Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 5 (2008) 299–306
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

Object detection in multi-epoch data

G. Jogesh Babu[a,*], A. Mahabal[b], S.G. Djorgovski[b], R. Williams[c]

[a] *Department of Statistics, 326 Joab L. Thomas Building, The Pennsylvania State University, University Park, PA 16802-2111, USA*
[b] *Division of Physics, Mathematics, and Astronomy, California Institute of Technology Pasadena, CA 91125, USA*
[c] *Center for Advanced Computing Research, California Institute of Technology Pasadena, CA 91125, USA*

Received 13 April 2007; received in revised form 23 January 2008; accepted 12 February 2008

Abstract

In astronomy multiple images are frequently obtained at the same position of the sky for follow-up coaddition as it helps one go deeper and look for fainter objects. With large scale panchromatic synoptic surveys becoming more common, image co-addition has become even more necessary as new observations start to get compared with coadded fiducial sky in real time. The standard coaddition techniques have included straight averages, variance weighted averages, medians etc. A more sophisticated nonlinear response chi-square method is also used when it is known that the data are background noise limited and the point spread function is homogenized in all channels. A more robust object detection technique capable of detecting faint sources, even those not seen at all epochs which will normally be smoothed out in traditional methods, is described. The analysis at each pixel level is based on a formula similar to *Mahalanobis distance*.

Keywords: Mahalanobis distance; Chi-square; Faint source detection; Coaddition; Gaussian noise

---

1. Introduction

Many major projects, ongoing and future synoptic surveys, Palomar-QUEST (http://palquest.org/), MACHO (http://www.macho.mcmaster.ca/), LSST (http://www.lsst.org/), OGLE (http://bulge.astro.princeton.edu/~ogle/) Pan-STARRS (http://pan-starrs.ifa.hawaii.edu/) etc., involve

---

[*] Corresponding author. Tel.: +1 814 863 2837.

E-mail addresses: babu@stat.psu.edu (G. Jogesh Babu), aam@astro.caltech.edu (A. Mahabal), george@astro.caltech.edu (S.G. Djorgovski), roy@caltech.edu (R. Williams).

URLs: http://www.stat.psu.edu/~babu, http://www.astrostatistics.psu.edu (G. Jogesh Babu).

1572-3127/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2008.02.003

repeated scans of large areas of the sky in several wavelength bands. Thus an important area of recent astronomical research has been the investigation of source detection in multi-epoch data. A question frequently asked is: What is the best way to combine image regions with low signal to detect faint objects? Historically, two basic methods have been used to search for faint astronomical objects:

(i) Use a larger telescope to collect a larger number of photons.
(ii) Stack a large number of registered images in order to improve the signal-to-noise ratio.

The former has engineering and monetary limitations while the latter may not work for transients that are seen only once and for very faint objects where the signal remains below detection threshold even after the pixel-to-pixel coadding of several images. Szalay et al. [11] had proposed a method that used chi-square coaddition by treating pixels from different images but at the same location to be uncorrelated. It is an improvement over standard coaddition but has its own limitations, as the pixels in different images at any location are in fact correlated. The procedure described here is a more robust technique to detect faint sources from multi-epoch data. It is designed not only to have high sensitivity, but to detect changes between images.

The analysis at each pixel level is based on a statistic similar to the measure of distance introduced a by P.C. Mahalanobis in 1936 [8,2]. The *Mahalanobis distance* $D_{\nu,\Sigma}$, of a multivariate row vector $\mathbf{x}$ from a group of values with mean vector ν and covariance matrix Σ, is defined as

$$D_{\nu,\Sigma}(\mathbf{x}) = \sqrt{(\mathbf{x}-\nu)\Sigma^{-1}(\mathbf{x}-\nu)^T}.$$

It is used in classical multivariate analysis and differs from Euclidean distance. It is scale-invariant, and is based on correlations between variables by which different patterns can be identified and analysed. If the covariance matrix Σ is the identity matrix, then the Mahalanobis distance reduces to the Euclidean distance. If Y is a Gaussian random vector with mean ν and covariance matrix Σ, then $D^2_{\nu,\Sigma}(Y)$ is distributed as χ^2 with p degrees of freedom, where p is the rank of Σ. It is a useful way of determining similarity of an unknown sample set to a known one [3,2]. Mahalanobis distance is available in **R** Stats package (http://www.r-project.org/). **R** is a free public domain software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. Tutorials for **R** are available at http://astrostatistics.psu.edu/datasets/ and at http://www.sr.bh am.ac.uk/˜ajrs/R/.

For the methodology described in this paper, it is required that the pixel size for different images is invariant. However, the method does not depend on the point spread function (PSF) since signal from genuine sources spreading to different extent in different images will still lead to significant correlations between different images at the pixels with the source than if there were no source. Of course it is best if the PSFs are fairly similar. In the following sections we provide the details of the method as well as the tests we have run so far and future plans for large-scale implementation.

2. Methodology

We start with N images of a given region of the sky. We first use standard techniques to ensure that all images are of the same size in terms of area covered as well as the pixel dimensions. This involves using a tool like *Swarp* (available from TERAPIX web site http://terapix.iap.fr/), which can extract only the common overlapping regions from all the images. In addition the objects in the images can be matched to a known catalogue for ensuring that the images have accurate