LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-PROC-410045

Detecting Rare Events in the Time-Domain

A. Rest, A. Garg

January 22, 2009

Classification and Discovery in Large Astronomical Surveys
Ringberg Castle, Germany
October 14, 2008 through October 17, 2008

Disclaimer

Detecting Rare Events in the Time-Domain

Armin Rest*, Arti Garg† and SuperMACHO/ESSENCE collaborations

*Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA
†Lawrence Livermore National Laboratory, 7000 East Avenue, L-413, Livermore, CA 94550, USA

Abstract. One of the biggest challenges in current and future time-domain surveys is to extract the objects of interest from the immense data stream. There are two aspects to achieving this goal: detecting variable sources and classifying them. Difference imaging provides an elegant technique for identifying new transients or changes in source brightness. Much progress has been made in recent years toward refining the process. We discuss a selection of pitfalls that can afflict an automated difference imagine pipeline and describe some solutions. After identifying true astrophysical variables, we are faced with the challenge of classifying them. For rare events, such as supernovae and microlensing, this challenge is magnified because we must balance having selection criteria that select for the largest number of objects of interest against a high contamination rate. We discuss considerations and techniques for developing classification shemes.

Keywords: Galactic Halo; Dark Matter; Supernovae; Time series analysis
PACS: 98.35.Gi 95.35.+d 97.60.Bw 95.75.Wx

INTRODUCTION

The field of astronomy is being revolutionized by wide-field ground-based optical and infrared surveys. The days when astronomical research was done using data taken over the course of a few nights on a single, large, classically scheduled telescope is giving way to research experiments done with data from large surveys in archives and databases. In the 1990's, several microlensing surveys towards the LMC (e.g. MACHO, OGLE, EROS) provided the first coherent wide-field *and* time-domain datasets. Such wide-field surveys have been made possible not only by the tremendous progress in instrumentation, but also by technological and data management advances. Future surveys like Pan-STARRS, and LSST promise to have a nightly data yield in terabytes and total accumulated data in petabytes. As a community our challenge is to develop systems that can handle this flood of data. In particular, in the time-domain, it is important that the detection and classification of a transient has rapid turn-around in order to trigger follow-up observations. In this paper we discuss how to extract a few rare events from a sample of millions. The first important step is to minimize false detections due to instrument or reduction artifacts. Once we've identified true astrophysical transients, we are faced with the challenge of classifying them. Throughout this paper, we will use examples from the ESSENCE[1] and SuperMACHO[2] projects, two time-domain surveys, to illustrate the concepts.

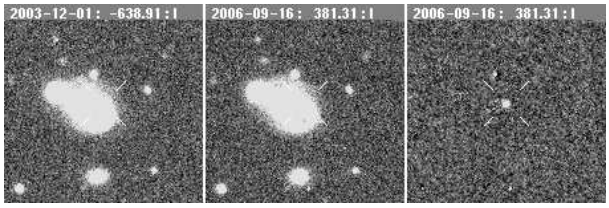

FIGURE 1. ESSENCE difference image: The left panel is the template in the *I* band from 2003-12-01, the middle panel from 2006-09-10, and the right panel is the difference image. The difference image shows a supernova that is buried in the galaxy in the original images.

ELIMINATING FALSE DETECTIONS IN DIFFERENCE IMAGE ANALYSIS

Difference imaging is most commonly used to detect transients and determine their light curves. If difference imaging is not done carefully, however, the number of false detections due to instrument and reduction artifacts can outnumber real detections by several orders of magnitude. In the following, we describe the difference imaging process and how false detections can be minimized.

Difference Image Analysis

The principle of difference image analysis (DIA) is straight-forward: Subtracting one epoch (denoted as template) from another gives the difference between the two images, the signal of interest for time-domain surveys. In practice it is rather complicated. The first step is to reduce all of the images in a standardized way (bias subtraction, flattening). The next steps are the more difficult ones: Before subtraction, the images need to be aligned, and their PSFs matched. These last steps have rapidly evolved in the last few years. The first implementation was by [3] who introduced a method that registered images, matched the point spread function (PSF), and matched the flux of objects in order to detect transients. Derivatives of DIA have been widely applied in various projects (e.g., SuperMACHO[2, 4]; ESSENCE[1]; CHFT SNLS[5]). Since the PSF varies over the field-of-view due to optical distortions or out-of-focus images, for example, it is essential to use a spatially varying kernel [6, 7].

DIA is most useful for detecting and analizing transients against complex backgrounds such as galaxies or crowded fields. Figure 1 shows the power of DIA on an example from the ESSENCE SN search: The difference image (right panel) shows a SN which would have been very difficult to detect in the original image (middle panel) since it is buried within the galaxy. DIA can similarly be used to detect transients in crowded fields and against a background of unresolved stars.

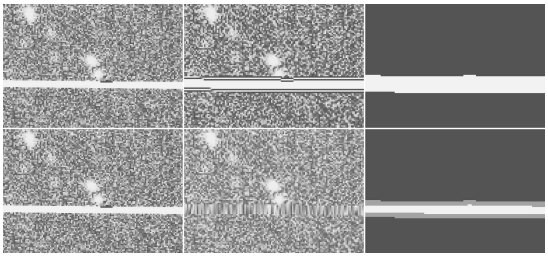

FIGURE 2. An example of a bleed of saturated pixels from a bright star. The top panel show the original image, the deprojected image, and the resulting difference image after subtracting the template. Notice that the saturated pixels spread, increasing the number of "bad" pixels. By using simple, linear interpolation to more accurately estimate the sky flux at these pixels (lower panel, left image), we can reduce the contribution from the saturated pixels and render more pixels usable (lower panel, middle and right image).

Mask Spread due to Image Alignment and PSF matching

Before the images can be subtracted, they need to be aligned. Regardless of the method, alignment always involves the convolution of at least one of the images, i.e. the output pixel is a weighted sum of the input pixels within a kernel of a certain size. In theory, any output pixel which has a bad input pixel (e.g. a saturated pixel or a pixel from a bad column) within its kernel is unreliable and must be masked. As a result the number of masked, unusable pixels increases; the masks "spread". The upper panel of Figure 2 shows a cutout from an image (left panel), its deprojection so that it is aligned with the other image (middle panel), and the mask image of the deprojection (right panel). The white horizontal line is a bleed from a saturated star. The image convolution used for alignment introduces a "ringing" at the edge of the bleed. This ringing is due to the fact that these saturated pixels have flux values much greather than the value corresponding to the number of photons incident from the sky; consequently, even assigning smaller weights to these "bad" pixels does not adequately mitigate their effect. The contribution from their pixel values is simply too large and they are masked out (white in upper right panel).

The lower panels show the same deprojection, but this time the pixel values of the bleed are interpolated prior to the convolution. This interpolation does not need to be perfect or sophisiticated, in this case we just linearly interpolate between the upper and lower pixel value. The "ringing" then disappears (middle panel lower row). Since the pixel values of the originally saturated pixels now have more realistic values, output pixels with contributions from these "bad" pixels can now be used if the weight of the contributions is small. In the right lower panel, pixels for which the contribution from bad pixels is smaller than 10% are light grey and can still be used, whereas pixels with a larger contribution are white. These images demonstrate how interpolating can minimize the spreading of the masks.

Since the subsequent PSF matching is another convolution that spreads the mask even more, interpolating bad pixel values can greatly reduce the number of unusable pixels in the final difference image. Figure 3 shows an example from the SuperMACHO survey. The upper panels show an image from 2003-12-01 and 2002-12-10 from left to right,

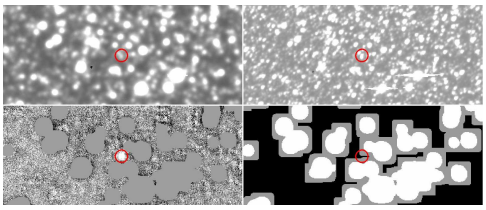

FIGURE 3. The upper panel shows SuperMACHO images from 2003-12-01 and 2002-12-10 from left to right, respectively. In the lower panel, the corresponding difference and mask images are shown from left to right, respectively.

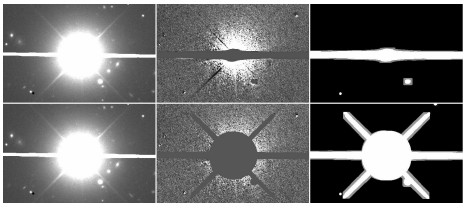

FIGURE 4. The upper panel shows an image of a very bright, saturated star (left panel), the difference image (middle panel), and the mask image (right panel). In the lower panel, the bright star and its spikes are masked, greatly reducing false detections.

respectively. Note that the image from 2003-12-01 has significantly worse "seeing" than the image from 2002-12-10. This leads to a significant spreading of the masks (grey and white pixels in lower right panel) since the size of the convolution kernel is proportional to the larger PSF size of the two images. Interpolating the saturated pixels before aligning the images renders a greater fraction of these pixels usable (grey in lower right panel). Without the interpolation, the transient indicated with the red circle could not have been detected.

Saturated Stars

As described above, interpolation to estimate saturated pixel values enables a reduction in the number of masked pixels in the difference image. Conversely, in some cases extra masking can all but eliminate false detections. An example is shown in Figure 4. The upper panel shows an image of a very bright, saturated star (left panel), the difference image (middle panel), and the mask image (right panel). The difference image in the middle panel shows the halo of the saturated star as well as the spikes, both which are only incompletely subtracted. These residuals often trigger low S/N detections. In general, such bright stars are rare, and masking both the star and its spikes, as shown in the lower middle panel, greatly reduces false detections at the small cost of increasing the fraction of masked pixels by less than 1%.

After aligning the images and masking saturated pixels, we can apply standard difference imaging techniques to obtain difference images that can then be used to detect transients.

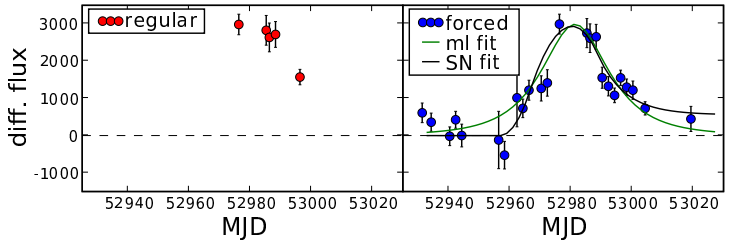

FIGURE 5. The left plot shows the difference image lightcurve of a transient from the SuperMACHO survey. The right plot shows the corresponding forced photometry lightcurve, with microlensing and SN fits overplotted.

Detecting Point Sources in Difference Images

One of the main problems with the image differencing approach is that there are more residuals, e.g. cosmic rays and bleeds, than genuinely variable objects in the difference image. As a result a standard profile-fitting software package, like DoPHOT [8], has problems determining the proper PSF to use for photometry in the difference image. Instead we use a customized version of DoPHOT on the difference images that forces the PSF to be the one determined for the original, flattened, unconvolved image. Applying this *a priori* knowledge of the PSF also helps to guard against bright false positives, such as cosmic rays and noise peaks, which generally do not have a stellar PSF.

The left panel of Figure 5 shows an example of a difference image lightcurve of a transient from the SuperMACHO survey (filled red circles). The trigger in the SuperMACHO survey is 3 detections with "signal-to-noise" (S/N)$\geq$ 5. With only five detections having a S/N$\geq$ 5, however, one can see that not much can be said about the nature of the event. To improve the light curve, we note that unlike standard images, difference images have two important properties:

- Difference images are largely empty of sources as most of the objects are non-varying. The likelihood that two sources are close enough that they need to be fitted simultaneously is very small.
- The event position can be determined from the trigger detections.

This opens up the possibility to do "forced" photometry, i.e. photometry is forced in all difference images at the event position. The blue circles in the right panel of Figure 5 show the forced photometry of the example event. Clearly, the event lightcurve is now much more constrained, giving rise to the difference imaging expression, "Zero flux is not zero info!"

Detecting Extended Sources in Difference Images

Nearly all astrophysical transients are point sources. One of the rare exceptions are scattered-light echoes, which is light of a variable source reflected by dust. The study of such light echoes, in particular light echoes of SNe, provide a host of newly-recognized

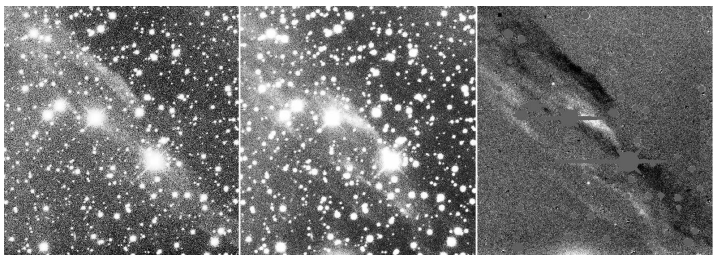

FIGURE 6. Light echo arclets associated with Tycho. The orientation is N up and E to the left and the images are 325 x 250 arcseconds. The left and middle panels show the first epoch image from 20 October 2006, and the second epoch image from 13 December 2007, respectively. The right image is the difference image. Areas of negative flux, where the echo has moved from its location in the earlier October image, appear dark; and areas of positive flux, representing the December 2007 location, appear white.

observational benefits which have only just begun to be exploited including: the capacity to understand the connection between remnant properties and the outburst spectral type, access to observables related to asymmetric explosion properties, and geometric distances [9, 10, 11]. The challenge is to find these light echoes, since they are in general faint ($V > 22.5 mag/arcsec^2$) and often at angular distances to the source event of several degrees. Figure 6 shows one of the brightest light echoes associated with the Tycho SN [11].

The current searches for light echoes [11] use visual inspection to find light echoes. This is of course not doable in future large scale surveys like PanSTARRS, Sky Mapper, and LSST. The problem is that standard photometric packages are optimized to find centrally concentrated objects like stars and galaxies. However, light echoes often do not have such a peak, but they rather are extended arclets. In the case of light echoes from ancient Galactic SNe, these arclets can have width's and length's of tens or arcseconds. Therefore what is needed is a photometry software that internally either bins or smoothes the data (binning or smoothing the images is prohibitive for large surveys due to the disk space and CPU requirements). If the source event position is known, then the smoothing kernel optimally has the shape of an arclet of a circle whose center is at the source event position.

LIGHT CURVE FITTING AND CLASSIFICATION

The light curve provides the essential information for classifying events in the time domain. While ideally the light curve will contain data from many bands which provides sensitivity to changes in the SED, multiple images must be traded off against imaging more sky or going deeper to detect more and fainter sources. For this reason, the SuperMACHO survey was primarily conducted in a single band. For SuperMACHO in particular, and time domain surveys in general, analysis of the light curve *shape* provides the primary parameters for classification. At it's most basic, this analysis requires fitting the light curve to analytic models or templates describing the transient of interest. In many cases, however, either a model is unavailable, or because of low signal-to-noise or sparse temporal coverage the fit is not sufficient to discriminate between different

classes of transients. The lower panel of Figure 5 shows a microlensing and SN fit to a transient detected in the SuperMACHO survey, which have very similar χ^2. In these cases parameters describing the light curve (e.g. rate-of-rise, time-to-maximum) can provide a suite of attributes for developing classification schema (see [12] for a more detailed discussion).

The advent of large data sets has spurred an explosion in development of classification schemes over the past decade. This conference has demonstrated the usefulness of many of these techniques. For classification of rare events, however, these techniques are not always appropriate. Support Vector Machines, for example, require an *a priori* knowledge of the relative frequencies of the events being studied. Often the event rate or the full composition of the background population is unknown. Principal Component Analysis can be used to define a parameter space where the events of interest reside. If these is strong overlap with another transient, however, this can make classification difficult.

Another concern is that when classifying rare events, the "standard of proof" must be higher. This is particularly true when the rate of the rare event is much lower than similar objects (e.g. type Ia SNe vs. microlensing in the SuperMACHO survey). In this scenario we must balance loose selection criteria that will keep the maximum number of the events of interest, "high efficiency", against swamping the events of interest with a more frequent event type, "high contamination". In addition, if the event is particularly rare (e.g. tidal disruption events), then selection criteria that intuitively match our understanding of the underlying physical properties can bolster acceptance of the results. For example in the SuperMACHO project, because SNe Ia have asymmetric light curves while microlensing is a symmetric phenomenon, selection criteria that explore light curve symmetry provide both a robust and intuitive method for event classification.

Ultimately any classification scheme will not provide 100% accuracy. How one should proceed depends upon the scientific question being asked. For a project such as SuperMACHO where the goal is to understand the rate of microlensing, the scientific result depends critically upon a quantitative understanding of the detection efficiency and contamination rate. For a project such as ESSENCE, the rate of SNe Ia does not significantly impact the result; however, contamination from other types of SNe or transients does. In this instance additional follow-up for classification, particularly spectroscopic, is warranted.

CONCLUSIONS

Difference imaging can provide a powerful tool for identifying rare transients, even against complex backgrounds. Increasing the reliability of difference image detections, however, is critical for large surveys due to the large volume of data and need for fast turn-around. We discuss several techniques that both decrease the number of false detections from difference imaging and also increase the usable area of the images. For time-domain surveys, difference image analysis provides only the first step for identifying and classifying transients. In addition, obtaining reliable and complete light curves through specialized photometry and classifying objects based on their variability

are critical to extracting the science from these projects. Simple techniques such as using a fixed-PSF profile with a forced centroid can greatly improve difference flux light curves. Developing classification schemes based on these light curves remains an area of active research. We find that often the science goals will dictate which classification sheme is most appropriate.

ACKNOWLEDGMENTS

AR thanks the NOAO Goldberg fellowship program for its support. The SuperMACHO and ESSENCE surveys was undertaken under the auspices of the NOAO Survey Program. AG's work prepared by LLNL under Contract DE-AC52-07NA27344. SuperMACHO is supported by the STScI grant GO-10583 and GO-10903.

REFERENCES

1. G. Miknaitis, G. Pignata, A. Rest, W. M. Wood-Vasey, S. Blondin, P. Challis, R. C. Smith, C. W. Stubbs, N. B. Suntzeff, R. J. Foley, T. Matheson, J. L. Tonry, C. Aguilera, J. W. Blackman, A. C. Becker, A. Clocchiatti, R. Covarrubias, T. M. Davis, A. V. Filippenko, A. Garg, P. M. Garnavich, M. Hicken, S. Jha, K. Krisciunas, R. P. Kirshner, B. Leibundgut, W. Li, A. Miceli, G. Narayan, J. L. Prieto, A. G. Riess, M. E. Salvo, B. P. Schmidt, J. Sollerman, J. Spyromilio, and A. Zenteno, *ApJ* **666**, 674–693 (2007), `arXiv:astro-ph/0701043`.
2. A. Rest, C. Stubbs, A. C. Becker, G. A. Miknaitis, A. Miceli, R. Covarrubias, S. L. Hawley, R. C. Smith, N. B. Suntzeff, K. Olsen, J. L. Prieto, R. Hiriart, D. L. Welch, K. H. Cook, S. Nikolaev, M. Huber, G. Prochtor, A. Clocchiatti, D. Minniti, A. Garg, P. Challis, S. C. Keller, and B. P. Schmidt, *ApJ* **634**, 1103–1115 (2005).
3. A. C. Phillips, and L. E. Davis, "," in *ASP Conf. Ser., 77, Astronomical Data Analysis Software and Systems IV*, 1995, vol. 4, pp. 297–+.
4. A. Garg, C. W. Stubbs, P. Challis, W. M. Wood-Vasey, S. Blondin, M. E. Huber, K. Cook, S. Nikolaev, A. Rest, R. C. Smith, K. Olsen, N. B. Suntzeff, C. Aguilera, J. L. Prieto, A. Becker, A. Miceli, G. Miknaitis, A. Clocchiatti, D. Minniti, L. Morelli, and D. L. Welch, *AJ* **133**, 403–419 (2007).
5. P. Astier, J. Guy, N. Regnault, R. Pain, E. Aubourg, D. Balam, S. Basa, R. G. Carlberg, S. Fabbro, D. Fouchez, I. M. Hook, D. A. Howell, H. Lafoux, J. D. Neill, N. Palanque-Delabrouille, K. Perrett, C. J. Pritchet, J. Rich, M. Sullivan, R. Taillet, G. Aldering, P. Antilogus, V. Arsenijevic, C. Balland, S. Baumont, J. Bronder, H. Courtois, R. S. Ellis, M. Filiol, A. C. Gonçalves, A. Goobar, D. Guide, D. Hardin, V. Lusset, C. Lidman, R. McMahon, M. Mouchet, A. Mourao, S. Perlmutter, P. Ripoche, C. Tao, and N. Walton, *A&A* **447**, 31–48 (2006), `arXiv:astro-ph/0510447`.
6. C. Alard, and R. H. Lupton, *ApJ* **503**, 325–+ (1998).
7. C. Alard, *A&A* **144**, 363–370 (2000).
8. P. L. Schechter, M. Mateo, and A. Saha, *PASP* **105**, 1342–1353 (1993).
9. A. Rest, N. B. Suntzeff, K. Olsen, J. L. Prieto, R. C. Smith, D. L. Welch, A. Becker, M. Bergmann, A. Clocchiatti, K. Cook, A. Garg, M. Huber, G. Miknaitis, D. Minniti, S. Nikolaev, and C. Stubbs, *Nature* **438**, 1132–1134 (2005).
10. A. Rest, T. Matheson, S. Blondin, M. Bergmann, D. L. Welch, N. B. Suntzeff, R. C. Smith, K. Olsen, J. L. Prieto, A. Garg, P. Challis, C. Stubbs, M. Hicken, M. Modjaz, W. M. Wood-Vasey, A. Zenteno, G. Damke, A. Newman, M. Huber, K. H. Cook, S. Nikolaev, A. C. Becker, A. Miceli, R. Covarrubias, L. Morelli, G. Pignata, A. Clocchiatti, D. Minniti, and R. J. Foley, *ApJ* **680**, 1137–1148 (2008), `0801.4762`.
11. A. Rest, D. L. Welch, N. B. Suntzeff, L. Oaster, H. Lanning, K. Olsen, R. C. Smith, A. C. Becker, M. Bergmann, P. Challis, A. Clocchiatti, K. H. Cook, G. Damke, A. Garg, M. E. Huber, T. Matheson, D. Minniti, J. L. Prieto, and W. M. Wood-Vasey, *ApJL* **681**, L81–L84 (2008), `0805.4607`.
12. A. Garg, "," in *this proceedings*, 2008.