RevMexAA (Serie de Conferencias), **48**, 33–37 (2016)

SHORT PERIOD VARIABLE STARS RECOGNITION BY USING MVA METHODS IN PI OF THE SKY EXPERIMENT

L. Obara[1]

RESUMEN

Pi del Cielo es un sistema de telescopios robóticos de amplio campo de visión, para la búsqueda de fenómenos astrofísicos que ocurren en escalas de tiempo cortos, especialmente para la detección de la contraparte óptica de la emisión de ráfagas de rayos gama. El sistema fue diseñado para operar autónomamente en el seguimiento de una gran parte del cielo para objetos con magnitudes límite en el intervalo de 12^m–13^m y una resolución temporal de 1–100 segundos. LUIZA es un programa, basado en lenguaje C++, para el procesamiento eficiente de datos de Pi del Cielo. Se implementó en LUIZA un algoritmo de fotometría, basado en la fotometría de la Busqueda Automatizada de Todo el Cielo (ASAS), y se comparó con los algoritmos basados en la reconstrucción de cúmulos de píxeles y el algoritmo sencillo de fotometría de apertura. Este algoritmo optimizado de fotometría fue entonces aplicado a una muestra de imágenes de prueba, las cuales fueron modificadas para incluir diferentes patrones de variabilidad de un conjunto de estrellas en una muestra de entrenamiento. Diferentes estimadores estadśticos fueron considerados para el desarrollo del algoritmo para la identificación general de estrellas variables. El algoritmo se utilizará en datos reales para buscar estrellas variables de corto plazo.

ABSTRACT

Pi of the Sky is a system of wide field–of–view robotic telescopes to search for short timescale astrophysical phenomena, especially for prompt optical GRB emission. The system was designed for autonomous operation, monitoring a large fraction of the sky with limit magnitude in the range 12^m–13^m and time resolution on the order of 1–100 seconds. LUIZA is a dedicated framework developed for efficient off–line processing of the Pi of the Sky data, implemented in C++. The photometric algorithm based on the All Sky Automated Survey (ASAS) photometry was implemented in LUIZA and compared with the algorithm based on the pixel cluster reconstruction and simple aperture photometry algorithm. Optimized photometry algorithms were then applied to the sample of test images, which were modified to include different patterns of variability of the stars in training sample. Different statistical estimators were considered for developing the general variable star identification algorithm. The algorithm will be used to search for short–period variable stars in real data.

Key Words: stars: variable — techniques: image processing — telescopes

1. PI OF THE SKY PROJECT

Pi of the Sky is a robotic telescope designed for observations of short timescale astrophysical phenomena, especially for prompt optical counterparts of Gamma Ray Bursts (GRBs). Other scientific goals include searching for novae and supernovae stars and monitoring of interesting objects such as blasars, AGNs, variable stars or possible sources of gravitational vaves. Apparatus design allows to monitor of a large fraction of the sky with range 12^m - 13^m and time resolution of the order of 1 - 100s.

Pi of the Sky project involves two observatories The prototype device was installed in 2004 in Las Campanas Observatory (LCO), Chile, and moved to San Pedro de Atacama in 2011. It consists of 2 CCD cameras on single mount. First detector unit of the final system was installed in October 2010 at INTA El Arenosillo near Huelva, Spain, additional 3 units were installed in July 2013. Each unit consists of 4 custom designed cameras with 2000x2000 pixel matrix and Canon lenses (f =85 mm, f/d = 1.2). All cameras in Spain and one camera in Chile have only standard UV- and IR-cut filters, whereas the second camera in Chile has an R Johnson-Bessel filter installed.

2. LUIZA FRAMEWORK

LUIZA is a dedicated framework developed for efficient off-line processing of the Pi of the Sky data, implemented in C++. It's based of a following assumtions:

- each data (image) analysis can be divided into small, well defined steps, implemented as so called processors,

[1]Faculty of Physics, University of Warsaw, Pasteura 5, 02-092 Warsaw, Poland (Lukasz.Obara@fuw.edu.pl).

- each step has to have well defined input and output data structure,
- by defining universal data structures we make sure that different processors can be connected in a single analysis chain,
- processor configuration and their parameters can be set by user at run time in a simple steering file.

Two basic data formats implemented in LUIZA, required for image analysis are:

- **GloriaFitsImage** - class for storing FITS images, which uses FITSIO library for reading and storing images, and basic methods for image manipulation implemented,
- **GloriaFitsTable** - class for storing other data structures, which can be described as tables, including lists of reconstructed objects (stars), reference star catalogues, calibration corrections etc.

To use the LUIZA framework user has to create a simple XML stering file, by selecting active processors and defining their order in analysis chain. In the same steering file user can also define input-output streams and set the processor parameters.

3. PHOTOMETRY

All basic steps required for the standard image analysis chain, including frame stacking, image normalization (dark subtraction, flat correction), photometry, astrometry and calibration to reference stars, were previously implemented in LUIZA. For photometry (star identification and instrumental brightness determination) two different algorithms were defined.

As a part of the presented study, new processors were added to LUIZA framework, implementing aperture photomety algorithm adopted from ASAS. The new algorithm was divided into three analysis steps (processors):

- **CalculateGaussPSF**
  Processor calculates point spread function (PSF) shape parameters for stars found on the CCD image. For each star found in the image (after background subtraction) following parameters of its profile are calculated from the selected pixels: RMS in X, RMS in Y, XY-correlation, FWHM, Elongation and orientation. Median of the obtained parameter values is calculated in a given number of frame sectors and the resulting map PSF shape parameter is stored in the output table. It can then be used for calculating Gaus kernel for given image.
- **ApplyGausKernel**
  This processor can be used to convolute sky image with Gaus kernel based on the input PSF shape parameters (calculated in the previous step or loaded from the file). Kernel size can be set by user or calculated automatically from the shape parameters. Convolution with the Gaus kernel supresses the noise contribution and improves signal to noise ratio for dim stars. After convoluting the sky frame with the kernel, the processor searches for stars in the image and calculates their position (center of gravity). On its output the processor produces a list of objects with their positions on CCD (X, Y).
- **ApertureMagnitudoASAS**
  Processor implementing the actual photometry algorithm. The instrumental brightness of identified objects is calculated as a weighted sum of pixels inside the aperture. Weights correspond to the fraction of the pixel within an aperture. The size of aperture is not fixed, but follows the size of star PSF.

To verify the accuracy of the photometry algorithm, instrumental brightness had to be converted to absolute magnitude scale by comparing identified objects to the set of reference stars. Most precise calibration of the Pi of the Sky images is obtained when normalizing star brightness to Tycho-2 reference stars with brightness in the range 7^m - 9^m. After calibrating brightness measurements to reference stars, photometry precision can be estimated by looking at the distribution of brightnesses measured for given stars for a larger sample of images. Shown in Fig. 1 is the spread of the brightness values obtained with the described algorithm photometry (after calibration to reference stars), as a function of the average star magnitude, for 10s exposures. Left plot shows RMS values calculated for single stars while the right plot shows the estimated photometry uncertainty averaged in magnitude bins. Highest precision of brightness measurement, of the order of 0.02^m, is obtained for stars 6.5^m - 9^m.

4. SEARCHING VARIABLE STARS

A significant fraction of stars visible in the sky (1-2%) is variable, changing its brightness with time. These variations can be of internal origin (due to processes inside the stars, as in case of Cepheids, Flare

IV Workshop on Robotic Autonomous Observatories (Torremolinos, Málaga, Spain, September 28- October 2, 2015)
Editors: M. D. Caballero-García, S. B. Pandey, D. Hiriart & A. J. Castro-Tirado

IV Workshop on Robotic Autonomous Observatories (Torremolinos, Málaga, Spain, September 28- October 2, 2015)
Editors: M. D. Caballero-García, S. B. Pandey, D. Hiriart & A. J. Castro-Tirado

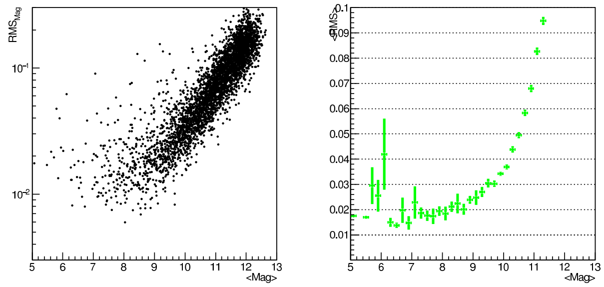

Fig. 1. Estimated error of the ASAS photometry algorithm as a function of the star magnitude, for single stars (left) and averaged in magnitude bins (right).

stars, Novae or Supernovae) or due to environmental reasons (eg. in case of binary systems). In the past, Pi of the Sky searched for variable stars using data collected with the propotype detector in Chile, based on stacked frames corresponding to 200s exposure. Periods in the range from 0.1 to 10 days were considered for stars with a number of observations larger than 200. In the presented study we look for stars with shorter variability scales, of the order of tens of minutes or hours. This requires using single 10s exposures for the analysis, instead of stacked frames. However, with the lower photometry accuracy for single exposures, new variable star selection criteria have to be developed.

To identify a variable star we need to measure it on many (hundreds of) images. Final identification is always based on the reconstruction of the star light curve, but the preselection can be based on the measured magnitude distribution. For constant stars, measured in good conditions, brightness measurements should form a normal distribution with a width corresponding to the photometry uncertainty. The estimated uncertainty of magnitude is due to statistical signal fluctuations, background and CCD noise, but can be also affected by obsevational conditions (clouds, Moon background, readout errors). Still, we expect that the magnitude distribution for constant stars should be much narrower and more gaussian-like, than for the variable stars. Compared in Fig. 2 are the magnitude distributions for constant star (blue plot) and variable star (red plot). For constant star, distribution is much narower with only a small tail, while for the variable star the distribution is much wider and the tail towards larger magnitude is more visible.

To understand the differences in magnitude distribution for constant and variable stars, we selected a set of variable stars visible in the considered sam-

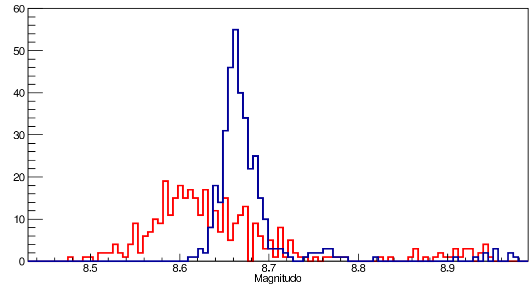

Fig. 2. Distribution of magnitude for selected constant star (blue) and variable star (red).

ple of images based of the GCVS catalog. Fig. 3 shows the RMS of the magnitude distribution when comparing the selected variable stars (red symbols) and for constant stars (black symbols), as a function of object magnitude. For bright stars (below 9.5^m),

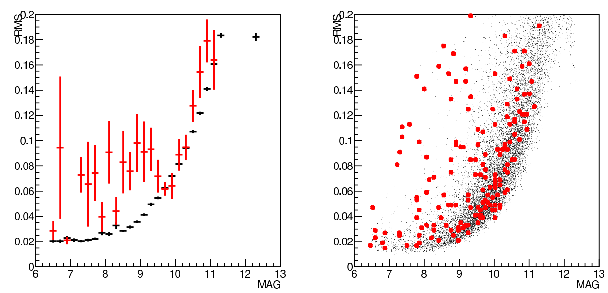

Fig. 3. Spread of the magnitude distribution for constant stars (black) and variable stars (red).

average magnitude spread for variable stars is significantly higher than for constant stars (left plot). For dimmer stars the spread for variable and constant stars is similar and corresponds to (increasing) photometry uncertainty. However, when we look at single stars (right plot), we see that bright variable stars can also have small spread of magnitude distribution. This can be due to small amplitude of variability (below our sensitivity) or to the small fraction of measurements influenced by variability (eg. for eclipsing binary with very narrow eclipses). Therefore, we should not only look at the distribution spread (measured in terms of the RMS) but consider also other parameters related to the shape of the magnitude distribution. We decided to include higher moments of the distribution (kurtosis, asymetry, skewness) in the procedure of variable object preselection. The algorithm was based on the following variables:

- RMS of the magnitude distribution of the considered star relative to the average magnitude

IV Workshop on Robotic Autonomous Observatories (Torremolinos, Málaga, Spain, September 28- October 2, 2015)
Editors: M. D. Caballero-García, S. B. Pandey, D. Hiriart & A. J. Castro-Tirado

uncertainty for this magnitude

- kurtosis of the magnitude distribution
- asymmetry of the magnitude distribution
- skewness (AG) of the magnitude distribution
- output of the Shapiro-Wilk test comparing magnitude distribution to the normal distribution.

For constant stars we expect RMS ratio close to 1 and values close to zero for all other variables. For variable stars both positive and negative values of higher moments are possible, depending on the light curve shape.

To define the best selection algorithm for selection of variable star candidates, we use Multi-Variable Analysis (MVA) approach, as implemented in the TMVA package for CERN root. In a first step (training) algorithm has to "learn" the functional dependencies between different variables for signal (variable stars) and background (constant stars) events. As the sample of stars selected based on GCVS catalog was too small and it could not guarantee covering different variability types with equal probability, we decided to prepare simulated sample of variable stars. In each CCD frame considered, pixels corresponding to 150 selected stars (6.5^m - 12.5^m) were scaled in such a way as to model variability with amplitude between 0.08^m and 0.5^m, period between 6 hours and 3 days, and different lightcurve shapes. These stars were then tagged as "signal sample" in the MVA while all other stars found in CCD images were used as "background". The input data set was divided into two parts: one used for algorithm training, and the second one used to test the result. As the two samples are statistically independent, we can make sure that the algorithm was not overtrained and can properly separate signal from background. TMVA package includes many different algorithms of event clasification. In our case we decided to use the so called Boosted Decision Tree (BDT) approach to separate variable stars. For each star (magnitude distribution) it returns the value corresponding to the probability that the considered star is variable.

5. RESULTS

As mentioned above, high photometry accuracy on single frames is possible only for bright stars. Therefore, we focused on the algorithm optimization in the magnitude range 6.5^m - 9.5^m and the results presented in this section refer to these stars only. Shown in Fig. 4 are the distributions of input variables used for BDT analysis. Compared are the distributions of variable values for signal (blue histogram) and background (red histogram) events. Clear differences are observed between variable star

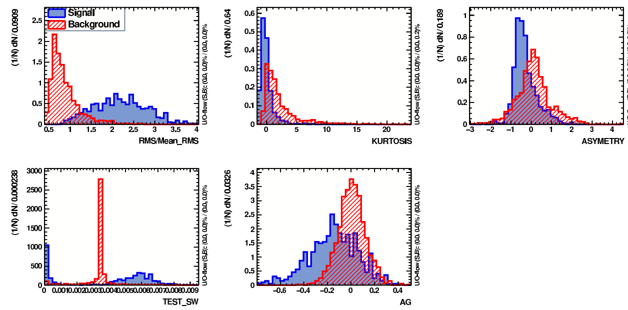

Fig. 4. Input variables for the BDT algorithm. Compared are the distributions for signal (blue) and background (red) events. Stars between 6.5^m and 9.5^m are considered.

(signal) and constant star (background) distributions.

As the expected number of variable stars is low, we require very high background rejection efficiency ($\sim$ 99%). This will still result in few dozen of candidates in each field. The final decision whether the star is variable will need to be taken after a visual inspection and light curve analysis. Efficiency of bacground rejection based on the BDT algorithm is presented on figure 5. For the bright stars considered 99% background rejection corresponds to signal selection efficiency of about 75%. This efficiency loss is mainly due to variable stars with small variability amplitude. Results are preliminary and still need to be verified.

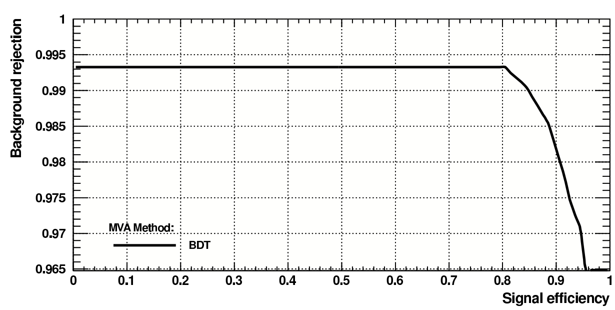

Fig. 5. Background rejection versus signal selection efficiency for variable star selection based on BDT analysis.

Compared in Fig. 6 is the BDT response for the statistically independent samples of training and test events. The distributions for both samples are very similar, which confirms that the algorithm was not overtrained.

IV Workshop on Robotic Autonomous Observatories (Torremolinos, Málaga, Spain, September 28- October 2, 2015)
Editors: M. D. Caballero-García, S. B. Pandey, D. Hiriart & A. J. Castro-Tirado

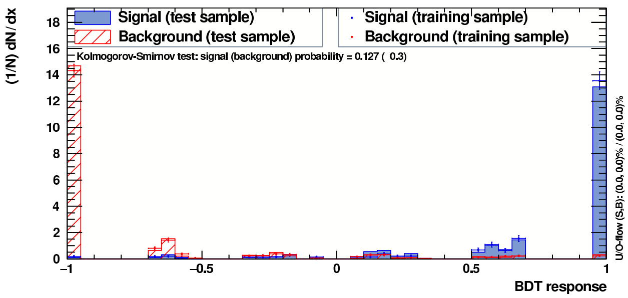

Fig. 6. Distribution of BDT classifier for training samples (points) and test samples (histogram) of signal (blue) and background (red) events.

Number of decision trees contributing to the final event clasification is a parameter entered by user.

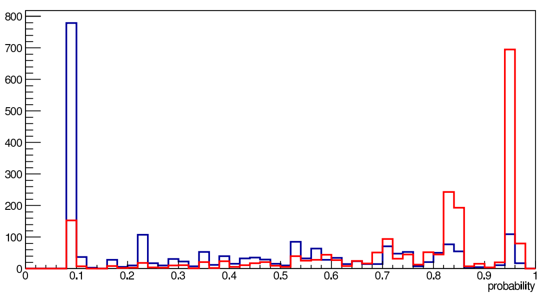

Fig. 7. Distribution of likelihood for signal (red) and background (blue) events.

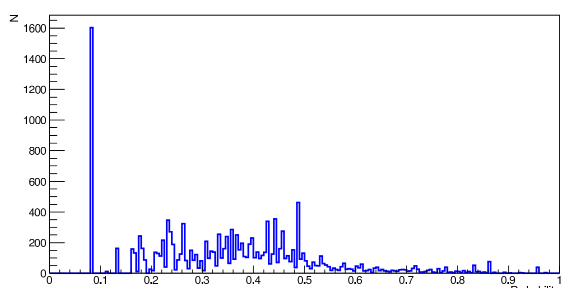

Fig. 8. Distribution of likelihood for sample from Pi of the Sky data.

The result of the analysis is a probability, whether the analyzed star is variable or constant. Compared in Fig. 7 is the estimated likelihood for the simulated stars (red) and constant stars (blue). Probability is higher than 65 percent for majority of signal stars.

The plot in Fig. 8 presents the probability of variability for clusters recognized on Pi of the Sky data (sample of CCD frames). A very small part of the stars in the plot 8 shows a probability higher than 65 percent and they are selected for further analysis. On the last step we have to perform light curve check for these candidates.

6. CONCLUSIONS

ASAS photometry algorithm was implemented in LUIZA Framework. Tests showed its high precision in the analysis of single sky frames. The algorithm was used to search for the variable objects on the single CCD frames. The MVA algorithm used for analysis of magnitude measurement distributions gives promissing results. Bright variable stars can be selected with about 75% efficiency.

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 283783. The research work is also financed by the Polish Ministry of Science and Higher Education in 2011-2014, as a co-funded international project.

REFERENCES

Burd, A., et al. 2005, NewA, 10, 5, 409-416
Cwiok, M., et al. 2007, ASS, 309, 531-535
Malek, K., et al. 2010, AdAst, 22, Article ID 194946
Cwiek, A., et al. 2014, SPIE 92900T
Mankiewicz, L., et al. 2014, RMxAC, 45, 7-11
Zarnecki, A. F., et al. 2012, SPIE 845408
Obara, L., Żarnecki, A.F., Zadrożny, A., Opiela, R. 2014, SPIE 9290, 92900K