Digitisation, archiving and dissemination of the 275 Gbytes of the UKST/AAO H-alpha survey data

Harvey T. MacGillivray

Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, United Kingdom
hmg@roe.ac.uk

Abstract

We plan to digitise the UKST/AAO H-alpha survey of the Southern sky using the SuperCOSMOS facility at the Royal Observatory Edinburgh. The digitisation will be performed at 10 microns resolution (0.67 arcsec), and will result in a database of 275 Gbytes of pixel data for the 200 fields of the survey. These data will be processed to produce also a point source catalogue. As the H-alpha survey is an official AAO sky survey, both the pixel data and the point source catalogue will be made available to the community as a service facility. We give details of the digitisation process, archiving of the resulting data and the plans for dissemination of the database.

Keywords: surveys

1 Introduction

The UKST/AAO H-alpha survey on Tech Pan emulsion provides an excellent opportunity for recording detail of the Galactic Plane and Magellanic Clouds at unprecedented resolution in one of the most interesting regions of the visible spectrum. The availability of such high quality imaging demands that we ensure there is access to the survey in digital form, also of high quality and at high resolution, in order to maximise the scientific return from the undertaking. To this end, we intend to digitise and distribute the survey material using SuperCOSMOS.

SuperCOSMOS is the latest in a line of photographic plate digitisers designed and constructed at the Royal Observatory Edinburgh (ROE). Its predecessors were GALAXY and COSMOS. With SuperCOSMOS we have built upon experience gained over the past 30 years to produce a new state of the art facility. SuperCOSMOS is an order of magnitude more powerful than its most recent predecessor, COSMOS. One of the primary reasons for the construction of SuperCOSMOS was to allow digitisation of the major photographic sky survey material (from the UK and Palomar Schmidt telescopes) at high resolution and with high accuracy. This is currently underway.

2 SuperCOSMOS

The main components of SuperCOSMOS consist of a high-precision granite, air-bearing xy table which provides the high accuracy position measurements, a linear CCD array (of 2048 pixels) which provides the high speed scanning capability, and transputers and DEC Alpha workstations for rapid data processing and analysis. The entire system is housed in an environmental chamber operating under class 100 clean room conditions and with strict thermal stability, thereby ensuring integrity of the data during the scan process. SuperCOSMOS has an absolute positional accuracy of 0.25 microns, with a repeatability of better than 0.1 microns (Hambly *et al* 1998). Both glass plates and films can be accepted for scanning.

3 Digitisation and archiving

The photographic material is digitised to 15 bits (32768 grey levels) at a resolution of 10 microns (0.67 arcsecs on UKST plates). The central 320 x 320 mm (6 x 6 square degrees) of a UKST plate is scanned in a timescale of 2 hours, producing over 2 Gbytes of pixel data. These data are processed in a routine manner using the COSMOS crowded-field algorithm (Beard *et al* 1990) to produce a catalogue of the sources (stars, nebulae, etc) detected above a pre-selected isophotal threshold level.

Thirty two parameters are determined for each source, including coordinates on the plate, size, shape, orientation and integrated brightness. Transformation of plate positions to coordinates on the sky is undertaken using the PPM catalogue at present. The Tycho/Hipparcos catalogues will be used once fully interfaced into the SuperCOSMOS system. Typically, plate solutions with residuals in Right Ascension and Declination of order 0.20-0.25 arcsecs are obtained currently, from cross-referencing with PPM.

The survey itself will consist of 200 films with field centres separated by 4 degrees. The central 5 x 5 square degrees on each film will be digitised, thereby allowing overlap from field to field, and resulting in 1.37 Gbytes of data from each. The total pixel data from the 200 field set will occupy some 275 Gbytes. All data (both pixels and point sources) will be archived on high capacity (20 Gbyte) exabyte tape. For safekeeping the "master" backup of the H-alpha survey scan data will be held at the ROE, while a copy will be shipped to the AAO.

4 Dissemination

As the H-alpha survey is now an official sky survey to be undertaken for the community by the AAO, the digital data will be made available for general access. Some aspects of the data distribution, however, need to be finalised; for example are the pixel data conducive to compression, or is detail lost and the data should be distributed in raw uncompressed form only?

One thing is certain, however, that developments in storage technology will ensure that distribution of the full uncompressed data set is not out of the question. Already, we are seeing the arrival of Digital Video Disks (DVDs) on the scene. At present, these are single-sided, single-layered with a maximum capacity of 4.7 Gbytes. However, the planned development of these is for double-sided, double-layered with 4 times the above capacity. And it is certain that the future of high capacity storage on CD-type technology will not stop there.

At the present time, our planning assumption is that the H-alpha survey in digital form will be available on the higher capacity DVDs. The full survey will, therefore, require 15 such media. If compression were used, then it is envisaged that the entire H-alpha survey of the Galactic Plane and Magellanic Clouds could be made available on typically a single disk.

As well as distributing the pixel data on digital video disk, we plan to allow the pixel data to be accessible via the World Wide Web serviced from the ROE and also the AAO. We have already developed a JAVA-based interface which will allow the user to view areas of survey pixel data, with the coordinate selection driven by the user. The browser allows not only the pixel data to be viewed but also simultaneously the relevant region of the point-source catalogue. Objects can be selected either from the point source database (which will allow the corresponding pixel data to be viewed) or alternatively, given a selection from the pixel data, the corresponding source in the database is highlighted. The DVD version of the database will be distributed with similar access software.

In addition to the above, there are two other plans currently being considered for allowing ease of use of the pixel data. Firstly, we are intending to put the pixel data on the same xy grid system as the J, R and I survey pixel data (with J as 'master'). This will facilitate comparison of the H-alpha data with broadband data in the other passbands, and allow colours to be intercompared. Secondly, we intend to produce "averaged down" versions of the full pixel data for each plate (compressed by a factor of 8 in each dimension) as TIFF files of size 3312 x 3312 pixels. This will allow the data to be easily manipulated in order to aid the construction of mosaics of large regions (or even the whole Southern sky part) of the Milky Way. The TIFF files will be distributed with the full data set, or can be acquired separately as desired.

Our aim is to have the entire database ready for rapid dissemination very shortly after the final photographs have been taken by the UKST (i.e. in about 2-3 years time). However, as there will also be strong interest in selected areas of the Galactic Plane and the Magellanic Clouds, distribution of large, contiguous, high-interest regions of sky will be performed in a timely manner.

Acknowledgements

A number of people have contributed in a major way to the SuperCOSMOS project in recent years:- Nigel Hambly, Lance Miller, Bill Cormack, Jim Herd, Dennis Kelly, Stephen Beard, Clive Davenhall and Bernard McNally. I am grateful to them.

References

Beard, S.M., MacGillivray, H.T. and Thanisch, P.F. 1990 MNRAS, 247, 311.

Hambly, N.C., Miller, L.A., MacGillivray, H.T., Cormack, W.A. and Herd, J.T. 1998 MNRAS, submitted