Astronomical Data Analysis Software and Systems XIV **P2.26**
ASP Conference Series, Vol. XXX, 2005
P. L. Shopbell, M. C. Britton, and R. Ebert, eds.

VISTA Data Flow System: Pipeline Processing for WFCAM and VISTA

James R. Lewis, M.J. Irwin, S.T. Hodgkin, Peter S. Bunclark, D.W. Evans

Cambridge Astronomy Survey Unit, Institute of Astronomy, Madingley Road, Cambridge CB3 7XT, UK

R.G. McMahon

Institute of Astronomy, Madingley Road, Cambridge CB3 7XT, UK

Abstract. The UKIRT Wide Field Camera (WFCAM) on Mauna Kea and the VISTA IR mosaic camera at ESO, Paranal, with respectively 4 Rockwell 2kx2k and 16 Raytheon 2kx2k IR arrays on 4m-class telescopes, represent an enormous leap in deep IR survey capability. However with an expected data rate of an image of the sky every 5-30s and combined nightly data-rates of typically 1TB, automated pipeline processing and data management requirements are paramount.

Pipeline processing of IR data is far more technically challenging than for optical data. IR detectors are inherently more unstable, while the sky emission is over 100 times brighter than most objects of interest, and varies in a complex spatial and temporal manner.

The pipelines are designed around a selectable modular scheme, driven by processing recipes for maximum flexibility. Our general philosophy is that all fundamental data products are in multi-extension FITS files with headers describing the data taking protocols in sufficient detail to trigger the appropriate pipeline processing components. All derived information, DQC, photometric and astrometric calibration and processing details are also incorporated into the FITS headers. Generated catalogues are stored in FITS binary tables. The headers provide a basis for ingest into databases for archiving, real time monitoring of survey progress and survey planning.

To reduce the data storage I/O overheads and transport requirements, we intend to use, as much as possible, the lossless Rice tile compression scheme as used transparently, for example, in CFITSIO. For this type of data (32 bit integer) the algorithm typically gives a factor of 3-4 compression

1. WFCAM Pipelines

There will be essentially three separate pipelines for WFCAM. A summit pipeline will assess data quality on-the-fly, provide feedback to survey planning and

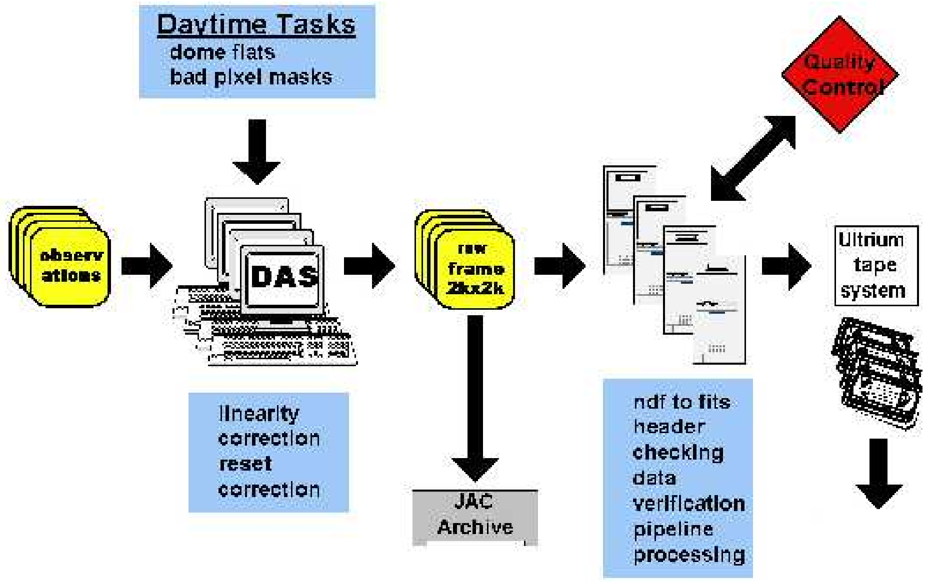

Figure 1. An overview of the summit processing setup for WFCAM. Each detector feeds a separate data acquisition PC, which then forwards the images in parallel to the summit data processing PCs. Data transport to the UK for science processing is via Ultrium LTO-I tapes

progress tools and provide first pass science products. Figure 1. schematically shows the corresponding data flow operations at the summit on Mauna Kea.

A standard calibration pipeline to do full instrumental signature removal, astrometric and photometric calibration and catalogue generation will be run in Cambridge. Data compression will be absolutely vital for these projects as, with compression, we anticipate the yearly storage requirements for WFCAM to be ≈10 Tbytes per year and for VISTA ≈50 Tbytes per year. Transfer of processed data from Cambridge to the WFCAM and VISTA Science Archives in Edinburgh will be via the Internet. The Cambridge pipeline setup is illustrated in Figure 1..

A third pipeline to generate advanced data products (e.g. deep stacking) will be run in Cambridge and Edinburgh.

2. Image Processing

The first stage of image processing will be done in the DAS before the data is written out. This includes combining individual reads into a single exposure and removing the reset (bias) frame. By and large data obtained from NIR arrays are strongly non-linear. This effect will be removed in the DAS for WFCAM but not for VISTA.

The main reduction steps and their approximate location in the pipeline are:

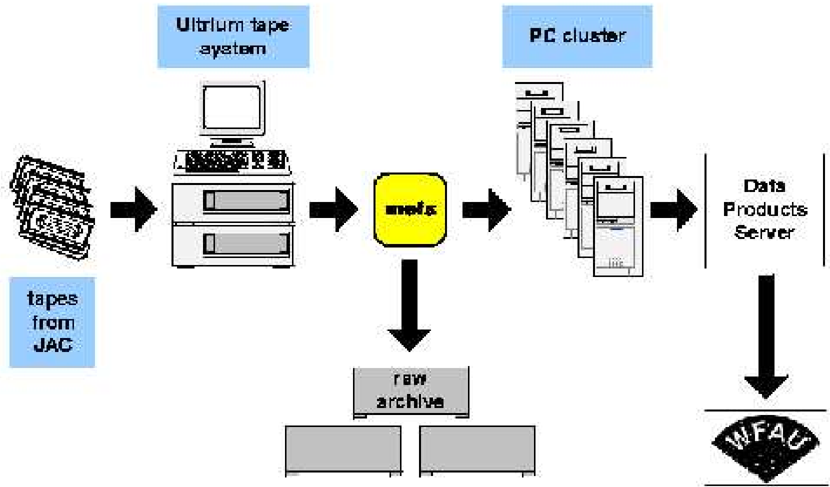

Figure 2. A schematic of the data processing setup for WFCAM in Cambridge. Incoming tapes are ingested, the data verified, converted to MEF format and fed to the processing cluster. A raw data archive is held on-line, whilst processed products are automatically transferred via the internet to the Wide Field Astronomy Unit (WFAU) in Edinburgh.

- Linearity correction: if it has not been done in the DAS.
- Dark and reset anomaly correction: removes two additive effects caused by dark current and residual bias structure
- Flat field correction: to remove pixel-pixel sensitivity differences, gain differences between data channels and between detectors.
- Confidence map generation: we define a confidence map as a normalised inverse variance weight map denoting the 'confidence' associated with the flux value in each pixel.
- Defringing: If the fringe spatial pattern is stable, it is possible to decouple sky correction from fringe subtraction.
- Sky subtraction: This can be done using the dither sequences themselves or with observations of offset sky regions.
- Image Persistence and Detector Crosstalk: These are electronic effects which can be modelled and removed.
- Interleave: To recover lost resolution when observations are undersampled (as will be the case with WFCAM and, in best seeing conditions, VISTA) observations may be 'microstepped' into a series of exposures with sub-pixel shifts. The resulting image pixels are interwoven to form a single output image.

- Dithering: Shift and average all the exposures in a dither sequence into a single output image.
- Catalogue generation: Objects are extracted from the images and catalogued to FITS binary tables. Standard object descriptors include assorted aperture flux measures, intensity weighted centroid estimates, and shape information, such as intensity-weighted 2nd moments to encode the equivalent elliptical Gaussian light distribution.
- Astrometric calibration: Catalogued objects are matched to astrometric standards to define a WCS for each image/catalogue.
- Photometric zeropoint: for the purposes of quality control a photometric zeropoint will be determined for each observation by direct comparison of instrumental magnitudes with 2MASS. The complete night's observations will allow for a more accurate value to be determined later.

3. Trial Commissioning

We have tested the WFCAM and VISTA pipeline modules on infrared data from several existing cameras. This includes the NIR ISAAC camera on the VLT. As a test of our pipeline and commissioning procedures, we downloaded the raw data for the 'Faint InfraRed Extragalactic Survey' (FIRES) from the ESO archive. These data have already been reduced and presented by Labbé etal (2003) and therefore make an excellent benchmark against which to test our reduction procedures. A result of this testing is shown in Figure 3..

References

Labbé, I., *et al*, 2003, AJ, 125, 1107

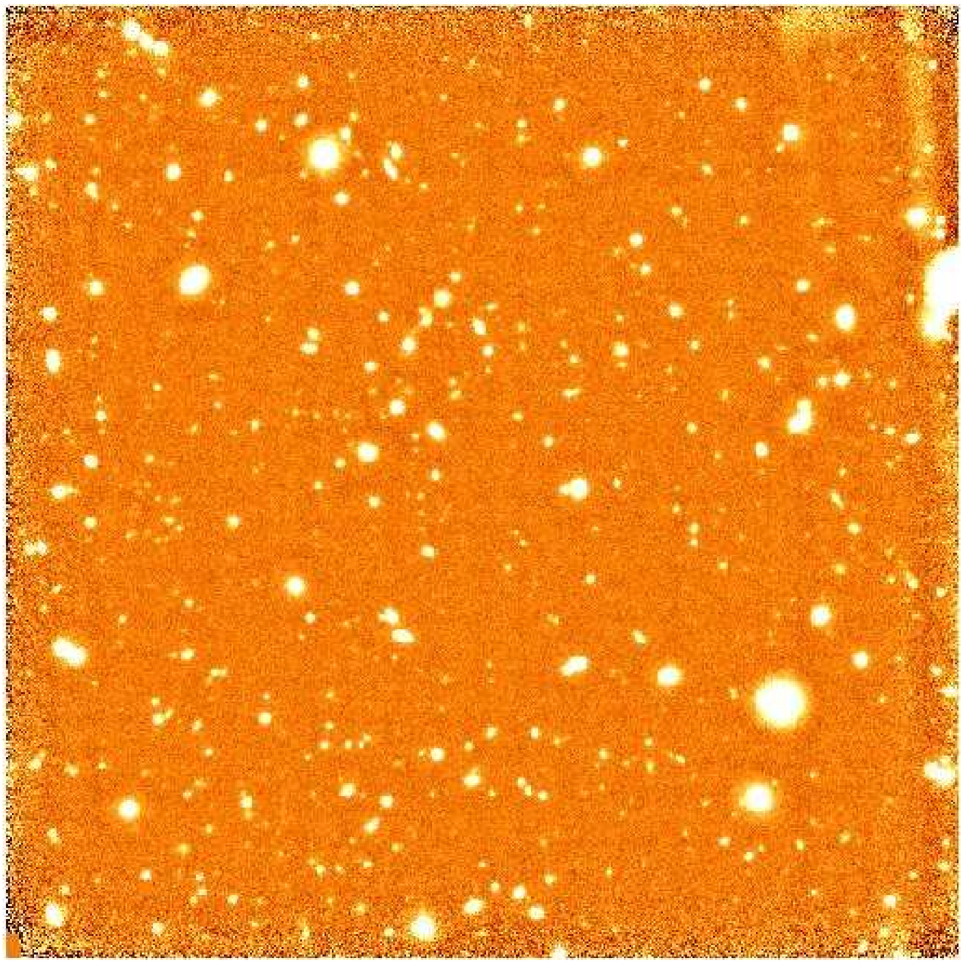

Figure 3. The result of processing and stacking 1760 individual 120 K_s band images (44 observing blocks)