A Pipeline for the ROTSE–IIIdArchival Data

B. B. Güçsav · C. Yeşilyaprak · S. K. Yerli · N. Aksaker · U. Kızıloğlu · D. Çoker · E. Dikiciolu · M. E. Aydın

Received: date / Accepted: date

Abstract We have constructed a new, fast, robust and reliable pipeline to detect variable stars from the ROTSE-IIId archival data. Turkish share of ROTSE-III archive contains approximately one million objects from a large field of view (1.85°) and it considerably covers a large portion of northern sky (δ > −25°). The unfiltered ROTSE-III magnitude of the objects ranges from 7.7 to 16.9. The main stages of the new pipeline are as follows: Source extraction, astrometry of the objects, light curve generation and inhomogeneous ensemble photometry. A high performance computing (HPC) algorithm has also been implemented into the pipeline where we had a good performance even on a personal computer. Running the algorithms of the pipeline on a cluster decreases analysis time significantly from weeks to hours. The pipeline is especially tested against long period variable stars with periods of a few hundred days (e.g Mira and SR) and variables having periods starting from a few days to a few hundred days were detected.

Keywords Software: Pipeline · Methods: data analysis · Telescopes: ROTSE–IIId

PACS (PACS codes)

1 Introduction

Robotic Optical Transient Search Experiment (ROTSE - Akerlof et al, 2003) is a network of telescopes located all around the world. The primary goal of the ROTSE–IIIproject is to observe Gamma-Ray Bursts (GRB) in optical light. Each ROTSE–III telescope consists of 45 cm with a wide field of view (1.85°). The telescopes were built to respond rapidly to GRBs (<10 s) which are triggered by satellites such as Swift, Integral and HETE. The ROTSE–III system runs unattended with fully automated observation, data acquisition and analysis (see Rykoff et al (2005) for the pipeline).

The ROTSE–IIId collaboration uses 70% of each ROTSE–III telescope’s observation time. The rest of the time is allocated for discretion by the local organization. The ROTSE–IIId telescope is located at TÜBİTAK National Observatory (TUG)², Bakırlıtepe, Antalya, Turkey. Scheduled observations were started in May 2004 and they were distributed among the Turkish astronomers by TUG. In this work, all of the public Turkish observations have been used. Some of articles which made of using the ROTSE–IIId data of Turkish share are as follows Baykal et al (2005), Bilir et al (2006) and Kızıloğlu et al (2009).

The aim of this work is to detect variability as well as finding new variables using the ROTSE–IIId archival data. To achieve this main goal a multistage algorithm was developed and then they are crafted into a series of routines: the pipeline. In section 2.1 ROTSE–IIId observations and structure of the data, in section 2.2 the summary of ROTSE–IIId pipeline are given. The data handling, our pipeline and its structure are given in section 3. The main stages of the pipeline are as follows: source extraction (§3.2), astrometry of field stars (§3.3). light curve generation (§3.4) and inho-
mogeneous ensemble photometry (§3.5). We conclude our work with results and some suggestions for future work.

2 ROTSE Data

2.1 ROTSE–IIId observations and structure of the data

The ROTSE–IIId has an 45 cm primary mirror and is outfitted with a 2k \times 2k TE cooled, CCD camera with 3.3"/pixels, making a 1.85\textdegree field of view. QE (Quantum Efficiency) of the CCD peaks at 550 nm. The telescope and CCD were described in Akerlof et al (2003).

The CCD observations have been carried out in three different exposure times: 5 (\textasciitilde 54\%), 20 and 60 seconds. There are two important magnitude limits for these exposures given in Table 1 saturation (mean of maximum brightness of stars having no saturated pixels in their Full Width at Half Maximum - FWHM) and limiting (mean of the faintest star’s magnitudes).

Table 1 Magnitude limits of ROTSE–IIId exposures.

| Exposure Time | Saturation Magnitude | Limiting Magnitude |
|---------------|----------------------|--------------------|
| 5 s           | 7\pm7                | 15\pm6             |
| 20 s          | 9\pm0                | 16\pm2             |
| 60 s          | 10\pm0               | 16\pm9             |

The unfiltered ROTSE–IIImagnitude of the objects, described in section Akerlof et al. (2003). The telescope and CCD were described in Akerlof et al (2003).

The CCD observations have been carried out in three different exposure times: 5 (\textasciitilde 54\%), 20 and 60 seconds. There are two important magnitude limits for these exposures given in Table 1 saturation (mean of maximum brightness of stars having no saturated pixels in their Full Width at Half Maximum - FWHM) and limiting (mean of the faintest star’s magnitudes).

The journal of observations used in the pipeline ranges between May 2004 and June 2010 and it consists of 234,764 frames from 645 different pointings. This size of the archived data from the Turkish share (regardless of the pointings) since 2004 is approximately 2 TiB.

266 pointings were observed less than 100 times and they were not included in the pipeline. This limiting value is an arbitrary choice. However, to have statistically reliable data sets we had to implement this minimum lower limit. This number were also used in other large volume ROTSE–IIIdanalyses (Wozniak, 2004).

2.2 The summary of ROTSE–IIId pipeline

The data acquisition system is constructed on top of several daemons (e.g. weather, clamshell, camera). The telescope is operated under two modes: alarmed and scheduled. The latter mode is used in the entire Turkish share and thus, in our pipeline. The ROTSE–IIId telescopes have a well designed data reduction pipeline allowing the near real-time processing of the CCD images taken by the entire ROTSE–IIId network. Regardless of types of observation carried out on the telescope, all images are fed into the ROTSE–IIId pipeline. Furthermore, CCD images of all observations (taken with either alarmed or scheduled mode) were automatically processed (bias subtracted, flat fielded and fringe corrected) immediately after the frame has been download to the disk (Rykoff et al, 2005). SExtractor software (Bertin and Arnouts, 1996) is then used to detect objects, to measure centroid positions and to determine instrumental magnitudes (using 5 pixels aperture). In addition to centroid positions, roundness and sharpness values are also used to eliminate non-star like objects. Instrumental magnitudes of each object in the frame is calibrated by comparing all the field stars against the “USNO–A2.0 R–band catalog” (Monet, 1998) (because the R filter is the nearest to the QE maximum) to obtain ROTSE–IIImagnitudes. By simply using triangulation method from the USNO catalog each object’s coordinate and instrumental magnitude are calibrated at the same time. The pipeline finally outputs files of calibrated object catalogs which are tabulated data in the form of binary FITS tables. The algorithm of the pipeline were coded in IDL and schematic representation is shown in Figure 1 (blue colored blocks). Thus, the frame (and each detected object in the frame) can now be used in light curve analysis using object’s \( \alpha, \delta, \) magnitude and error in magnitude.

3 The New Pipeline

3.1 Data Handling

Input data for the pipeline can either be a calibrated object file or any corrected image file of the ROTSE–IIId. This is required if the frame is not valid i.e the WCS (World Coordinate System) headers were wrong or there were recorded bad weather conditions, or technical problems. Otherwise the frame is valid and it enters to the stage-C of the pipeline. Thus, using these filtering methods, at least 80\% of the archive was handled with the pipeline. The algorithm of the pipeline shown in Figure 1 (yellow colored blocks). It seems that the first two stages of our pipeline duplicates the ROTSE–IIId pipeline (see Figure 1). However, since a) SCAMP has been used in astrometry calibration (see §3.3) and b) the parameter set that we use in our pipeline was not produced by the ROTSE–IIId, we had to introduce our own stages in the pipeline.
3.2 Stage-A: Source Extraction

ROTSE–IIIdcorrected images are used as INPUT data in this stage. A catalog of stars with their instrumental magnitude and frame coordinates is produced as an OUTPUT.

SExtractor code (Bertin and Arnouts, 1996) is used to identify (and to perform an aperture photometry) the stars in the frame. Four types of magnitudes are calculated by SExtractor where only one of them is used in the pipeline, namely MAG_APER (magnitude from aperture). A classical value of 5 pixels ROTSE–IIIdaperture (e.g. Kızıloğlu et al, 1995) was used in the SExtractor settings. The output of SExtractor consists of instrumental magnitude, frame coordinates and many statistical calculations for each star. By using the SExtractor’s statistical output we have improved the reliability of magnitudes calculated which wasn’t possible using ROTSE–IIIdpipeline; it contained only a few columns of information.

As an example; for the ROTSE–IIIdpointing of 0006+4305 there were 12,037 sources in “USNO–B1 R” catalog (hereafter USNO–B; Monet et al, 2005) for $m_R < 18$. When the SExtractor is applied to this pointing, depending on atmospheric conditions, number of sources varied between 279 and 11,399. Thus, the stage misses only 5% of the USNO-B sources.

3.3 Stage-B: Astrometry of Field Stars

The output of stage-A is used as an INPUT. Similar to the ROTSE–IIIdpipeline, the calibrated object catalogs are calculated as an OUTPUT.

The SCAMP (Bertin, 2006) software package is used to map USNO–B stars with the inputted field star coordinates so that a transformation matrix can be calculated. Instead of USNO–B catalog, GSC, UCAC or 2MASS catalogs could also be used.

SCAMP has been improved for CCD frames taken from large FOV. The calculated astrometrical error of the ROTSE–IIIdis approximately 1-2" for its FOV (less than its pixel size). Therefore, SCAMP seems to be a suitable tool for astrometric calibrations in our pipeline.

The pointing accuracy (or error) of the ROTSE–IIIdtelescope is given as approximately $1'$. (Akerlof et al, 2003). The mean of differences between “targeted frame centers” and “frame centers recorded in frame headers” gives the pointing accuracy of ROTSE–IIId. These values are approximately $\Delta \alpha = 5.1'$ and $\Delta \delta = 1.0'$. Since unaligned frames within a few arcmin (Bertin, 2006) can be handled with SCAMP, this error doesn’t effect the resultant astrometry. Due to the telescope optics (Güçsav, 2010), the best astrometric alignment of the frames were done within the central 1.79°diameter of the whole ROTSE–IIIdframe. Thus, stars falling only into this region were used (see Fig. 2).

Since, SCAMP cannot handle frames having excessive image deformations (e.g. bad focusing, weather effects, telescope instability) they are automatically discarded and no output is produced. This feature of SCAMP is also used as a “filtering” algorithm for faulty frames.

3.4 Stage-C: Light Curve Generation

The output of stage-B or the output of ROTSE–IIIdpipeline is used as an INPUT.

The output contains a list of equatorial coordinates and instrumental magnitudes of sources, namely it is called “calibrated object catalog” (see Fig. 1 grey colored...
box at the middle). Each inputted star’s light curve (instrumental magnitude vs. JD) is created as an OUTPUT.

In this stage, the aim is to follow each star’s equatorial coordinates in each frame throughout the whole time span and then collect the corresponding instrumental magnitudes at that coordinate. As in the stage-B, coordinates of USNO–B field stars are used to match calibrated catalogs. In order to do this, a fixed circular aperture (namely “matching aperture”; hereafter $\phi_m$) has to be chosen to scan through each catalog. The $\phi_m$ can be neither a low value (which increases the chance to miss the target star) nor a high value (which increases the chance to multi-match the target stars). Therefore, $\phi_m$ was chosen according to ROTSE–IIIId pixel scale (namely 3.3”/pixel) and to achieve a standard Gaussian photon distribution, 3 pixels range were taken. Thus, 10” was chosen as an optimum $\phi_m$ value ($\phi_{m_{opt}}$) which will be used throughout the pipeline [Güçsav, 2010].

As a side effect of the matching algorithm, especially in crowded fields, there might still be multi-matches in the field. In such cases, the nearest star in the calibrated catalog to the USNO–B star was chosen as the match. The pipeline can be fine-tuned manually for some of the overcrowded fields by decreasing $\phi_m$ to 6” to be able to decrease multi-matches. In a future version, this fine-tuning can be integrated into the pipeline by marking each field with a roundness value which would make it possible to apply different $\phi_m$ values to each field whenever it is necessary.

3.5 Stage-D: Inhomogeneous Ensemble Photometry

The output of stage-C is used as an INPUT. Cleaned light curves are created as an OUTPUT.

In classical photometry, a single comparison star is used effectively for many years by astronomers (see Henden and Kaitchuck, 1990; Young et al., 1991). However, in recent years differential photometry with “many comparison stars” has increased both reliability and accuracy of the light curves. Ensemble photometry is also a new kind of differential photometry which works with inhomogeneous CCD data sets (e.g. Honeycutt, 1992; Saesen et al., 2010). We have also implemented the ensemble photometry in the pipeline to decrease statistical errors in instrumental magnitude of stars. For example, light curve of a 10$^{\text{th}}$ magnitude star was found to be almost constant. According to SExtractor, mean of RMS flux errors of the PSF fitting to the star was calculated to be 0.0006. However, with the implemented ensemble photometry, we have reached to a scatter value of 0.0002. Note that, this level of noise is also related to the other frame statistics (see scatter-error graph of the pointing that this star is located - Figure 3), as the magnitude of the source decreases, scatter increases and therefore goodness of the PSF fitting decreases. As can be seen in the figure, while scatter of a 10$^{\text{th}}$ magnitude star is around 2 mmag, it increases to 130–200 mmag at 16$^{\text{th}}$.

The main aim of the technique is to find non-varying stars (i.e. reference stars) throughout the time span of the light curves. The objects that have non-star like shapes are all ignored by using SExtractor’s roundness and sharpness analysis (see §2.2). Main criteria in choosing the reference stars are (1) to choose the stars far enough from edges of the frame, (2) to have the star isolated from the others, (3) to have no flags set in the SExtractor’s output, (4) to have roundness value close to zero, and (5) to have sharpness value close to one. In order to not to foul the technique, light curves of mostly non-variable stars have to be used. Therefore, rough variability detection has to be applied to the light curves; namely ‘scatter-and-error’ relation of the light curve has to be calculated. According to the result of this relation the star can now be accounted as a reference. Afterward, a mean reference level is calculated from all chosen reference stars and it is used to calculate the relative (or differential) magnitude of other stars.

By applying the technique second time (starting from the scatter-and-error relation), reliability of reference stars is increased: stars of the first run is inputted as the star list in the second run. With this run, variation of reference stars is decreased and therefore the accuracy of the final cleaned light curves is increased. As an example, a reference star’s light curves before and after ensemble photometry applied are given in Fig. 4. As seen in the figure, after ensemble photometry (stage-D) applied, scatter of the light curve (i.e. sigma) decreased.
Fig. 3 Mean magnitude (in magnitude) versus mean of RMS flux errors (in mmag) of sources in an example pointing which contains 31,984 sources is given (see §3.5). The bright end of the graph is extended at upper left as an inset with the same units. Similarly, dim end of the graph is marked with grid lines.

This technique has the advantage of removing frame to frame background variations due to Moon light and unstable weather conditions. As a disadvantage of this technique, the absolute magnitude of stars cannot be calculated. A sample of the final light curve of a variable from our pipeline is given in Fig. 5.

3.6 The Pipeline Structure

The new pipeline works on a small cluster called Infinitus which made use of 8 different computers with 36 cores each having 8 GHz CPU speed. GNU/Linux operating system and Lustre 1.6.7.4 file system is used on all computers. Computers are interconnected with a gigabyte ethernet. The pipeline is mainly written in C language. Parallel algorithms have been used in every step of the pipeline which made use of MPICH library. In order to balance the CPU load between cores the same amount of star’s data (both light curves and star’s information) are stored.

The prototype of the pipeline was started with IDL. However, to be able to have a robust and fast pipeline, they are converted and parallelized into C-language. With the prototype, it took approximately one week to create light curves of a single medium crowded pointing with a 4 cored PC. This duration decreased to about 2 hours with the pipeline.

4 Conclusion

We have constructed a new, fast, robust and reliable pipeline to detect variable stars from the ROTSE–IIId archival data. The main stages of the pipeline were as follows: Object identification, astrometry of the objects, light curve generation and inhomogeneous ensemble photometry. For the first time in the ROTSE–IIIdarchive, a high performance computing (HPC) algorithm has been implemented into the pipeline. Depending on the data quality, either corrected CCD images (stage-A) or the calibrated object catalogs produced by the ROTSE–IIIdpipeline (stage-C) could be used in the pipeline. The last stage of the pipeline (stage-D), namely inhomogeneous ensemble photometry (implemented to the ROTSE–

---

3 http://www.mcs.anl.gov/research/projects/mpich2/
IIIarchive for the first time), gives relative light variations of the measured stars with high precision. Within 645 pointings (see section 2.1) of observations in 2004–2010, light curve of approximately one million stars are produced.

Work on identifying new variables and searching for periods of known variables are still in progress. The following statistical tests to detect variable stars have been applied to the light curves: Scatter-Error Analysis (commonly used), Abbe Index (Saesen et al., 2010) and Analysis of Variance (Wozniak, 2004). For period hunting the following techniques have been used: PDM (Phase Dispersion Minimization: Stellingwerf, 1978), Lomb–Scargle periodogram (Lomb, 1976; Scargle, 1982) and SIGnificance SPECtrum (SigSpec: Reegen, 2007).

A few thousands of light curves have passed from all three statistical tests mentioned above. These light curves most probably are un-identified variable stars.

Light curves of approximately ten thousands stars from 4 pointings are converted into phase-magnitude table. These phase graphs were then visually inspected for repeating patterns. Among these reduced list, 152 stars show variability and according to SIMBAD (Güçsav, 2010), 20 of them are unknown variables which are not classified yet.

As a result of an early version of the pipeline, Mira and SR stars known in SIMBAD have been examined. New actual periods of 78 Mira variables have been identified and 18 of them have a period for the first time (Yeşilyaprak et al., 2011) with early version of the pipeline. Also, approximately 300 SR stars are under investigation and 15 of them were examined in details by Dikicioğlu (2011). Tabulated values of minimum and maximum values of both amplitude and period are given in Table 2.

As a by product of the pipeline, deformation and/or degeneration of frames could also be detected; e.g. background variations due to Moon light and unstable weather condi-
Table 2: Quantitative limits of light curves created by the pipeline.

| Amplitude | Period (days) | Reference |
|-----------|---------------|-----------|
| Min. 0 | 0.17±0.02 | Güçsav (2010) |
| Max. 5 | 720.0±43.1 | Yesilvaprak et al (2011) |

The main aim of the pipeline usage is to find new variables, periods of known variables, and to classify these variables using the above period hunting techniques.

The pipeline can easily be adapted to other ROTSE–III telescopes which will increase the sky coverage and number of detected unknown variables, transients etc.

Acknowledgements This project utilizes data obtained by the Robotic Optical Transient Search Experiment (ROTSE). ROTSE is a collaboration of Lawrence Livermore National Lab, Los Alamos National Lab and the University of Michigan (http://www.rotse.net).

All observations were made with the ROTSE–IIId telescope and the archival data of ROTSE–IIId obtained at the TÜBITAK (Turkish Scientific and Research Council) National Observatory (TUG), so we thank to ROTSE–IIId Collaboration and TUG for the optical and archival facilities (TUG - ROTSE–IIId projects of Turkish observers).

We also thank Prof. Dr. Ü. Kızıloğlu for consulting, suggestions and helps.

This study was supported by TÜBITAK with the project TBAG-108T475.

This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France and cdsclient tool located at CDS and NASA Astrophysics Data System Bibliographic Services.

References

Akerlof CW, Kehoe RL, McKay TA, Rykoff ES, Smith DA, Casperson DE, McGowan KE, Vestrand WT, Wozniak PR, Wren JA, Ashley MCB, Phillips MA, Marshall SL, Epis HW, Schier JA (2003) The ROTSE–III Robotic Telescope System. The Publications of the Astronomical Society of the Pacific 115:132–140

Baykal A, Kızıloğlu Ü, Kızıloğlu N, Balman Ş, İnam SC (2005) X-ray outburst of 4U 0115+634 and ROTSE observations of its optical counterpart V635 Cas. Astronomy and Astrophysics 439:1131–1134

Bertin E (2006) Automatic Astrometric and Photometric Calibration with SCAMP. In: Astronomical Data Analysis Software and Systems XV, Astronomical Society of the Pacific Conference Series, vol 351, pp 112–115

Bertin E, Arnouts S (1996) SExtractor: Software for source extraction. Astronomy and Astrophysics Supplement 117:393–404

Bilir S, Güver T, Aslan M (2006) Separation of dwarf and giant stars with ROTSE–IIId. Astronomische Nachrichten 327:693–697

Dikicioglu E (2011) Multiple Periods for Semiregular (SR) Variable Stars. Master’s thesis, Ataýurt University, Erzurum, Turkey

Güçsav BB (2010) Detection of Different Types of Celestial Objects from Robotic Telescope Archives. Master’s thesis, Ankara University, Ankara, Turkey

Henden AA, Kaitchuck RH (1990) Astronomical photometry: a text and handbook for the advanced amateur and professional astronomer

Honeycutt RK (1992) CCD ensemble photometry on an inhomogeneous set of exposures. Astronomical Society of the Pacific 104:435–440

Kızıloğlu Ü, Kızıloğlu N, Baykal A (1995) ROTSE Observations of the Young Cluster IC 348. The Astronomical Journal 130:2766–2777

Kızıloğlu Ü, Ozbilgen S, Kızıloğlu N, Baykal A (2009) Optical and X-ray outbursts of Be/X-ray binary system SAX J2103.5+4545. Astronomy and Astrophysics 508:895–900

Lomb NR (1976) Least-squares frequency analysis of unequally spaced data. Astrophysics and Space Science 39:447–462

Monet DG (1998) The 526,280,881 Objects In The USNO-A2.0 Catalog. Bulletin of the American Astronomical Society 30:1427

Monet DG, Levine SE, Canzian B, Ables HD, Birn AR, Dahn CC, Guetter HH, Harris HC, Henden AA, Leggett SK, Levison HF, Lugtenbuhl CB, Martini J, Monet AKB, Munn JA, Pier JR, Rhodes AR, Riepe B, Sell S, Stone RC, Vrba FJ, Walker RL, Westerhout G, Brucato RJ, Reid IN, Schoening W, Hartley M, Read MA, Tritton SB (2005) The USNO-B Catalog. The Astronomical Journal 125:984–993

Reegen P (2007) SigSpec. I. Frequency- and phase-resolved significance in Fourier space. Astronomy and Astrophysics 467:1353–1371

Rykoff ES, Aharonian F, Akerlof CW, Alatalo K, Ashley MCB, Güver T, Horns D, Kehoe RL, Kızıloğlu Ü, McKay TA, Özel M, Phillips A, Quimby RM, Schaefer BE, Smith DA, Swan HF, Vestrand WT, Wheeler JC, Wren J, Yost SA (2005) A Search for Untriggered GRB Afterglows with ROTSE–III. The Astrophysical Journal 631:1032–1038

Saesen S, Carrier F, Pignolksi A, Aerts C, Handler G, Narwid A, Fu JN, Zhang C, Jiang XJ, Vanautgaerden J, Kopacki G, Steslicki M, Acker B, Poretti E, Uyttewelhoeven K, Gielen C, Ostensen R, De Meester W, Reed MD, Kolarczkowski Z, Michalska G, Schmidt E, Yakut K, Leitner A, Kalomeni B, Cherix M, Spano M, Prins S, van Helshoecht V, Zima W, Huygen R, Vandenbuss-
Scargle JD (1982) Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. Astrophysical Journal 263:835–853
Stellingwerf RF (1978) Period determination using phase dispersion minimization. Astrophysical Journal 224:953–960
Wozniak WSJVWTGV P R (2004) Identifying Red Variables in the Northern Sky Variability Survey. The Astronomical Journal 128:2965–2976
Yeşilyaprak C, Yerli S, Güçsav B, Aksaker N, Dikicioğlu E, Helvacı M, Çoker D, Aydın M, Dinçel B, Uzun N (2011) Long-Term Variations and Periods of Mira Stars from ROTSE–IIIId. Accepted by New Astronomy
Young AT, Genet RM, Boyd LJ, Borucki WJ, Lockwood GW, Henry GW, Hall DS, Smith DP, Baliumas SL, Donahue R, Epand DH (1991) Precise automatic differential stellar photometry. Astronomical Society of the Pacific 103:221–242