Mem. S.A.It. Vol. 74, 989

Memorie della

Real-time Variability Studies with the NOAO* Mosaic Imagers

R. Chris Smith[1] and Armin Rest[1,2]

[1] NOAO/Cerro Tololo Inter-American Observatory, Colina El Pino S/N, La Serena, Chile; e-mail: csmith@ctio.noao.edu
[2] University of Washington, Dept. of Astronomy, Seattle, WA, USA

Abstract. Wide field imaging with modern digital detectors is opening up a new horizon in astronomical research – the exploration of the time domain in astrophysical systems on large scales. Several synoptic survey telescope projects promise to provide the datasets to fuel such exploration. But to fulfill this promise, the projects must design and develop data management systems on a much larger scale (many Terabytes per day *continuously*) than has previously been achieved in astronomy. Working together, NOAO and the University of Washington are developing pipeline systems to explore the issues involved in real-time variability analysis. These efforts are not simply theoretical exercises, but rather are driven by NOAO Survey programs which are generating large data flows. Our survey projects provide a science-driven testbed of data management strategies needed for future initiatives such as the Large Synoptic Survey Telescope and other large-scale astronomical data production systems.

Key words. Astronomical Databases – Microlensing – Stars: imaging – Stars: oscillations – Supernovae – Tecniques: photometric –

1. Introduction

Wide field digital imaging is changing the way much of astronomical research is being done. While still valid, the era of the science served by a one to two night run is giving way to scientific analysis of large, often archival, datasets. While the archives of space-based missions led the way in such archival studies, the large, and more importantly, homogeneous, datasets from survey projects such as SDSS, 2MASS, and DPOSS are providing new views of the wide areas of the sky, and in the process new ways of doing astronomy on large scales. And building upon these new resources, yet more avenues of exploration will be opened through the international Virtual Observatory efforts to tie these and other archives together.

Send offprint requests to: R. Chris Smith

* NOAO, the U.S. National Optical Astronomy Observatory, is operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation.

Correspondence to: Colina El Pino S/N, La Serena, Chile; csmith@ctio.noao.edu

While many of these new large datasets are ideal to study the *spatial* distribution of objects across the sky, the next wave of large scale wide field projects promises to enable studies of the *temporal* dimension. The so-called "Synoptic Survey Telescopes" (SSTs), such as Pan-STARRS, LSST, and SNAP, will open up the time domain on large scales. Although other planned wide field imaging telescopes, such as VISTA and VST, are not dedicated to synoptic observing, they too have the potential to make substantial contributions to this exciting opportunity to study variability, both of large statistical samples of objects and of those most interesting rare objects which are only discovered and characterized when they stand out from large samples.

These wide field time-domain survey projects, both on SSTs and elsewhere, pose new and significant challenges in their design and construction, challenges which can be broken down into roughly three areas: the telescope, the camera, and the data. Many of the telescopes represent innovational concepts to obtain the widest of fields (e.g., the multiple telescope design of Pan-STARRS and the three-mirror design of LSST). The cameras attached to these telescopes will include extremely large mosaics of digital detectors (3 to 5 GIGApixels), and new technologies such as orthogonal transfer arrays and CMOS imaging devices are being explored to meet this challenge.

However, among the most significant challenges is that of the data management for these large scale projects. The simple volume of the planned data flow is daunting, amounting to of the order of 5 Terabytes (TB) per night, leading to archives totalling several Petabytes (PB). For example, projects such as Pan-STARRS and LSST will produce more data *per week* than the Sloan survey will over its full lifetime! And the volume of the data produced is less than half of the challenge.

In order to really take advantage of the time domain aspects of these projects, one must plan for *near real time* reduction and analysis of these large, continuous data flows. Variable and moving objects must be identified automatically and robust classification must be provided to discriminate true objects from artifacts in the data (e.g., cosmic rays, internal reflections, satellite trails). Large scale databases must be employed and optimized for time-domain queries. Detection efficiencies must be calculated in real time as well. And all of this information needs to be made widely available to astronomers so that they can follow-up on the subset of objects which interest them.

2. Time-Domain Astronomy at NOAO

NOAO has identified the challenges represented by the Virtual Observatory efforts and the SST datasets as opportunities to provide new facilities to support astronomical research. In the past NOAO has focussed on providing access to telescopes, together with software to reduce and analyze the data from those telescopes (e.g., IRAF[1]). In order to provide access to, and enable science from, large datasets, NOAO initiated an open program of "survey project" opportunities, providing to up to 20% of the telescope time on NOAO telescopes to large scale projects with the caveat that the data be made publicly available in a timely manner. To manage and promote the use of these data, the NOAO Data Products Program (DPP) Shaw, Boroson, & Smith (2002), has been established. The DPP is charged with developing the infrastructure, including pipelines, archives, and interfaces, to support and encourage the use of the data products from ground-based optical and infrared observations.

The NOAO Surveys program has already provided the foundation for many SST

[1] IRAF, the Image Reduction and Analysis Facility, is distributed by the National Optical Astronomy Observatories

"precursor" projects exploring the time-domain on moderately large scales, reducing and analyzing Mosaic datasets in near real time, albeit usually only over periods of several nights. These projects include the Deep Lens Survey Wittman et al. (2002), the High-z SN Search Schmidt et al. (1998), and the Nearby Galaxies Supernova Search Strolger et al. (2003), to name just a few. In particular, these three projects share a common pipeline software heritage, although each has developed independent solutions to the individual challenges of each survey. The DPP group is collaborating with the NOAO survey projects to develop and incorporate the tools and experience generated from these surveys into services for the larger community. Together with colleagues from the University of Washington (UW), the DPP has identified two particularly scientifically interesting and technically challenging projects to work with in developing the tools necessary to support time domain astronomy in near real time. These two, the SuperMacho and ESSENCE projects, provide a science driven approach to the challenge of preparing for SST data management.

2.1. Probing Dark Matter with Microlensing

One of the foremost outstanding problems in the physical sciences is the nature and distribution of the "dark matter" that is the gravitationally dominant component of the mass in all galaxies, including our own Milky Way. One way to search for a class of astrophysical dark matter objects called MAssive Compact Halo Objects, or MACHOs, is to search for the transient brightening of background stars due to the gravitational lensing by foreground MACHOs. This "microlensing" signature has been detected by several projects Alcock et al. (2000); Udalski et al. (1997); Renault et al. (1997), but the nature of the lensing objects remains a mystery. The SuperMacho project is designed to detect an order of magnitude more microlensing events toward the Large Magellanic Cloud than previous surveys ($\sim$12 per year over five years), which will both (a) move the analysis out of the realm of small number statistics and (b) provide a greater number of "exotic" events (e.g., binary lenses Alcock et al. (1998)) which can lift the degeneracy between the mass, location, and velocity of the lens.

2.2. Probing Dark Energy with Supernovae

Perhaps the most surprising cosmological result in the last five years has been the credible evidence for cosmic acceleration Riess et al. (1998); Perlmutter et al. (1999) based on observations of high-redshift Type Ia supernovae. These observations indicated that the SNe were about 0.25 magnitudes fainter than they would have been in a matter-dominated universe with $\Omega_M = 0.3$ and no cosmological constant. Faint SNe imply that the expansion of the universe from the time of the Big Bang to the present is larger than in any decelerating model. This requires *acceleration* which could have its origin in the negative pressure associated with "dark energy." In order to determine the properties of this dark energy, and in particular how its effects have varied with cosmic epoch, the "ESSENCE" Supernova project aims to discover and follow $\sim$200 Type Ia SNe distributed evenly over the redshift range [0.15, 0.75] during the five year lifetime of the project.

2.3. Science-driven Requirements

While these two projects seek widely differing scientific goals, the observing requirements for both are extremely similar due to the fact that both seek to detect and follow faint transient sources with significant variability over timescales of days. To achieve this, both are designed around an observing cadence of a half night every other night. In addition, both projects have selected the NOAO/CTIO Blanco 4m telescope combined with the Mosaic 8Kx8K camera to achieve the photometric depth and sky coverage necessary for each project. These commonalities have allowed us to merge the data management efforts into one "SM+SN" reduction and transient detection pipeline system. Indeed, it is this sort of synergy of differing science drivers satisfied by common observational parameters which provides the foundation upon which the SSTs can build their science cases.

The design of data management system for the SM+SN projects is constrained by the data rates and the need for rapid transient detection. During one half-night's observations, each project produces roughly 10 GB of raw data, giving a total of $\sim$20 GB per combined SM+SN night. Both projects observe every other night during dark time ($\sim \pm 10$ nights from new moon) for three consecutive months, creating a total of 0.6 TBs of raw data per year. Both projects must process this data and automatically detect faint transient sources on complicated backgrounds. These detections must be cataloged, matched against previously known variable sources, and if new, announcements must be generated.

The timescale for these announcements is one of the driving issues of the data management system design. Although the typical LMC microlensing event to be studied by SuperMacho lasts of order 80 days, the system must be prepared to alert on much shorter timescales, within a day or two of an event detection, to trigger studies of the important exotic events. Supernova have a much shorter rise time (usually $\sim$10 days from detection), and spectroscopy of these objects during the rise can be critical to the classification of relatively uncommon types of SNe Ia such as the 1991T-like events. Given these needs, the transient analysis system should produce announcements, or at least events to be reviewed, within at most 12 hours of the observations to allow for the possibility of follow-up observations on other telescopes the subsequent night.

3. The Prototype: SM+SN PIPELINE

These requirements have been translated into a data management system which had its first beta testing in October 2001 at CTIO during the first run of the SuperMacho project, and was run in full form during the October through November SuperMacho and ESSENCE runs in 2002. It is currently implemented in a combination of IRAF routines, C code, Perl and Python scripts tied together to provide an integrated but modular environment. The design can be broken down into four major sections: data reduction, transient detection, database, and data products. The actual data processing is broken down into a modular structure to allow for changes in the specific modules for different uses of the pipeline.

Our initial stage of processing involves standard data processing to remove instrumental signatures, such as bias subtraction and flat fielding. Accurate astrometric solutions, to $<0.1''$ RMS in the case of SuperMacho, are generated in this stage as well, since the identification of transient objects is tied to linking multiple detections at the same precise location. An important component of this standard processing is the propagation of noise arrays, which again are required downstream in the transient detection to understand the significance of any given peak in the data. Finally, some limited automated data quality assessment is built in to provide a check on the data flowing through the pipeline and attempt to detect problems as soon as they occur.

The heart of this time-domain driven pipeline is the transient detection and analysis. The SM+SN transient detection system depends on very accurate difference imaging techniques Alard & Lupton (1998). One of the main problems with the image differencing approach is that there

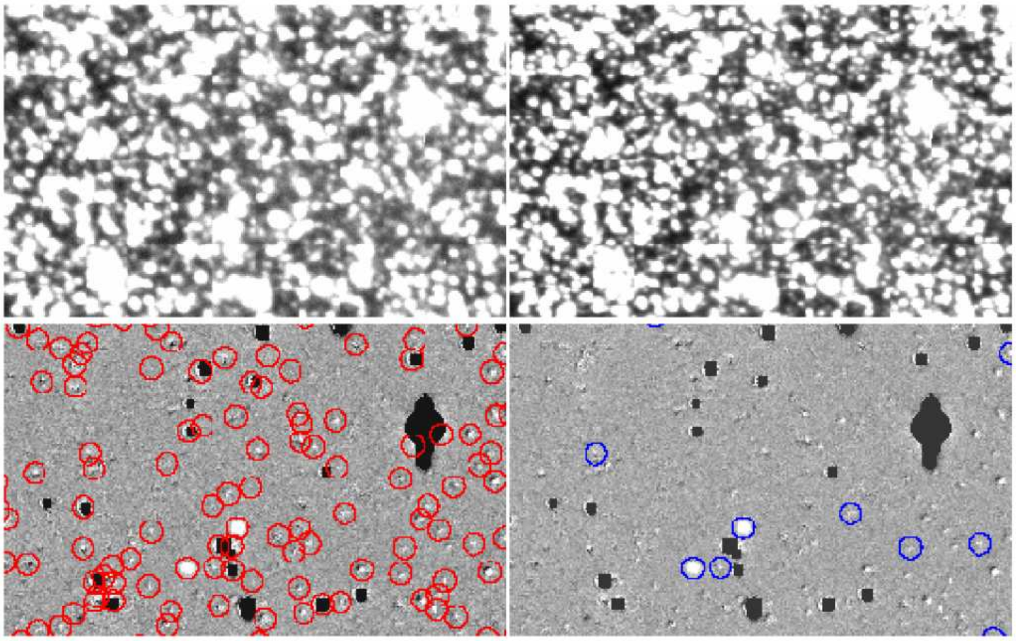

Fig. 1. The upper two panels show example stamps of an image (left) and the associated template (right) in the bar of the LMC (close to a "worst case" for crowding). The lower left panel shows the corresponding difference image, with the objects identified with DoPHOT marked with red circles. The cores of saturated stars are masked out (black spots). Objects that passed our cuts are indicated with blue circles in the lower right panel. Note the two genuine variable objects close to the center of the stamps.

are sometimes more residuals than genuinely variable objects in the difference image. We apply a first cut to these residuals by comparing them to the expected PSF, derived from the same image immediately before image subtraction. The majority of false positives, however, are the faint residuals of cores of stars. In general, they look like "dipoles", i.e. they have positive and negative flux, concentrated into two neighboring areas (see Figure 1). However, by comparing the number of pixels with positive flux and their cumulative flux to the ones with negative flux, we can identify and reject the residuals with dipolar properties.

The transient detections which pass these filters are ingested into a modern fast relational database (PostgreSQL in our case). This database keeps track not only of the transients, but of all of the data, as it flows off of the telescope and through the pipeline, so that any given transient can be traced back to investigate possible problems in the pipeline or observations. The database also serves as the analysis engine, providing transient light curves to external visualization GUIs and storing the classification of transients as variable stars, AGN, moving objects, microlensing events, supernovae, or simply unknown types. Finally, the database provides for relations to external databases, such as variable star catalogs, so that the newly identified transient objects can be linked to previous observations.

The final products of the pipeline are the reduced images and, more importantly, the detections of transients in those images. The SM+SN projects have waived the proprietary period on the images data taken during the search, so both the raw and reduced images are made publicly available soon after the observations. Transfer time to the archive in Tucson is currently a major bottleneck in this process, one which we are working to improve before the next year's observations. The transients are announced immediately upon discovery, through web pages, mailing lists, and IAU circulars. Finder charts and "postage stamps" of the data are also provided to enable and encourage those interested in these events to follow them with other resources. This near real time release of data and data products provides a realized prototype of the LSST observation model.

4. NOAO Pipeline and Archive Development

Significant research and detailed designs are still needed to meet the needs of these and future wide field time-domain surveys. Completely automated data quality monitoring, in real time, is clearly a requirement to generate quality TB to PB datasets. Automated transient classification is also a necessary feature of time domain pipelines for such large scale projects. In addition, modern pipeline infrastructure, based on computer industry standards (e.g., C++, CORBA, and Java), is a key component necessary to scale up to SST data volumes. Finally, these data must flow smoothly into well engineered archives, designed not only to store the data but also to provide tools to analyze the data through interfaces which allow access to grid-based computing resources.

The NOAO DPP and UW collaboration is carrying these concepts forward, together with experiences from other pipeline and archive programs across the world, into the design and implementation of services for the general astronomical community. The two key goals of these efforts are to serve as a significant partner in the international Virtual Observatory, and to meet the challenges in data management posed by the LSST. The results of these efforts will enable the astronomical community to fully realize the scientific potential of exciting new datasets opening up the time domain over large areas of the sky.

Acknowledgements. The ESSENCE project has been funded in part by the National Science Foundation through grant AST-0206329.

References

C. Alcock et al., 1998, BAAS, 30, 1415
C. Alcock et al., 2000, ApJ, 542, 281
C. Alard and R. H. Lupton, 1998, ApJ, 503, 325
S. Perlmutter et al., 1999, ApJ, 517, 565
C. Renault et al., 1997, A&A 324, L69
A. G. Riess et al., 1998, AJ 116, 1009
B. Schmidt and et al., 1998, ApJ 507, 46
D. Shaw, T. Boroson, and C. Smith, 2002, in *Observatory Operations to Optimize Scientific Return*, P. J. Quinn et al., eds., *Proc. SPIE*, 4844, 2002.
L. Strolger et al., 2003, *in preparation*
A. Udalski, M. Kubiak, and M. Szymanski, 1997, Acta Astronomica, 47, 319
D. Wittman, et al. 2002, in *Astronomy with Large Telescopes: Survey and Other Telescope Technologies and Discoveries*, J. A. Tyson and S. Wolff, eds., *Proc. SPIE*, 4836