HEIDI: An Automated Process for the Identification and Extraction of Photometric Light Curves from Astronomical Images

M. Todd\textsuperscript{1}*, H. U. Wallon Pizarro\textsuperscript{2}, P. Tanga\textsuperscript{3}, D. M. Coward\textsuperscript{4} and M. G. Zadnik\textsuperscript{1}

\textsuperscript{1}Department of Imaging and Applied Physics, Bldg 301, Curtin University, Kent St, Bentley, WA 6102, Australia
\textsuperscript{2}IBM Research - Australia, 1060 Hay St, West Perth, WA 6005, Australia
\textsuperscript{3}Laboratoire Lagrange, UMR7293, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d’Azur, BP 4229, 06304 Nice Cedex 4, France
\textsuperscript{4}School of Physics, M013, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

ABSTRACT

The production of photometric light curves from astronomical images is a very time-consuming task, taking several hours or even days. Larger data sets improve the resolution of the light curve, however, the time requirement scales with data volume. The data analysis is often made more difficult by factors such as a lack of suitable calibration sources and the need to correct for variations in observing conditions from one image to another. Often these variations are unpredictable and corrections are based on experience and intuition.

The High Efficiency Image Detection & Identification (HEIDI) pipeline software rapidly processes sets of astronomical images, taking only a few minutes. HEIDI automatically selects multiple sources for calibrating the images using a selection algorithm that provides a reliable means of correcting for variations between images in a time series. The algorithm takes into account that some sources may intrinsically vary on short time scales and excludes these from being used as calibration sources. HEIDI processes a set of images from an entire night of observation, analyses the variations in brightness of the target objects and produces a light curve all in a matter of minutes.

HEIDI has been tested on three different time series of 50 images each of asteroid 939 Isberga and has produced consistent high quality photometric light curves in a fraction of the usual processing time. The software can also be used for other transient sources, e.g. gamma-ray burst optical afterglows, Gaia transient candidates.

HEIDI is implemented in the programming language Python and processes time series astronomical images in FITS format with minimal user interaction. HEIDI processes up to 1000 images per run in the standard configuration. This limit can be easily increased, with the only real limit being system capacity, e.g. disk space, memory. HEIDI is not telescope-dependent and will process images even in the case that no telescope specifications are provided. HEIDI has been tested on various Linux systems and Linux virtual machines. HEIDI is very portable and extremely versatile with minimal hardware requirements.

Key words: methods: analytical – methods: data analysis – methods: numerical – techniques: image processing – techniques: photometric – astrometry

1 INTRODUCTION

The process of analysing a time series of astronomical images to measure the variations in brightness of an object over time and to calibrate those variations against reference sources is a very time-consuming task. An accurate understanding of the variation in brightness of an object over time, i.e. its photometric light curve, allows modelling of key characteristics of the object (e.g. the shape of an asteroid, the type of a supernova or gamma-ray burst optical afterglow).

The measurement process often varies only little depending on the type of object, whether it is an asteroid

* E-mail: michael.todd@icrar.org (MT)
tumbling through space, a supernova or the rapidly fading optical afterglow from a gamma-ray burst. The process usually involves examining the images in the time series for at least one suitable reference source in order to calibrate the brightness. Some processes require the reference sources to have particular characteristics, e.g. a specific spectral class or magnitude. Other processes may need sources that have been identified as a photometric standard. If more than one calibration source is available, the precision can be improved enormously; however, this is also much more time-expensive. The process must be repeated for each image and corrections made from one image to another. These corrections can incorporate predictable adjustments such as extinction due to changing air mass. In addition, often the corrections that need to be made in order to obtain a consistent light curve are variable, non-uniform and unpredictable, and are based on experience and intuition.

The ability to quickly analyse a set of images and to produce a high quality light curve is particularly valuable on automated telescopes, e.g. the fully robotic Zadko Telescope in Gingin, Western Australia (Coward et al. 2010). Participation in programmes like the prompt follow-up of gamma-ray burst alerts for the detection of transient optical emissions as part of the Télescope à Action Rapide pour les Objets Transitoires (TAROT) network (Klotz et al. 2008) and the Gaia Follow-Up Network for Solar System Objects (Gaia-FUN-SSO, http://www.imcce.fr/gaia-fun-ssso) (Todd et al. 2013) results in the creation of data sets containing transient sources that need rapid analysis.

We have developed a High Efficiency Image Detection & Identification (HEIDI) pipeline software for processing images and analysing the variations in brightness of optical transients. We take the approach that all sources in an image are potential reference sources. The variations in the brightness of the sources in each image of a time series are assessed automatically. Corrections are made to compensate for these variations while also taking into account that some sources may intrinsically vary on short time scales (e.g. short period variable stars). To facilitate production of photometric light curves Solar System objects are identified automatically during the image processing. As a result the automated analysis of an image set and the production of a calibrated light curve is achieved in a matter of minutes. HEIDI thus fulfils an emerging need for the rapid analysis of time series image data.

2 DESCRIPTION OF THE AUTOMATED PIPELINE SOFTWARE HEIDI

HEIDI is implemented in the programming language Python (http://www.python.org) for Linux platforms. It processes astronomical images in FITS format and requires only a minimal amount of user-supplied detail. It can also be installed on Linux virtual machines. This makes the software extremely versatile.

Preprocessing tasks such as astrometric calibration and image alignment are accomplished using existing applications as described in §2.1. The main functions of identification and selection of calibration sources, identification of targets and production of light curves are accomplished using newly developed Python code as described in §2.2 and §2.3. HEIDI has been benchmarked on a Pentium class system and a virtual machine, i.e. low-performance, low-cost systems.

2.1 Astrometric calibration and image alignment

The first step in the image analysis process is to perform the astrometric calibration. The amount of information contained in FITS image headers depends heavily on the system in place at the observing site. While an integrated system would typically produce FITS images which include calibration information in the image headers, some sites may not even include the pointing information. We have therefore adopted a standard approach of uploading one image from the set to the astrometry.net server in order to obtain a new astrometric calibration (Lang et al. 2010). By using the blind astrometric calibration system provided by astrometry.net, HEIDI thus removes any reliance on the images having already had astrometric calibrations or having to verify the pointing information, orientation or scale.

After obtaining the astrometric calibration, HEIDI assumes that all the images in the time series have the same pointing and updates the FITS image headers. The World Coordinate System (WCS) information from astrometry.net is included in the headers of the updated FITS images. The images are then processed through SExtractor (Bertin & Arnouts 1996) in order to build a catalogue of all objects found in the image set. The SExtractor catalogue is subsequently processed through SCAMP (Bertin 2006) which produces image headers that are ready to be used by SWarp (Bertin et al. 2002). SWarp aligns the images and includes the appropriate header information from any translation or rotation applied by SWarp during the alignment process. Most of the processing time is taken up with the SExtractor-SCAMP-SWarp workflow. On a single-CPU Pentium-class system, the astrometric calibration and image alignment stage for a set of 50 images takes about 10 min (vs. 1.25 min for the automated production of the photometric light curve by HEIDI, §2.2, §2.3).

2.2 Correlation of sources

In the correlation stage the images from the astrometric calibration and alignment stage are reprocessed through SExtractor to produce catalogues of all the sources that have been detected across the entire image set. These catalogues are correlated by HEIDI to identify the sources which are common to every image in the time series. These common sources are the candidate reference sources for the photometric calibration of the image set.

HEIDI calculates the natural logarithm of the flux value for each candidate source as a representation of relative (instrumental) magnitude. We refer to the natural logarithm of the flux hereafter simply as log flux. Examination of the log flux value for each candidate source may show some variation across the image set. HEIDI tests the log flux variation of each candidate source and rejects sources as candidates if the variation is greater than a predetermined threshold. An examination of the log flux values for the candidate sources...
HEIDI: An Automated Process for Light Curves

in a time series of 50 images for asteroid 939 Isbergaj shows a variation common to all candidate sources (Figure 1). As this variation appears to affect every source in a consistent manner, we conclude that this variation can be analysed and corrected.

For a given time series of N images, HEIDI finds all sources that are common to all N images in that time series. These M sources are selected as candidates for photometric calibration and are tested for variability over the time series.

STEP 1: For the N images and M sources, HEIDI reads the flux values \(\phi_{ki}, k \in \{1, \ldots, N\}, i \in \{1, \ldots, M\}\) from the SExtractor catalogues and calculates the natural logarithm of the flux. We will denote the natural logarithm of the flux as \(\phi_{ki}, k \in \{1, \ldots, N\}, i \in \{1, \ldots, M\}\) and refer to it hereafter as log flux. The initial log flux set is hence \(\langle \phi_{ki} \rangle_{k \in \{1, \ldots, N\}}\), \(\phi_{ki} = \ln \phi_{ki}\) and contains \(N \cdot M\) values.

STEP 2: For each image \(k \in \{1, \ldots, N\}\) in the time series, HEIDI calculates \(M\) log flux means: For each source \(i \in \{1, \ldots, M\}\), the source \(i\) is excluded (denoted by \(\setminus\)) and the log flux mean of the remaining sources \(j \in \{1, \ldots, M\} \setminus \{i\}\) is determined:

\[
\langle \phi_{ki} \rangle = \langle \phi_{kj} \rangle_{j \in \{1, \ldots, M\} \setminus \{i\}} = \frac{1}{M-1} \sum_{j=1, j \neq i}^{M} \phi_{kj}, \quad k \in \{1, \ldots, N\}.
\] (1)

STEP 3: HEIDI determines the global maximum of the log flux:

\[
\varphi_{\max} = \max_{k,i} \varphi_{ki} = \max_{k \in \{1, \ldots, N\}} \varphi_{ki}.
\] (2)

For all \(k \in \{1, \ldots, N\}, i \in \{1, \ldots, M\}\) the difference between the global maximum log flux \(\varphi_{\max}\) and the log flux means \(\langle \varphi_{ki} \rangle\) (determined in STEP 2) are calculated:

\[
\Delta \varphi_{ki} = \varphi_{\max} - \langle \varphi_{ki} \rangle, \quad k \in \{1, \ldots, N\}, i \in \{1, \ldots, M\}.
\] (3)

This will be our correction factor.

STEP 4: To smooth out the variations across the image set, the log flux values \(\varphi_{ki}, k \in \{1, \ldots, N\}, i \in \{1, \ldots, M\}\) are adjusted, yielding an adjusted log flux set:

\[
\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\}} = \frac{1}{N} \sum_{k=1}^{N} \varphi_{ki}^*, \quad i \in \{1, \ldots, M\},
\]

\[
\max_k \varphi_{ki}^* = \max_{k \in \{1, \ldots, N\}} \varphi_{ki}^*, \quad i \in \{1, \ldots, M\},
\]

\[
\min_k \varphi_{ki}^* = \min_{k \in \{1, \ldots, N\}} \varphi_{ki}^*, \quad i \in \{1, \ldots, M\}.
\] (4)-(7)

STEP 6: For each source \(i \in \{1, \ldots, M\}\) HEIDI tests whether the deviation of the adjusted log flux values \(\langle \varphi_{ki}^* \rangle_{k \in \{1, \ldots, N\}}\) from the adjusted log flux mean \(\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\}}\) is within a predetermined tolerance \(tol\). This test rejects sources that are varying in brightness on short time scales (e.g. short period variable stars) or are otherwise not varying in accordance with the other candidates in the set (e.g. edge effects or vignetting). Only sources that pass this test will remain in the candidate list:

\[
| \max_k \varphi_{ki}^* - \langle \varphi_{ki}^* \rangle_{k \in \{1, \ldots, N\}} | \leq tol \quad \wedge \quad | \min_k \varphi_{ki}^* - \langle \varphi_{ki}^* \rangle_{k \in \{1, \ldots, N\}} | \leq tol, \quad i \in \{1, \ldots, M\}.
\] (8)

Sources that do not pass this test will be excluded from the candidate list. If \(X\) is the number of excluded sources then the reduced candidate list consists of \(M - X\) candidates. This concludes the first pass and reduces the original candidate list \(\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\}}\) to \(\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\} \setminus \{i_1, \ldots, i_X\}}\):

\[
\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\} \setminus \{i_1, \ldots, i_X\}} = \langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\} \setminus \{i_1, \ldots, i_X\}}, \quad i \in \{1, \ldots, M\} \setminus \{i_1, \ldots, i_X\}.
\] (9)

STEP 7: Repeat STEP 2 to STEP 6 for the reduced candidate list \(\langle \varphi_{ki} \rangle_{k \in \{1, \ldots, N\} \setminus \{i_1, \ldots, i_X\}}\).

This process of filtering variable sources is applied iteratively. The tolerance \(tol\) for the deviation test of the log flux (STEP 6) is 0.5, 0.2 and 0.1 for the first, second and third
pass, respectively. We found that additional iterations with finer tolerances do not significantly improve the result as one reaches the limit of measurement precision. The candidate stars that survive this filtering process are used in the photometric calibration of the image set for producing the light curve for the target as described in §2.3. In addition, the candidates that fail the test are recorded for further study.

Figure 2 shows a sample image from a time series of \( N = 50 \) images for asteroid 939 Isberga. This image contains exactly \( M = 100 \) sources that are common to all images in the time series, marked with a cross (\( \times \)). These sources are selected as candidate sources and are tested for their suitability to be used as calibration sources. Circles (\( \odot \)) indicate those sources that have been selected as calibration sources. The position of the target, asteroid 939 Isberga is indicated by a square (\( \square \)).

2.3 Identification of target sources and production of the light curves

HEIDI has been designed to facilitate the production of photometric light curves for Solar System objects (target sources). Following the selection of calibration sources as described in §2.2, the target sources are identified and selected. HEIDI uses the Institut de Mécanique Céleste et de Calcul des Éphémérides (IMCCE) Sky Body Tracker (SkyBoT) Cone-Search and Resolver services (Berthier et al. 2006) to quickly identify the known Solar System objects in the time series, our target sources. HEIDI processes the time series very quickly: For a set of 50 images it takes HEIDI only approximately 15 s to identify one target source, to test for its presence in the image set, to compile the list of the corresponding flux values and to produce the photometric light curve for this object, i.e. \( \approx 0.3 \text{ s per object per image} \). As a consequence we have implemented HEIDI to produce light curves for all of the objects found in the time series rather than just the primary target.

The IMCCE SkyBoT Cone-Search service identifies which Solar System objects lie within a specified field of view at a given epoch: HEIDI calls the Cone-Search service and obtains a list of the objects that may be found in the time series by providing the centre coordinates in right ascension and declination and the epoch of observation.

For each object returned by the Cone-Search service, HEIDI calls the IMCCE SkyBoT Resolver service to obtain accurate coordinates: HEIDI provides the name of the object and the epoch for the first image in the time series to the Resolver service. If known, HEIDI also provides the International Astronomical Union (IAU) Observatory Code for the site at which the images were recorded. The Resolver service then returns accurate coordinates for the specified object. If the IAU Observatory Code is provided then the Resolver service returns coordinates in the topocentric frame, otherwise the coordinates are given for the geocentric frame.

To determine the object coordinates in the remaining images of the time series, HEIDI finds the record for each object in the IAU Minor Planet Center Orbit (MPCORB) database (http://www.minorplanetcenter.net). The coordinates for each object are computed from the orbital elements in the MPCORB database on a per-image basis across the time series. These computations are much faster than obtaining coordinates from the Resolver service. It takes only a fraction of a second to calculate the coordinates for an object for all images in the time series, compared with several seconds per image when retrieved from the Resolver service. However, due to rounding errors in the computation, the coordinates provided by the Resolver service may be more accurate than the ones computed by HEIDI. HEIDI will thus determine the difference between the coordinates provided by the Resolver service and the ones computed locally and apply this offset as a correction to all computed coordinates across all images.

The respective coordinates in each image of the time series are then tested to determine the presence of the target sources. To minimise false detections without introducing false non-detections, this is done in a two-step process: In a first step, an area of 10 x 10 pixels around the coordinates of the target is tested for the presence of a source; if no source is found, the area is expanded to 15 x 15 pixels and the test repeated. It is not possible to determine whether a failure to detect a target source is due to the faintness of the object or an error in the position calculation. We thus assume that the primary target for the measurement and production of a light curve has a well-known orbit and also a good signal-to-noise ratio in the time series. If the object is detected in the images of the time series, the log flux of the object is recorded for each image. Plotting the log flux of the object and the log flux of the calibration sources without correcting for the variations between images
HEIDI: An Automated Process for Light Curves

Figure 3. Uncorrected log flux measurements from calibration sources (·) and target (+) in the time series for asteroid 939 Isberga. The calibration sources can clearly be seen to vary in accordance with each other. The target, asteroid 939 Isberga, follows a similar trend. The light curve is not readily apparent.

in the time series does not give a clear indication of the light curve of an object. Figure 3 shows the uncorrected log flux measurements from calibration sources and the target in a time series for asteroid 939 Isberga.

The log flux mean of the calibration sources identified in §2.2 is calculated for each image (similar to Equation 1 but this time over the entire candidate source set):

$$\langle \phi_{ki} \rangle_{i \in \{1, \ldots, M\}} = \frac{1}{M} \sum_{i=1}^{M} \phi_{ki}, \quad k \in \{1, \ldots, N\}. \tag{10}$$

The global log flux mean is also determined:

$$\langle \phi_{ki} \rangle_{i \in \{1, \ldots, M\}} = \frac{1}{NM} \sum_{k=1}^{N} \sum_{i=1}^{M} \phi_{ki}, \quad k \in \{1, \ldots, N\}. \tag{11}$$

This is used for reference and correction of variations across the time series.

For each image $k \in \{1, \ldots, N\}$ in the time series, the log flux values of the calibration sources are adjusted by adding the difference between the global log flux mean and the log flux mean for that image. For all $k \in \{1, \ldots, N\}$, $i \in \{1, \ldots, M\}$ the difference between the global log flux mean $\langle \phi_{ki} \rangle_{k \in \{1, \ldots, N\}}$ and the log flux mean $\langle \phi_{ki} \rangle_{i \in \{1, \ldots, M\}}$ is calculated:

$$\Delta \phi_k = \langle \phi_{ki} \rangle_{k \in \{1, \ldots, N\}} - \langle \phi_{ki} \rangle_{i \in \{1, \ldots, M\}}, \quad k \in \{1, \ldots, N\}. \tag{12}$$

For each image $k \in \{1, \ldots, N\}$ in the time series, the log flux value of the target object $\phi_{kt}, k \in \{1, \ldots, N\}$ (where $t$ denotes the target source) is adjusted by the same correction factor $\Delta \phi_k$. To smooth out the background variations across the image set, the log flux values $\phi_{ki}, k \in \{1, \ldots, N\}$ for the target object $t$ are adjusted:

$$\phi_{kt}^* = \phi_{kt} + \Delta \phi_k, \quad k \in \{1, \ldots, N\}. \tag{13}$$

Applying this correction to the time series for asteroid 939 Isberga shows a light curve that is readily apparent (Figure 4). This light curve for asteroid 939 Isberga produced by HEIDI clearly shows a primary period consistent with the expected rotation period of 2.9173 h (0.12155 d) (Molnar et al. 2008).

The consistency of the results achieved by HEIDI is further demonstrated by extending the analysis to multiple time series for asteroid 939 Isberga that were obtained on three separate dates. The light curves determined by HEIDI for these three time series are shown superimposed in Figure 5. It can be easily seen that the three curves follow a very similar pattern. These curves can also be overlaid on one another to

Figure 4. Corrected log flux measurements for calibration sources (·) and target (+) in the time series for asteroid 939 Isberga. The variation in the calibration sources is minimal ($< \pm 0.05$). The light curve of the target is clearly apparent.

Figure 5. The light curves produced by HEIDI for asteroid 939 Isberga using time series that were obtained on three separate dates show the consistency of the results. It can be easily seen that the three curves follow a very similar pattern. A primary periodicity consistent with the expected rotation period of 2.9173 h is evident.
show that both the period and the amplitude are consistent (Figure 6).

3 FUTURE DEVELOPMENTS

The current version (as of November 2013) has good astrometric accuracy but performs only relative photometry, i.e. does not produce calibrated magnitudes. We plan to add the capability for the production of light curves calibrated to a standard photometric system.

The preprocessing tasks of astrometric calibration and image alignment described in §2.1 are independent of the type of object being studied. The candidate sources that are rejected by the filtering process described in §2.2 warrant further study to determine whether they were excluded due to some intrinsic variability. These excluded sources can include non-moving sources that have characteristic light curves, e.g. variable stars, supernovae and gamma-ray burst optical afterglows. In the next version of HEIDI, we will include the ability to also produce photometric light curves for all the excluded objects as a standard function.

Uncertainties on the produced light curves can be inferred from the dispersion of the magnitudes of the calibration sources and will be used to include error bars in the output. If the target source is near, i.e. within 2 to 3 arcsec, a star of similar or greater brightness, it is possible that the star is selected instead of the target source. The next version of HEIDI will include a step to automatically exclude detections in which the position information falls outside a predetermined tolerance. In addition, an algorithm to detect moving objects will be implemented for targets with less accurate position information or unknown targets.

4 CONCLUSIONS

The High Efficiency Image Detection & Identification (HEIDI) pipeline software rapidly processes sets of astronomical images. It is very fast and reliable. HEIDI takes a set of images from a night of observing and produces a light curve in a matter of minutes.

HEIDI is implemented in the programming language Python. It has been tested on various Linux systems and Linux virtual machines. It has minimal hardware requirements and the installation process is relatively straightforward. HEIDI runs in virtual machines without impacting performance and can thus be installed on non-Linux systems (if a Linux virtual machine is available). The High Efficiency Image Detection & Identification (HEIDI) pipeline software we developed has a small footprint yet is very powerful and extremely versatile.

Given programmes like the prompt follow-up of gamma-ray burst alerts for the detection of transient optical emissions and the Gaia Follow-Up Network for Solar System Objects, which require prompt follow-up and rapid analysis of time series data, HEIDI will be very useful and provide a service and a speed that have not been available.

Testing HEIDI on multiple fairly large time series for asteroid 939 Isberga yielded extremely satisfying results. The light curves produced using time series obtained on three separate dates showed consistency with the three curves following a very similar pattern.

HEIDI will be installed at the Zadko Telescope in Gingin, Western Australia. This will allow us to analyse very rapidly all the images in the data archive of the Zadko Telescope and search for interesting variable sources. It will also allow prompt analysis and rapid issuing of results for time critical events such as gamma-ray burst optical afterglow emissions.

The HEIDI package is available from the first author on request.

ACKNOWLEDGMENTS

MT thanks Frédéric Vachier for assistance with photometry. MT thanks Andrew Williams for assistance with processing FITS image files. The work reported on in this publication has been supported by the European Science Foundation (ESF), in the framework of the GREAT Research Networking Programme. DMC is supported by an Australian Research Council Future Fellowship.

REFERENCES

Berthier, J., Vachier, F., Thuillot, W., Fernique, P., Ochsenbein, F., Genova, F., Lainey, V., Arlot, J.-E., 2006, Astronomical Data Analysis Software and Systems XV, 351, 367
Bertin, E., Arnouts, S., 1996, A&AS, 117, 393
Bertin, E., Mellier, Y., Radovich, M., Missonnier, G., Dide- lon, P., Morin, B., 2002, Astronomical Data Analysis Software and Systems XI, 281, 228
Bertin, E., 2006, Astronomical Data Analysis Software and Systems XV, 351, 112
Coward, D. M., et al., 2010, PASA, 27, 331
Klotz, A., Boër, M., Eysseric, J., Damerdji, Y., Laas- Bourez, M., Pollas, C., Vachier, F., 2008, PASP, 120, 1298
Lang, D., Hogg, D. W., Mierle, K., Blanton, M., Roweis, S., 2010, AJ, 139, 1782
Molnar, L. A., et al., 2008, Minor Planet Bulletin, 35, 9
Todd, M., Coward, D. M., Tanga, P., Thuillot, W., 2013,
PASA, 30, 14