National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh

DEPARTMENT OF INFORMATION TECHNOLOGY

technical report NUIG-IT-261104

A pipeline for automatically processing and analysing archival images from multiple instruments

Seathrún Ó Tuairisg (NUI, Galway)
Aaron Golden (NUI, Galway)
Raymond Butler (NUI, Galway)
Andrew Shearer (NUI, Galway)
Bruno Voisin (NUI, Galway)

itreports@nuigalway.ie http://www.it.nuigalway.ie/

A pipeline for automatically processing and analysing archival images from multiple instruments

Seathrún Ó Tuairisg[a], Aaron Golden[a], Raymond Butler[b], Andrew Shearer[a], Bruno Voisin[a]

[a]Dept of Information Technology, National University of Ireland, Galway, Ireland [b]Dept of Experimental Physics, National University of Ireland, Galway, Ireland

ABSTRACT

To take advantage of the recent upsurge in astrophysical research applications of grid technologies coupled with the increase in temporal and spatial coverage afforded to us by dedicated all-sky surveys and on-line data archives, we have developed an automated image reduction and analysis pipeline for a number of different astronomical instruments. The primary science goal of the project is in the study of long-term optical variability of brown dwarfs, although it can be tailored to suit many varied astrophysical phenomena. The pipeline complements Querator,[1] the custom search-engine which accesses the astronomical image archives based at the ST-ECF/ESO centre in Garching, Germany. To increase our dataset we complement the reduction and analysis of WFI (Wide Field Imager, mounted on the 2.2-m MPG/ESO telescope at La Silla) archival images with the analysis of pre-reduced co-spatial HST/WFPC2 images and near infrared images from the DENIS archive. Our pipeline includes CCD-image reduction, registration, astrometry, photometry, and image matching stages. We present sample results of all stages of the pipeline and describe how we overcome such problems as missing or incorrect image meta-data, interference fringing, poor image calibration files etc. The pipeline was written using tasks contained in the IRAF environment, linked together with Unix Shell Scripts and Perl, and the image reduction and analysis is performed using a 40-processor Origin SGI 3800 based at NUI, Galway.

Keywords: Data reduction, data archives, image processing, optical astronomy

1. INTRODUCTION

Recent technological advances in observational astronomy have revolutionised the manner and depth at which astronomers image the night sky, and improvements in electronic image acquisition and reduction have increased the information potential which can be gleaned from such exposures. Systematic sky surveys have seen the rise of image archives stored at multiple sites and which can be accessed by astronomers across the globe via high-speed networks. Observed through pre-arranged strips of the sky through different broad-band wavelength filters, the accumulated data is quickly reduced and stored, and can be trawled through in search of many different astronomical phenomena. The concept of a virtual observatory to exploit this increase in the spatial and temporal coverage of the sky has gained currency in recent years and is now being independently implemented by several different groups.[2–5] Indeed, studies of several different astrophysical phenomena, including the characterisation of various types of variable stars, can only benefit from an increased source of image data. Here we discuss the development of a fully automated pipeline to reduce and analyse a stream of archival image data, observed using several different instruments, both ground- and space-based data. Also included are options to accommodate non-archival generic CCD data obtained from ground-based telescopes. Automation of such an image-processing pipeline is not only necessary, because of the sheer volume of data streaming in each night; it is also beneficial, as it ensures that this consistency is maintained in the data reduction.

Further author information: (Send correspondence to S.Ó T.)
S.Ó T.: E-mail: seathrun@itc.nuigalway.ie, Telephone: +353 (0)91 524411 3180
A.G.: E-mail: agolden@itc.nuigaway.ie, Telephone: +353 (0)91 524411
R.B.: E-mail: rbutler@nuigalway.ie, Telephone: +353 (0)91 524411
A.S.: E-mail: shearer@itc.nuigalway.ie, Telephone: +353 (0)91 524411
B.V.: E-mail: bvoisin@itc.nuigalway.ie, Telephone: +353 (0)91 524411

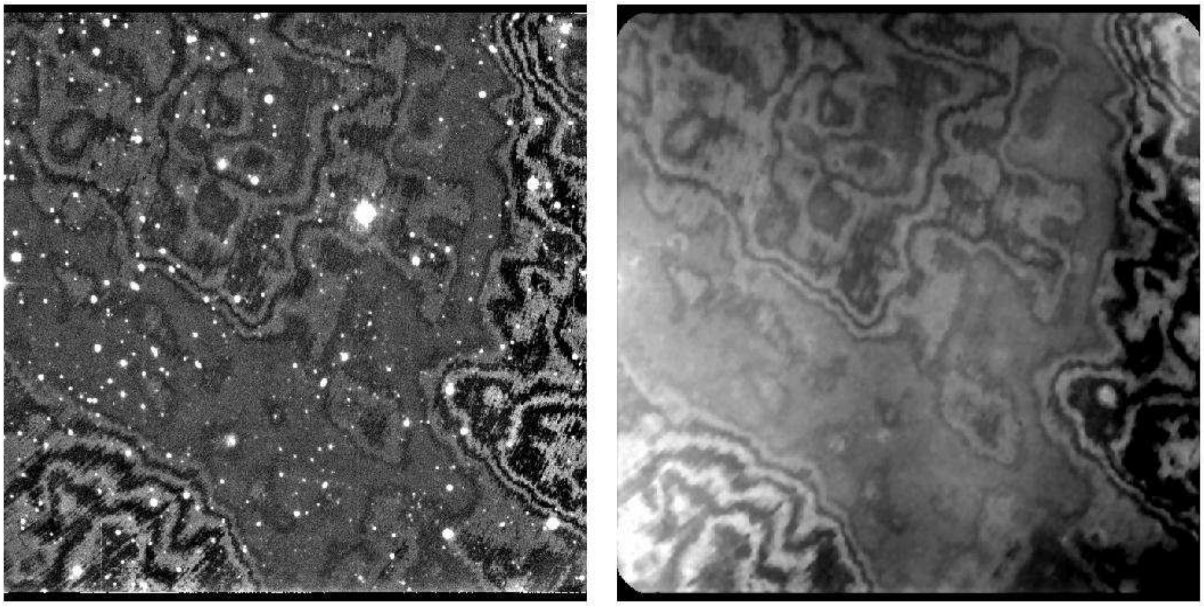

Figure 1. The fringe template image (right) which the pipeline uses to correct the interference fringing present in I-band CCD BFOSC data (left) obtained from the observatory at Loiano, Italy. Unlike this one, which we created from a series of images offset from one another, fringe correction templates for larger, more widely used instruments, are usually available from archival sites.

2. THE INSTRUMENTS

The reduction and analysis pipeline is currently able to handle archival data-sets from three different astronomical instruments. The Hubble Space Telescope's Wide Field Planetary Camera 2 (WFPC2) delivers three 800×800 pixel images with an plate scale of 0.1" (arcsec) per pixel and one 800×800 pixel image with an plate scale of 0.046" per pixel, together comprising an L-shaped mosaic. Its image characteristics are well known, and include a slight charge transfer efficiency problem which is correctable. Its geometric distortion is also quite well-mapped, which allows the pipeline to use WFPC2 images, where available, as astrometric templates when working with co-spatial image-sets from other instruments. The Wide Field Imager (WFI) is a focal reducer-type camera which is permanently mounted at the Cassegrain focus of the 2.2-m MPG/ESO telescope at La Silla (Chile). Its eight 2000×4000 pixel CCD mosaic affords a field of view of 34×33 arcminutes and a plate scale of 0.238" pixels. The instrument's wide field-of-view compensates for the occasionally relatively sparse temporal coverage for a given object. DENIS (**DE**ep **N**ear **I**nfrared **S**urvey) is a project to survey the southern sky simultaneously in three wavelength bands (Gunn-i, 0.82 μm; J, 1.25 μm; and Ks, 2.15 μm) with limiting magnitudes 18.5, 16.5 and 14.0, respectively. The observations are performed with the 1m-ESO telescope, also at La Silla. The survey is carried out by observing strips of 30° in Declination and 12 arcminutes in Right Ascension with an overlap of 2 arcminutes between consecutive strips. The archive also provides a point-source catalogue, with the position of a general extracted point source being delivered with an accuracy better than 1 arcseconds and its magnitude to better than 0.1 magnitudes.

Also incorporated into the pipeline is the ability to reduce generic CCD data from other instruments - including BFOSC* CCD data from the 1.52m telescope at Loiano, Bologna. Thus the pipeline has been used to complement our own observations of brown dwarf species with archival (both targeted and serendipitous) observations of the same object.

---

[*] Bologna Faint Object Spectrograph and Camera

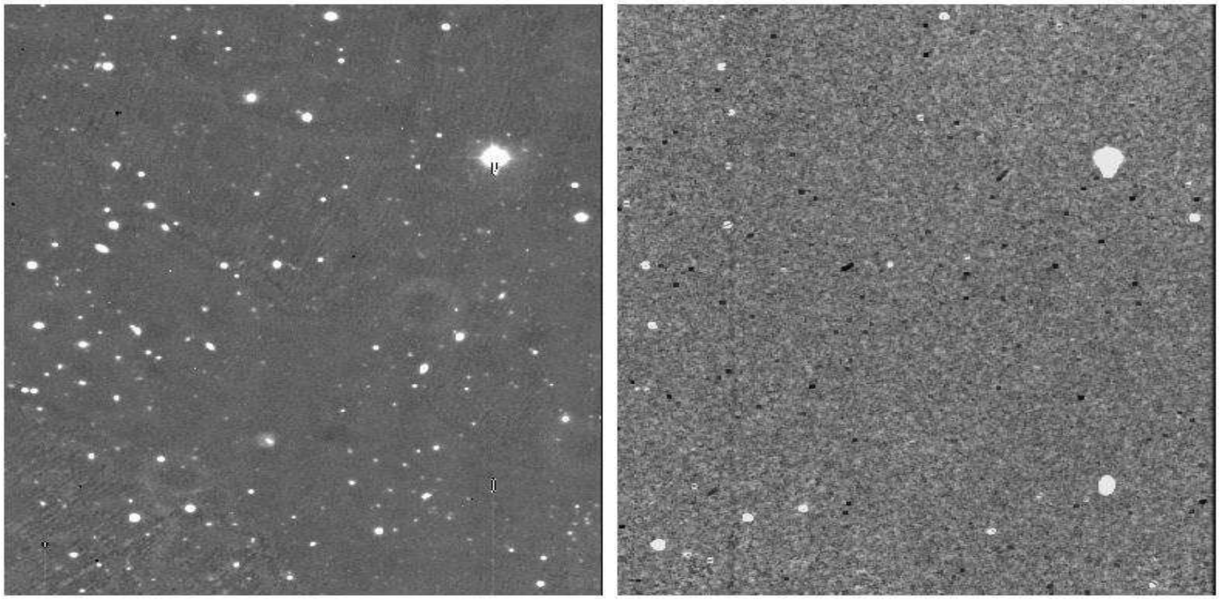

Figure 2. Results from the image matching stage of the analysis pipeline when applied to a series of images of the brown dwarf 2MASS 0036+18, which is the bright central star of the group of three in a horizontal line in the centre of the image on the left. The difference image, as explained in the text, is shown on the right.

3. PIPELINE STRUCTURE AND OVERVIEW

The pipeline was constructed using the tasks and scripting language contained in the IRAF (Image Reduction and Analysis Facility) package. These were complemented by Unix shell scripts (invoked directly from within the IRAF environment). For more complicated or unorthodox tasks FORTRAN code can be linked to the IRAF environment. Access to external catalogues can be provided using Perl. Various external applications (for example image matching/subtraction algorithms[6,7] were also be incorporated into the pipeline. We have chosen this form of pipeline implementation (performed on an 40-processor SGI Origin 3800) rather than coding directly from first principles because of the obvious gains in development speed and flexibility. Although not the most computationally efficient solution, computational speed is not a bottleneck for our purposes. For simplicity, the pipeline is currently non-interactive, and any specific reduction processes required must be flagged as such in the code.

4. PRELIMINARY DATA REDUCTION

The preliminary reduction steps of the pipeline are greatly simplified by the pre-reduced state of the archival images, which are usually bias-subtracted, dark-subtracted, flat-fielded and, in some cases, cosmic-ray corrected. The quality of the pipeline results is crucially dependent on ensuring consistent meta-data (stored as keywords in the FITS header) between the different images and image-sets. These keywords describe filter information, data acquisition modes, World Coordinate System (WCS) for astrometric purposes and any number of different instrumental parameters. Updating missing or incorrect keywords, where necessary, is the first step in the pipeline. Often meta-data from one image-set (for example the well-described HST/WFPC2 image headers) can be 'exported' to other data-sets, taken using smaller telescopes, and whose image headers are not as well-described. For example by correctly utilising the registration stage of the pipeline (outlined below) to align data from other instruments to, for example WFPC2 images, the necessary WCS information is automatically included in the finished products.

Since most archival data is usually pre-reduced either upon or during retrieval, the pipeline checks the presence of the necessary header keyword to confirm this fact. With non-archival data (data obtained ourselves on-site)

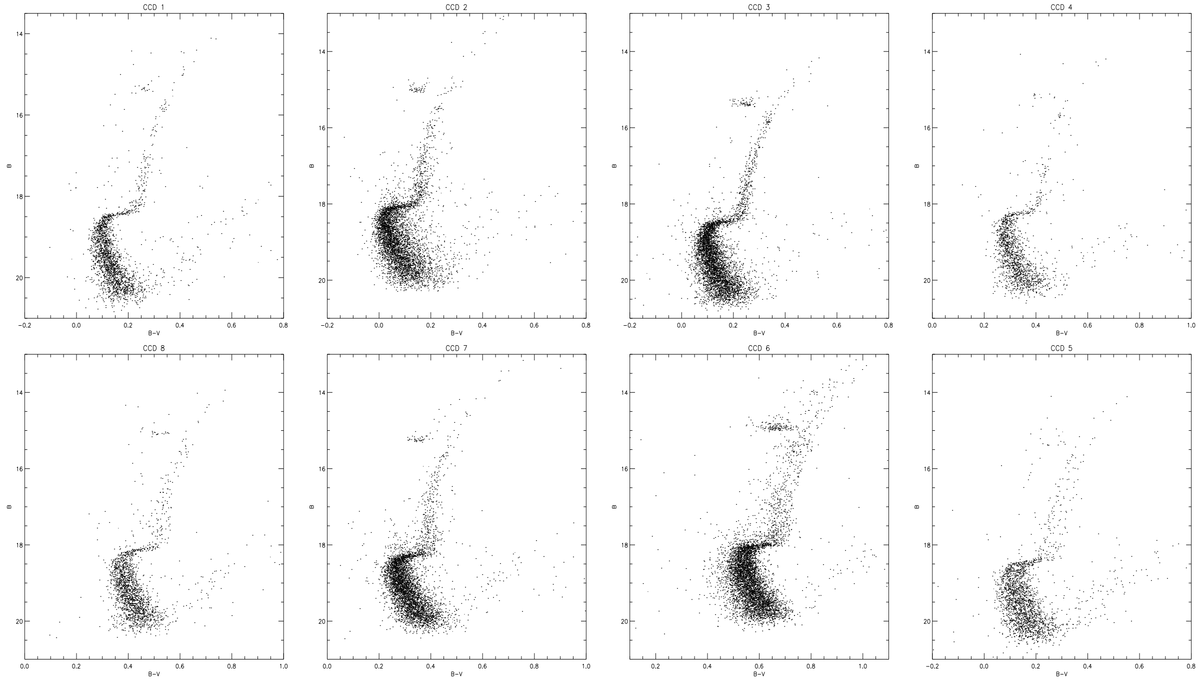

Figure 3. Colour-magnitude diagrams (CMD) obtained from the automatic photometric analysis of WFI 47 Tuc images. Each CMD represents one chip of the 4×2 detector mosaic.

the usual CCD reduction processes are carried out. This entails optimally combining calibration frames such that the time delay between their acquisition is minimised. Images should ideally be reduced using calibration frames (with identical instrumental configurations - readout speed, filter, etc.) taken on the same night. Otherwise the "next-best" calibration frames are used, and the resultant images flagged as such.

The next steps are the subtraction of a bias offset (applied to ensure positive-valued pixels) and, occasionally, a thermally-induced dark current, which scales with exposure time. These are carried out using standard IRAF tasks. Flat-fielding, involving the correction of time-dependent pixel-to-pixel sensitivity variations across the chip, is carried out for each filter using a source of constant illumination across the chip. Where both sky flat-fields (a more accurate reflection of the CCD's response to the night sky) and dome flat-fields (with a deeper count to give a more accurate reflection of each pixel's sensitivity) are taken, the latter, after correction for their different global structure, are combined with the former. Defective pixel lists are usually supplied by the observatory and can easily be corrected by either interpolating the nearest neighbours or by flagging them as defective. There also exist algorithms to deal with extrinsic or transient sources of bad pixels, such as cosmic ray hits.

Fringing, as can be seen in Figure 1, is caused by multiple reflections from penetrating long-wavelength photons in thinned CCDs and is corrected by subtracting a fringing template, modeled for each instrument instrument, although automatically accessing the pattern on the science image in order to determine the correct scaling is non-trivial.

Obtaining the geometrical transformation necessary to register multiple images of the same field but of different spatial alignment is crucial if photometry or astrometry is to be performed on a set of images. An adequate representation of the World Coordinate System, which defines the relationship between pixel coordinates in the image and sky coordinates, resides as meta-data within most archival images. In other cases, we resort to pattern matching under the assumption that several stars are common to the images in question. This is an iterative process which aids the identification and exclusion of incorrectly matched tie-points from the fit. Upon successful registration the proper WCS information can be stored in the headers of the newly registered images.

5. ADVANCED DATA PROCESSING

5.1. Photometry and Astrometry

Following image registration, options for both aperture and crowded-field photometry (e.g. the DAOPHOT algorithms[15]) are also included in the pipeline. These stages give an alternative choice to image matching (see below), especially when analysing specific astronomical objects, as opposed to looking for serendipitous intensity variations in a large image-set. These processes also have the added advantage of being less computationally intensive than image matching and, crucially, are not as reliant on a large degree of spatial overlap between a set of images. In such a case the resulting light-curve has fewer points. A master star-list is created by adding all the images in the combined image-set and applying a star-finding algorithm to the resultant image. Once this star-list is created, the next stage in performing crowded-field photometry is to produce an initial list of stellar magnitudes. These can then be used as first estimates in producing a point-spread[†] function. Once the profile of each star has been accurately fit with the PSF model it is a simple matter to remove each star and examine the residuals of the fit visually. The process can then be reiterated by using successfully subtracted stars as input models for an improved point-spread function. The inclusion of an accurate astrometrical analysis stage in the pipeline is dependent on the FITS headers containing proper WCS information. Some instruments also require a distortion correction to be applied to the images (for example the WFPC2 images described earlier). A long baseline between successive archival images of the same object would be conducive to astrometric studies, although we have yet to implement this in our pipeline.

5.2. Image Matching

Since the primary scientific goal of this project is to detect photometric variations in brown dwarfs, we needed a robust system to differentiate between spurious intensity variations (possibly due to unrejected cosmic-rays or flat-fielding errors) and bona-fide stellar intensity variations. An integral part of the analysis pipeline is the inclusion of the ISIS image-matching software package,[6] originally developed as an alternative approach to difference imaging as a method of optimal image subtraction. The method has since gained favour in a number of different fields, for example in the OGLE[‡] project,[8, 9] in globular cluster research[10–13] and in planetary searches of open clusters.[14] After image registration, a selection of the best images is combined to produce a high signal-to-noise reference image (an example of which can be seen in Figure 2. The convolution kernels needed to transform the reference frame to each science frame is then obtained. The convolved reference frame is then subtracted from each science frame, giving a series of difference images. Any residual in a difference image is indicative of a flux variation between the two images. The flux deviations in such an image can be due either to authentic flux variations (due to variable stars) or to spurious fluctuations due to some systematic source (e.g. flat-fielding anomalies or errors related to the point-spread function). To allow a greater chance of detecting variable stars, a median difference image (see Figure 2), in which each pixel is the median of the set of corresponding pixels in the difference image set, is created. Variable candidates are visible above the local background.

6. INITIAL SCIENTIFIC RESULTS

We have mainly used the pipeline for studying possible photometric variations in brown dwarfs, although we have also tested the photometric stage using archival images of the globular cluster 47 Tuc. This cluster's dense and compact core provides a good model for assessing the quality of the crowded-field photometry stage of the pipeline. Figure 3 shows a series of colour-magnitude diagrams in B and V (using instrumental magnitudes). Each plot represents the data taken from each CCD of the 4×2 mosaic which comprises the WFI detector. The alignment of the cluster's core, with an increased stellar density, with the centre of the mosaic results in a wider spread in the plot for the corresponding chips.

Figure 2 displays the reference image (the combination of all the input images) used in the image subtraction and the resultant image when difference images are combined. The brown dwarf 2MASS 0036+18 is the bright

---

[†] The point-spread function (PSF) is the 2-dimensional brightness distribution function produced on a detector from a point source.

[‡] Optical Gravitational Lensing Experiment.

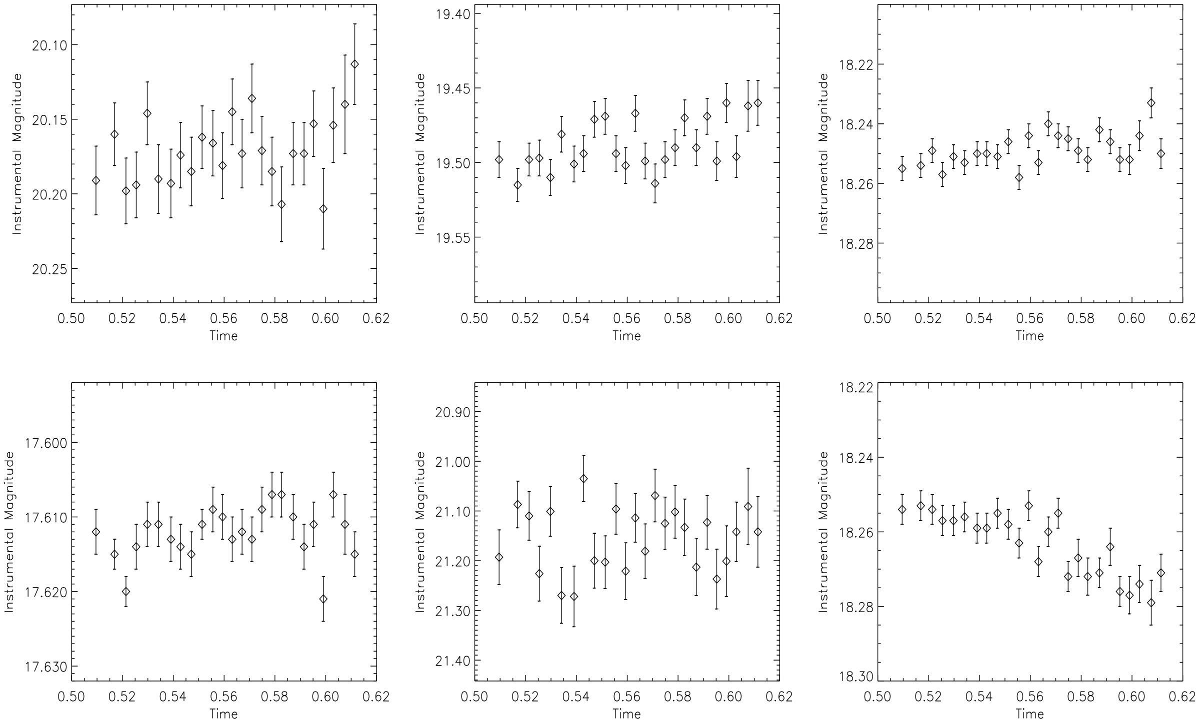

Figure 4. Preliminary results after applying the image reduction and analysis pipeline to a series of images (taken through the I-band filter) centered on the brown dwarf 2MASS 0036+18. The plots show the time-series of the instrumental magnitudes, with error-bars, for five field stars and the brown dwarf (shown bottom-right), whose plot suggests a decrease in brightness.

central star of the group of three in a horizontal line in the centre of the field. The combined difference image suggests some variability on its part. The two bright objects on the right of the frame are image artifacts. The upper one is the signature of a saturated star, as is evident from the reference image; the lower one is caused by bad pixels. The presence of several objects to the left of the difference image suggests possible variations in the point-spread function across the image.

Figure 4 shows photometric results by applying the pipeline to a series of 25 images centered on the brown dwarf 2MASS 0036+18. The plots show the instrumental magnitudes, with error-bars, of five bright stars in the field. The sixth plot (bottom right) is that of the brown dwarf. These preliminary results indicate possible intensity variations in this brown dwarf larger than the uncertainty in the magnitude estimates. However, close examination of the plots reveal slight systematic fluctuations common to all plots which need to be corrected before any concrete conclusions can be drawn from the analysis.

7. CONCLUSIONS AND FUTURE WORK

We have described a pipeline which can process and analyse data from a number of different astronomical instruments, and have shown how such a pipeline can greatly enhance the scientific return from these instruments. Although the primary science goal of this project is to assist in the detection of optical variations in brown dwarfs (which could be indicative of atmospheric activity), it could provide solutions to a wide array of other astrophysical problems provided the robust automated pipeline is supplied with relevant data. Neverthless, targeting a specific stellar object or objects for analysis using the pipeline presents us with one of its main drawbacks, that of the occasional lack of common spatial overlap between images from different instruments. In light of this, the success of the pipeline depends more on serendipitous variable discoveries given a random, but co-spatial, set of sky images.

REFERENCES

1. F. Pierfederici, “Querator: an advanced multi-archive data mining tool”, in *Astronomical Data Analysis, Jean-Luc Starck; Fionn D. Murtagh; Eds. 2001, Proc. SPIE* **4477**, pp. 246-253, 2001.
2. R.G. Mann et al., “AstroGrid: the UK’s Virtual Observatory Initiative”, in *Astronomical Data Analysis Software and Systems XI, ASP Conf.Ser.* **281**, p. 3, 2002.
3. F. Pierfederici, P. Benvenuti, A. Micol, B. Pirenne and A. Wicenec, “ASTROVIRTEL: Accessing Astronomical Archives as Virtual Telescopes”, in *Astronomical Data Analysis Software and Systems X, ASP Conf. Ser.* **238**, p. 141, 2001.
4. P.J. Quinn, P. Benvenuti, P.J. Diamond, F. Genova, A. Lawrence and Y. Mellier, “Astrophysical virtual observatory (AVO): a progress report”, in *Virtual Observatories. Edited by Szalay, Alexander S. Proc. SPIE* **4846**, pp. 1-5, 2002.
5. A.S. Szalay, “The National Virtual Observatory”, in *Astronomical Data Analysis Software and Systems X, ASP Conference Proceedings* **238**, p. 3, 2001.
6. C. Alard and R. H. Lupton, “A Method for Optimal Image Subtraction”, in *The Astrophysical Journal* **503**, 325, 1998.
7. A.B. Tomaney and A.P.S. Crotts, “Expanding the Realm of Microlensing Surveys with Difference Image Photometry”, in *The Astronomical Journal* **112**, p. 2872, 1996.
8. L. Eyer and P.R. Woźniak, “Photometric detection of high proper motions in dense stellar fields using difference image analysis”, in *The Monthly Notices of the Royal Astronomical Society* **327**, pp. 601-609, 2001.
9. I. Soszyński et al., “Optical Gravitational Lensing Experiment: Difference Image Analysis of OGLE-2000-BUL-43, a Spectacular Ongoing Parallax Microlensing Event”, in *The Astrophysical Journal* **552**, pp. 731-737, 2001.
10. A. Olech, J. Kaluzny, I.B. Thompson, W. Pych, W. Krzeminski, and A. Shwarzenberg-Czerny, “RR Lyrae Variables in the Globular Cluster M55. The First Evidence for Nonradial Pulsations in RR Lyrae Stars”, in *The Astronomical Journal* **118**, pp. 442-452, 1999
11. J. Kaluzny, A. Olech, I. Thompson, W. Pych, W. Krzeminski and A. Schwarzenberg-Czerny, “RR Lyrae variables in the globular cluster M 5”, in *Astronomy & Astrophysics Supplement Series* **143**, pp. 215-226, 2000.
12. G. Kopacki, “Variable stars in the globular cluster M 92”, in *Astronomy and Astrophysics* **369**, pp. 862-870, 2001.
13. J. Strader, H.O. Everitt and S. Danford, “Variable stars in the core of the globular cluster M3”, in *Monthly Notices of the Royal Astronomical Society* **335**, pp. 621-627, 2002.
14. B.J. Mochejska, K.Z. Stanek, D.D. Sasselov and A.H. Szentgyorgyi, “Planets in Stellar Clusters Extensive Search. I. Discovery of 47 Low-Amplitude Variables in the Metal-rich Cluster NGC 6791 with Millimagnitude Image Subtraction Photometry”, in *The Astronomical Journal* **123**, pp. 3460-3472, 2002.
15. P.B. Stetson, “DAOPHOT - A computer program for crowded-field stellar photometry”, in *Astronomical Society of the Pacific, Publications* **99**, pp. 191-222, 1987.