Astronomical Image Processing

A1

Head of Experiment: Sergey Lebedev

The following experiment guide is NOT intended to be a step-by-step manual for the experiment but rather provides an overall introduction to the experiment and outlines the important tasks that need to be performed in order to complete the experiment. Additional sources of documentation may need to be researched and consulted during the experiment as well as for the completion of the report. This additional documentation must be cited in the references of the report.

Imperial College
London

RISK ASSESSMENT AND STANDARD OPERATING PROCEDURE

1. PERSON CARRYING OUT ASSESSMENT					
Name	Geoff Green	**Position**	Chf Lab Tech	**Date**	16/09/08

2. DESCRIPTION OF ACTIVITY
A1 Astronomical Image Processing

3. LOCATION					
Campus	SK	**Building**	Huxley	**Room**	403

4. HAZARD SUMMARY			
Accessibility	X	**Mechanical**	
Manual Handling		**Hazardous Substances**	
Electrical	X	**Other**	X
Lone Working Permitted?	Yes ☒ No ☐	**Permit-to-Work Required?**	Yes ☐ No ☒

5. PROCEDURE	PRECAUTIONS
Use of Computer Display	Avoid prolonged sessions; Take Breaks
Use of 240v Mains Powered Equipment	Isolate Socket using Mains Switch before unplugging or plugging in equipment
Accessability	All bags/coats to be kept out of aisles and walkways.

6. EMERGENCY ACTIONS
All present must be aware of the available escape routes and follow instructions in the event of an evacuation

Computational Image Processing: A Deep Galaxy Survey

A 3rd Year Lab Module

Version 1.2: Revised Sept. 2013

Dr. D.L. Clements and N.Skrzypek

Astrophysics Group

Imperial College

1 Overview

The aim of this experiment is to use an optical CCD image of the extragalactic sky to determine the number counts of galaxies in the local universe (ie. number of galaxies brighter than a given magnitude as a function of magnitude). To do this you will have to write a computer programme that will read in the CCD image, and then detect individual galaxies within this image and determine their brightness. This brightness can then be converted into an apparent magnitude and the number counts can be derived. You should compare these results with those obtained by other astronomers, and discuss any differences or disagreements between your data and the data in the literature.

2 Introduction

This experiment involves the use of a deep optical image of an extragalactic field (shown in Figure 1) to obtain information about the population of galaxies in our universe. The image was taken at the Kitt Peak 4m telescope using the CCD mosaic camera and using a Sloan r band filter, which has a central wavelength at about 620nm. More information on the filter, camera and telescope can be found at http://www.noao.edu/kpno/manuals/dim/. The image was taken as part of the optical observations supporting the SWIRE infrared survey using the Spitzer space telescope. More information about this project and the Spitzer Space Telescope can be found at their relevant websites, detailed in the references.

Much work has already been done on the optical images obtained at the telescope. Many separate exposures with different pointing have been combined and various instrumental effects have been removed. The image you see in Figure 1 is thus a fully reduced image ready for scientific analysis, though there are still various features about it, such as the bright star in the upper part, that are perhaps not ideal.

2.1 Apparent Magnitudes

The brightness of astronomical objects as observed from Earth is typically measured in what are called 'apparent magnitudes' (see eg. Walker, 1989, p. 12). This system derives from observations made by early Greek astronomers who classified stars into several different magnitude

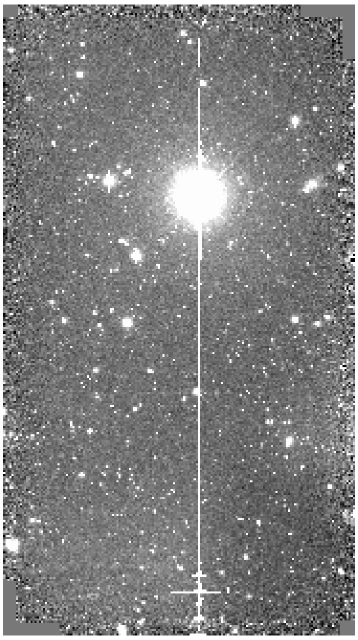

Figure 1: The optical survey field used in this experiment.
Note the many objects, almost all of which are galaxies, the bright star in the upper centre with CCD bleeding affecting nearby pixels and the associated column of pixels, and the variability of noise with position

classes. Since the eye's reaction to brightness is logarithmic, the magnitude system is itself logarithmic. The system was formalised in historical times using the standard definition relating the difference in magnitude between two objects to the flux (eg. Wm^{-2}) received from them:

$$m_1 - m_2 = -2.5log_{10}(f_1/f_2) \tag{1}$$

where m_1 and m_2 are the apparent magnitudes of the two objects, and f_1 and f_2 are the fluxes of the two objects. This definition has two important implications about astronomical magnitudes:

- an object 2.5 magnitudes brighter than another is emitting 10 times as much flux, while a difference of 5 magnitudes implies 100 times as much flux, and
- an object with a smaller magnitude is *brighter* than an object with a larger magnitude

2.2 Magnitude Systems

The standard definition is entirely relative. It merely addresses the magnitude difference between two objects. To allow sources observed at different telescopes to be compared, we must go further than Pogson's definition. Firstly, the wavelengths of light received by the telescope need to be standardized. This leads to the use of specific sets of filters in the optical and near-infrared part of the spectrum. These include the Johnson filters (eg. Walker, 1989; Bessel, 1990) - UBVRI - and the more modern SDSS filters (see the Sloan Digital Sky Survey website for more details), one of which is used in this investigation. In addition to the filter set used, we must also choose a flux against which to compare the fluxes received in our observations. This is done through the so-called zero point of a magnitude system. There are two magnitude systems commonly used in astronomy today. The Vega system, where the brightness of the star Vega (also known as α-Lyrae) is defined as being zero magnitude in whatever filter is being observed. There is also the AB magnitude system, where a specific flux density, $3631 \times 10^{-26} Wm^{-2}Hz^{-1}$ is defined to be zero magnitudes. In the sometimes strange systems of units used by astronomers, this is 3631 Janskys, since 1 Jy is defined to be 10^{-26} $Wm^{-2}Hz^{-1}$.

There are various advantages and disadvantages between these two systems. For the current project, the provided astronomical images already contain a calibration factor in their 'header' that allows calibration to AB magnitudes.

3 CCD Astronomy

Most optical astronomy today relies on the use of Charge Coupled Devices (eg. Walker, 1989) - CCDs. These devices have considerable advantages over traditional photographic techniques, but there are also some limitations as well.

The key advantages of CCD detectors are the following:

- Much greater detector quantum efficiency. This means that a CCD detector detects a much greater fraction of the photons falling on it than a photographic plate, or the eye. A typical CCD can detect as many as 90% of the photons falling on it, while a photographic plate, at best, is 10% sensitive. Thus switching a telescope from a photographic detector to a CCD, you are effectively multiply its photon collecting capabilities by nearly an order of magnitude. What increase in mirror size would this improvement in performance be matched by, if you continued to use a photographic plate, but instead made the mirror of the telescope larger?

- CCDs are very linear, photon counting detectors

  For each photon detected by a CCD, an electron is produced. These electrons are then counted in the readout electronics. There is thus a clear relationship between the signal of an object ie. the number of 'counts' in the CCD image, and its brightness. Measurements of the brightness of sources using CCDs is thus much easier than with photographic plates which are inherently non-linear detectors, and where relationships between image size, due to scattering of light in the photographic emulsion, and source brightness must be very carefully calibrated.

  CCDs are not perfectly linear, though. Each detector pixel has a maximum size - it can only collect so many electrons. Above a certain number of counts the simple linear relationship breaks down. The CCD used in this experiment uses a 16bit counter, so for counts-per-pixel approaching 2^{16} ie. 65536 the linearity breaks down. The practical aspect if this is that sources brighter than $\sim$50000 counts per pixel will have unreliable fluxes.

  For very bright objects things can get worse than this. A CCD pixel that is overfilled will 'bloom' - with the electrons it has collected effectively overflowing and spreading to nearby pixels. These can then overfill themselves and contaminate still further pixels. This overflowing takes place preferentially in specific directions on the chip (generally along the rows and columns of the CCD), leading to the sort of line going from top to bottom in Figure 1, associated with the bright star in the upper central region of the image.

- A CCD is a digital detector

  This means that all the data can be handled by computer, and stored directly to tape or disk at the telescope and easily copied and distributed across the world. With photographic plates there were many additional stages including developing and copying, and the original media itself needs to be treated very carefully. Its thus not possible to provide students with access to original photographic plates - we just use copies here, which age over time. We can, however, give you access to genuine CCD data in exactly the same form used by professional astronomers. There are however various additional processing stages before an image like the one in Fig. 1 is obtained. This includes calibration, and the combination of many different subexposures into the one image. You may wish to think about what instrumental effects need to be corrected before a final image such as this is produced, and discuss with the demonstrator why numerous subexposures might be needed.

4 Galaxy Surveys

Galaxy surveys (see eg. chapter 13 of Cosmological Physics by Peacock) are one of the main tools of extragalactic astronomy. They can have many goals, but a central one is to examine the rate and way in which the galaxy population evolves. They can be conducted at any number of wavelengths, from radio to x-ray, but the optical waveband remains our main window on galaxy evolution. Such observations consist of deep exposures of a 'blank' area of sky - ie. one in which no prominent nearby stars or galaxies are to be found. This does not mean that there will be no nearby objects at all – there will always be 'foreground' objects of one sort or another on any image – but the vast majority of objects seen in a deep survey field will be galaxies.

The most basic way of studying the galaxy population is to examine the number of objects detected at each given brightness.

4.1 What is Expected from a Galaxy Survey?

Consider a population of objects uniformly distributed in space with a given distribution of luminosities, and a given density. The received flux from any of these objects is given by the inverse square law. Using these assumptions, and the definition of magnitude given above, it can easily be shown that the number of sources detected to a given flux limit goes as:

$$logN(m) = 0.6m + constant \tag{2}$$

where N(m) is the number of sources detected brighter than a magnitude limit m.

- Derive this result
- How many more objects are seen if you go one magnitude deeper?

This result provides the basic expectation for any uniformly distributed, unchanging population of objects. It applies just as well to stars in our own Galaxy (except where the disk-like structure of the Galaxy breaks the assumption of uniformity) as to galaxies in the universe. Indeed, since galaxies are distributed much more uniformly than stars in our galaxy, their distribution conforms to this relation more closely than stars.

However, there are two ways in which galaxies might deviate from this relation:

- **Cosmology** Equation 2 assumes that a normal Euclidean volume ($4/3\ \pi r^3$) applies to the distribution of galaxies. The tacit assumption here is that space is flat. General relativity, when applied to the universe as a whole (see eg. Peacock, 1999; Narlikar, 1983) shows that this is not necessarily the case. As well as a flat universe, described by a Euclidean geometry, other possibilities include an open universe, in which the volume rises faster with distance than the Euclidean case, or a closed universe, in which the volume increases more slowly. By counting the number of objects at ever increasing distances, corresponding to ever fainter fluxes, it was originally hoped that the volume element of the universe could be measured, and different cosmological models tested. This was the original motivation for galaxy 'number count' studies. Unfortunately, this has not proved to be an achievable goal since the second possible way in which galaxies deviate from the relation in equation 2 proves to be a much stronger effect.
- **Evolution** Galaxy evolution is the name given to the way in which the properties of galaxies change over time. Because of the finite speed of light, the further away we look, equivalent to observing to fainter magnitude limits, the further back in time we are looking. The galaxy population could, in principle, be different at these earlier stages of the universe — they might be brighter individually or have a higher number density — so the assumptions behind the derivation of equation 2 might be incorrect. Deviations from the simple form of equation 2 thus provide indications of the evolution of the galaxy population. Think what extra information you might need to use to be able to say something specific about the form of galaxy evolution eg. the population as a whole becoming brighter or their number density increasing, and discuss this with your demonstrator.

There are also a number of experimental effects that will determine to what extent your data matches equation 2, and which you need to be aware of.

- **Incompleteness** As you observe to fainter and fainter fluxes, the objects clearly become more difficult to detect. This leads some fraction of the sources at any given magnitude to be missing from your survey. The survey is, in this sense 'incomplete'. The level of incompleteness is eventually 100% at the detection limit of the survey. A plot of the total number of sources detected brighter than a given magnitude (the Number Count plot) will thus roll over from the gradient given in equation 2, to being a flat horizontal line, since, when your incompleteness is 100% you detect no more sources at all, no matter how faint you go. This roll over to a horizontal line in the number counts can be a fairly gentle process, with incompleteness rising from 0 to 100% over a range of magnitudes.

- **Bright Sources** While faint sources lead to incompleteness, bright sources can also lead to problems. This comes from two reasons. Firstly, bright sources are rare, and thus you will have relatively few of them in your catalog. Simple Poisson statistics apply to counting galaxies, so the error on the number counts are simply $\sqrt(N)$. Consideration of these errors will allow you to account for the statistical effects of the rare bright sources. However, a second factor complicates matters for bright sources, which will be particularly noticeable for simple methods of measuring the flux from these objects. Bright galaxies are usually nearby, and they will thus be extended in the images. If you are using a simple fixed aperture for flux measurement, as is recommended below, then some of the flux for a bright extended object will be missed. This will make the brighter objects appear to be fainter in the catalog than they really are. What effect do you think this will have on the number counts? Think about this, and discuss it with a demonstrator.

- **Foreground Contamination** While the field chosen for these observations is well away from the galactic plane, there will still be some level of contamination by galactic stars. Bright ones can be clearly picked out, but there are many faint stars that will also contaminate the galaxy counts. You will need to keep the issue of contamination in mind when you discuss your number count results. What shape are the galaxy counts likely to have as a function of magnitude, bearing in mind that we can clearly see beyond the stars in our own galaxy in these observations? What other observations might you be able to make, or processing you might be able to apply to the data, to be able to separate stars from galaxies?

Now that you have some idea of what to expect from a galaxy survey it's time to discuss the experiment, and how you'll produce galaxy catalogs from the data provided.

5 The Experiment

The goal of this experiment is to take a deep optical image of the sky, and to develop code that will produce a catalog of all the objects in this image, together with their photometry, allowing studies of the galaxy population to be undertaken. To achieve this, software that performs astronomical image processing must be written.

To be able to do this, your code must go through a number of steps:

1. Read the astronomical image into the computer so that further processing steps can be applied

2. Detect objects in the image

   This step is formally known as 'image segmentation' since you are dividing the image up into the segments you're interested in - the objects - and the segments you're not interested in - the rest.

3. Extract the amount of light coming from each of these objects

   In this step you'll measure the number of counts associated with each image. This process is termed photometry. There are many different ways to do this, but we will adopt a simple 'aperture photometry' approach in the first instance.

4. Produce a catalog of the objects you have detected and the amount of light emitted by them

5. Analyse the data from this catalog to see how well your image processing code is performing, and to see what you have found out about the galaxy population.

5.1 The Data

The data for this experiment can be downloaded from:

`http://astro.ic.ac.uk/public/mosaic.fits.zip`

5.2 Reading the data

Astronomical images are usually stored using the FITS standard (Flexible Image Transport System). A FITS file will contain not just the image data, stored in a machine independent form, but also a set of header data which describes the data, how it was obtained, and, in some cases, how it was processed. This 'metadata' is stored as KEYWORD VALUE pairs, where a keyword is a description of the parameter. For example you might find the keyword OBSERVAT and the value KPNO indicating that the data was taken at the Kitt Peak National Observatory (KPNO). For the purposes of this experiment most of the contents of the FITS header are not needed. One item you will need, though, is the instrumental zero point for these observations. This is stored in the FITS keyword MAGZPT.

You will need to read the contents of the FITS data file into memory so that your programmes can be applied to the data. There are several libraries of software to do this for different data analysis systems. You should pick the language and data analysis system you're going to use that best matches your skills. MatLab is available on the machines in 3rd year lab and is a popular and well-supported system. You can also use the more general programming language C++ that you've been taught in the 1st and 2nd years. Other systems like IDL can also be used if you're familiar with them.

5.2.1 In Matlab

The standard MatLab command fitsread can be used to read the data from the FITS file:

```
image = fitsread(filename)
```

This will put the contents of the image into the MatLab variable 'image'. More information on this command can be found in the MatLab documentation and help files.

Additional information on the use of MatLab is available as a supplement to the Waves on a String experiment.

5.2.2 In IDL

IDL is one of the standard data analysis systems in use in astronomy. To read FITS files use the astrolib collection of routines for IDL which can be found at

`http://idlastro.gsfc.nasa.gov/`

along with the necessary documentation.

5.2.3 In C++

The most commonly used libraries for handling FITS files in are CFITSIO, which is written in C but callable by C++ and other languages, and CCFITS, which is a native C++ interface to the CFITSIO library. This software can be downloaded from:

`http://heasarc.gsfc.nasa.gov/docs/software/fitsio/ccfits/`

Documentation and demonstration code is available on that website to get you started with reading in the astronomical data.

5.2.4 Visualizing the Data

When processing image data it is always useful to see what is going on. For this reason there are several visualisation tools available which you will find useful. You can view FITS files on Windows using software such as FITSVIEW (http://www.nrao.edu/software/fitsview/) or the more sophisticated ds9 (http://hea-www.harvard.edu/RD/ds9/). As with CFITSIO you will need to download and install this software. Unlike CFITSIO these are not libraries, but stand-alone applications. They will display FITS files and a number of other formats. To display a FITS image that you've generated you will have to write it out as a FITS file using CFITSIO and then read it into the visualisation programme.

5.3 Source Detection

The first step towards producing a galaxy catalog is to detect sources in the image. This can be done in a number of ways, but the simplest is to look for the highest pixel value in the image. This will obviously find the brightest source. The photometry of this source can be extracted and saved to your catalog. You then need to indicate that this bright source has been dealt with so that you can move on to the next brightest. One way of doing this is to use a 'mask' image. This is an 'image' matching the telescope image in size which flags areas that have been dealt with (eg. each source as it has been added to the catalog) or areas which are deemed 'bad' such as those associated with a bright star.

5.4 Source Photometry

Once you have found a source you need to measure how bright it is. With CCD data this is simply a matter of determining the total value of pixels contributing to the source, minus any background light. Subtracting off the total background contribution to the counts is very important because you will otherwise get a value for the brightness of your source which is much too high. There are many ways to estimate the flux of a source and to subtract off the background to such a measurement. One simple way to do this is to count the total value of all the pixels within a fixed distance from the source. This is so-called 'aperture photometry', because it is equivalent to observing the source with a fixed aperture size. A reasonable value to pick for this is a diameter of 3" (equivalent to about 12 pixels given the 0.258" per pixel scale of the image). To calculate the contribution of the background, an annular reference aperture can be used to calculate the local background to the object, in terms of counts-per-pixel. You can then multiply the background in counts per pixel by the number of pixels in your source aperture to work out the total background contribution to the aperture flux. This can then be subtracted from the aperture flux to give the flux due solely to the object.

Using a fixed aperture does, as noted above, cause problems for extended objects, many of which will be the brighter objects in the survey. Once your program is running you may wish to experiment with different, larger apertures to see what effect this has on your catalog and number counts. What problems come from using ever larger apertures? Can you think of alternative approaches to aperture photometry that might avoid any of these problems?

5.5 Calibrating the Fluxes

You need to convert from instrumental counts to source magnitudes. This usually involves comparison of the number of counts obtained in a standard star with the known magnitude of that star. Fortunately this has already been done for you. The result of this calibration is stored in the MAGZPT keyword in the FITS header, and the error on this calibration is stored in the MAGZRR keyword. You must then use these values to convert from observed number of counts in to source magnitude. To do this you first convert your counts into instrumental magnitudes:

$$mag_i = -2.5log_{10}(counts) \tag{3}$$

Instrumental magnitudes are then converted to calibrated magnitudes by adding the instrumental zero point:

$$m = ZP_{inst} + mag_i \tag{4}$$

or equivalently:

$$m = ZP_{inst} - 2.5log_{10}(counts) \tag{5}$$

5.6 Producing the Catalog

The results that you calculate for each object detected need to be stored in a catalog. This can be a simple ASCII file that can be read into another programme for data analysis. You need to store all the information you think might be useful in this file, including the (x,y) position of the sources, the total counts associated with them, the error on these counts, the background measured, and any other information that might be useful. You might, for example, wish to store the photometry for several different aperture sizes for use in later processing. In storing the data you will wish to convert the counts in the pixels into astronomical magnitudes using the instrumental zero points given in the FITS header. This will calibrate your data and allow comparison with existing and future astronomical observations.

5.7 Analyzing the Data

The basic result you are looking for in this experiment is the number count plot - the number of sources per square degree brighter than a given magnitude as a function of that magnitude (ie.

N(<m) as a function of m), with appropriate error bars. If you plot $log(N(m))$ as a function of m, you can then compare this to the expected theoretical relation given in equation 2.

6 Comparison with Existing Results

Number counts are a standard tool in astrophysics, so the results you produce here can be compared with any number of existing papers. The paper by Yasuda et al. (2001, Astronomical Journal, 122, 1104) provides an ideal basis for comparison with your results, but feel free to search for other, similar papers using the bibliographical capabilities of the NASA ADS system (URL given below).

Compare your results with the existing papers and with the predictions from equation 2. Where are there disagreements between your results and those in the literature? Where do you think these come from? Have you detected any signs of galaxy evolution? Discuss your conclusions in your report.

7 References and Useful Documents

Walker, G., 1989, Astronomical Observations, pub. Cambridge University Press
Sloan Digital Sky Survey: http://www.sdss.org
Bessel, M.S., 1990, Publications of the Astronomy Society of the Pacific, 102, 1181
Peacock, J.A., Cosmological Physics, 1999, pub. Cambridge University Press
Narlikar, J.V., 1983, Introduction to Cosmology, pub. Jones & Bartlett
Kitt Peak Mosaic Camera Manual: http://www.noao.edu/kpno/manuals/dim/
Spitzer Space Telescope Website: http://www.spitzer.caltech.edu/
SWIRE Survey Website: http://swire.ipac.caltech.edu/swire/swire.html
CFITSIO: http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html
CCFITS: http://heasarc.gsfc.nasa.gov/docs/software/fitsio/ccfits/
SAO/ds9: http://hea-www.harvard.edu/RD/ds9/
FITSVIEW: http://www.nrao.edu/software/fitsview//
ADS: http://cdsads.u-strasbg.fr/abstract_service.html
Yasuda et al., 2001, Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data, *The Astronomical Journal*, vol. 122, pp 1104 - 1124