VIRTUAL OBSERVATORY: Plate Content Digitization, Archive Mining & Image Sequence Processing
edited by M. Tsvetkov, V. Golev, F. Murtagh, and R. Molina, Heron Press, So£a, 2005

The Harvard Plate Scanning Project

D. Mink[1], **A. Doane**[2] **R. Simcoe**[2] **E. Los**[2] **J. Grindlay**[3]

[1]Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, USA

[2]Harvard College Observatory, Cambridge, Massachusetts, USA

[3]Department of Astronomy, Harvard University, Cambridge, Massachusetts, USA

Abstract.

There are over 500,000 glass photographic plates in the Harvard Plate Stacks, exposed in both the northern and southern hemispheres between 1885 and 1993. This 100 year coverage is a unique resource for studying temporal variations in the universe. We realized that it would be even more useful if the images were all available in digital form over the Internet. We report on progress in the design and construction of a high resolution, large field scanner, which we are building after discovering that no commercial model could adequately deal with our largest, most useful plates. In addition to plate scanning, we discuss our progress toward digitizing the catalog of plates, constructing an automatic pipeline to make the scanned images useful, and the scanning of the handwritten observing logs which provide the most complete information about the conditions under which the plates were taken.

1 What We Have

Harvard College Observatory has over 500,000 photographs of the sky, taken by a variety of telescopes in both the northern and southern hemispheres. The earliest plates in the collection are some of the oldest glass astronomical photographic plates, 1100 of them, taken by Benjamin Apthorp Gould, Jr. from 1872 to 1882, mostly on a collodion (wet) plates [1]. Harvard has by far the largest collection of astronomical plates in the world, 25% of the world's total. Telescopes in the United States, South America, New Zealand, and South Africa gave complete sky coverage over much of the long time period, 1872 to 1989, during which the plates were taken. Over 350,000 should be useful for photometry and astrometry. Table 1 shows the major series of plates, with plate

Table 1. Harvard Plate Collection

Series	Ap (in.)	Scale ”/mm	+-dec	+-ra dec=0	N/S	Total	Years	Limit	Q
A	24	60	2.9	14.3	N-S	27504	1893-1950	18	5
ADH	30	68	2.4	9.6	S	7067	1950-1963	18-19	4
AM,AC	1.5	600	21.1	67.7	N/S	75000	1898-1957	13-14	3
B	8	179	6.3	20.2	S	76874	1885-1954	17	4
BR	8	209			S	4176	1938-1944	17	4
C	11	84	2.4	5.9	N	23270	1886-1947	–	2
DNB,DSB	1.6	580	20.5	65.4	N/S	9000	1962-1989	15	5
FA,AI	1.5	1200	—-	—-	N	70000	1901-1958	–	2
H	24	60	0.9	3.6	N	6644	1906-1953	–	3
I	8	163	5.7	18.2	N	59246	1889-1946	17	4
IR	8	162	5.6	18.0	N	12798	1934-1976	17	4
J	24	98	2.7	10.8	N	4770	1942-1957	–	3
MA	12	97	3.4	10.9	N	11737	1905-1983	17-18	5
MB	4	193	6.8	22	N	2722	1914-1932	–	2
MC	16	98	2.7	13.6	N	40596	1909-1992	17-18	5
MD	4	193	6.8	21.7	N	30000	1911-	11	2
MF	10	167	5.8	18.7	N− S	40897	1915-1955	17	4
RH,RB	3	391	13.8	44.2	N/S	33000	1928-1963	15	3
RL	4	290	8.2	32.8	N	5062	1933-1962	–	–
SH,SB	60	26	.36	.45	N-S	20000	1934-1989	–	–
Schmidt	12	1200	27	108	N	30000	1953-1968	–	–
X	13	42	1.2	5.8	S	19090	1888-1951	–	–

scales, field sizes in degrees, hemisphere, number of plates, duration of series, magnitude, and plate quality on a scale of 1 through 5, with 5 being best.

Figure 1 shows the temporal distribution of the three plates series which have been completely catalogued so far.

2 What We Have Done

Over the past 10 years, we have manually created digital catalogs of about 135,000 plates from five of the most useful plate series. Those for the A series (mostly 14x17-inch, primarily southern hemispheres), MC (mostly 8x10-inch plates, northern hemisphere), and MF (mostly 8x10-inch plates, southern hemisphere) are searchable online (See Figures 2, 3, and 4)

We have tested two commercial scanners using 8x10-inch plates. The CreoScitex EverSmart, which costs between $35,000 and $50,000, has a very high resolution (4800 dpi). Because of its high native resolution, the scanner makes several passes and stitches the scans together atuomatically. We scanned a plate at 1200 dots per inch and were satis£ed with the resulting image, though the

20 minutes it took was longer than our limits and the system crashed after each image acquisition.

We then tested and bought a UMAX PowerLook 3000 for $5,000. It had adequate resolution (1200 dpi), was reliable (for a while), and scanned an 8x10 plate in under 10 minutes. As a trial run, we scanned about 100 plates, 40 of them from the MC series containing the M44 cluster. Those plates are now

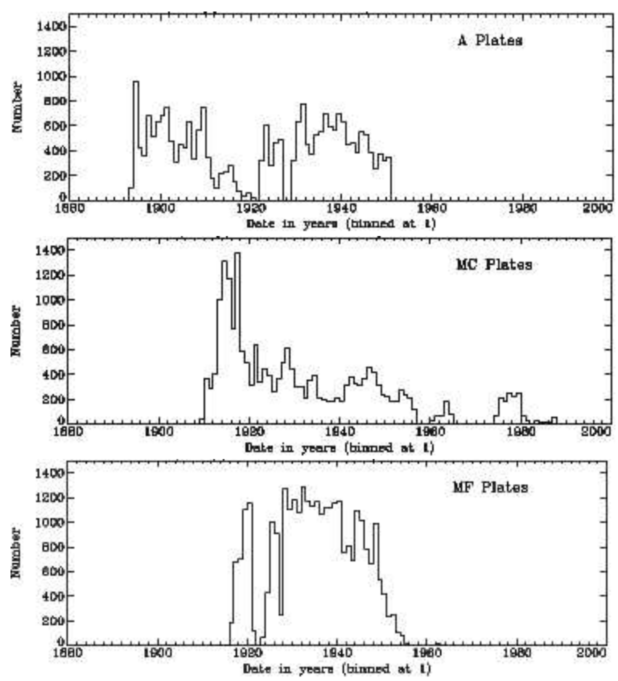

Figure 1. Distribution of plates in the A, MC, and MF series by decade

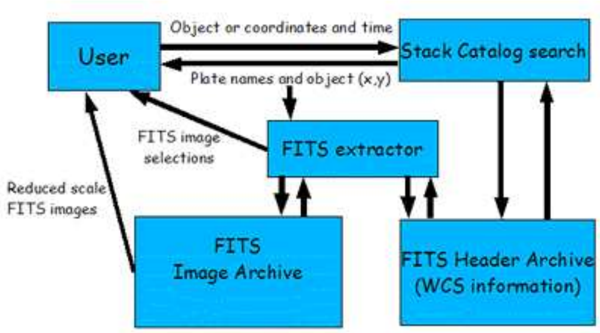

Figure 2. Access to images over the World Wide Web

accessible over the World Wide Web and can be displayed as in Figure 5.

3 What We Are Doing

To scan Harvard's archive of historic plates in a 3-5 year time frame, we needed a machine that can scan 200 times faster than machines designed 20+ years ago, such as the USNO PMM [3] , which took 1- 4 hours to scan a single 14 x 14 inch plate. To meet astrometric, photometric, and archival goals, the machine needs sub-micron positional accuracy, at least 12 bits of photometric density

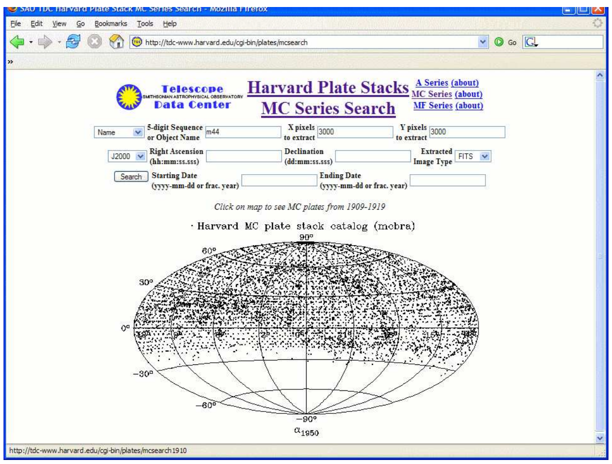

Figure 3. Web page for searching the MC Series plate catalog

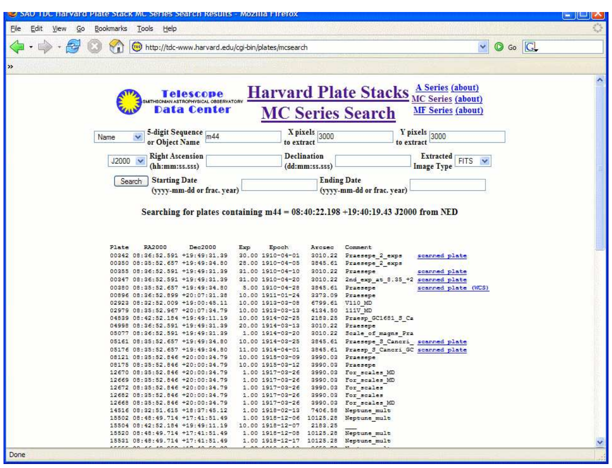

Figure 4. Result of MC Series search for M44

range, and a scan speed that allows human handling to limit the average plate processing time.

Using technology common to semiconductor wafer and flat panel display inspection stations, a machine can be built today that can do ultra fast, ultra precise scanning. A drawing of the machine we are putting together is shown in Figure 6. The 4Kx4K digitizing camera with 16bit, 11 μm pixels will be moved across the image, taking slightly overlapping exposures. A single 8 x 10 inch plate will be scanned in about 20 seconds (though we will probably do 2 at once). A 14 x 17 inch plate will be scanned in a little over a minute, generating enough data in that time to fill a DVD (3 Gigabytes). Figure 7 shows a plate being digitized.

Figure 8 shows how the digitized images will be processed for archiving. As each plate is scanned, the positions of sources in the image will be catalogued and a world coordinate system will be fit to the entire plate by matching them to a catalog using the IMWCS program from the WCSTOOLS package [2]. An image of each plate will be saved, binned so that it can be displayed at full resolution over the World Wide Web as a JPEG or FITS file. Sky positions and photometric information will then be attached to the image position catalog and the resulting catalog will be accessible over the Web, too. A database of sources will be developed so that individual objects can be tracked over time.

To ease plate processing, we need to have digital catalogs before we scan plates. Existing digital catalogs have been created very slowly over the past 15 years by part-time plate stack staff. Digital catalogs are necessary for the plate digitizing pipeline. We have a 135,000-plate head start, but the rest of the collection needs to be digitally catalogued in a more timely way.

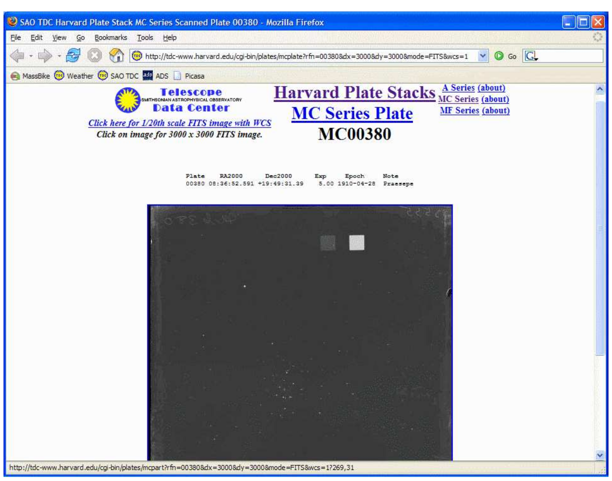

Figure 5. Image of M44 plate displayed over the web

To outsource catalog transcription, we will need to convert existing observing log books to digital images. At the Center for Astrophysics, we have the staff and equipment to quickly scan single sheet logs, but less than one-third of our 910 log books are in binders from which pages can be extracted, and most of those are from the less-interesting last half of the twentieth century when other image collections are more likely to overlap with Harvard's. We plan to start in-house log book scanning while we are developing our new plate scanner.

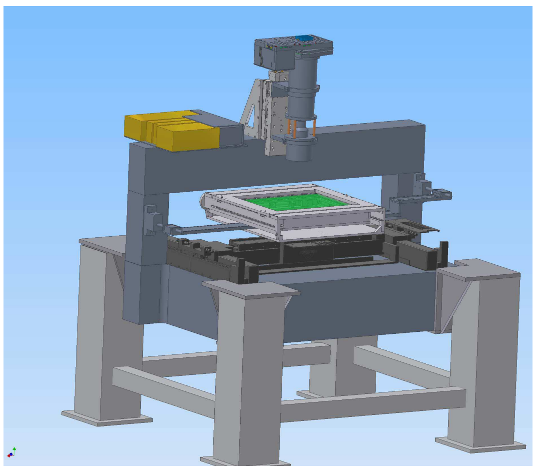

Figure 6. Drawing of digitizer based on semiconductor scanning technology

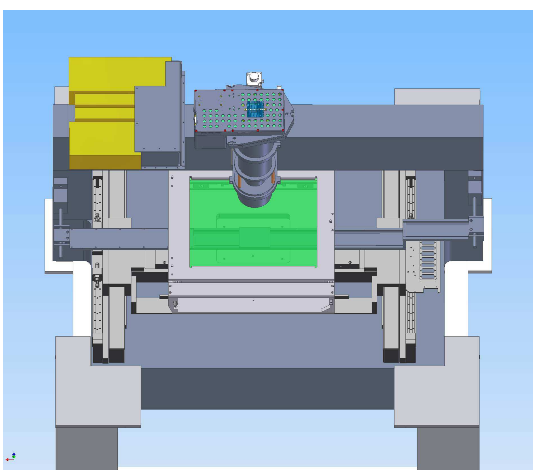

Figure 7. Position of an astronomical plate being scanned on digitizer

4 What We Plan to Do

As money becomes available, we will continue to scan our plates, producing an online image archive and a catalog of objects found in those images. After several years of scanning, our digital archive will total approximately 500 terabytes. However, we are expecting disk prices to drop and disk capacities to increase as we scan so that we can incrementally increase our hardware over time.

We are looking at additional locations to duplicate the 300 Terabytes of images we will generate, though we intend to serve it all, too.

Since many of the older log books are too fragile to travel and transcription in Cambridge is likely to be too expensive, they should be imaged into a digital format. We are applying for a grant to set up and staff a high resolution camera to turn the rest of the log books into digital images. Log books contain information which may not fit or be processed into existing digital catalogs, so all scanned log book pages will be put online linked to the images they describe. Log book images can then be transcribed anywhere. We are applying for a second grant to fund the transcription of handwritten logs into digital text.

Acknowledgments

We are currently in the first year of a National Science Foundation grant which is funding the design, construction, and trial use of a scanner which will rapidly scan plates up to 14x17 inches in dimension. The project is directed by PI Josh Grindlay and supported by NSF grant AST-0407380, which we gratefully acknowledge.

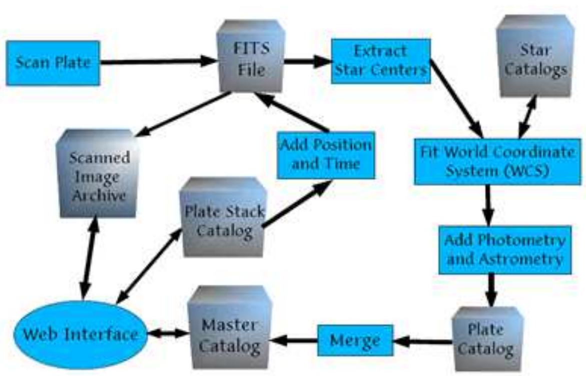

Figure 8. The pipeline for processing digitized plate images produces a catalog of objects, too.

References

[1] M. L. Hazen (1991) *Astron. J.* **101** 1.

[2] D. J. Mink (1997) *Astronomical Data Analysis Software and Systems VI, A.S.P. Conference Series, Vol. 125, G. Hunt and H. E. Payne, eds.* 249.

[3] D. G. Monet, S. E. Levine, B. Canzian, H. D. Ables, A. R. Bird, C. C. Dahn, H. H. Guetter, H. C. Harris, A. A. Henden, S. K. Leggett, H. F. Levison, C. B. Luginbuhl, J. Martini, A. K. B. Monet, J. A. Munn, J. R. Pier, A. R. Rhodes, B. Riepe, S. Sell, R. C. Stone, F. J. Vrba, R. L. Walker, G. Westerhout, R. J. Brucato, I. N. Reid, W. Schoening, M. Hartley, M. A. Read, S. B. Tritton (2003) *Astron. J.* **125** 984.