DIGITAL ACCESS TO A SKY CENTURY AT HARVARD: INITIAL PHOTOMETRY AND ASTROMETRY

S. Laycock\(^1\), S. Tang\(^2\), J. Grindlay\(^2\), E. Los\(^2\), R. Simcoe\(^2\), and D. Mink\(^2\)

\(^1\) Gemini Observatory, 670 North A’O’Hoku Place, Hilo, HI 96720, USA; silas@head.cfa.harvard.edu
\(^2\) Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA

Received 2008 September 26; accepted 2010 July 7; published 2010 September 10

ABSTRACT

Digital Access to a Sky Century at Harvard (DASCH) is a project to digitize the collection of \(\sim500,000\) glass photographic plates held at Harvard College Observatory. The collection spans the time period from 1880 to 1985, during which time every point on the sky was observed from 500 to 1000 times. In this paper, we describe the DASCH commissioning run, during which we developed the data-reduction pipeline, characterized the plates and fine-tuned the digitizer’s performance and operation. This initial run consisted of 500 plates taken from a variety of different plate series, all containing the open cluster Praesepe (M44). We report that accurate photometry at the 0.1 mag level is possible on the majority of plates, and demonstrate century-long light curves of various types of variable stars in and around M44. DASCH will generate a public online archive of the entire plate collection, including images, source catalogs, and light curves for nearly all astronomical objects brighter than about 17th magnitude.

Key words: astrometry – stars: variables: general – techniques: photometric

Online-only material: color figures

1. INTRODUCTION

Astronomical source variability is poorly explored on decadal and longer timescales, where large-scale systematic surveys are generally lacking. The Digital Access to a Sky Century at Harvard (DASCH) project (Grindlay et al. 2009) is designed to open the window on variability studies over a very wide range of timescales (\(\sim1\) day to \(\sim100\) years) as recorded on the \(\sim500,000\) glass photographic plates archived at Harvard College Observatory (HCO). The plates were exposed over the entire sky for a century (1880–1985) from small telescopes in the southern and northern hemispheres. This mammoth undertaking began at the instigation of the then director of HCO, E. C. Pickering, and has yielded a large number of landmark astronomical discoveries (e.g., Cepheid variables, quasars). Traditional visual inspection of the plates has only scratched the surface of this vast archive.

With our recent development of a very high-speed plate scanning machine (Simcoe et al. 2006), it has become practical to digitize the full collection. In this paper, we report on the initial digitization and analysis software as developed from scans of a field centered on the Praesepe (M44) open cluster. The \(\sim500\) plates scanned to include the cluster cover a \(50^\circ \times 50^\circ\) field (with radially decreasing coverage). Table 1 gives information on size, field-of-view, telescope aperture, observing location, and date range of these plates. From this sample we have developed the DASCH astrometry and photometric analysis pipeline. We describe the key features of this analysis system and present initial results for astrometric and photometric accuracy and completeness that can be achieved.

Large time domain surveys such as Pan-STARRS, Palomar Transient Factory, and the Large Synoptic Survey Telescope (LSST) are poised to revolutionize time domain astronomy. DASCH can explore much of the same science now, by opening HCO’s treasure trove of century-long observations to modern analysis methods.

The fundamental objective of DASCH is a precision (\(\leq0.1\) mag) database cataloging the changes in brightness of every resolved star in the sky brighter than about 17th magnitude, over 100 years (Grindlay et al. 2009). Such a goal requires careful control of many potential sources of inconsistency. Following the assignment of an accurate astrometric solution to each plate, the transmitted intensity scale must be calibrated to yield stellar magnitudes on a standard system. Photometric calibration entails three distinct challenges: (1) detecting all the stars recorded on the plate (without introducing spurious objects); (2) measuring the brightness of every star in a consistent and optimized manner; and (3) removing the effects of uneven illumination (vignetting), development, and emulsion irregularities. In addition, the calibration must be consistent between separate plates and between plate series, despite large variations in plate scale, exposure time, development, emulsion characteristics, telescope optics, and atmospheric conditions at the time of observations. In this paper, we describe how the above variables are accounted for in the DASCH photometry pipeline, followed by the extraction of light curves from the database.

We present light curves for interesting objects in the M44 field, selected to demonstrate the value of DASCH to a wide variety of science. The examples include an eclipsing binary (RY Cancri), an RR Lyrae pulsating variable (CQ Cnc), a Mira-type system (U Cnc), and the famous cataclysmic variable (CV) U Geminorum. Exciting new science is already being done with DASCH as demonstrated by the discovery of a new class of variable stars reported in Tang et al. (2010). A subsequent paper in preparation will report on a generalized large-amplitude variable search of the M44 region, and extends the pipeline with additional processing steps.

The entire data set of digitized images, scanned logsheets, metadata, star catalogs, and light curves is intended as a public resource. Details of data access and project status can be found at the DASCH Web site.

2. DIGITIZATION

The DASCH digitizer (Simcoe et al. 2006) is capable of scanning one 11 \(\times\) 17 inch or two 8 \(\times\) 10 inch plates in less than
a minute. The digitizer output is a set of 64 or 32 partially overlapping frames which are then mosaiced to form a single image of each plate. Mosaic generation uses the X-Y table encoder positions to align the images, which are averaged in the overlap region. The individual exposures are stored alongside the mosaics for users who may wish to examine them. The design and workings of the digitizer are fully described by Simcoe et al. (2006), its essential features are a precise (0.2 μm) X-Y stage carrying the plate, which is illuminated from below by a red LED array and imaged from above by a 4k × 4k CCD camera at 1:1 scale via a telecentric lens assembly, the exposure is controlled by the LED pulse duration (7 ms average). Digital images are captured via a 12 bit analog-to-digital converter (ADC), which thanks to sensitive control of the exposure level is sufficient to record the full dynamic range of the photographic negative. The transmitted intensity scale is then inverted to give a positive image for further analysis. The camera pixels measure 11 μm which due to the telecentric (zero-magnification) lens are able to capture details finer than the emulsion grains. The design specifications for the digitizer are such that every salient detail of the original data is recorded, and therefore the job of scanning the plates will not need to be repeated. We have already completed the job of digitizing the logbooks for the entire plate collection, and their transcription to electronic metadata is in progress.

3. ASTROMETRIC SOLUTION

An automated procedure based upon routines in WCSTools (tdc-www.harvard.edu/software/wcstools/) first determines an approximate world coordinate system (WCS) for the plate. This determines the correct orientation of the plate, the image scale (" pixel\(^{-1}\)) and fits a tangent plane projection (TAN) to produce a rough (10 ~ 20 arcsec) celestial coordinate frame. The TAN solution is used a starting point for SExtractor (Bertin & Arnouts 1996, hereafter SE) to detect all objects on the plate and assign initial R.A., decl. coordinates. The SE object table is matched against the Hubble Space Telescope Guide Star Catalog (GSC2.2; Lasker et al. 1990) with a 20" matching radius, keeping only the closest match to each star. A rough photometric calibration is derived for the entire plate by fitting a quadratic to the GSC2.2 (B) versus instrumental magnitude, this calibration is used to reject incorrect matches (disagreement greater than 1 mag) to produce a clean calibration catalog. Finally, the calibration catalog is used as input to the IRAF task CCMAP, which fits a more accurate plate solution to the (X,Y) versus (R.A.,decl.) data. We use the TNX plate solution (http://iraf.noao.edu/projects/ccdmosaic/tnx.html) which consists of a tangent plane projection plus polynomial terms to model geometric distortions in the image. Distortions originate primarily in the original telescope/camera optics with only minor contributions from the DASCH scanner. Our lens introduces distortions of order 1 pixel at the corner of the CCD frame. In the present work, this is mitigated by overlapping the frames and averaging to create the full-plate mosaic image. Following extensive testing of astrometric plate solutions it was found that a sixth-order polynomial distortion model in X and Y is required to produce a solution free of systematic residuals, while increasing the number of terms produces no further improvement. Examples of actual astrometric residuals for different plate series are shown in Figure 1, plotted as a function of distance from the center of the plate. The residual scatter (rms) of the final plate solutions is of order 1 pixel, which corresponds to between 1 and 5 arcsec depending on the plate scale. The final astrometric solution was written into the FITS headers and used for subsequent matching of objects against the photometric calibration catalog.

4. PHOTOMETRIC CALIBRATION

4.1. The Photographic Calibration Curve

The sensitivity of photographic materials is traditionally described by the characteristic curve of log(Exposure) versus log(Density) where exposure is modulated through a reciprocal arrangement of exposure time and incident light intensity; density refers to a measure of number of developed grains per unit area, and hence the degree of darkening of the emulsion. The characteristic curve is normally “S” shaped with a linear (in log–log space) portion in the middle, over which the aforementioned reciprocal relation is valid. At higher exposure values the linear regime breaks down, with increased exposure leading to an ever diminishing increase in density, a situation referred to as reciprocity failure. The combined effects of telescope optics and photographic material on the stellar point-spread function (PSF) have been discussed extensively elsewhere (e.g., Moffat 1969; Bunclark & Irwin 1984; Lasker et al. 1990).

Given the heterogeneous nature of the plate collection, and paucity of detailed records on emulsions, developers, and procedures at the different observing stations, it is safest to assume that the characteristic curve is different for every single plate. Contemporary photometric calibration was performed in a variety of ways. These included the following. (1) Secondary exposures of standard star fields. (2) Multiple exposures obtained by opening and closing the camera shutter for precisely timed intervals, with the telescope drive turned off in-between so as to produce calibration sequences from the brighter stars on the plate, to be compared against the fainter science-program stars. (3) The use of glass wedges and wire diffraction gratings to produce calibration sequences. These devices were placed over the objective lens to produce multiple images of the brightest stars on a plate. The diameters of these images are in an exact ratio to the primary images providing a calibration, as described by Chapman & Melotte (1913). (4) Calibration wedges (small sections of the plate exposed by a lamp, through filters of known density) are present on a few percent of the plates. None of these ingenious approaches are suitable for furnishing a uniform photometric calibration of the entire plate collection.

The obvious approach then is to bypass the characteristic curve altogether and calibrate instrumental magnitudes measured from the plates directly, against a catalog of known magnitudes, via a non-linear fitting function. Previously this method was used with plate data only for small regions of the sky where accurate magnitudes had been laboriously determined. However, with the advent of all-sky (or large area) star catalogs such as the Hubble Guide Star Catalog (GSC), Sloan Digital Sky Survey (SDSS), All Sky Automated Survey (ASAS), Two-Micron All-Sky Survey (2MASS), and others, reference stars are available everywhere, with a sufficient density to guarantee coverage of any particular plate. In fact direct calibration via on-plate reference stars is superior to all other methods one could imagine. Provided the reference stars are densely distributed and cover the full range of star brightnesses recorded on the plate, local variations in sensitivity, illumination, and PSF can all be mathematically fitted and corrected for. The fact that the reference stars lie on the plate means their light has followed exactly the same optical path and development conditions as their neighbors, removing all of the systematics inherent in using
separate standard stars. In particular, the effects of atmospheric absorption and atmospheric differential refraction are implicitly included in the calibration curve, provided the calibration star magnitudes are known in the same photometric band as recorded on the plate; or can be transformed into that band, as we describe below.

Following extensive research, the most accurate and complete reference catalog available for the full sky is the GSC2.2. Although itself derived from photographic plates, the underlying photometric system of the GSC2.2 is tied to a set of photoelectric magnitude calibrators (the Guide Star Photometric Catalog; GSPC) which were observed for every plate in the POSS-II survey (Reid et al. 1991; we cannot use the GSPC directly because it is very sparse). The GSC2.2 magnitudes are provided for all stars in a blue ($B$) passband and for nearly all stars a red ($R$) passband also. The GSC2.2 covers a large magnitude range $B = 1–19$ with quoted uncertainties of $\sim 0.2$ mag (Russell et al. 1990) and the bright end ($B < 10$) coming from the Hipparcos satellite catalog (Høg et al. 1997). The GSC-$B$ magnitude system turns out to be very close to the natural system of the majority of Harvard plates, which are primarily unfiltered blue-sensitive plates. The two GSC passbands also enable us to determine the wavelength response of each Harvard plate by fitting a color-dependent term to the calibration curve. The SDSS turned out not to be ideal for two reasons: DASCH requires a uniform calibration over the entire sky; the SDSS (currently and may be improved in future data releases) has a bright limit at $r \sim 14$ (Scheider et al. 2003) below which the magnitudes are not accurate due to detector saturation, making SDSS unsuitable for the magnitude range of the Harvard plates.

4.2. Object Detection

The first step in our photometry pipeline is to run SE (Bertin & Arnaults 1996; version 2.5.0) on the astrometrically calibrated digital plate image. SE has many features that make it ideal for our purposes. It is fast, a crucial factor when one considers the software pipeline should eventually keep pace with the scanning: SE takes 2–6 minutes for each 750 MB, 380 megapixel image on a single 3GHz CPU depending on the number of stars. Instrumental magnitudes are computed in fixed aperture, isophotal, and Kron flexible aperture forms. Source radii at specified flux levels are written, providing an analog of the methods used by photometrists during the photographic era. Sky background is modeled and available for inspection: a useful feature for examining the degree of vignetting and large-scale emulsion defects. Deblending is performed to account for contamination of magnitudes by neighboring stars. We run SE with a low detection threshold (1.5 $\sigma$) in order to be sure of detecting all potentially real objects. Due to the grainy nature of photographic emulsion and strong spatial gradients, standard image-noise models are not appropriate for detection thresholding, so we select the real objects after the fact, using a robustly derived limiting magnitude cut (described below). Some of our key SE parameters are provided in Table 2, the final output is relatively insensitive to most of the SE options since we determine measurement uncertainties by reference to a large catalog of calibration stars. The SE output catalog forms the basis of the actual measurements that go into the database. It is matched against the GSC2.2 catalog using a 5 arcsec matching radius to create the calibration sequences for each plate.

4.3. Fitting the Calibration Curve

The methods in use for CCD photometry all rely on the summation of pixel values belonging to each star. Under the assumption of a linear detector the derived flux ($F$) is directly proportional to the actual photon flux received from the star. Calibration of CCD measurements is therefore simply a matter of determining the constant of proportionality, which takes the
form of a zero-point offset \((Z)\) when the expression is formulated in magnitude units, as in Equation \((1)\):

\[
M_{\text{CCD}} = Z - 2.5 \log_{10}(F).
\] (1)

As described in Section 4.1, photographic magnitudes follow a non-linear calibration curve, the shape of which depends (in addition to the physical properties discussed in Section 4.1) upon the measurement technique. For DASCH commissioning purposes we selected three methods: fixed aperture, isophotal, and image size. A more sophisticated profile-fitting technique will likely be implemented in future. However, dealing with the spatially variable and peculiarly shaped PSFs found on the plates means that it will be limited to a subset of the data, while our present aim is to utilize as much of the collection as possible. An example of DASCH calibration curve is shown in Figure 2.

In fixed aperture photometry, \(F\) is the sum of all pixel values lying in a circular aperture centered on the star’s center of light (centroid). The functional form of this particular calibration curve has been determined by Bacher et al. (2005), in terms of four parameters which are in principle measurable from the image, or may be fitted for. The Bacher function enables very precise calibrations using only a limited number of reference stars, which need not extend much fainter than the magnitude at which there cease to be saturated pixels in the cores of the stars. Unfortunately, fixed apertures mean that whatever aperture size is chosen will only be the optimum choice for a single point along the curve. Apertures which are too large are dominated by sky noise and stellar crowding, while at the bright end too small an aperture will capture only saturated pixels, leading to a flat (and therefore useless) calibration curve.

The most widely used method in the photographic era was measurement of the image diameter through a microscope, fitted with a reticle eyepiece or iris diaphragm. The diameters of standard stars would be measured and compared to those of the program stars. This was (and still is; for example, Berdnikov et al. 2007) performed by eye. Interpolating between closely spaced standard magnitudes, a level of accuracy approaching 0.1 mag is achievable by the best practitioners, although the work is painfully slow and requires a high degree of skill and experience. For comparison purposes we created an analog of this method, by calibrating our reference stars against their image diameters (in fact image area, to account for the non-roundness and spatial variation of the PSF), measured at a specified threshold. Previous examples of microscope-measured

| Parameter Name | Value | Definition |
|----------------|-------|------------|
| DETECT_THRESH | 1.5   | Detection threshold in units of \(\sigma\) above local sky |
| DETECT_MINAREA| 5     | Minimum number of contiguous pixels above threshold to qualify as an object |
| DETECT_TYPE   | CCD   | DASCH images are captured in units of transmitted intensity, not photographic density |
| MASK_TYPE     | CORRECT | In computing source flux, pixels shared with a neighboring star(s) are excluded |
| CLEAN         | N     | Objects in the wings of bright stars are not excluded |
| FILTER        | Y     | Convolve image by a smoothing kernel to suppress noise before detecting objects |
| BACKPHOTO_TYPE| LOCAL | Use a background annulus around each source to compute its net flux |

**Notes.** Only parameters that affect the output are given. Most others were left at defaults.
image diameters calibrated by on-plate standard stars via a fitting function are provided by Hortnagl et al. (1992), Tinney et al. (1993), who used splines to parameterize the curve. Our method is similar but of course the image diameters are measured by software (SE) rather than micrometer.

Isophotal photometry (ISO) can be thought of as a combination of the preceding methods: a threshold is specified in terms of the local sky background, and all contiguous pixels above this threshold are assigned to the star. This technique is implemented in SE and was inherited from the previous generation of digitizers such as the SuperCosmos Sky Survey (Hambly et al. 2001a). A complete discussion of DASCH crowded field photometry is given by S. Tang et al. (2010, in preparation).

For our sample plates, we show the rms distributions resulting from each of the three techniques in Figure 3. In order to calibrate the instrumental magnitudes we need a fitting algorithm: in the case of fixed aperture there is a known functional form (Bacher et al. 2005), but for ISO and image size there appears to be no universal function. The literature contains many approaches, ranging from polynomials and splines (e.g., Tinney et al. 1993). A very sophisticated algorithm is described by Russell et al. (1990) for the GSC, derived from photographic plates of the Second Palomar Schmidt Survey (POSS-II) that attempts to linearize the response by determining the characteristic curve from the star images themselves.

After extensive testing with simple fitting forms, we found that no polynomial model works consistently well. We therefore turned to interpolation schemes which require no assumed function yet can be very robust against outliers. After some experimentation with box-car average and kernel density estimator approaches which allow the code to interpolate smoothly across gaps in the calibration sequence, we identified the statistical-analysis function \texttt{rlowess} (Cleveland. 1981), variants of which are incorporated in several widely used packages including “R” and Matlab. The algorithm takes a series of \((X, Y)\) points such as the calibration data \((B_{GSC} \text{ versus instrumental DASCH magnitude})\) and derives a smooth curve through the points. The only control parameter is a smoothing scale that defines the proportion of input points used to determine each output value; we chose 0.2. The algorithm is robust against outliers, applying an iterative sigma-rejection routine to identify and eliminate bad points from the fit. The (tabulated) value of the \texttt{rlowess} curve is then used to convert (by interpolation from a very finely spaced grid) the instrumental ISO magnitude into a calibrated magnitude in the natural system of the plate. An example of a typical \texttt{rlowess} fit to ISO magnitudes is given in Figure 2, illustrating several important features: the quality of fit is seen in the residuals (lower panel), where the residuals are used to define an error envelope in order to attach uncertainties to each measurement. Extrapolation of the fit at the bright end is sometimes required if the plate contains too few bright stars to define a reliable curve. Due to the grain inherent in photographic materials, which is often visible at size scales similar to faint stars, the noise models used internally in SE (and indeed all other CCD photometry routines) are not strictly applicable and we therefore run the star-detection stage with a very low threshold to be sure of picking up all potentially real objects. False detections are dropped at the calibration stage, as most are not matched with reference stars and hence get ignored. Remaining false detections are outliers to the calibration curve and are removed by the \texttt{rlowess} algorithm. All uncertainties are robustly derived by comparison of the final results against the reference magnitudes, by which we mean the fit residuals define the error distribution. The limiting magnitude is determined by locating the point at which the scatter \((\sigma)\) of the fit residuals begins to dominate over the slope of the curve; in practice, we chose the point that is \(2\sigma\) (two times mean standard deviation of the residuals) brighter than the faintest non-flagged (see Section 4.6) stars that survived \texttt{rlowess} rejection. Histograms of limiting magnitude for five different plate series are provided in Figure 4.

4.4. Annular Photometry—Correcting for Radially Symmetric Vignetting and Point-spread Function Degradation

The calibration curve is derived as a function of radial distance from the optical axis (physical center) of the plate. This enables allowance for the effects of vignetting and radially dependent PSF variations. The different plate series exhibit these effects...
to varying extents, which are greatest on those with a large plate scale. Vignetting cannot be removed by flat fielding as with CCD images because the plate collection does not contain contemporary flat-field exposures, and the non-linear response precludes using flats constructed from the images themselves.

The design of the telescopes and cameras provided a wide area of full illumination on the plates. Figure 5 shows the average illumination pattern for two plate series, MC and RH, which are representative of the moderate-field (5°–7°) and wide-field (20°–25°) plate series. Each illumination pattern was generated by median combining (with rejection of 3σ outliers) about 100 digitized plate images after first normalizing them by their mode pixel-brightness values. Thus, stars, defects, clouds, and uneven development were averaged out. An approximately radially symmetric vignetting pattern is evident in the wide-field plate series. The illumination falls off by only 5% from the center to the edge of the minor axis, and by 10% along the major axis.

This pattern bears extremely favorable comparison with modern instrumentation, especially when one considers the huge field of view (25° across). The moderately wide field (e.g., MC plate series) exhibits a perfectly flat illumination pattern, which enables second-order effects to be visible. In particular, there is a narrow border around the image which is slightly brighter, and is probably an artifact of development or ageing. A checker pattern due to the digitizer frames is also barely visible in the figure, contributing a variation of less than 0.5%.

PSF distortion is generally minor over most of the plate. It increases with distance from the plate center, being most pronounced in the plate corners, as can be seen in Figure 6. The effect on photometry is to change the fraction of light from a star that falls within a fixed aperture, such that fluxes tend to be underestimated as one moves off-axis. Isophotal photometry alleviates the effect somewhat because all pixels above the threshold are included in the flux measurement, regardless of the shape of the profile. Using a low threshold ensures that flux loss occurs far out in the wings of the profile where it is less significant. Both these effects change the shape of the calibration curve, and also the limiting magnitude, since reduced effective exposure leads to diminished sensitivity.

The combined radial effects on photometry can be seen in Figure 7 which plots the rms distributions of light curves constructed from measurements located in specific radial bins. For this analysis we used M44 as a probe. The plates are all centered at different positions on the sky, so the cluster moves around in plate coordinates—enabling one to select points from each star’s light curve containing only measurements made when it lay within a specific distance of the plate center. Using only the M44 cluster-members meant we could calibrate our DASCH magnitudes against the WebDA catalog (www.univie.ac.at/webda) which provides photometric magnitudes for 300 stars (all confirmed cluster members) with errors of ~0.02 mag. This enables
Figure 8 shows the photometric errors are dominated by the outer regions, compared to those for the entire plate. It is evident from calibration curves for a plate divided into four equal-area annular regions centered on the optical axis. Within each region the illumination and PSF are assumed constant. In addition, a rectangular border around the plate is used to flag stars likely to be affected by edge damage and uneven developer action (within a few millimeters of the plate edge, developer can penetrate horizontally into the emulsion, resulting in increased development). The annular calibration scheme is demonstrated in Figure 8 which shows the calibration curves for a plate divided into four equal-area annular regions, compared to that for the entire plate. It is evident from Figure 8 that the photometric errors are dominated by the outer region of the plate, and are reduced by the annular approach. In this example, the mean photometric residual rms = 0.473 mag for the whole plate; however, its value for the annular bins is as small as 0.206 for the inner 25% of the plate by area, and is 0.4 mag for the outer 25% which is still an improvement over the whole-plate result.

Following experimentation separately with the number and placement of the annular calibration regions we found that eight concentric equal-area bins are optimum in reducing the rms. The outer border of the plate forms a ninth bin whose width is 10% of the plate’s minor-axis length, stars in bin 9 are flagged for caution in subsequent analysis.

4.5. Local Calibration—Correcting for Localized Spatial Variations in Plate Sensitivity

In addition to vignetting (dealt with in Section 4.4), there are sometimes irregularities in the emulsion on scales of ~2 cm. Such features are evident to the eye and can be more clearly seen in the background maps and source-subtracted images output by SE. A possible origin is insufficient agitation during development and fixing of the plate. Agitation involves rocking the plate or the chemical tank either by hand or mechanically, in order to continuously redistribute the chemicals and prevent exhausted developer from remaining in contact with the emulsion. Any spatial differences in development result in changes to the characteristic curve, and hence the sensitivity and contrast, while incomplete fixing leads to foggy areas. On wide-field plates, local variations in calibration can also arise from partial cloud cover during the exposures as well as differential airmass.

By generating a map of the calibration residuals for each plate, we apply a third level of correction to the calibration process to remove irregular spatial variations. A regular grid of 50 × 50 positions in plate coordinates is defined covering the plate. At each position, the average calibration residual is computed for all stars within a “cell” of 0.5 radius. The exact statistic is a clipped median (three iterations, removal of 3σ outliers) including only stars brighter than the local limiting magnitude. The chosen reference catalog (GSC2.2) and sample radius provide ~100 stars per grid cell in the vicinity of M44, enabling residual maps to be generated on the required scale. The cell radius and grid spacing are chosen to oversample the spatial variations evident in typical images and error maps. Once the map is generated, each star’s magnitude is corrected by the clipped-median residual of the nearest grid point. From the above description it should be clear that the maps are inherently smooth due to the cell radius being larger than the grid spacing, which ranges from 0.1 on MC plates to 0.5 on RH plates. The physical spacing of the cell centers is 0.2 inches on the 8 × 10 inch plates. Cells containing fewer than 10 stars are replaced by the mean of the eight neighboring cells. Cells whose centers lie within a cell radius of the plate edge are more likely to be underfilled; this affects only the outer region of bin 9 which is already flagged (Section 4.6) in the photometry.

Maps of the calibration residual, its inherent uncertainty, and number of points per cell are shown in Figure 9, which also includes a map of the final residuals, following correction. It can be seen that the annular calibration described in Section 4.4 has removed all trace of vignetting, and the local calibration then drives the remaining spatial variations down by about 50%. An analysis of light curves generated from this photometry (see Section 6, Figure 13) shows the same level of improvement, demonstrating that spatial variations are handled effectively by the DASCH pipeline. A future refinement will be accounting for local variations in the slope of the calibration curve in addition to the magnitude-independent shift handled here.

4.6. Flags

Certain plate defects can be difficult to completely correct for, but fortunately only a small fraction of stars are affected. Examples include emulsion irregularities, dirt, and optical distortions. Fortunately these effects tend to be concentrated in the outer edges of the plates, which are most susceptible to handling, and lay furthest from the optical axis (relevant to distortion) of the telescope. In our photometric pipeline, we flag stars for the following factors: distance from plate center, proximity to plate edge, quality of rlowess calibration-curve fit, quality of local-map correction (σ and number of local reference stars used), local limiting magnitude, SE flags (blend, neighbor contamination), existence of known near neighbors. All of these conditions are evaluated on a plate-by-plate and star-by-star basis, enabling very specific filtering of the DASCH database to investigate
Figure 8. Improvement in photometric accuracy with radial correction. By fitting the calibration curve in a series of concentric annular regions on each plate, radially dependent effects (primarily vignetting) are greatly reduced. The calibration curves shown are four concentric annular equal-area bins. For comparison, the top panel shows the calibration curve for the whole plate if radial effects are ignored. The improvement affected by the annular approach is evident from the visible scatter and rms values, and it is clear that the errors are dominated by the outer regions of the plate (bin 4).

(A color version of this figure is available in the online journal.)

Figure 9. Local photometric calibration map correction of irregular spatial variations. The top left panel shows a map of calibration residuals (GSC2.2 − DASCH) following annular-region fitting of the calibration curve; each pixel shows the clipped median of the residuals for stars within a 0.5 radius. Bottom left panel shows the number of stars included in the median calculation, while the bottom right panel shows the rms of the residuals. Finally, the result of applying the calibration map is shown in the top right panel.

Systematic effects and to include/exclude data points for scientific analysis. Further advances in filtering will be described in a future paper, for now we are conservative in flag setting, such that any unflagged point is overwhelmingly likely to be a real star with reliable properties. In this paper, most numerical results include only data points meeting our default quality thresholds: (1) star not within 10% minor-axis distance of plate edge; (2) total rms of calibration-curve residuals less than 1 mag; (3) magnitude brighter than 2σ limiting magnitude for relevant annular bin; and (4) plates without multiple exposures, gratings, prisms, or filters. Generally in plots all points are shown, where possible smaller plot symbols indicate flagged points.
5. COLOR RESPONSE OF PHOTOGRAPHIC EMULSION

The majority of the Harvard plate collection are blue-sensitive, unfiltered plates. A small fraction used filters to produce red- and yellow-sensitive measurements, although details of the wavelength response of the various emulsions are not available. In order to generate consistent and precise magnitudes for temporal analysis it is essential that changes in color response from one plate to another be well understood, otherwise small brightness variations could be masked (or mimicked) by calibration shifts.

Any given plate will yield magnitudes on a unique “natural system” defined by the combined wavelength-dependent transmission properties of the atmosphere, telescope, and camera optics, and the emulsion response. Fortunately all of this can be reduced to a single co-efficient: the color term \( C \), which appears in Equation (2) relating magnitudes on the standard system, \( M \), to the plate’s system \( m \), where \( M_1 \) and \( M_2 \) are two standard passbands close to \( m \). This simple form assumes that a given star’s incident spectrum is approximately linear over the effective bandpass of the system, and can therefore be completely described by the photometric color index \( (M_1 - M_2) \).

Since real stars have spectra of differing shapes, the chosen standard passbands should ideally bracket \( m \) in effective wavelength. This discussion is equally applicable to CCD photometry and explains why the best results are obtained when standard stars are chosen to be of similar color to program stars:

\[
m = M_1 + C(M_1 - M_2). \tag{2}
\]

For every digitized plate then, the first task is to measure how its natural magnitude system compares to the standard one, by determining \( C \). Our procedure was to step through a series of closely spaced \( C \) values. For each trial value, we converted the reference stars’ magnitudes according to Equation (2) and then fitted the instrumental plate magnitudes against \( M \). For example, when using reference stars from the GSC2 catalog, \( M_1 \) and \( M_2 \) are Johnson \( B \) and \( R \) magnitudes, respectively, and we have Equation (3). When the correct color term is applied, the scatter of residuals (standard deviation of \( \Delta m \) computed for every star) about the calibration curve reaches a minimum, as demonstrated in Figure 10:

\[
\Delta m = m_{\text{DASCH}} - B = C(B - R). \tag{3}
\]

The color term for each plate is determined using a region near the center where the calibration curve is most cleanly defined. Color response may vary slightly across a plate due to differential airmass, which would be most pronounced on wide-field plates. The effect is dealt with in the next release of the pipeline (S. Tang et al. 2010, in preparation). During this commissioning project we analyzed \(~500\) mostly blue plates from six different plate series, spanning 100 years, purposely including a few yellow and red filtered plates. Reference stars were drawn from the GSC2.2 which provides \( B \) and \( R \) magnitudes accurate to \(~0.2\) mag. For a more precise comparison analysis over a small area centered on M44, we used the WebDA (www.univie.ac.at/webda) catalog which provides precise \((0.01)\) magnitudes in \( B \) and \( V \) for about 300 M44 cluster members. The results of the color analysis show systematic differences in \( C \) between the different blue-sensitive plate series, at the level of about \(~0.1\) mag. This translates to an order of magnitude smaller systematic shifts between the natural magnitude systems, since the majority of M44 cluster members have \((B - V)\) colors less than \(~0.5\) mag, resulting in typical \( \Delta m \) \(~0.05\). These results are displayed in Figures 11 showing how the color terms are distributed according to plate series. It was initially expected that early plates might be significantly bluer as panchromatic emulsions became widely available only in the 1900s, and that different batches might exhibit different color sensitivities, not to mention the largely unknown issue of filters.

It comes therefore as a surprise that histograms of the number of plates versus \( C \) (Figure 11) show only small differences between plate series. The width of each histogram is of order \(~0.1\) mag, which is comparable to the differences between their modes. The small shifts are real however and the RH histogram shows an intriguing “shoulder” which upon closer examination is due to a change in color response in 1932. In Figure 12, we compare the scatter in DASCH magnitudes resulting from three different color-term correction strategies. For most stars (and plates) variations in color response contribute only a few hundredths of a magnitude, but for very blue stars in particular, the effects can be large.

6. LIGHT-CURVE EXTRACTION

Light curves are the ultimate purpose of DASCH and also provide an independent way to verify the quality of the astrometry, photometry, and metadata. The photometric calibration pipeline described above produces a catalog of stellar magnitudes and positions over time. To generate the light curve of a given star, its brightness history as recorded in the set of plate catalogs is ordered against observation time as recorded in the original logbooks. The information held in the logbooks consists of a one-line handwritten entry for each plate, noting the date and time of observation, exposure time, R.A. and decl. of the plate center, hour angle (HA), plus comments on any filters, prisms or gratings used, weather conditions, and so on. From 1880 until 1953 the start and stop time of every exposure was recorded in local sidereal time, after which a mixture of local time (e.g.,
were included for comparison; these appear at a distinct distribution which is well described by a single Gaussian; parameters from these fits are given in the legend. Five DN plates observed through a red filter.

Differences in color response are not a significant source of scatter in light plate-dependent; (2) series-dependent; and (3) constant.

In principle the values of time, R.A./decl., and HA should all agree, so the software makes a comparison and if a conflict exists, “votes” the most probable solution and flags the entry for human investigation.

EST) and UT was used. Prior to 1925 dates were recorded as local calendar date at the beginning of the night. From 1925 to 1953 two local dates are given, corresponding to the start and end of the night. In order to use the logbook records, they were digitally photographed and the images transcribed by a data entry company. Finally, the metadata files were converted to a common system; the Modified Julian Date (MJD) by software designed to determine the time system in use, and to which date longitudes, and finding the correct date requires calculation of the sidereal time of local midnight. This is a fairly complex problem because the observations were made at several different longitudes, and finding the correct date requires calculation of the sidereal time of local midnight. In principle the values of time, R.A./decl., and HA should all agree, so the software makes a comparison and if a conflict exists, “votes” the most probable solution and flags the entry for human investigation.

For the purposes of temporal analysis it is more useful to compute the Heliocentric Julian Date (HJD), which corrects times to an imaginary clock located at the solar-system barycenter. This is required to correct for a shift in photon arrival time of up to ±8 minutes, which would be a significant source of error in the study of short period phenomena such as compact eclipsing binaries.

The DASCH pipeline captures the magnitude and position of every detected star in every plate. The resulting database is then used to construct light curves for any subset of these objects. For the following analysis, light curves for every resolved GSC2.2 star brighter than $B = 17$ in the region of sky covered by the M44 plates were extracted using a highly optimized piece of software based on the Starbase package (a command-line driven relational database for astronomy: http://cfa-www.harvard.edu/john/starbase/starbase.html). The code relies on pre-sorting and indexing the photometry catalogs; each one is read only once by the search tool, which identifies all target stars that lie on the plate and takes advantage of the index to rapidly locate them (without having to examine every entry). If an expected target star is not found in a particular plate catalog, then the local limiting magnitude for its position is returned, to be treated as an upper limit. Execution time for 500 plates, each containing an average of 42,000 stars, is $\sim$1 s per light curve for a single CPU thread. Further optimizations are in development which enable rapid processing of thousands of plates. Analysis of large numbers of light curves is used in Figure 13 to estimate photometric performance as a function of magnitude.

7. COMPARISON BETWEEN DASCH AND GSC2.2 MAGNITUDES

Careful examination of the outliers in our photometric calibration curves revealed three classes of astronomical objects: variable stars, galaxies, and blends. Of these, variable stars obviously do not fall on the calibration curve unless they happen by chance to be at their catalog brightness on a particular plate. Galaxies do not conform because of their non-stellar profiles; the calibrations for aperture and image-size magnitudes do not apply to extended objects. Isophotal and Kron (flexible aperture) magnitudes formally apply to any object; however, changes in image-scale and photometric depth cause the galaxy...
magnitudes derived from different plates to disagree with one another. In deeper images, more of the galaxy gets above the detection threshold, and hence the isophotal area expands and observed flux increases. Blended stars and those with near neighbors can yield unreliable magnitudes that may mimic real variables as seeing, plate scale, and depth change between plates. Fortunately there are existing all-sky catalogs (e.g., GSC2.2) that go deeper than most of the Harvard plates and are largely complete down to our typical limiting magnitudes. Prior knowledge of positions and magnitudes enable one to predict which stars will suffer from neighbor effects in a given plate.

8. THE DASCH PHOTOMETRIC CATALOG

We can take advantage of the hundreds of independent measurements of each star to average out systematic and statistical errors. In this way, for non-variable stars DASCH has the potential to create an all-sky photometric catalog. For variable stars the median brightness will be more representative than that measured at a single instant, and the variability amplitude will be known. Among the applications of such a catalog are an improved GSC for observatory and spacecraft operations, as well as a list of standard stars for general calibration use. It has been recently discovered that a significant fraction (5%) of the standards in current use are in fact long-term variables (Vogt et al. 2004); DASCH will define a large set of constant stars proven to be good standards over our 100 year baseline.

For the DASCH commissioning project we computed the median magnitudes for 75,000 GSC2.2 stars within 10° of M44 in order to refine our own calibrations. Median magnitudes were computed using only good measurements (flag specifications given in Section 4.6) for stars having at least 10 such points in their light curve. As a benchmark we compared the DASCH (median) and GSC2.2 magnitudes against WebDA photoelectric $B$ magnitudes which have a precision of order 0.01 mag. The overall rms of $B_{\text{WebDA}} - B_{\text{DASCH}}$ is 0.08 compared to a much larger value of 0.22 for $B_{\text{WebDA}} - B_{\text{GSC}}$; in both cases after removing obvious outliers, this result is shown in Figure 14. A large improvement in plate calibration is possible by using the DASCH median magnitudes in place of GSC2.2 to refine the individual curves, as demonstrated in Figure 14. Two improvements are evident: first, the scatter of points about the lowess fit is dramatically tightened; second, the large-amplitude outliers are almost completely removed. It is clear therefore that iterating our calibration pipeline leads to impressive gains in both consistency and absolute precision. Further development along these lines will be reported in a future paper (S. Tang et al. 2010, in preparation).
Figure 15. Light curve of a constant star demonstrating about ±0.1 mag photometry over 400 plates that span 100 years and six different plate series. The star’s magnitude is plotted against time (upper panel) and chronological plate sequence (middle panel). Large solid circles denote good data points (of which there are 302), while open circles indicate a larger-than-usual scatter in the calibration for that point. Arrows denote upper limits derived for plates that did not go deep enough to detect the star. Smaller versions of these symbols denote observations where the star was very close to the plate edge. Two versions of the light curve are shown: black points denote a calibration that accounts for radially dependent effects such as vignetting, while the red points incorporate an additional step correcting for irregular spatial variations in plate sensitivity. The bottom panel shows for each measurement how far the star was from the plate center, and the number of local calibrators used to correct for spatial variation.

(A color version of this figure is available in the online journal.)

Figure 16. Light curve of the eclipsing binary RY Cancri. This light curve contains 417 measurements of brightness obtained over 88 years. Note the occasional faint points below the rest, which correspond to eclipses of the primary star by its fainter companion.

(A color version of this figure is available in the online journal.)
9. DEMONSTRATION SCIENCE LIGHT CURVES

Many variable stars are known within a few degrees of the well-known open cluster M44, including dwarf novae (DNe), eclipsing binaries, and various classes of pulsating stars. The examples in this section are intended to demonstrate the value of DASCH in a range of applications. In Figures 15–22 the star’s magnitude is plotted against time (upper panel) and chronological plate sequence (middle panel). Large solid circles denote good data points, while open circles indicate a larger-than-usual scatter in the calibration for that point. Arrows denote upper limits derived for plates which did not go deep enough to detect the star. Smaller versions of these symbols denote observations where the star was very close to the plate edge.

Two versions of each light curve are shown: black points denote a calibration that accounts for radially dependent effects such as vignetting, while the red points incorporate an additional step correcting for irregular spatial variations in plate sensitivity. The bottom panel shows for each measurement how far the star was from the plate-center (points) and how many local calibrators were used (red line).

The first example is a star that does not appear to vary. Such constant-brightness stars enable sensitive determination of various systematic effects and provide a completely independent measure of uncertainties. In Figure 15 is the light curve of U Cancri, a semi-regular variable of the Mira type. This is an example of a red giant star that is slowly pulsating. The star brightens and fades by over 4 mag roughly every 300 days. During its minima U Cancri is not visible on most of our plates because it gets too faint, as can be seen from the many upper limits (arrow symbols).

(A color version of this figure is available in the online journal.)
Figure 20. U Geminorum, the most famous recurrent DN. The DASCH light curve shows a scattering of bright points corresponding to DN outbursts. Upper limits appear for plates that did not reach deep enough to detect the system in its quiescent state. (A color version of this figure is available in the online journal.)

Figure 21. CQ Cancri DASCH light curve. (A color version of this figure is available in the online journal.)
show in Figure 21. Folding at 0.524 days, which is the pulsation
measure distances. The century-long DASCH light curve is
luminosity obey a relation enabling RR Lyrae to be used to
white giants pulsating with a constant period. The period and
progress.

comparison with AA VSO records. An in-depth analysis of
were missed by the gaps in plate coverage, as verified by
DN outbursts, each few years apart. Others occurred but
visible in Figure 19 beginning at phase 0.5. We were able to
do not line up exactly. At least two distinct rising portions are
driven by specific science goals. Thus, the project will
presented in Tang et al. (2010). Further comprehensive analysis
highlights the potential of DASCH in a number of active research
areas. Detailed examination of the DASCH light curves of
stars in the Praesepe commissioning field has already led to
the discovery of a new class of variables, K III giants with
large amplitude multi-year dimming caused by dust ejection, as
presented in Tang et al. (2010). Further comprehensive analysis
of the ~70,000 light curves in this field is ongoing. DASCH
proceeds in a systematic fashion with field (and hence plate)
selection driven by specific science goals. Thus, the project will
produce a flood of scientific advances during its expected ~5
year scanning phase, without waiting for the entire collection to
be digitized. The DASCH Web site4 provides many additional
examples of scanned images, photometry, light curves, and
additional background to the project. DASCH progress and
the latest scientific results are also available on the site, which
will eventually serve the entire database (images, catalogs, light
curves) to the community.

Time domain astronomy is undergoing a revolution, with
new dedicated facilities such as ASAS (Pojmanski 1997),
Pan-STARRS (Kaiser et al. 2002), and the Palomar Transient
Factory (Law et al. 2009) recently coming on line, and LSST
Claver et al. (2004) on the near horizon. DASCH is leading the
way at unexplored timescales, which will not be accessible to
these new facilities for another hundred years.

We thank our colleagues at Harvard and beyond who have
offered advice and expertise in photographic astronomy, dig-
itization, and software development. In particular, we thank
A. Doane, G. Champine, and A. Sliski for their inestimable
contributions. DASCH has been supported by NSF grant AST-
0407380 for which we are grateful. S.L. acknowledges the gener-
ous support of Gemini Observatory.

REFERENCES
Bacher, A., Kimeswenger, S., & Teutsch, P. 2005, MNRAS, 362, 542
Berdnikov, L., Pastukhova, E. N., Gorynya, N. A., & Zharova, A. V. 2007, PASP, 119, 82
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

4 http://hea-www.harvard.edu/DASCH

Figure 22. CQ Cnc Period = 0.524654 days

Figure 22. CQ Cancri DASCH light curve folded at the pulsation period of
0.5 days.

of such a star demonstrating about +/−0.1 mag photometry
from 400 plates that span 100 years and six different plate
series. Comparing the rms for the black (0.154 mag) and red
(0.105 mag) set of points, the effect of correcting the irregular
plate variations gives about a 50% improvement over radial
correction alone.

The eclipsing binary RY Cancri. The light curve of RY
Cancri contains 417 measurements of brightness obtained over
88 years. In Figure 16, one can see a scattering of faint points
below the rest, which correspond to eclipses of the primary
star by its fainter companion. The eclipses are very narrow as
is typical for compact binaries. The light curve of RY Cancri
is shown folded in Figure 17, at its known binary period of
1.092943 days, to show the eclipse. The magnitudes plotted
are corrected for spatial variations in the different plates (the
red points in previous plot), and the symbols are as described
above. The light curve is plotted over two cycles, the first cycle
shows estimated error bars for each point.

U Cancri, a semi-regular variable of the Mira type. Figures 18
and 19 provide an example of a red giant star that is slowly
pulsating. The star brightens and fades by over 4 mag roughly
every 300 days. During its minima U Cancri is not visible on
some of our plates, as can be seen from the upper limits (arrow
symbols). Because the star’s pulsations are not perfectly regular,
when we fold the light curve the points from different cycles
do not line up exactly. At least two distinct rising portions are
visible in Figure 19 beginning at phase 0.5. We were able to
measure a period of 304 days from the DASCH light curve
using the Lomb–Scargle periodogram.

U Geminorum, a CV. U Geminorum is a famous CV that
shows regular DN outbursts. Outbursts (due to an instability
in the accretion disk) produce a 2–5 mag brightening, lasting
for several days. In the ~22 year light-curve segment shown in
Figure 20; we see the quiescent luminosity (B ~ 14) interspersed
by DN outbursts, each few years apart. Others occurred but
were missed by the gaps in plate coverage, as verified by
comparison with AAVSO records. An in-depth analysis of
all CVs in the Praesepe DASCH commissioning field is in
progress.

CQ Cancri, an RR Lyra variable. RR Lyrae stars are hot
white giants pulsating with a constant period. The period and
luminosity obey a relation enabling RR Lyrae to be used to
measure distances. The century-long DASCH light curve is
shown in Figure 21. Folding at 0.524 days, which is the pulsation
period of the star, reveals the characteristic RR Lyrae profile
Figure 22.

10. CONCLUSIONS

DASCH is well on the way to digitizing the HCO photo-
graphic plate collection, the largest time domain all-sky imag-
ing survey in existence. The survey covers the entire celestial
sphere to 17th magnitude, with 500–1000 measurements per sky
position, spanning a century (1880–1980s). This paper reports
the successful commissioning of the DASCH plate digitizer,
using ~500 plates selected to cover the Praesepe star cluster.
We have demonstrated the astrometric and photometric quality
of the digitized images, and presented a software pipeline for
reduction and analysis. The pipeline quantifies and corrects for
a wide range of instrumental signatures, including geometric
distortion, vignetting, emulsion sensitivity variations, and at-
otmospheric effects. Systematic photometric variations between
plate series, originating in the different emulsions and telescopes
used over the years at the various observing sites, are reduced
to below the 0.1 mag level. Thus, DASCH achieves the basic
prerequisites for time domain astronomy.

The sample of representative light curves in this paper
highlights the potential of DASCH in a number of active research
areas. Detailed examination of the DASCH light curves of
stars in the Praesepe commissioning field has already led to
the discovery of a new class of variables, K III giants with
large amplitude multi-year dimming caused by dust ejection, as
presented in Tang et al. (2010). Further comprehensive analysis
of the ~70,000 light curves in this field is ongoing. DASCH
proceeds in a systematic fashion with field (and hence plate)
selection driven by specific science goals. Thus, the project will
produce a flood of scientific advances during its expected ~5
year scanning phase, without waiting for the entire collection to
digitized. The DASCH Web site4 provides many additional
examples of scanned images, photometry, light curves, and
additional background to the project. DASCH progress and
the latest scientific results are also available on the site, which
will eventually serve the entire database (images, catalogs, light
curves) to the community.

Time domain astronomy is undergoing a revolution, with
new dedicated facilities such as ASAS (Pojmanski 1997),
Pan-STARRS (Kaiser et al. 2002), and the Palomar Transient
Factory (Law et al. 2009) recently coming on line, and LSST
Claver et al. (2004) on the near horizon. DASCH is leading the
way at unexplored timescales, which will not be accessible to
these new facilities for another hundred years.
