1999). The main target areas were divided into three categories: Patches, Deep Fields, and Pre-Flames. There are 4 Patch areas which are galaxy-crowded fields (Patches A, B, C, & D). In the Deep Fields, there are 6 areas; 3 Deep fields (Deep-1, 2, & 3), the HDF (Hubble Deep Field)-South, and Chandra. The targets of Pre-Flames are 160 stellar fields of clusters and galaxies, including the LMC and SMC, which would be extracted for follow-up observations with the VLT instrument, FLAMES (Fiber Large Array Multi-Element Spectrograph) (Momany et al. 2001). The selected areas and targets are mapped by myself and shown in Figure 1. From the year 2002, the GALEX field was added to the Deep fields (http://www.eso.org/science/eis/surveys/strategy_GALEX.html). Observations have been conducted with the imager: WFI (Wide Field Imager) on the ESO 2.2m telescope; EMMI (ESO Multi-Mode Instrument), SUSI2 (Superb Seeing Imager-2), and SOFI (Son of Isaac) on the NTT (New Technology Telescope). The details of these imagers are listed in Table 1.

As a result of two dozen nights of observations each year, a large amount of imaging data has been accumulated. Moreover, the various targets and the multiple stages of the data reduction processes add complexity to the handling of the data. Therefore, automatic data manipulation by using pipeline software has been developed, which can handle the data from the preparation for observations, to the observations and data reductions, and then to the final stage of data distribution via Internet. In the next two sections, the overall scheme of the data processing is outlined and the

Table 1. Imagers used for EIS.

Imager	Filters	Size[arc min2]	Telescope	CCD chips	Remarks
WFI	UBVRI	34 x 33	2.2m	4x2 x 2k x 4k	
EMMI	BVI	9.2 x 8.6	NTT	2k x 2k	Red arm
SUSI2	UBVRI	5.5 x 5.5	NTT	2 x 2k x 4k	
SOFI	JHK	4.9 x 4.9	NTT	1k x 1k	

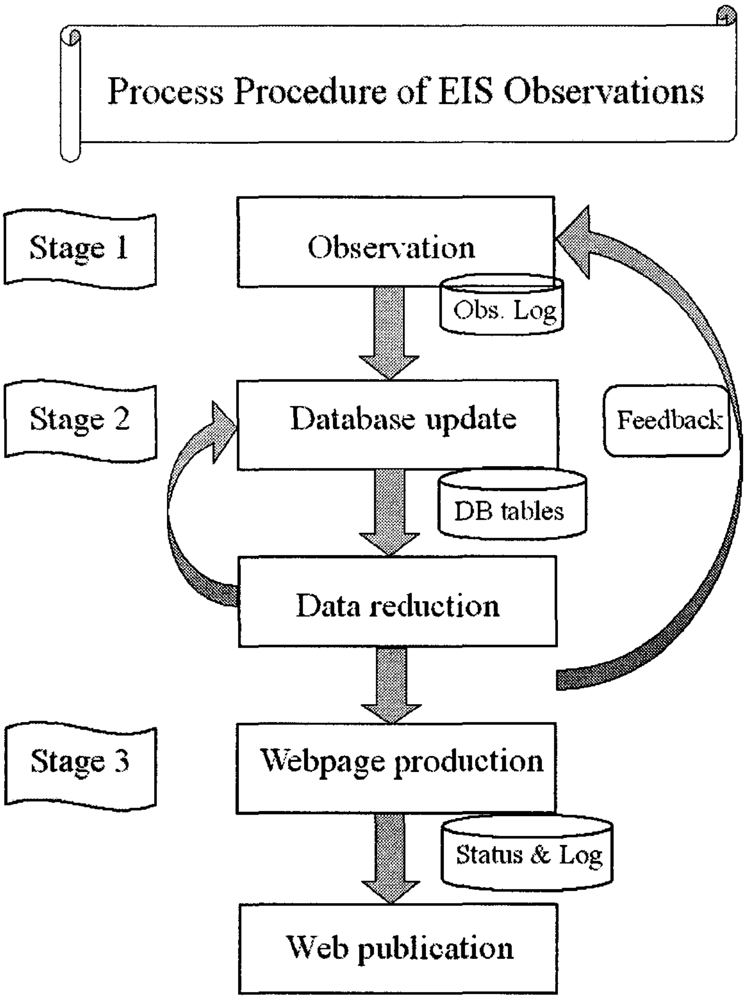


Figure 2. Overall process procedure of the EIS project.

procedures for data handling are analyzed, which have been conducted by myself including all the figures. Conclusion and discussions are followed.

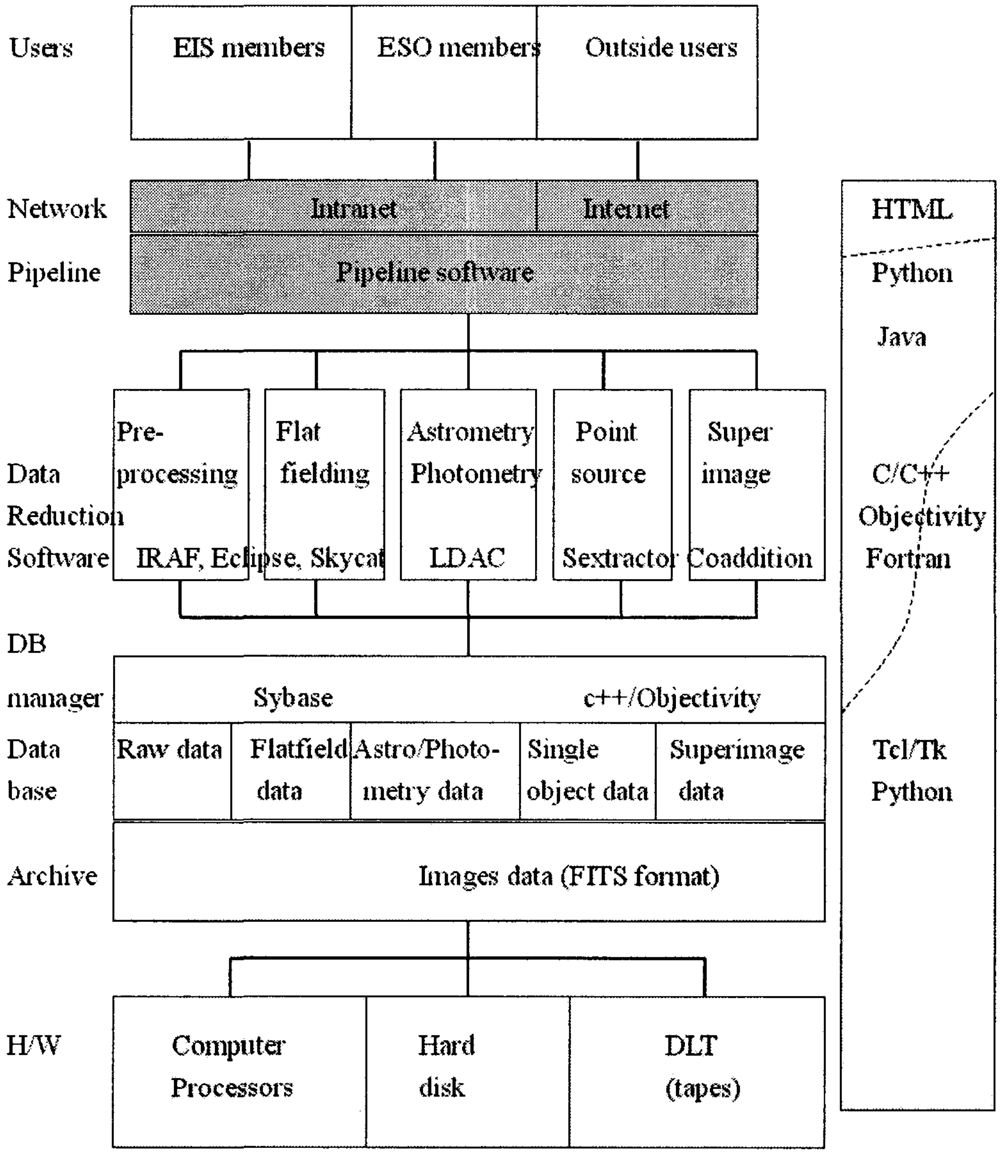


Figure 3. Overall structure of EIS data processing to be implemented with pipeline software.

2. SCHEME AND STRUCTURE OF EIS PIPELINE SOFTWARE

In order to develop a pipeline software, the EIS project are identified and the scheme of the EIS pipeline software is established. The aim and guideline of the pipeline software are defined first. Six aspects are considered in the aim. They are:

- fast processing;
- automatic processing;
- smart methods for handling huge amounts of data;
- user-friendly interface, i.e. easy human interaction;
- easy management of data, files, software, and hardware; and
- feedback within the process, for the purpose of accuracy assessment of the image data and for the advance scheduling of the observations.

The guidelines for the development of the pipeline software were set in the following four goals. The first is to develop the pipeline software as quickly as possible. The second is the maximum utilization of ready-made and currently-used programs and databases. Moreover, current methods of manual data processing should continue to be used, so that newly developed pipeline software can be verified by comparing the results produced by the old and new processes. The third goal is the consideration of the behavior of users. By understanding the philosophy of user behavior, a more sophisticated level of human interaction may be achieved. Implementing Graphical User Interfaces (GUIs) is a good way to achieve this. The last goal is to keep track of the processes via feedback. For each process, a log file should be generated, so that each activity can be traced back for review. Statistics of data size, processing time, etc. should also be assessed by users. The pipeline software would be strengthened if some useful practices of data management are employed, such as providing statistics, diagnostics, quality assessments, and catalogues.

As a next step, the overall procedure of the EIS project was identified. It may be divided into three stages: target preparation and observation; data reduction; and data distribution, as depicted in Figure 2. When an observation is conducted, the observing log is recorded into a database. The data reduction process is performed interactively with the help of the database, by updating the database at each step. After completing the data reduction, the data are distributed to users via the Internet, i.e. the data are published on the EIS web pages (http://www.eso.org/science/eis). New targets for the next observation are prepared, after reviewing the results of the observed data.

The overall structure of the EIS processing was analyzed, and the framework to be constructed by using pipeline software is presented in Figure 3. Though it is drawn hierarchically for easy understanding, actual processing by the pipeline software would be performed interactively across the layers. Users can access the software and data on the intranet or Internet via the pipeline software. The pipeline software manages and handles all the data reduction software, and directly accesses the database and the storage media.

The major tasks are the five processes of data reduction: preprocessing, flat-fielding, astronomy & photometry, point source extraction, and superimaging. These processes are closely related to the database, and each process produces resultant data. The observed raw data are preprocessed and flat-fielded to give the flat-fielded data. Astrometry and photometry data are produced after the processes of astrometry and photometry, and single object data and catalogs result from the point source extraction. Overlapped superimage data are generated after coaddition of images.

Various programming languages are employed for optimal use of each part. The main part of the pipeline software, concatenation between data reduction programs, were coded by Python script (http://www.python.org). Python came out of the C-language, and was designed to have easy programming and handy GUI (Graphical User Interface) designing. It also has good capability to be embedded within other applications, which is important for this project, as the data processing software are coded in many languages, such as C/C++, Fortran, Tcl/Tk, and other scripts.

While the observed and reduced data are stored in an archive, the key content of the data are indexed in a database and managed by Sybase. Currently, the database tables based on Sybase are accessed by long command-lines. In order to handle the database in a user-friendly way and to manage it consistently, the Sybase commands will be accessed by GUI features. This will be presented in a separate paper. The resultant catalogs are interfaced by Objectivity, and the network interface has been programmed by Java with HTML. All these processes are performed on an Unix platform.

3. PROCESSES OF EIS SOFTWARE

The key parts of the pipeline software, the processes of data reduction and data flow, have been investigated. As is shown in Figure 3, various software are employed for the data reductions. Some of them are in-house software for astronomical data reduction, while others are adopted from other organizations and applications. Preprocessing and flat fielding are performed mainly by using IRAF, with complementary usage of Eclipse and Skycat. LDAC, which performs astrometric (geographic) correction and photometric (radiometric) correction of images, is adopted from Leiden Observatory (http://www.strw.leidenuniv.nl/denis/Bluebook/section2.7.2.html). The source (target) extraction software, SExtractor, was employed from Paris Observatory (Bertin 1998). Coaddition (overlap) was developed from the algorithm for the Hubble Space Telescope (Hook & Fruchter 1997).

The processes of preprocessing, flat-fielding, astrometry, and photometry are analyzed in depth, to define the dataflow precisely. As SExtractor and Coaddition are off-the-shelf software, the handling of the data input and output may be adequately controlled by the pipeline software. The flowchart in Figure 4 shows preprocessing and flat-fielding, which are performed in 6 consecutive stages - ingestion, master bias, master dome, master skyflat, master thermal bias, and science frame correction. Though the input and output data are closely related to each other, they are presented separately in this figure. The input data are shown on the left-hand side of the main flow, and output data shown on the right, which makes the figure simple and easy to understand. At first, the headers of the observed frames are read into a mainframe computer (ingestion). Then, frame data are loaded one after another, and processed to make the master bias, dome, skyflat, and thermal bias (in case of IR image frames). Finally, the science frames are loaded, and then corrected with the bias, domeflat, skyflat, and fringe image frames. The corrected science frames are handed over to the astro-photometric process, after constructing the weight maps and updating the reduction tables.

Figure 5 presents the flowcharts of the astrometry and photometry, which consists of 4 stages - dithering, dither pre-coadd, dither coadd, and post coadd processes. On the left side of the figure, the names of the processes (subroutines) are described. By this process, each input frame is mapped to a flux-preserving conic which is on a projected equal-area grid, so that distortions in the area and in the shapes of objects can be minimized. The flux of each pixel in the input frame is redistributed over a superimage, and coadded on a pixel-by-pixel basis. In the first stage, the dithering process, the position of each object on an image frame is identified and corrected (astrometry). In case of WFI and SUSI2 imagers, they are not mounted on a single chip, but on sets of mosaic chips, as shown in Table 1. The mosaic images are processed one after another at first, treating each chip image as an individual frame. In the second stage, instrumental magnitudes are derived from a pair of overlapping frames in the same area, and then configuration files for astrometric coaddition are prepared in the third stage. In the last stage, a superimage is generated by deriving the instrumental magnitudes, and combining each frame; and a catalog of the detected objects in the imaged area is produced after astrometric and photometric corrections.

It is found that data are frequently accessed to or from storages and in updating. These data are reused and updated in other places. Therefore, data should be handled carefully with consideration, so that it may be managed in a uniform and consistent way, allowing the processes to flow smoothly

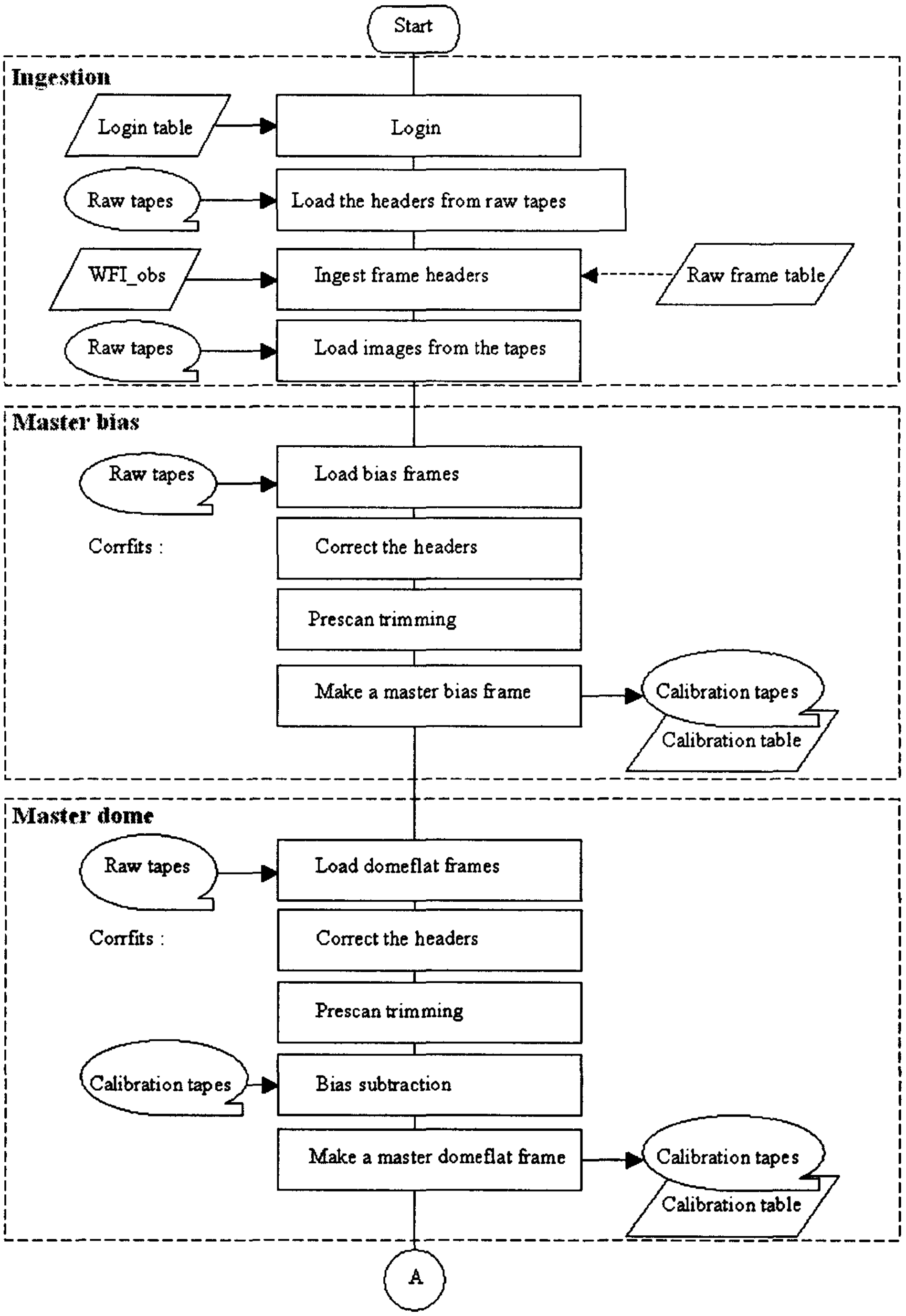


Figure 4. System flowchart of preprocessing and flat fielding.

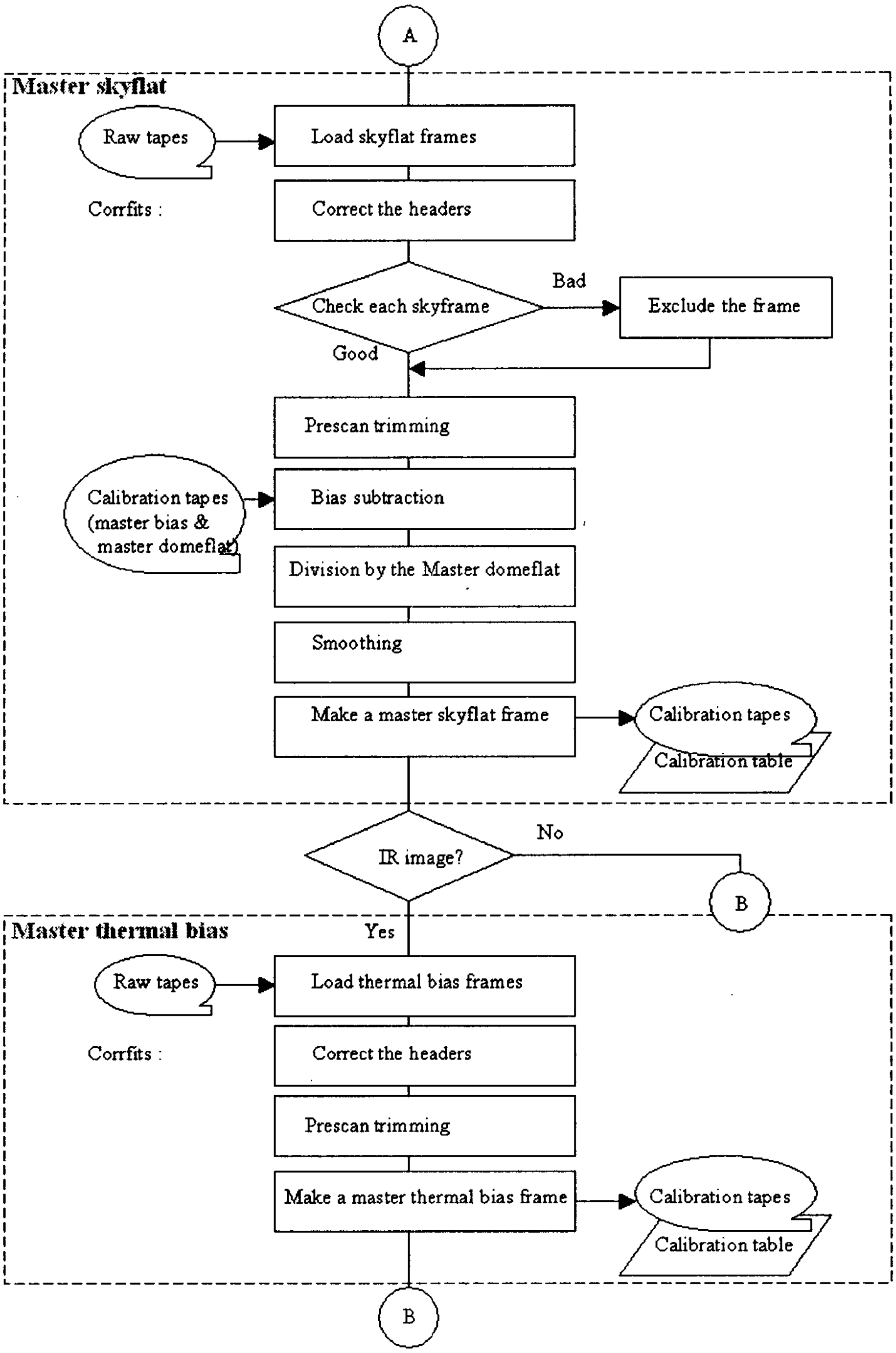


Figure 4. (Continued)

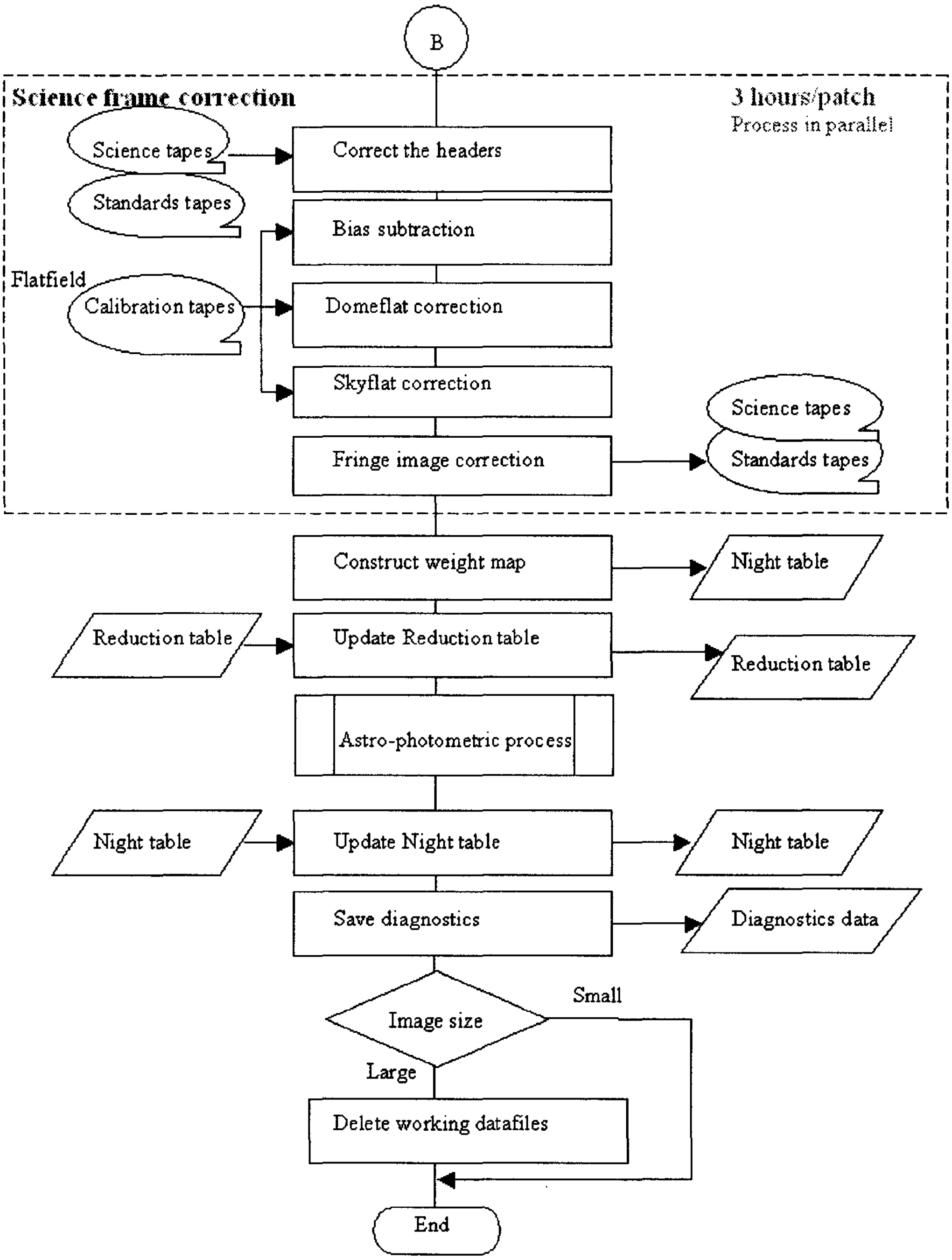


Figure 4. (Continued)

1. **Dithering process**: Actual astrometric calibration.
Primitive photometric calibration.
Script: eis_do_dither_pipeline

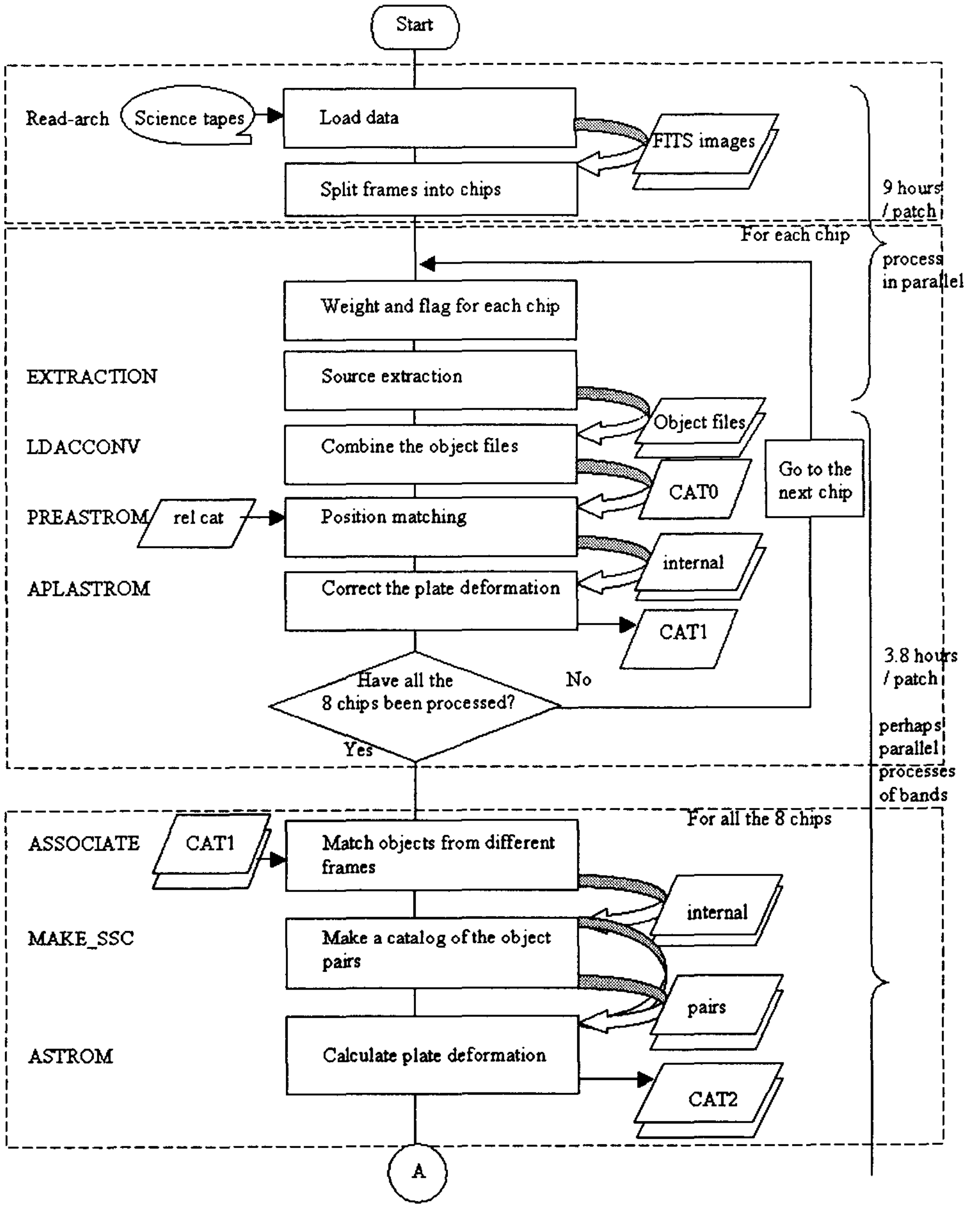
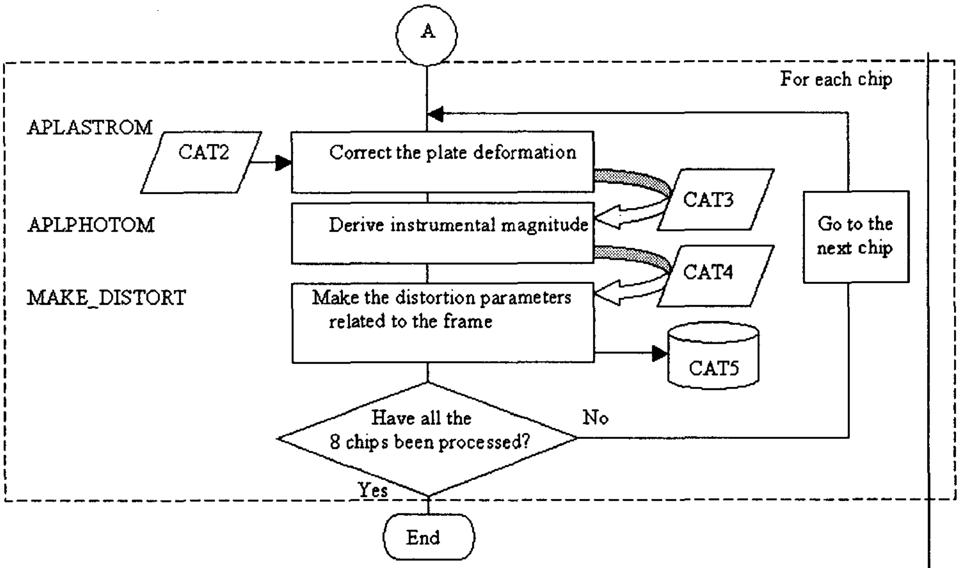


Figure 5. Flowchart of the astrometry and photometry.



2. **Dither Pre-Coadd processing** : Actual relative photometric calibration.
Prepare the FITS image files for coaddition.

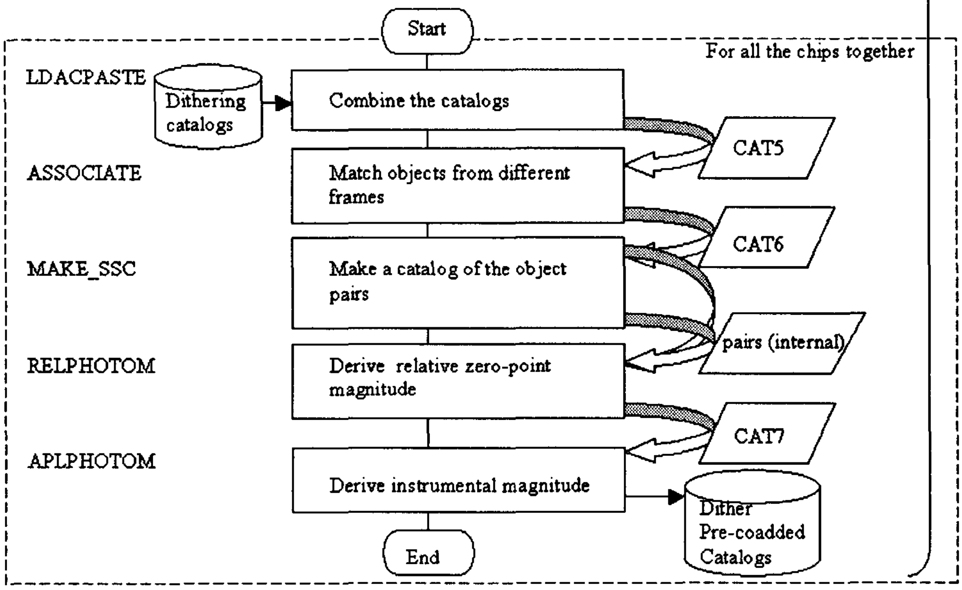
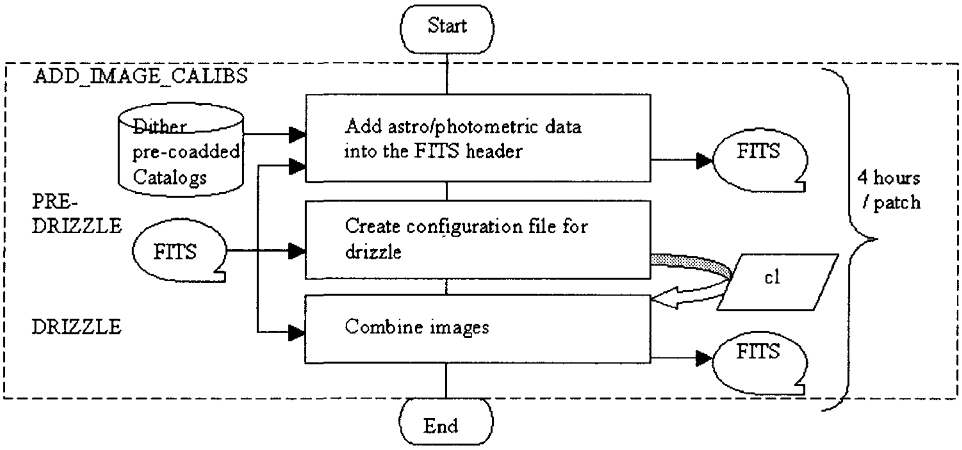


Figure 5. (Continued)

3. **Dither Co-add processing** :

Background subtracted images can be created in parallel



4. **Post Co-add processing** :

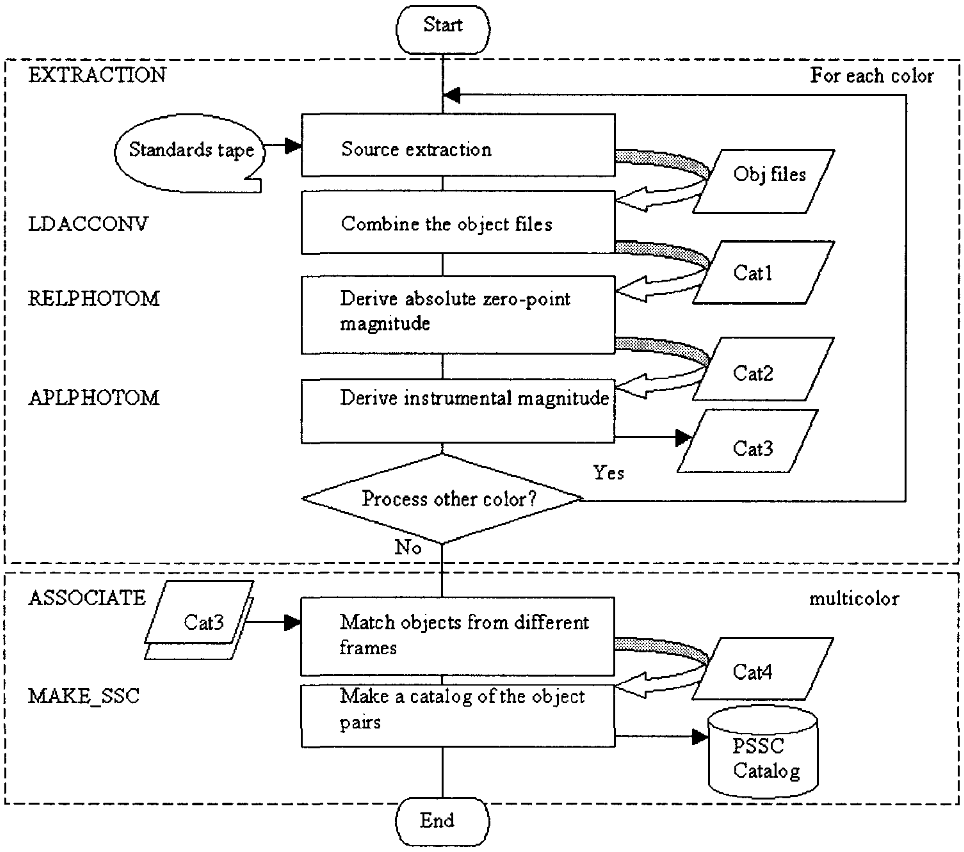


Figure 5. (Continued)

and quickly without interruption.

4. CONCLUSION AND DISCUSSIONS

Imaging surveys are popular in ground-based observations, space observations, and in remote sensing. There are common features in the data processing amongst these applications. Handling large amounts of data is also a common issue, which may be solved by implementing pipeline software. As a case study, the development of the EIS pipeline software has been analyzed and assessed. The EIS observation and data processing must handle various target areas, with several imagers. The scheme for developing the EIS pipeline software was defined, and the overall structure of the EIS works was identified. The flow of the EIS process was traced, from observations of the targets to the distributions to the user community via the Internet. The process of data reduction was investigated, and flowcharts were presented. From the analysis, it was found that data flow and data management are important parts of the whole project. As a result, a customized tool for data management has been developed, which will be presented elsewhere soon. The analyzed flowcharts may be useful in other fields concerned with the massive data processing of images, as the basic principles of the image processing are more or less the same.

ACKNOWLEDGEMENTS: I acknowledge Andre B. Fletcher (KAO) for editing this paper.

REFERENCES

Benoist, C., da Costa, L., Olsen, L. F., Deul, E., Erben, T., Guarnieri, M. D., Hook, R., Nonino, M., Pradoni, I., Scodeggio, M., Slijkhuis, R., Wicenec, A., & Zaggia, S. 1999, A&A, 346, 58

Bertin, E. 1998, SExtractor 2.0, User's Manual (Paris: Paris Observatory)

da Costa, L., Arnouts, S., Benoist, C., Deul, E., Hook, R., Kim, Y.-S., Nonino, M., Pancino, E., Rengelink, R., Slijkhuis, R., Wicenec, A., & Zaggia, S. 1999, The Messenger, 98, 36

Hook, R. N., & Fruchter, A. S. 1997, ASP Conference Series, 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. E. Payne (San Francisco: ASP), p.147

Momany, Y., Vandame, B., Zaggia, S., Mignani, R. P., da Costa, L., Arnouts, S., Groenewegen, M. A. T., Hatziminaoglou, E., Madejsky, R., Rite, C., Schirmer, M., & Slijkjuis, R. 2001, A&A, 379, 436

Nonino, M., Bertin, E., da Costa, L., Deul, E., Erben, T., Olsen, L., Pradoni, I., Scodeggio, M., Wicenec, A., Wichmann, R., Benoist, C., Freudling, W., Guarnieri, M.D., Hook, I., Hook, R., Mendez, R., Savaglio, S., Silva, D., & Slijkhuis, R. 1999, A&AS, 137, 51

Olsen, L.F., Scodeggio, M., da Costa, L., Benoist, C., Bertin, E., Deul, E., Erben, T., Guarnieri, M.D., Hook, R., Nonino, M., Pradoni, I., Slijkhuis, R., Wicenec, A., & Wichmann, R. 1999a, A&A, 345, 681

Olsen, L. F., Scodeggio, M., da Costa, L., Slijkhuis, R., Benoist, C., Bertin, E., Deul, E., Erben, T., Guarnieri, M. D., Hook, R., Nonino, M., Pradoni, I., Wicenec, A., & Zaggia, S. 1999b, A&A, 345, 363

Pradoni, I., Wichmann, R., da Costa, L., Benoist, C., Mendez, R., Nonino, M., Olsen, L. F., Wicenec, A., Zaggia, S., Bertin, E., Deul, E., Erben, T., Guarnieri, M. D., Hook, I., Hook, R., Scodeggio, M., & Slijkhuis, R. 1999, A&A, 345, 448

Renzini, A., & da Costa, L. 1999, The Messenger, 98, 33

Scodeggio, M., Olsen, L. F., da Costa, L., Slijkhuis, R., Benoist, C., Deul, E., Erben, T., Hook, R., Nonino, M., Wicenec, A., & Zaggia, S. 1999, A&AS, 137, 83

Zaggia, S., Hook, I., Mendez, R., da Costa, L., Olsen, L. F., Nonino, M., Wicenec, A., Benoist, C., Deul, E., Erben, T., Guarnieri, M. D., Hook, R., Pradoni, I., Scodeggio, M., Slijkhuis, R., & Wichmann, R. 1999, A&AS, 137, 75