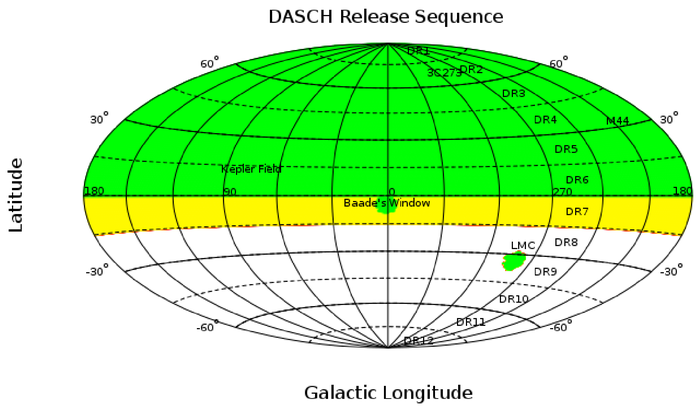
DASCH PROJECT UPDATE

Edward J. Los

The "Digital Acesss to a Sky Century @ Harvard" (DASCH) project is an effort to scan all of the glass plate negatives in the archive of the Harvard College Observatory for the purpose of generating lightcurves over a span of 100+ years from 1884 to 1992. Scanning began in 2006 after construction of a special purpose scanner capable of scanning ~50 plates per hour with a resolution of 11 microns. This report describes the latest DR6 public release, the ongoing challenges of working with this unique archive, and possible future directions.

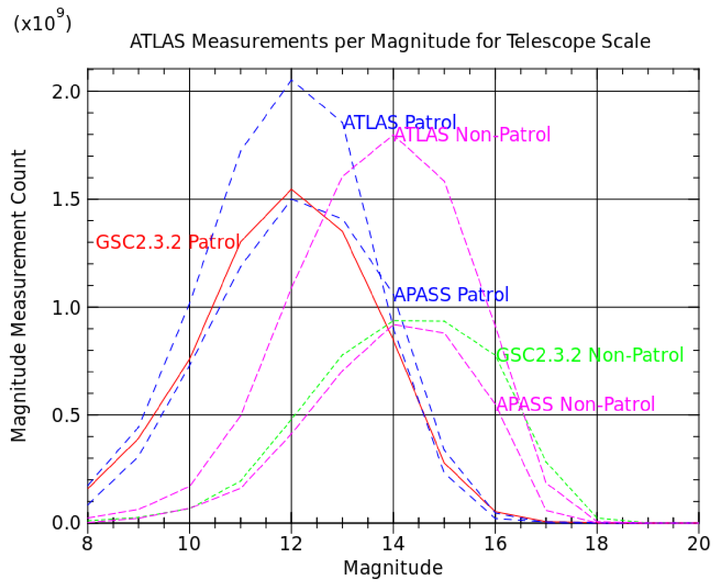
DATA RELEASE 6



Data Release 6, which became public on January 31, 2019, covers galactic latitudes between 0 and 15 degrees. This release represents a change in the original plans to avoid the disk of the Milky Galaxy until the end of scanning. The release area covers 5,200 square degrees of sky, more than that of any previous release. The release includes 54,000 new plates, an increase of 48% from the previous total. The number of images with calibrated photometry nearly doubles from 5.0 billion to 9.4 billion. Contributing to this increase is the decision to scan all of the 1200 arcsec/mm meteor patrol images instead of waiting until the end of the release cycle.

ATLAS refcat2 PHOTOMETRIC CALIBRATION

We are nearing completion of an eight month effort to reprocess of the DASCH data with the ATLAS refcat2 catalog (Tonry 1809.09157v1). This catalog uses PAN-STARRS data in the Northern Sky supplemented with recalibrated APASS data and other catalogs. As of May 22, 2019, a total of 292,348 plates have been processed with ATLAS refcat2 photometry and inserted into the database. These plates yielded 23,674,000,000 magnitudes matched with 287,229,000 stars to generate 78,091,716 lightcurves with more than 10 points.



The above preliminary graph shows that this new calibration catalog significantly increases the photometric yield of the DASCH images compared with data using the GSC2.3.2 guide star catalog and the APASS DR8 catalog. Note: GSC2.3.2 and APASS data uses Johnson B and V magnitudes while ATLAS refcat2 uses Sloan g and r magnitudes. Images calibrated with ATLAS refcat2 show a median lightcurve RMS of ~0.15 between mag 10 and 17, a result comparable to that of the other two catalogs.

RELEASE STRATEGY

Because of the magnitude of the scanning effort and the many project challenges discussed below, the DASCH project uses a monthly update cycle to make available any new or revised data in the released area. These revisions include new plates scanned, correction of errors, improved algorithms, and new calibration catalog releases. Each observation is tagged with a "versionId" which increments monotonically whenever revised data is inserted for any previously scanned plate.

The quality of the data varies dramatically across a wide range of telescopes and observing practices. Each observation includes astrometry and photometry accuracy estimates and a series of flags ("AFLAGS", "BFLAGS", and "QUALITY") for any data that is questionable or incomplete. Depending on individual research needs, these flags can be used to filter the data. Thumbnail fits files are available for a final check on image quality, and astrometric accuracy.

SCANNING STATUS

A flood from a broken Cambridge water main interrupted scanning in January, 2016. We resumed scanning in September, 2016 after replacing a major part of the scanner and all of the computers. Over the next two years 63,179 plates were conserved and returned to the stacks. Many of these plates suffered emulsion damage which will reduce the depth of the limiting magnitude of the photometry.

In early 2018, a low level background oscillation appeared in the scanned tiles. We traced this problem to a failure in the scanner CCD's thermoelectric cooler. A new cooler and heat sink configuration shows promise in eliminating the noise. Fortunately, the regularity of the noise will allow us to subtract the noise from the tiles and reprocess the mosaics.

Since the flood, we have been scanning 1500 to 2000 plates per week and hope to finish scanning by the end of 2020.

ARCHIVE CURATION STATUS

The scanner project depended on a parallel curation project. Because an unknown number of plates had been discarded in the 1950's there was no accurate count of the number of available plates. Only 130,000 card catalog records had been digitized and the best estimate of 530,000 plates was ~5% accurate. A crowd sourced project to digitize the logbooks ultimately produced a listing of 540,000 plates, but this total keeps evolving as errors and missing gaps are found and corrected. Finally, staff was tasked with going through the entire archive to add barcodes and record what was actually on the shelves. The latest total is 548,036 logbook entries for which 341,234 have been scanned. About 100,000 plates are pending, including 17,197 large format 14" x 17" plates and 2,349 5" x 7" or smaller plates. We do not plan to scan the 53,422 spectra plates and have rejected 18,602 additional plates not suitable for photometric calibration. We believe that 28,704 plates are missing and/or discarded.

FUTURE WORK

In parallel with the actual scanning we have implemented numerous projects to improve the accuracy, yield and access to the DASCH data. The primary challenges of the archive are (1) availability of suitable calibration catalogs, (2) a wide range of telescope fields of view, (3) the nonlinear photographic media, and (4) multiple exposure plates.

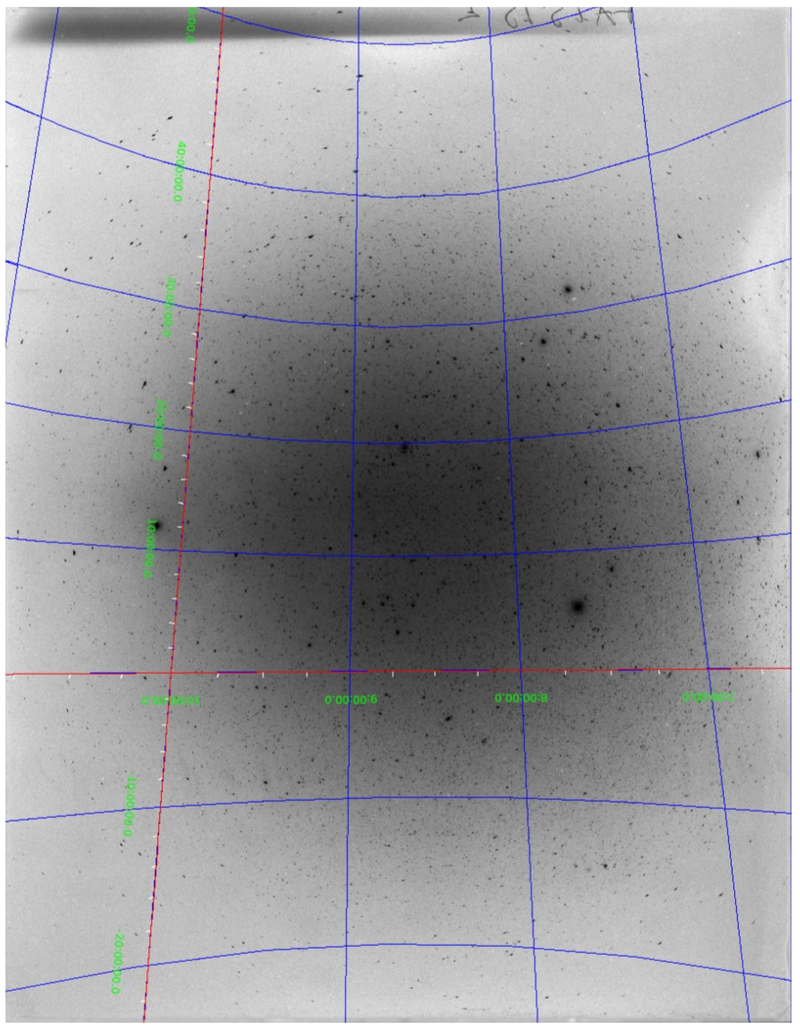
ASTROMETRIC AND PHOTOMETRIC CALIBRATION CATALOGS

An independent check of the scanner accuracy was provided by the opportunity to scan 840 Lowell Observatory plates of Pluto to assist the New Horizons mission. The residuals reported by the JPL are approximately 11 microns: one pixel of the DASCH digitizer camera. For a 600 arcsec/mm patrol plate, this error translates to 6 arcsec, which is enough to identify all of the stars on the plate. The GSC2.3.2 and Tycho-2 catalogs met this accuracy requirement.

The DASCH plates, however, required proper motions good to 6 arcsec over 115 years or 50 mas/yr. We found in retrospect that the Hipparchus proper motion data was not sufficiently accurate to meet this requirement. Only with the Tycho-GAIA release (which is included in UCAC5) did we have sufficiently accurate proper motions. The ATLAS refcat2 catalog uses GAIA Release 2 (Luri 1804.09376v1) proper motions which have enabled significant progress in understanding the evolution of the Milky Way.

The GSC2.3.2 catalog prevented accurate color correction of the DASCH photometry because it is a combination of multiple sources which were not properly calibrated with each other. We found that the Kepler Input Catalog improved our color corrections significantly and began working with APASS data as soon it became available. We believe that the ATLAS refcat2 release discussed above has the best combination of all-sky coverage and accuracy in SDSS g and r magnitudes. We look forward to upgrading the APASS photometry to APASS DR10 or later.

WIDE FIELD ASTROMETRY

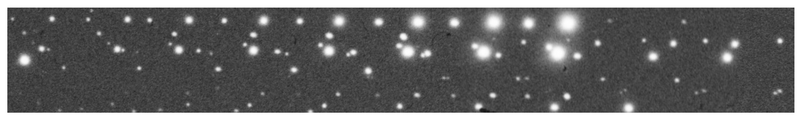


The pipeline was originally developed for the 5 degree x 7 degree 16 inch Metcalf Doublet plates which could be fitted with a 6th order polynomial algorithm provided by IRAF. The workhorse of the archive, however are the 600 arcsec/mm patrol plates covering 34 degrees x 42 degrees. Even larger are the meteor patrol plates such as fa13752 shown above at 1200 arcesc/mm. Most of these patrol plates had simple doublets which generated third order comatic images with high positional distortion at the edges of the plates.

NONLINEAR PHOTOGRAPHIC MEDIA

The nonlinear nature of photograph emulsions precludes the use of many standard procedures such as PSF fitting and deblending. We recognize that the required photometric accuracy depends on the type of variable being studied. We believe that the current piecewise linear fitting approach provides the best combination of accuracy and magnitude range. Since the lowess technique used by this approach is computationally expensive, pipeline performance might be improved by better statistical algorithms. (Steinhardt et. al., 2018).

MULTIPLE EXPOSURE PLATES



The Harvard Plate archive contains about 122,000 multiple exposure images such as plate mc05077 shown above. These multiple exposures were used to extend the dynamic range of the film, observe short-term variability, and to overlay calibration fields. Some multiples were generated by single exposures with the use of coarse gratings or sub-aperture wedges. We found that astrometry.net could separate these exposures providing that the fields were widely spaced in the sky. The astrometry.net software, however, will confuse closely spaced fields. These latter fields can be detected by correlation of of 2000 pixel wide fields around bright stars. However, there is still a need to find the convolution kernel for each type of multiple exposure. There is also a need to identify the members of each successive fit by altering astrometry.net to perform iterative fitting without intervening executions of the DASCH pipeline. This modification would provide astrometry.net with zones of avoidance when vetting prospective fitting solutions.

For the 8-inch Draper Doublet, Voigtlander Reworked by Clark (I series) and the 8-inch Bache Doublet, Voigtlander, reworked by Clark (B series) the outer zone images split into two in a direction perpendicular to the radial direction. These plates will require special handling of the Sextractor source lists to avoid confusion.

ACKNOWLEDGEMENTS

The DASCH project at Harvard is grateful for partial support from NSF grants AST-0407380, AST-0909073, and AST-1313370; which should be acknowledged in all papers making use of DASCH data.

We acknowledge the one-time gift of the Cornel and Cynthia K. Sarosdy Fund for DASCH, and thank Grzegorz Pojmanski of the ASAS project for providing some of the source code on which the DASCH web-interface is based.

The ongoing AAVSO Photometric All-Sky Survey (APASS) has improved DASCH photometric calibration and is funded by the Robert Martin Ayers Sciences Fund.

WEBSITES

Plate Stacks website: http://platestacks.cfa.harvard.edu
DASCH project website: http://dasch.rc.fas.harvard.edu
Umbrella site under construction: https://projects.iq.harvard.edu/dasch

2019 DASCH team: Upper Left Image Top Row: Melena Hogan, Sarah Lavallee, Grace Johnson-DeBoufre.
Bottom Row: Josh Grindlay, Lindsay Smith Zrull, Anne Callahan, Jamison Cloud. Center Image: Lindsay Smith Zrull, Meta Partenheimer, Mollie Tobin, Thom Burns, Anne Callahan, Lauren Klotzman. Upper Right Image Top Row: Ed Los, Matthew De La Ossa, Bob Simcoe, Ray Kenison. Bottom Row: Olivia Ambo, Katie Frey, Steve Siok, Jamison Cloud. Not pictured: Alisia True, Caroline Wolinsky, Chase Green, and Helen Driscoll

V6.0 2019-07-03